USE OF EARTH OBSERVATIONS FOR ACTIONABLE DECISION MAKING IN THE DEVELOPING WORLD

EDITED BY: Niall Patrick Hanan, Ashutosh S. Limaye and Daniel Eric Irwin
PUBLISHED IN: Frontiers in Environmental Science

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88966-382-8
DOI 10.3389/978-2-88966-382-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

USE OF EARTH OBSERVATIONS FOR ACTIONABLE DECISION MAKING IN THE DEVELOPING WORLD

Topic Editors:
Niall Patrick Hanan, New Mexico State University, United States
Ashutosh S. Limaye, National Aeronautics and Space Administration (NASA), United States
Daniel Eric Irwin, National Aeronautics and Space Administration (NASA), United States

Citation: Hanan, N. P., Limaye, A. S., Irwin, D. E., eds. (2021). Use of Earth Observations for Actionable Decision Making in the Developing World. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-382-8

Table of Contents

frontiers in Environmental Science

EDITORIAL
published: 27 November 2020
doi: 10.3389/fenvs.2020.601340

Editorial: Use of Earth Observations for Actionable Decision Making in the Developing World

Niall P. Hanan[1]*, *Ashutosh S. Limaye*[2] *and Daniel E. Irwin*[2]

[1] *Plant and Environmental Sciences Department, New Mexico State University, Las Cruces, NM, United States,* [2] *National Aeronautics and Space Administration (NASA), Marshall Space Flight Center, Huntsville, AL, United States*

Keywords: Earth Observation, SERVIR, satellite data, applied research, sustainable development

Editorial on the Research Topic

Use of Earth Observations for Actionable Decision Making in the Developing World

INTRODUCTION TO THE SPECIAL TOPIC

The global community faces fundamental challenges related to natural resource management, particularly in low and middle income countries. Increasing populations and changing climate exacerbate the challenges of food security, water, and environmental management and sustainability faced by rural communities, cities, and governments alike. Sound environmental and risk management decisions must often be made with incomplete or inadequate information, particularly in the developing world. However, Earth Observations (EO) can bring objective data to developing-country land-users, land managers and policymakers by addressing key information gaps through consistent and repeatable observations (Zell et al., 2012) and through capacity development efforts (Kumar et al., 2020).

Technological innovations and proactive EO agendas developed by space agencies around the world are helping (Brown et al., 2013). In the United States, for example, the prioritization of key EO science and applications in the recent National Academies decadal survey (National Academies of Sciences, Engineering, and Medicine, 2019) will ensure continuity of orbiting sensors that measure critical Earth system parameters. Similarly, the European Space Agency's Copernicus program (Brachet, 2004) makes high quality and up to date satellite data freely available, and consequently the use of these data is expanding rapidly. There is an explosion of active, ongoing research in the US and in the developing world that can directly translate into working theories and operational systems for developing countries. Ensuring alignment between the research community and decision makers, and orienting the latest science from the United States to help solve local/regional problems, has been the primary objective of the NASA Applied Sciences Program (Friedl, 2017). The purpose of the articles in the Special Topic "Operational Earth Observation for Sustainable Development and Risk Reduction in the Developing World" is to focus on the challenges and solutions for use of EO data for actionable decision making around the world, particularly in developing countries.

SERVIR is a joint initiative of NASA and the U.S. Agency for International Development (USAID) which fosters applications of EO to help developing countries assess environmental conditions to improve planning and management interventions. Working with regional technical organizations ("regional hubs") around the world, SERVIR aims to improve resilience and sustainable resource management at local, national, and regional scales through the increased use of Earth Observations, Earth science, and technology. SERVIR hubs have regional mandates to work with national governments of their member countries, thus promoting strong pathways between

OPEN ACCESS

Edited and reviewed by:
André Mascarenhas, Humboldt University of Berlin, Germany

***Correspondence:**
Niall P. Hanan
nhanan@nmsu.edu

Specialty section:
This article was submitted to Land Use Dynamics, a section of the journal Frontiers in Environmental Science

Received: *31 August 2020*
Accepted: *23 October 2020*
Published: *27 November 2020*

Citation:
Hanan NP, Limaye AS and Irwin DE (2020) Editorial: Use of Earth Observations for Actionable Decision Making in the Developing World. Front. Environ. Sci. 8:601340. doi: 10.3389/fenvs.2020.601340

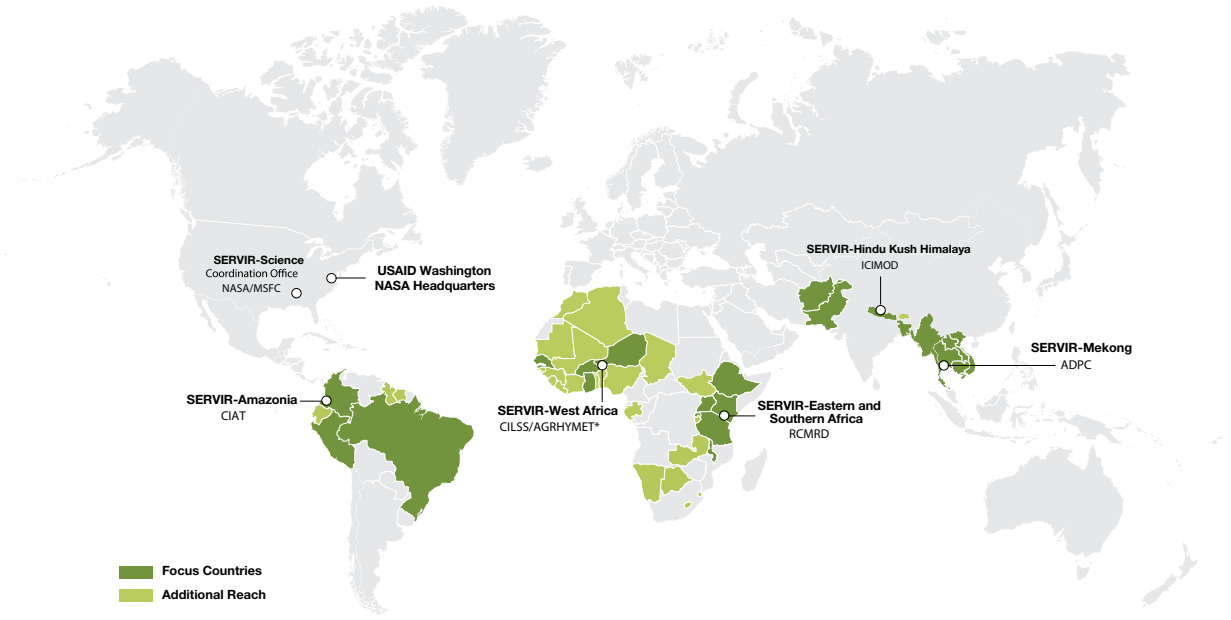


FIGURE 1 | Geographic regions and primary countries served by the five SERVIR regional hubs.

EO-based solutions and policy makers. SERVIR hubs are active in five regions (**Figure 1**), including Eastern and Southern Africa, the Hindu Kush Himalaya (HKH), lower Mekong, West Africa, and most recently in Amazonia. SERVIR focuses on four themes: land use, land cover, and ecosystems; food security and agriculture; water and related disasters; and weather and climate. For the land cover theme, SERVIR has ongoing activities in all regions, to strengthen the capacity of regional partners to link the latest science to promote sustainable use of natural resources. In agriculture, SERVIR promotes EO-based crop monitoring technologies to better analyze and predict food security conditions. Connecting satellite-derived rainfall and other datasets helps SERVIR improve the capacity of countries to forecast streamflow and identify regions where water-related disasters can occur. Weather and climate activities inform the other themes in ensuring the latest and robust availability of weather information, in addition to short- and long-term weather forecasts.

Articles in this Special Topic introduce a range of EO-based approaches for monitoring natural resources, risk, and disaster analysis and prediction in SERVIR regions of Africa, Himalaya-Hindu Kush, and Southeast Asia. These include the use of EO for hydrological monitoring, drought and flood prediction, land use, land cover and agricultural monitoring, and invasive species. Several papers focused on emerging remote sensing technologies with applications in less developed regions, while others focused on applications and stakeholders from local (village) scale to national, regional, and global scales.

Two papers focus explicitly on the critical phases of stakeholder engagement during design, development, and implementation of EO applications to enhance relevance, use and impact. The paper by Thapa et al. documents the steps involved in engaging relevant communities to define their needs and design appropriate EO applications. Saah et al. provide a specific example of the user engagement process during development of land cover mapping tools for diverse users in the Mekong Basin and Hindu Kush-Himalaya regions.

Several papers focused on use of EO as inputs for hydrological modeling applications. Purdy et al. focused on the use of EO for analysis of groundwater dynamics and agricultural sustainability in Bangladesh. Nelson et al. and Alcantara et al. demonstrate the utility of cloud-based hydrological modeling systems for stream-flow and flood prediction, with a focus on the transfer of model outputs into actionable information for regional and local decision makers. Sikder et al. analyze the impact of EO inputs, land surface model, and river routing schemes on river-flow predictions in the critical Ganges-Brahmaputra and Mekong river basins of South and Southeast Asia. Oddo and Bolten demonstrate that the integration of satellite data into a flood modeling system for the city of Bangkok, Thailand can be of immense value in potential lives and property saved through shorter emergency response times. Phongsapan et al. explore the utility of satellite data and improved hydrological forecasting for flood risk reduction in Myanmar.

An analysis by Spruce et al. for the Lower Mekong Basin revisits the difficult process of land use and land cover (LCLU) classification, particularly in the context of temporal analysis of change, where classification errors in individual time-periods can be compounded in change analysis. The LCLU maps made available for the region are important for a variety of EO applications, including hydrological and agricultural monitoring and modeling. Poortinga et al. analyzed long-term land cover

estimates for Vietnam, together with national agricultural yield statistics, to suggest significant increases in agricultural yield (per unit area) with relatively little change in total cropland area. Tiwari et al. presented methodologies using Earth Engine and satellite data to map irrigated and rain-fed wheat in Afghanistan, providing a tool for crop area estimation critical for yield forecasting and food security planning in remote regions. Analyses of this type can contribute significantly to understanding the spatial and temporal dimensions of land use and food insecurity in low income and often remote rural communities.

Three studies focused on rangeland monitoring. Ndungu et al. outline a comprehensive EO tool for rangeland monitoring tailored to stakeholder needs at local and regional scales in East Africa. The Rangeland Decision Support Tool provides easy access to satellite-based vegetation indices that can be integrated with ancillary GIS datasets specific to the needs of local communities and range managers. Anchang et al. developed a work-flow for mapping woody canopy cover using Google Earth Engine, local field data and expertise, machine learning, and data from the Sentinel satellites for Senegal, West Africa. The cloud-based computing approach reduces the computational barriers often faced by the EO community in countries with relatively poor internet bandwidth, and is a tool available for local calibration and application in other regions. Meanwhile, Ouko et al. demonstrated an approach to mapping and modeling invasive plants in the ecologically, economically, and culturally valuable rangelands of East Africa, using cell-phone based citizen scientist inputs on invasive plant locations, satellite, and GIS data, to map the presence and spread of invasive species. This study focused on an invasive shrub (*Acacia reficiens*) and prickly pear (*Opuntia* spp.), but similar approaches could be adopted in other regions where invasive plants are impacting ecosystem integrity and rural livelihoods.

One paper in this Special Topic, by Flores-Anderson et al. examined the critical, but less well-known, potential to use hyperspectral imagery for lake water quality assessment. While the review paper by Leibrand et al. examines how satellite data can contribute to all aspects of planning and management of rural electrification, renewable and sustainable energy planning and management.

CONCLUSION

The studies in this Special Topic share a major common theme—the use of EO to facilitate improved decision making by land users and managers at local, national, and regional scales. The value of EO for decision making in land cover monitoring, agricultural assessments, or streamflow predictions is high, given the spatial continuity, and the temporal return frequency afforded by space-based platforms. That value is particularly critical in developing countries where EO information is sometimes a key source for decision makers. Responsiveness to the specific needs of different stakeholders, based on local economic, cultural, and environmental conditions is especially important, because EO applications then have a higher chance of achieving real impact for local land users or via sustainable policy interventions. The result is a diversity of individual strands (approaches) to EO-informed solutions and it is up to scientific communities, such as SERVIR, to weave those solutions together in the hope and anticipation that these approaches will be replicated in other parts of the world. The scalability and replicability of activities in a global program such as SERVIR is critically dependent on these individual efforts. This compilation of articles provides greater insight into some of those activities.

AUTHOR CONTRIBUTIONS

All authors contributed equally to this article.

REFERENCES

Brachet, G. (2004). From initial ideas to a European plan: GMES as an exemplar of European space strategy. *Space Policy* 20, 7–15. doi: 10.1016/j.spacepol.2003.11.002

Brown, M. E., Escobar, V. M., Aschbacher, J., Milagro-Pérez, M. P., Doorn, B., Macauley, M. K., et al. (2013). Policy for robust space-based earth science, technology and applications. *Space Policy* 29, 76–82. doi: 10.1016/j.spacepol.2012.11.007

Friedl, L. (2017). "Chapter 4: Benefits assessment of applied earth science," in *Satellite Earth Observations and Their Impact on Society and Policy*, eds M. Onoda and O. Young (Singapore: Springer), 73–79. doi: 10.1007/978-981-10-3713-9_4

Kumar, S. A., Camacho, S., Searby, N. D., Teuben, J., and Balogh, W. (2020). Coordinated capacity development to maximize the contributions of space science, technology, and its applications in support of implementing global sustainable development agendas—a conceptual framework. *Space Policy* 51:101346. doi: 10.1016/j.spacepol.2019.101346

National Academies of Sciences, Engineering, and Medicine (2019). *Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space*. Washington, DC: The National Academies Press. doi: 10.17226/24938

Zell, E., Huff, A. K., Carpenter, A. T., and Friedl, L. A. (2012). A user-driven approach to determining critical earth observation priorities for societal benefit. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5, 1594–1602. doi: 10.1109/JSTARS.2012.2199467

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

REVIEW
published: 27 August 2019
doi: 10.3389/fenvs.2019.00123

Using Earth Observations to Help Developing Countries Improve Access to Reliable, Sustainable, and Modern Energy

Amy Leibrand, Natasha Sadoff, Tanya Maslak and Amy Thomas*

Battelle, Columbus, OH, United States

OPEN ACCESS

Edited by:
Ashutosh S. Limaye, National Aeronautics and Space Administration (NASA), United States

Reviewed by:
Laurence Delina, Boston University, United States
Justin Reedy, University of Oklahoma, United States

****Correspondence:***
Natasha Sadoff
sadoffn@battelle.org

Specialty section:
This article was submitted to Science and Environmental Communication, a section of the journal Frontiers in Environmental Science

Received: *02 May 2019*
Accepted: *10 August 2019*
Published: *27 August 2019*

Citation:
Leibrand A, Sadoff N, Maslak T and Thomas A (2019) Using Earth Observations to Help Developing Countries Improve Access to Reliable, Sustainable, and Modern Energy. Front. Environ. Sci. 7:123. doi: 10.3389/fenvs.2019.00123

In this review paper, the authors identify priority areas, and opportunities for electric utilities in developing and emerging economies to incorporate Earth observation (EO) data into rural electrification planning, renewable energy resource assessment, distributed generation, grid operation and reliability, and disaster risk reduction and recovery efforts. Using a methodological framework, the authors conducted a comprehensive literature review of primary and gray literature. This paper reviews the many existing applications for EO data, such as the use of nighttime lights imagery for estimations of rural electrification, EO-derived normalized difference vegetation index (NDVI) products for vegetation monitoring for overhead transmission line management, solar radiance data for renewable energy project planning, and nowcasting for extreme weather events and other disaster monitoring. These and other applications can enhance energy security through improved governance of and access to modern and reliable electricity, renewable energy management, and disaster risk assessment in developing nations, paving the way for more sustainable social and economic development. Real-world examples of EO data use by utilities in developing and emerging economies, as well as barriers and opportunities for EO technology transfer, are discussed. Recommendations for stakeholder engagement, future EO training opportunities, and human capacity building are also presented.

Keywords: earth observation, rural electrification, renewable energy, electric grid, disaster risk

INTRODUCTION

The 2030 Agenda for Sustainable Development, adopted by the United Nations (UN) Member States in 2015, provides a shared blueprint of action centered around the 17 Sustainable Development Goals (SDGs), and an urgent call for action by all countries—developed and developing—to act in a global partnership to promote peace and prosperity.

As one of the SDGs, UN SDG 7[1], "Ensuring access to affordable, reliable, sustainable, and modern energy for all" and electrification is a priority in many developing regions. Globally, about 1.1 billion people are without electricity, mostly in rural and remote locations in developing countries (International Energy Agency, 2017a). Lack of access to reliable, modern electricity is a significant obstacle to social and economic development, quality of life, and public safety, particularly within developing countries (International Energy Agency, 2014; Cook et al., 2015)[2]. SDG 7 plays a central role in supporting all other SDGs through direct and indirect contributions to environmental sustainability, human development, and sustainable economic growth.

Satellite-based Earth Observation (EO) data offer opportunities to advance understanding of both natural and human-induced global changes from which forecast models, information products and other tools could improve decision- and policy-making for various societal benefit areas (National Research Council, 2007; National Academies of Sciences, Engineering and Medicine, 2015; National Academies of Sciences, 2018). This is particularly important when addressing issues that are regional to global in scale or require resources that may necessitate transboundary cooperation. There are several distinct roles EO can offer when applied across atmospheric, marine, and terrestrial systems to assist with environmental governance, which could also potentially be applied to other societal benefit areas, such as the energy sector. The roles EO can offer center around the supply of information (including identifying data needed to address an issue, monitoring status of the issue or data trends over time, and assessing or evaluating success of applied solutions); assisting in early warning and/or disaster response; and documenting compliance with specific requirements or prohibitions (Young and Onoda, 2017). EO data also reduce the manual effort required to conduct feasibility assessments and increase accuracy. The combination of EOs and geospatial data with demographic and other sector-specific data offers an opportunity to provide the derived information needed by stakeholders to analyze and model potential resource related supply and demand scenarios, evaluate impacts across sectors and geographical regions, create maps and other visualizations to relay information, and otherwise assist end-users in making informed decisions that will contribute toward achieving the SDGs. Combined EO and demographic or socio-economic information can also provide a more nuanced analysis of environmental risks or management issues, allowing for improved governance of natural resources, energy resources, and populations. Applying this framework of EO applications to the energy sector offers a wide range of opportunities for facilitating developing countries' quest to improve access to reliable electricity and resilience infrastructure, incorporate renewable energy into the electric sector, and improve energy governance.

Public acceptance of new or unfamiliar technologies and infrastructure investments and siting is a crucial component in effectively addressing evolving energy needs and promoting good environmental governance (Haggett, 2009a; Devine-Wright, 2010). While regulatory frameworks vary drastically, in many countries the potential environmental, social, economic, and health impacts of large-scale projects, such as renewable energy installations, require a formal impact assessment process. For example, EO information can support environmental impact assessments (EIA) by providing historical, current, and projected environmental or human settlement-related information, contributing where knowledge and data gaps may otherwise occur and otherwise supporting consensus building between developers and community members (Maclean et al., 2014). Governing bodies must employ a constructive approach to obtaining public buy-in, particularly in renewable energy projects where community opposition to the unfamiliar and the Not-In-My-Backyard ("NIMBY") phenomenon often plagues operational equipment siting (Larson and Krannich, 2016). Public engagement mechanisms take many forms and at a minimum should allow the public some degree of influence over the decision-making process. For example, granting local actors the authority over organization activities, such as coordination of meetings or creation of informational leaflets, enables a sense of ownership of the process and the project, thus increasing the level of trust (Areizaga et al., 2012). These and other approaches contribute to public acceptance of new approaches to energy management. The availability of EO information can support public engagement processes, such as in the siting of new renewable energy development projects (Hindmarsh and Matthews, 2008). Similarly, EO tools can also serve to provide stakeholders such as potential project developers or funders, grid operators, or national or private scale transmission entities, with renewable energy studies that take into account site-specific, seasonal variability forecasts for improved financial analyses (Kosmopoulos et al., 2018). Governance and public engagement can also be improved with greater understanding of human and physical environments or geography through EOs of appropriate resolution (Butler et al., 2011). EO tools can help to inform the public on potential impacts and opportunities of renewable energy installations by allowing visualizations to graphically represent relevant data or forecasts contributing to better understand potential impacts, community benefits, or justification for a new project. In Morocco, for example, EO information was used to identify settlements, and the boundaries of communities when assessing the development of large-scale solar installations (Hanger et al., 2016).

Additionally, the opportunity to engage end users of data products, such as those tools and information products incorporating EO, can influence future satellite mission design and ensure that investments in future EO capabilities are applications-oriented and serve both the scientific and user community. The need to foster closer relationships between the science and end user community, as well as incorporate socioeconomic factors in the planning and implementation of EO missions, is well documented (National Research Council, 2007; Brown et al., 2013; National Academies of Sciences, Engineering and Medicine, 2015).

This article provides an overview of the existing gaps and challenges to the use of EO data, current state of research, and

[1]https://sustainabledevelopment.un.org/sdg7

[2]https://sustainabledevelopment.un.org/?page=view&nr=2749&type=13&menu=1634

examples of how electricity providers are incorporating EO data into rural electrification planning, renewable energy resource assessment, grid management, and disaster risk reduction and recovery efforts throughout the developing world. While there are myriad opportunities for improved utilization of EO by developing countries to meet their energy needs and goals, challenges to the successful use of EO remain and require careful consideration (Lahoz and Schneider, 2014). This review, based on a methodological literature search, is not meant to be exhaustive or encompassing of all applications, tools, research areas, or case studies, but instead sets a necessary foundation or baseline for understanding and addressing common barriers for EO uptake for the energy sector in the developing world. Toward that end, the article concludes with recommendations on how to maximize current applications and reduce barriers, particularly through stakeholder engagement, future tailored EO needs, and human capacity building.

METHODS

A literature search that followed a methodological framework was conducted using targeted vocabulary terms to query Google Scholar and Google Search. Google Scholar allows simultaneous searching of a large number of unique publication types, including peer-reviewed journal articles, theses and dissertations, books, abstracts, and technical reports from academic publishers, professional societies, online repositories, universities, and other web sites. Google Scholar was selected over other scholarly databases because of its breadth of publication types. In an exploratory scan of literature, the authors found that journal articles were primarily focused on highly technical research or one-off projects, whereas non-traditional documents provided a better representation of information about the existing use of EO in real-world energy management applications. A perceived limitation to using Google Scholar is that publication quality is lower compared with other scholarly databases. However, a recent comparison of Google Scholar, Scopus, and Web of Science concluded that Google Scholar found "nearly all" Scopus and Web of Science citations and that "most citations found only by [Google Scholar] were from non-journal sources" (Martín-Martín et al., 2018). Another potential limitation to using Google Scholar is that results are ordered largely by citation count, which may not be as robust as other scholarly databases that rank by relevance. To mitigate this limitation, the authors used focused search strings to limit results to those most relevant and performed a scan of all document titles returned to assess potential relevancy. Titles deemed potentially relevant were marked for further review.

The Google Scholar review was supplemented by a query of Google Search to identify gray literature, which are non-peer reviewed materials produced by government, business, and industry in a variety of formats including white papers, presentations, newsletters, and other non-traditional materials where "publishing is not the primary activity of the producing body" (Schöpfel, 2010). Gray literature was an important component of this review because of the authors' desire to highlight the practical applications of EO use outside of a research context. To ensure coverage of all major aspects of space-based remote sensing for energy applications, hand searches of individual websites of select electricity providers and consortiums were also conducted.

Search terms related to the use of EO data for energy management were used to query online databases and resources. Broad keywords used for the search are summarized in a table included as **Supplementary Material**. The keywords were identified in a preliminary review of recent government reports and similar documents to identify priority areas within the electricity sector. Search terms were augmented as appropriate to include plurals and variants, root terms, alternative spellings, and synonyms. Terms were combined as necessary using Boolean AND/OR/NOT logic to achieve a manageable number of search results. Due to the extensive body of research available, the search focused on resources published from 2010 to the present. Using this search strategy, 9,856 potential titles related to EO data for energy management were identified. To further narrow the focus, the titles were initially scanned for relevancy. A total of 2,763 documents were down selected for abstract/summary review. If an abstract or executive summary suggested that a document was relevant or did not make relevance clear, the document was prioritized for full text review. Of the 2,763 abstracts/summaries reviewed, 384 documents were prioritized for full text review according to the following categories and criteria[3]:

- **Key Studies (High Relevance):** Documents that described in detail the specific uses of satellite imagery or data products related to priority areas or documents that described sector needs directly related to priority areas. A total of 261 documents were identified as "Key Studies" based on abstract/summary review. An in-depth review of full text was performed for these documents.
- **Relevant Studies (Intermediate Relevance):** Documents that broadly described general uses of satellite imagery or data products directly related to priority areas or that were highly technical in nature. A total of 55 documents were identified as "Relevant Studies" based on abstract/summary review. An expedited review of full text was performed for these documents. Upon expedited review, documents deemed high relevance were recategorized as "Key Studies" and subject to an in-depth review.
- **Supporting Studies (Low Relevance):** Documents that broadly described uses of satellite imagery or data products tangentially related to priority areas. A total of 68 documents were identified as "Supporting Studies" based on abstract/summary review. An expedited review of full text was performed for these documents. Upon expedited review, documents deemed high relevance were recategorized as "Key Studies" and subject to an in-depth review.
- **Not Relevant:** Documents not meeting the criteria for any of the preceding categories. A total of 2,379 abstracts/summaries

[3]The complete list of references deemed relevant are included in **Supplementary Material**.

reviewed were considered "Not Relevant" for the purposes of this project and were excluded from full text review.

The bibliography sections of particularly useful or key articles/reports were also further examined for additional relevant references. These secondary references were then obtained and reviewed for relevance according to the above criteria. An exhaustive review of all available literature, however, was not conducted. The final results of the literature search included peer-reviewed literature as well as gray literature, such as industry/consortia reports, workshop/conference reports, and key research reports from federal government agencies and others.

REVIEW RESULTS

Results are presented in this section according to four key themes identified through the literature review: rural electrification, renewable energy, grid management, and disaster risk reduction, and response. For each theme, gaps and challenges are discussed, information regarding the current state of research is presented, and examples of the real-world use of EO data in developing countries are provided.

Rural Electrification

While significant progress toward UN SDG 7.1 "Ensuring access to affordable, reliable, sustainable and modern energy for all" has been achieved in certain regions of the world, a significant proportion of the global population, particularly in rural and remote locations, still lack this critical resource. For example, sub-Saharan African countries are among the least electrified in the world[4]. In Zambia, approximately 72% of a population of 17 million lacks access to electricity, with rates of inaccessibility as high as 95% in rural areas (Cader et al., 2018). Electrification rates in urban vs. rural areas may vary drastically even within a single country. For example, the electrification rate in the urban areas of Cambodia is 97% while only about 50% of Cambodia's rural population has access to modern electricity (International Energy Agency, 2017b). Cambodia plans to extend the national grid and build hydropower plants to improve the access of both urban and rural populations to electricity (International Energy Agency, 2017b).

In Indonesia and Myanmar, it is estimated that about 45 million people in total were without access to electricity in 2016 (International Energy Agency, 2017b). Improving access in these countries is a high governmental priority. Myanmar has developed a National Electrification Plan that includes expanding the centralized grid and financing DG/microgrid solutions. Similarly, electrification plans for Indonesia include extension of the existing grid, as well as microgrids that rely largely on renewable energy, such as hydro and solar.

The World Bank reported in 2014 that more than 300 million people in India are impacted by frequent outages or a complete lack of power due to insufficient infrastructure[5]. In 2015, the Government of India's Ministry of Power launched the "Deen Dayal Upadhyaya Gram Jyoti Yojana" scheme (DDUGJY) to reform the power sector, which claimed in April 2018 to have electrified all villages in India[6]. However, the Indian government considers a village "electrified" if it has basic electrical infrastructure, public places are electrified, and 10% of its households have electricity[7]. Given this interpretation, there appears to be ample opportunity to expand access to reliable electrification to most households in India.

Spatial information for energy access planning traditionally involves labor-intensive field studies and surveys to collect and analyze data. Freely available EO data have facilitated the development of spatial maps for developing countries, providing critical information for addressing the problem of electricity access. In areas where ground-based data are sparse, EO data can provide critical insight into environmental or energy-related parameters. Further, tools that combine satellite-based EO data with *in-situ* ground-based data can provide a valuable resource for understanding conditions for siting new projects or expanding grid infrastructure.

The application of EO resources offers the potential for electricity providers in developing nations to identify human settlement areas that may still be lacking the infrastructure and access to electricity and monitoring progress in electrification of remote, rural areas. While EO data cannot completely replace ground-based information, it provides an actionable, objective resource for identifying unelectrified areas, ideal siting locations for grid expansion, and areas suitable for self-sustaining microgrids to help in meeting electrification goals.

Gaps and Challenges

There are many barriers to electrification in developing countries, including inadequate policy or legal mechanisms, market drivers, governance structures, physical infrastructure, human capital, and data access or technological gaps (Urmee et al., 2009; Ahlborg and Hammar, 2014; Chauhan and Saini, 2015). In some cases, even land acquisition for project siting can be complicated due to protected land areas. Data on renewable resources is required but many developing countries are data poor in terms of ground-based data systems that monitor and map renewable resources. Often, data for electrification planning are difficult or costly to access because on-the-ground resources are required. EO data, particularly open access data, can be used to overcome the logistical and financial challenges of data acquisition in particular. The use of nighttime lights imagery is a well-established method for identifying unelectrified areas and the extent of rural settlements, although ground-based population data are often needed to supplement imagery-based population assessments to ensure accuracy (Cader et al., 2018). EO-based land use data have been used for the siting of grid infrastructure, although, without the use of data analytics, this method requires manual effort to assess aerial imagery for optimal locations. Further, internet access is required to obtain EO datasets, which can be a challenge in regions where accessibility is limited.

[4]https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS

[5]http://www.worldbank.org/en/news/feature/2014/06/24/switching-on-power-sector-reform-in-india

[6]https://twitter.com/narendramodi/status/990455176581517312

[7]http://www.ddugjy.gov.in/portal/definition_electrified_village.jsp

State of Research

To assess rural access to electricity in regions where census data may be lacking, data are needed to determine population density in relation to existing utility assets. Current research in this area is largely focused on the development of machine learning models that analyze high-quality EO data to reduce manual efforts involving field surveys and human interpretation of aerial images. Researchers have demonstrated the use of data analytics with EO data products to assess population density and village attributes. Settlement identification and the evaluation of density distribution using open access data from satellites such as Landsat, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Visible Infrared Imaging Radiometer Suite (VIIRS), and others have been validated by researchers as useful methods that can supplement survey data in regions where census data are incomplete or inaccurate. EO-derived data such as nighttime brightness, tree cover change, and thermal emission and reflection can be combined and modeled to predict likely population clusters (Stevens et al., 2015; Longwei and Dengsheng, 2016; Robinson et al., 2017; Engstrom et al., 2019). Researchers are also developing machine learning models with EO data to assess electric infrastructure assets, such as locations of power plants and overhead transmission lines, and to estimate asset reliability. Researchers at Duke University's Energy Data Analytics Lab recently developed a machine learning method that automates the evaluation of satellite imagery to determine electrification rates[8]. These and other models have demonstrated the ability to predict the current availability of electricity in remote areas and the quality of existing infrastructure using EO data from Landsat 8, Sentinel 1, VIIRS, Google Maps, and others (Varshney et al., 2015; Mann et al., 2016; Oshri et al., 2018). EO data and machine learning together will facilitate improved prediction and analytic capabilities for efficient grid planning and monitoring of electrification progress. The US National Aeronautics and Space Administration (NASA) introduced its Black Marble product in 2012 to replace the DMSP-OLS nighttime lights dataset. Black Marble provides images of Earth at night, taken from NASA's Suomi satellite using VIIRS, which offers a resolution significantly better than the previous system used to acquire nighttime lights data. Black Marble can be used to classify different types of population centers, such as urban, rural, suburban, villages, towns, etc. (Altaweel, 2018).

Real World Use of EO Data in Developing Countries

There are a limited number of current applications or real world uses of EO data in developing countries. Satellite imagery known as "nighttime lights" has been used to assess the expansion of electrification and electricity consumption over time and space. The National Oceanic and Atmospheric Administration (NOAA) maintains archived Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) nighttime lights data from 1992 to 2013; however, the ongoing value of the DMSP-OLS dataset is limited because the data are no longer current. Nonetheless, researchers at the University of Michigan and The World Bank used this dataset to develop a map of the nighttime brightness of India's 600,000 villages as an indicator of electrification[9]. The project, known as India Lights, has been used extensively to assess the value of EO data for rural electrification planning. An analysis of the India Lights platform concluded that EO data are "surprisingly accurate" for assessing rural electrification, less so when the power supply is unreliable, but GIS maps and processing tools can help to improve assessments of brightness. On its own, nighttime lights may not be adequate but can be complemented with field information to identify specific locations for electricity expansion or microgrid installations (Dugoua et al., 2018).

EO data can also inform grid infrastructure siting. The Power Grid Company of Bangladesh Limited (PGCBL) used high-resolution satellite images in the development of a transmission line to expand the distribution system and increase access to electricity. The EO data were used in conjunction with maps and consultation with local stakeholders to assess geographic land features and alternative routes as options for the transmission line. EO data were crucial in identifying an optimal route that minimized impacts to existing infrastructure, settlements, and the environment. Specifically, satellite imagery was used to assess fish habitats along the proposed transmission route, urban areas, connecting roadways, river crossings, and other areas of importance (Asian Development Bank, 2017).

Microgrids, a variety of distributed generation (DG) that can operate autonomously, are widely used in developing countries to provide access to electricity in rural areas where connection to the central grid is not feasible. Powerhive East Africa Ltd., a U.S.-based microgrid solutions provider that is licensed to generate, distribute, and sell electricity in Kenya, uses a proprietary tool called SWARM for microgrid site selection[10]. The SWARM tool uses EO data to "identify potential customers and create preliminary microgrid designs for viable sites"[11]. In particular, the tool identifies the extent of the existing electric grid by incorporating nighttime lights imagery derived from NASA's Black Marble product, which is a cloud-free, higher resolution replacement for DMSP-OLS nighttime lights imagery (Román et al., 2019). As noted earlier, EO data can be complemented by ground-based data to assess population clusters and unelectrified areas.

Renewable Energy

Renewable energy is typically used in developing and emerging economies to power DG or microgrid systems in areas where connection to the centralized grid is impractical due to lack of infrastructure or other barriers. Often, these systems are part of a utility's overall solution to improving access to electricity. There are excellent renewable energy opportunities in many developing regions; however, policy, capacity, or financial barriers may prevent large-scale deployment of renewable energy technologies. Conversely, renewables-based small-scale

[8]https://bassconnections.duke.edu/project-teams/energy-data-analytics-lab-electricity-access-developing-countries-aerial-imagery-2017

[9]http://india.nightlights.io

[10]http://www.powerhive.com/our-technology/

[11]Ibid.

microgrid systems have seen rapid expansion in developing nations and are a crucial component to providing electricity access to rural populations (Stiles and Murove, 2018) and making further progress toward SDG 7.1.

The primary renewable technologies used in developing regions are solar photovoltaics (PV), small wind turbines, small-scale hydroelectric power, and biomass combustion devices (Stiles and Murove, 2018). Growth and resource potential are strong, particularly for solar energy. In 2016, developing countries added about 34 GW of solar capacity (BloombergNEF, 2018). In sub-Saharan Africa, solar energy potential is tremendous, particularly for microgrid solutions. Studies suggest that there is a "theoretical annual electricity generation potential of 660,000 TWh for Solar PV in Africa" (Quansah et al., 2016). Significant solar resources are also noted in Mexico, regions of Central and South America, Southeast Asia, India, and the Middle East[12].

In addition to the potential presented with solar energy, the World Energy Council estimates that global wind capacity could grow "from 435 GW in 2015 to 977 GW in 2030" (World Energy Council, 2016), with China, India, and Brazil leading installed capacity and continued sector growth among developing economies (Global Wind Energy Council, 2017). There is also substantial wind energy potential in Chile, Argentina, Kenya and other parts of sub-Saharan Africa, the Middle East, Vietnam, and the Philippines[13]. The greatest potential for utilization of wind energy is in remote villages where connection to a centralized grid is not feasible, and many of these countries have administrative goals for wind energy projects, such as the Philippines with a target of 2.3 GW of capacity by 2030 (International Energy Agency, 2017b) and Kenya with a goal to add 0.4 GW of capacity by 2020 (International Energy Agency, 2014).

Hydroelectric power is "the leading renewable source for electricity generation globally" (World Energy Council, 2016) with China, Brazil, and India as the fastest growing countries for this sector in 2017 (International Hydropower Association, 2018). Although, Brazil has recently removed several hydropower projects in favor of decentralized energy, other South American countries are seeing large growth in the hydroelectric power sector with plans for future plants in Peru, Columbia, and Argentina. Planned and installed hydroelectric projects also dominate the renewable energy sector in sub-Saharan Africa, where there exists significant untapped potential for hydroelectric power development (Stiles and Murove, 2018). While the potential growth of the hydroelectric power industry offers substantial opportunity for new use of EO data-derived products in this sector, so does the potential decline of water resources in certain regions due to environmental changes such as climate variability and change; therefore, electricity providers must adopt resilience planning for water shortages and manage risk accordingly if pursuing this energy source.

Geothermal energy is feasible in countries located in the "Ring of Fire," such as Mexico, Indonesia, the Philippines, as well as other regions worldwide. Turkey and various Pacific coastal countries in Central and South America have seen strong growth in the geothermal sector largely due to supportive governmental policies (Stiles and Murove, 2018). Ethiopia and Kenya in Africa's Rift Valley also provide strong potential for geothermal power. Moreover, Costa Rica, El Salvador, and the Philippines derive more than 10 percent of their electricity from geothermal sources (Van Nguyen et al., 2015).

EO data can play an important role in assessing the feasibility of renewable energy systems and prospecting for resources. Solar energy resource assessment, management of PV installations, wind power estimation and wind farm siting, and environmental impact assessment are long established uses of EO data. Similarly, EO-derived thermal infrared data, surface wavelength measurements, and gravity anomalies have been used for more than a decade in the geothermal prospecting of large land areas. EO data are also commonly used to assess water quantity and availability for hydroelectric project planning and monitoring, determination of ideal dam locations and reservoir size, and environmental impact from rerouting or damming water.

For successful renewable energy projects, both government, and community expectations must be met. Participatory governance for consensus building among stakeholders should include participation mechanisms throughout the life of a project, from the early proposal stage to post-project. Public engagement plays a "crucial role" in the social acceptance of renewable energy technologies, and may involve collaborative decision-making, community shareholding, and public education and consultation (Langer et al., 2017). Such approaches provide the community some level of control, which facilitates trust between stakeholders, provided that the participation process is fair and just. However, research suggests that mandating engagement between developers and the public in the siting process is difficult, and without incentives, developers may not be motivated to address community concerns (Haggett, 2009b; Bell et al., 2013; Christidis et al., 2017). Furthermore, public concern over "not being heard" is a theme repeated in the literature (Walker et al., 2010; Natarajan et al., 2018). Thus, consideration must be given to all stakeholders to ensure effective collaboration and trust. The U.S. Department of Energy (DOE) maintains a toolkit for public involvement in renewable energy and infrastructure[14] and among its key strategies include identifying issues of concerns and communicating clear and concise messages. Using EO data in maps or visualizations can offer the public clear, unbiased information on environmental parameters such as land use, land change, and available resources.

Trust is a critical factor in community adoption of renewable energy and associated data for energy assessment, such as EO. Level of knowledge and experience pertaining to renewable energy varies widely between policymakers, researchers, developers, and the public. Trust between stakeholders is largely contingent on accurate and timely communication, which is an "essential component of [the] public engagement process"

[12]https://solargis.com/maps-and-gis-data/download
[13]https://www.globalwindatlas.info/
[14]https://openei.org/wiki/RAPID/Best_Practices/Public_Involvement_in_Renewable_Energy_and_Infrastructure Project_Development

(Pellizzone et al., 2015). Accurate information about proposed projects should be tailored to the community and disseminated using appropriate methods (e.g., use of leaflets instead of social media in regions where the internet is inaccessible; McNeish et al., 2019). Knowledge is a surmountable barrier to behavioral change provided that the public is confident in the information source and the information is relatable. For example, Liu et al. (2019) found that in many countries around the world, trust in policymakers or decision-makers has supported the successful development of renewable energy projects, including in China and the Netherlands. A similar study of renewable energy development in China found that the public was more open to project development when accurate and reliable data were presented, including maps and analyses (Guo et al., 2015). Encouraging learning through public participation is an effective tactic to changing public opposition of renewable energy (Parkins et al., 2018).

Gaps and Challenges

Challenges that have inhibited development of renewable energy-powered DG and microgrids in developing countries include difficulty accessing data, lack of tools for energy resource monitoring, and lack of accurate energy forecasts (Rinaldo et al., 2017). Lack of technical expertise can be a barrier to appropriately accessing and applying EO data products to renewable resource planning. To bridge this knowledge gap, there are a multitude of web-based viewers and mapping tools available to the public online. For example, the Global Solar Atlas provides yearly averaged solar resource and photovoltaic (PV) power potential data at a resolution of about 1 km in a user-friendly and intuitive color-coded map with optional data download[15]. Additionally, there remains an opportunity to provide capacity building efforts among electric utility providers or other key stakeholders to increase awareness of relevant tools and the benefits of utilizing EO data in this sector, and train technical staff in the proper access and application of EO data to meet specific needs they are facing.

Challenges such as resource monitoring and accuracy of energy forecasts may be mitigated through the application of EO data to assess renewable energy resources for siting and ongoing monitoring efforts. Although EO data have been used extensively for solar resource mapping, one pitfall is the spatial and temporal resolution for solar forecasts, which is much higher with ground-based images. Also, satellite instruments may have difficulty detecting cloud shadow, which can indicate sudden changes in solar irradiance, a possible issue for larger scale PV operations (Wan et al., 2015). EO-derived wind data are generally viewed as having good temporal and spatial coverage for wind resource planning, although care must be taken when selecting an appropriate dataset, as some wind data may be captured at a height suboptimal for comparison with wind turbines. For example, the NOAA High Density Infrared Cloud Drift Winds product[16] records winds at an approximate height between 3,000 and 13,700 m above ground, whereas today's wind turbines have a hub height that rarely exceeds 125 m above ground (Lantz et al., 2017). Research supports this concern and has suggested that "wind profiles from lidar and masts" are more appropriate for estimating wind resource potential than EO data in estimating hub height winds (Hasager et al., 2010).

In general, geothermal energy has higher installation costs than solar and wind installations. As a result, geothermal energy projects in developing countries typically rely on government incentives, which can be a roadblock for many countries. Freely available EO data can address some of the outstanding financial challenges associated with geothermal exploration and plant siting by providing a more accessible data source compared with ground-based exploration.

State of Research

Like other areas of research using EO data for energy solutions, emerging research largely involves the use of machine learning models to mechanize the analysis of satellite images. For example, Duke University Energy Data Analytics Lab researchers recently developed machine learning models that automate the identification of DG solar and wind resources using open access EO and other data[17]. Most other research in this field is primarily focused on the automation and validation of solar nowcasting and forecasting using machine learning methods based on satellite images to accurately predict fluctuating meteorological conditions and PV output (Eissa et al., 2013; Jang et al., 2016; Catalina et al., 2019; Cornejo-Bueno et al., 2019).

Real World Use of EO Data in Developing Countries

EO data are commonly used to assess water quantity and availability for hydroelectric project planning and monitoring, determine ideal dam locations and reservoir size, and assess environmental impacts from rerouting or damming water for hydroelectric projects. Electric utilities in India and Colombia have demonstrated the practical use of EO data for hydroelectric power projects, and India's Central Electricity Authority best practices for hydroelectric project development promote the use of satellite imagery for efficient interpretation of land use and geology for environmental impact of projects (Central Electricity Authority, 2002). For the Miel I hydroelectric power plant in Norcasia, Colombia, Isagen Productive Energy monitors and tracks landscape restoration efforts as part of the post-construction process. Satellite imagery and aerial photographs are used to assess changes in vegetation cover to understand the adaptation of hydrobiological communities to the reconstructed habitat. Isagen maintains a comprehensive environmental management plan for the dam and reservoir, and EO data are part of the restoration plan to address the effects of plant construction (Isagen Productive Energy, 2015).

EO-derived thermal infrared data, surface wavelength measurements, and gravity anomalies have been used for more than a decade in the identification of likely geothermal hotspots and exploration of geothermal energy potential of large land areas. To target ideal locations for geothermal drilling and

[15] https://globalsolaratlas.info/

[16] https://www.ospo.noaa.gov/Products/atmosphere/wind.html

[17] https://ssri.duke.edu/news/gauging-renewable-energy-generation-using-satellite-imagery

well siting, the Kenya Electricity Generating Company, also known as KenGen, uses a GIS system to map potential hotspots identified using a combination of EO data, seismic maps, surface geology maps, and other data inputs (Wekesa and Gichini, 2016). KenGen also operates hydroelectric and wind power plants as part of its electric generation portfolio. To facilitate effective management of its resources, KenGen recently proposed the development of an "Integrated Spatial Resource Management" GIS system that would incorporate EO data products and other data sources, such as aerial photography, for resource and asset management (KenGen, 2018). EO data provide a more cost-effective data source for renewable energy exploration and plant siting compared with ground-based exploration.

Grid Management

The success of grid management, i.e., grid infrastructure operation and reliability, is often dependent on environmental factors. It is estimated that about 90 percent of load forecasting errors are weather-related, and the matching of supply and demand necessitates accurate data, as inaccuracies can instigate brownouts that impact other sectors, such as manufacturing and health care (Global Science and Technology Inc, 2016). Unreliable electricity plagues many developing countries, and frequent and long outages negatively impact socio-economic growth. Freely available EO data provide an opportunity for decision makers in developing countries to improve their knowledgebase. EO-derived data on current weather conditions and historical climate trends can be used in energy forecasting models and for infrastructure planning to improve grid reliability. Outage magnitude can be assessed and monitored with nighttime lights EO data. Inadequate line maintenance is also a cause of outages. Overhead transmission lines require routine vegetation management to mitigate the potential for falling limbs. Many global EO-derived products that measure vegetation are freely accessible, such as Terra/Aqua MODIS Normalized Difference Vegetation Indices (NDVIs) that are available at variable temporal and spatial resolutions[18]. These and other EO data products offer electric utilities an opportunity to enhance grid operation and reliability.

Gaps and Challenges

One challenge to using EO data for grid management includes the need for technical expertise to access and use EO data products. To mitigate this challenge, NASA, NOAA, and other organizations have developed freely available land use, nighttime lights, vegetation index products, and other useful datasets that are available in simple web-based map viewers. One example is the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB) Nighttime Lights interactive map, which provides global, nightly imagery at a moderate spatial resolution[19]. However, it has been suggested that nighttime lights may be of limited value in areas where power is intermittent, depending on the temporal resolution of the EO data product (Dugoua et al., 2018).

Another potential barrier to adoption of EO data products by utilities is unfamiliarity that could lead to uncertainty or hesitation in using the data for decision making. Most electricity providers will need some level of training in the appropriate use and interpretation of EO data, even in its most simple form as a map viewer, which could potentially be another roadblock to use. Depending on the structure of the electricity sector in any given country or region, training may need to occur across administrative jurisdictions (National Research Council, 2003).

State of Research

The most notable current trend in grid management is using artificial intelligence (AI) to manage grid operations and improve grid resilience. EO data are being integrated into AI systems along with information on the electrical distribution system and utility operations to build a body of knowledge that, with the proper algorithms, allows the grid to operate autonomously or alerts operators to vulnerabilities. Such systems could identify fluctuations in power supply and automatically reconfigure loads and resources (Kosowatz, 2018).

Many recent feasibility studies have been focused on demonstrating the effectiveness of high-resolution EO data for monitoring transmission lines for vegetation that might disrupt distribution. Satellite images can be used in transmission line management, for example, to determine whether the average height of vegetation beneath a power line exceeds a specific threshold and should be trimmed (Häme et al., 2016). EO data could potentially replace the more costly methods of using observation helicopters and unmanned aerial vehicles for visual inspection of overhead lines. Furthermore, emerging research shows the applicability of machine learning algorithms using EO-derived NDVI data to effectively identify transmission lines at highest risk of being affected by vegetation (Klein et al., 2018).

Real World Use of EO Data in Developing Countries

The Central American Electrical Interconnection System, known as SIEPAC, is an interconnection of the power grids of six Central American nations: Panama, Costa Rica, Honduras, Nicaragua, El Salvador, and Guatemala. SIEPAC regularly employs EO data for environmental impact and geotechnical studies for electric infrastructure development. For example, SIEPAC used Landsat images to understand features of geologic importance for transmission line siting in Nicaragua. The satellite images were used to identify complex volcanic features, areas of lava flow, and permanently flooded areas of flat relief. These aspects were used in part to assess the geological risk of constructing a transmission line to expand access to reliable electricity in the region (SIEPAC, 2004).

Falling trees account for most overhead transmission line failures, and downed power lines can spark fires. Energisa, a Brazilian power company, uses EO data to track vegetation conditions to identify trees that need pruning near transmission lines. Satellite imagery is coupled with an algorithm and digital photogrammetry, packaged as an "intelligent tool," which can estimate the presence and height of trees near overhead lines,

[18]https://modis.gsfc.nasa.gov/data/dataprod/mod13.php

[19]https://maps.ngdc.noaa.gov/viewers/VIIRS_DNB_nighttime_imagery/index.html

enabling effective and efficient vegetation management with minimal to no field effort required (Energisa, 2018).

Disaster Risk Reduction and Response

Extreme weather events and other natural disasters that could impact the energy sector in developing countries may include hurricanes, flooding, volcanic eruptions, earthquakes, tsunamis, and landslides. The IPCC finds that weather- and climate-related disasters have resulted in increased economic loss and inequity exists among local adaptive capacity for disaster risk management, with developed countries often better equipped for disaster response (IPCC, 2012). The Sendai Framework for Disaster Risk Reduction targets to "substantially reduce disaster damage to critical infrastructure and disruption of basic services, among them health and educational facilities, including through developing their resilience by 2030[20]." EO data can be integrated into tools to improve forecasting, risk modeling, and recovery related to these extreme weather and natural disaster events and assist electric utility providers with resilience planning efforts to minimize impact to critical infrastructure and societal services.

Nighttime lights imagery is commonly used in the days immediately after a natural disaster to determine the extent of an electric outage. This information is crucial to identify neighborhoods in the greatest need of assistance. Precipitation, temperature, and other freely available remotely sensed weather data can lower costs and increase the accuracy of forecasting tools for natural disasters or extreme events. Similarly, high-resolution, near-real time daytime satellite imagery can be used to identify building damage, debris, and other obstacles that may impede rescue and recovery efforts after a disaster. Other research involves measurement of vegetation loss; monitoring of landslides; the application of machine learning and big data to improve tropical cyclone/wind forecasts and assess the likelihood of flooding as well as other types of disasters (Yu et al., 2018); and the use of differential interferometry from radar satellite data to monitor surface deformation in evaluating landslide and flooding hazards to electrical transmission lines (Li et al., 2018).

Gaps and Challenges

One challenge in using EO data to assess impacts to energy infrastructure and access after a natural disaster is that disaster response occurs across a wide array of sectors, with many moving parts, each uniquely important. In many countries, responding agencies may work in silos, which could result in a lack of information sharing or inefficient data gathering. These administrative difficulties could impede effective use of EO data across a multitude of sectors, including the power sector (Anderson et al., 2017). It may also be difficult to determine the best dataset or product resource in a time of crisis. Routine disaster monitoring may be useful in that, in the event of a disaster, systems would already be in place to inform decision making, communication methods could be established, and hazard mitigation plans would be developed and ready to implement at the first warning sign.

Other challenges include the availability of tools to handle large amounts of EO data and combine them with data from other sources, such as maps or measurement studies (Van Westen, 2000). EO data can also be also limited by an absence of global coverage and by dependence on weather and daylight conditions (Washaya et al., 2018).

State of Research

Recent research is largely focused on implementing EO data for disaster warning systems. For example, the Institute of Hydrology, Meteorology and Environmental Studies of the Republic of Colombia (IDEAM) embarked on a project to strengthen disaster risk management in Colombia. The organization demonstrated the value of incorporating seasonal precipitation forecasts derived from GOES16 satellite data into a warning system for climate-related disasters and disruptions to energy generation (IDEAM, 2018). Researchers are also focused on using EO data to assess post-disaster damage. Research spans the rapid assessment of building damage to inform emergency responders of areas most severely impacted (Ramlal et al., 2018) to the evaluation of vegetation loss to inform vegetation protection strategies (Long et al., 2016; Hu and Smith, 2018).

Other similar research ties the use of EO data to flood management, tsunami monitoring and forecasting, landslide and volcanic activity monitoring, and general disaster management. Concepts covered span such aspects as applications of EO data, accuracy of EO-derived data processing, integrated management platforms, and case studies (Alfieri et al., 2018; Novellino et al., 2018; Schwarz et al., 2018; Wang and Xie, 2018; Korup et al., 2019).

Emerging research involves applying machine learning to natural disaster assessment and susceptibility. Projects include NASA's Deep Learning-based Hurricane Intensity Estimator, which uses machine learning methodology to estimate wind speeds by monitoring EO data from NASA satellites in near-real time, decreasing the time required to update the forecasting model[21]. Other research is largely focused on flood analysis, such as applying machine learning informed by land use maps derived from satellite images to estimate flood probability for a region of interest (Mojaddadi et al., 2017) and using a machine learning approach to analyze satellite images to classify and predict whether roads in a flooded region are passable (Said et al., 2018), among other projects.

Real World Use of EO Data in Developing Countries

Brazil's Energisa uses satellite imagery to assess weather conditions that contribute to grid interruptions and infrastructure damage after weather events have occurred. In the case of an event where strong winds overturned five distribution towers near Canarana, Energisa used images from GOES 16 to analyze the temperature and cloud conditions, along with wind intensity data from other sources, on the day of the event. Energisa is using these data to predict future events and to assess the need to reinforce certain assets to withstand extreme weather events (Energisa, 2017).

[20] https://www.unisdr.org/we/coordinate/sendai-framework

[21] http://hurricane.dsig.net/

Empresa de Transmisión Eléctrica S.A. (ETESA), Panama's state transmission company, uses satellite imagery obtained from NOAA's geostationary satellite GOES-R in near-real time to "contribute to more accurate and reliable weather forecasts of severe weather perspectives." EO-derived climatic parameters, such as air and soil temperature, relative humidity, barometric pressure, wind speed and direction, and solar radiation have expanded automation of Panama's National Hydrometeorological Network and modernized measurement and data processing systems. With the acquisition of receiving stations to obtain meteorological data from GOES-R, ETESA is better prepared for severe storms and climate-related events that may affect electrical transmission and distribution (Jimenez, 2016).

DISCUSSION

A review of recent literature and relevant websites summarized current or ongoing research areas as well as identified key opportunities for electricity providers to incorporate EO data into rural electrification planning, renewable energy assessment, grid management, and disaster risk reduction and recovery efforts. As discussed in this article, EO data have the potential to be useful in several ways for improving governance of and access to reliable, affordable, resilient, and sustainable energy. More specifically, the use of "nighttime lights" imagery for estimations of rural electrification, EO-derived NDVIs for vegetation monitoring for overhead transmission line management, solar radiance data for renewable energy project planning, and weather event nowcasting for natural disaster monitoring have all demonstrated value added for the energy sector in developing countries. Broadly, these and other applications of EO data can improve access to electricity, renewable energy management, and disaster risk management. This review paper provides an important resource in understanding the broad applications and future possibilities for improved energy access and management in the developing world.

This article intends to provide a global overview and baseline of gaps, challenges, and opportunities for using EO data in the energy sector in developing countries, as well as providing a sample of specific examples of real-world application already occurring in developing nations. However, it is important to note that such gaps, challenges, and opportunities may vary between and within regions and countries, and that our focus was specifically on developing countries, while developed countries, particularly in Europe, North America, and parts of Asia, may have additional examples of real-world application of EO data for energy access, renewable energy management, and disaster management. Due to the nature of the literature review and the keywords utilized, no particular region or country was the subject of in-depth search or analysis. Therefore, this review provides a high-level overview in developing countries only and may not include all relevant current applications in the energy sector worldwide. Further research into country specific needs in the developing world is needed to uncover additional, nuanced barriers to uptake of EO data in the energy sector given particular local context.

Nonetheless, the results of this review outline several important challenges, gaps, and opportunities for the continued and expanded utilization of EO data for the energy sector in developing countries. For example, though myriad EO information is available for free online, internet access can be unreliable or unavailable altogether, especially in remote areas. Some models require automated access to data continuously so that they can run with the most recent data outputs. Without consistent access to the internet, obtaining and/or applying EO data, models, tools, or other resources will remain difficult. To mitigate this risk, EO data could be accessed using infrastructure in urban areas where accessibility may be higher. Data could also be offered via data CDs or external hard drives where access is limited. Or, electricity providers could partner with other organizations, stakeholders, or value-added providers to access information or EO data in alternative arrangements. However, open data sharing is uncommon in several regions of the world, potentially making these types of arrangements complicated. For example, in Kenya, a qualitative study found that although study participants recognized the importance of sharing data, additional trust was needed between researchers before open data sharing policies would be followed (Jao et al., 2015). A study in Vietnam showed that while many Western countries share consensus around policy, infrastructure, and best practices in data sharing, these practices are less often found in areas where institutions or universities face funding limitations (e.g., increasing a feeling of competition between potential collaborators) or a lack of technical, governance, or practical solutions for data sharing (Merson et al., 2015). Pursuing a culture of transparency and data sharing, alongside concrete policies, best practices, and other drivers, is important within the research and applied management or decision-making communities.

Even with the ability to access EO data online, analysts or other electricity provider staff may lack general awareness of relevant EO resources, not to mention how to appropriately apply such resources to their planning and decision-making efforts. NASA, NOAA, and other organizations have developed freely available land use, "nighttime lights," vegetation index products, and other useful datasets that are available in simple web-based map viewers or data interfaces that are meant to be easily accessible for end users. However, without capacity building efforts, through technical trainings, demonstrations or provision of case studies, understanding where and how to obtain EO resources–and how to appropriately apply them—will remain a barrier. This can lead to uncertainty or hesitation in using EO data for improved decision making and management. Understanding data or model uncertainty or other technical limitations is especially important, particularly with regard to climate information and projections, and training should accompany initial application of related datasets. Training on topics could be as preliminary as data management using statistical or geospatial software to advanced statistical or AI techniques.

If existing, such training is not always available in the language, time zone, or context needed for a developing country.

Researchers and data providers should ensure that trainings or capacity building is available and adapted for audiences around the world. Trainings should take place over extended periods of time so that time-constrained staff are able to accommodate the additional workload. Focusing on several staff increases the likelihood of technical capabilities remaining despite turnover. Documentation of procedures including analyses, modeling, or other activities should be detailed and maintained over time.

Limitations in staffing is a common but serious practical human capacity challenge. Many electricity providers (as well as government ministries working in environmental planning or energy) are short-staffed and face technical expertise gaps. If management is aware of the benefit of using EO for energy management, resources could be committed over time to increase analysis and planning utilizing EO resources. In addition, collaboration between the private sector and government or research bodies, who may have resources or technical expertise, could benefit electricity providers. While consultants are often able to provide these services, relying on third party expertise is expensive and unsustainable.

Instead, further expanding knowledge of EO data applications through case studies, training, and technical assistance specific to electricity providers can facilitate access to reliable electricity in developing countries and lead to increased use of renewable energy and microgrids, and natural disaster risk reduction. There are resources available for training, such as NASA's Applied Remote Sensing Training (ARSET) program[22] which provides online trainings (guided and self-instructed) on several EO applications. NOAA's Climate Resilience Toolkit (CRT)[23] includes case studies and other resources on EO applications across many sectors. Other organizations such as IRENA or the International Energy Agency (IEA) similarly offer resources for EO application in the energy management sector. While several key resources for energy management using EO exist (Eckman and Stackhouse, 2012), additional work is needed to determine the use of those resources in developing countries. Further, additional attention should be paid to if applications are specifically tailored and adapted for electric utilities or electricity providers in the developing world.

There are also several opportunities for the research community to address these barriers and contribute to the uptake of EO data for utilities in the developing world. Data providers could provide or compile case studies on how their datasets have been utilized by other users. Case studies of existing applications could highlight leaders who have demonstrated EO data applications. Data providers can seek to make their datasets available in intuitive, easily accessible formats, providing ample documentation, and guidance on possible applications. User communities have noted in previous studies that datasets should be tailored to user requirements in easily accessible formats that reflect the software used for analysis (Lautenbacher, 2006).

Communities of practice or other data end user networks offer opportunities for knowledge exchange and other mechanisms for data users to share information and learn from each other. South-South information exchange, collaboration, and communication also allow for sharing of common ideas, needs, challenges, and innovative solutions within and between developing countries. The United Nations (UN) has advocated for South-South exchange in tackling issues such as climate change, inequality, and energy access[24]. In fact, several UN partnerships exist based on several themes relevant to sustainable development; data applications for energy access and management should be added as a key topic of interest[25]. The UN Framework Convention of Climate Change (UNFCCC) has similarly called for South-South cooperation and "triangular cooperation" on technologies for adaptation to climate change on topics such as technology transfer and capacity building[26].

Nonetheless, through broad multi-directional dialogue exchange to understand end user needs and opportunities from the earth observation research community, refinement of existing tools and resources, or the development of new tools may be able to continue to progress efforts to improve electrification, renewable energy management and disaster management in the energy sector among developing nations. There may be the opportunity to leverage existing venues and networks that bring the EO research community together, such as the Group on Earth Observations (GEO) or internationally attended scientific meetings such as the American Geophysical Union (AGU) or American Meteorological Society (AMS). Participation in these organizations can allow for the development of innovative solutions and set new research strategies for future efforts. Government science agencies as well as other private data providers could have a role in facilitating those linkages. For example, the SERVIR program[27], jointly funded by NASA and the US Agency for International Development (USAID), offers a unique opportunity for funded research teams to work together with data users to apply EO datasets to development-related questions including governance of resources and communities. Similar efforts could encourage a focus on energy management, providing scientific expertise to practical applications brought forward by practitioners, decision-makers, or other stakeholders in developing countries. Increased attention, in the form of funding, research, technology transfer, and dialogue, should be paid to energy applications given the central role that energy access plays in so many SDGs as an enabler of sustainable development[28].

AUTHOR CONTRIBUTIONS

AL conducted the literature review and was the primary author of the manuscript. AT, NS, and TM assisted in writing. AT submitted the manuscript. All provided review and revision.

[22] https://arset.gsfc.nasa.gov/

[23] https://toolkit.climate.gov/

[24] https://www.un.org/press/en/2019/dev3388.doc.htm

[25] http://www.expo.unsouthsouth.org/resources/gssd-expo-solutions/

[26] http://unfccc.int/ttclear/misc_/StaticFiles/gnwoerk_static/tn_meetings/18e2ee898379443c85397bd1b3d210a4/ee99ac~7846a44132b3bd6c3dca058c5a.pdf

[27] https://www.nasa.gov/mission_pages/servir/index.html

[28] https://sustainabledevelopment.un.org/?page=view&nr=2749&type=13&menu=1634

FUNDING

This work was funded through NASA Grant No. 80NSSC18K0325.

ACKNOWLEDGMENTS

The authors acknowledge the support of funding from NASA in making this research possible, as well as the contributions of Dr. Richard Eckman and Dr. Paul Stackhouse. The authors also acknowledge the support of Dr. Craig Zamuda of the US Department of Energy.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2019.00123/full#supplementary-material

REFERENCES

Ahlborg, H., and Hammar, L. (2014). Drivers and barriers to rural electrification in Tanzania and Mozambique–Grid-extension, off-grid, and renewable energy technologies. *Renew. Energy* 61, 117–124. doi: 10.1016/j.renene.2012.09.057

Alfieri, L., Cohen, S., Galantowicz, J., Schumann, G. J.-P., Trigg, M. A., Zsoter, E., et al. (2018). A global network for operational flood risk reduction. *Environ. Sci. Policy* 84, 149–158. doi: 10.1016/j.envsci.2018.03.014

Altaweel, M. (2018). *Black Marble: Nighttime Lights Data from NASA, GIS Lounge.*

Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., and Friedl, L. (2017). Earth observation in service of the 2030 Agenda for Sustainable Development. *Geo Spatial Inform. Sci.* 2, 77–96. doi: 10.1080/10095020.2017.1333230

Areizaga, J., Sano, M., Medina, R., and Juanes, J. (2012). Improving public engagement in ICZM: a practical approach. *J. Environ. Manage.* 109, 123–135. doi: 10.1016/j.jenvman.2012.05.006

Asian Development Bank (2017) *Bangladesh: Bangladesh Power System Enhancement and Efficiency Improvement Project: Initial Environmental Examination*. Manila: ADB.

Bell, D., Gray, T., Haggett, C., and Swaffield, J. (2013). Re-visiting the 'Social Gap': public opinion and relations of power in the local politics of wind energy. *Env. Polit.* 22, 115–135 doi: 10.1080/09644016.2013.755793

BloombergNEF (2018). *Climatescope: Emerging Markets Outlook 2018. Energy Transition in the World's Fastest Growing Economies.* London: BloombergNEF.

Brown, M. E., Escobar, V. M., Aschbacher, J., Milagro-Pérez, M. P., Doorn, B., Macauley, M. K., et al. (2013). Policy for robust space-based earth science, technology and applications. *Space Policy* 29, 76–82. doi: 10.1016/j.spacepol.2012.11.007

Butler, C., Parkhill, K., and Pidgeon, N. (2011). "From the material to the imagined: public engagement with low carbon technologies in a nuclear community," in *Renewable Energy and the Public: From NIMBY to Participation*, ed P. Devine-Wright (London; Washington, DC: Earthscan), 301–316.

Cader, C., Radu, A., Bertheau, P., and Blechinger, P. (2018). "Remote sensing techniques for village identification: improved electrification planning for Zambia," in *Africa-EU Renewable Energy Research and Innovation Symposium 2018*, eds M. Mpholo, D. Steuerwald, and T. Kukeera (Cham: Springer), 91–96. doi: 10.1007/978-3-319-93438-9_7

Catalina, A., Torres-Barrán, A., Alaíz, C. M., and Dorronsoro, J. R. (2019). Machine learning nowcasting of PV energy using satellite data. *Neural Proc. Lett.* 1–19. doi: 10.1007/s11063-018-09969-1

Central Electricity Authority (2002). *Best Practices on Survey and Investigations of Hydro Electric Projects*. New Delhi: Central Electricity Authority.

Chauhan, A., and Saini, R. P. (2015). Renewable energy based off-grid rural electrification in Uttarakhand state of India: technology options, modelling method, barriers and recommendations. *Renew. Sustain. Energy Rev.* 51, 662–681. doi: 10.1016/j.rser.2015.06.043

Christidis, T., Lewis, G., and Bigelow, P. (2017). Understanding support and opposition to wind turbine development in Ontario, Canada and assessing possible steps for future development. *Renew. Energy* 112, 93–103. doi: 10.1016/j.renene.2017.05.005

Cook, N., Campbell, R. J., Brown, P., and Ratner, M. (2015). *Powering Africa: Challenges of and U.S. Aid for Electrification in Africa.* Washington, DC: Congressional Research Service.

Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., and Salcedo-Sanz, S. (2019). Machine learning regressors for solar radiation estimation from satellite data. *Solar Energy* 183, 768–775. doi: 10.1016/j.solener.2019.03.079

Devine-Wright, P. (2010). Public engagement with large-scale renewable energy technologies: breaking the cycle of NIMBYism. *Wiley Int. Rev. Clim. Change* 2, 19–26. doi: 10.1002/wcc.89

Dugoua, E., Kennedy, R., and Urpelainen, J. (2018). Satellite data for the social sciences: measuring rural electrification with night-time lights. *Int. J. Remote Sens.*, 39, 2690–2701. doi: 10.1080/01431161.2017.1420936

Eckman, R. S., and Stackhouse, P. W. Jr. (2012). CEOS contributions to informing energy management and policy decision making using space-based Earth observations. *Appl. Energy* 90, 206–210. doi: 10.1016/j.apenergy.2011.03.001

Eissa, Y., Marpu, P. R., Gherboudj, I., Ghedira, H., Ouarda, T. B. M. J., and Chiesa, M. (2013). Artificial neural network based model for retrieval of the direct normal, diffuse horizontal and global horizontal irradiances using SEVIRI images. *Solar Energy* 89, 1–16. doi: 10.1016/j.solener.2012.12.008

Energisa (2017). *Relatório de Interrupção em Situação de Emergência (ISE)*. Mato Grosso: Energisa.

Energisa (2018). *Satellite Imagery and Pruning Management: A Revolutionary Approach*. Tulsa: Distributech/Clarion Energy.

Engstrom, R., Newhouse, D., and Soundararajan, V. (2019). *Estimating Small Area Population Density Using Survey Data and Satellite Imagery: An Application to Sri Lanka.* Washington, DC: The World Bank. doi: 10.1596/1813-9450-8776

Global Science and Technology Inc (2016). *NCEI Climate Products and Services Market Analysis: Power Sector Engagement*. Oxford: Acclimatise.

Global Wind Energy Council (2017). *Global Wind Report: Annual Market Update 2017*. Brussels: GWEC.

Guo, Y., Ru, P., Su, J., and Anadon, L. D. (2015). Not in my backyard, but not far away from me: Local acceptance of wind power in China. *Energy* 82, 722–733. doi: 10.1016/j.energy.2015.01.082

Haggett, C. (2009a). "Chapter 2: Planning and persuasion: public engagement in renewable energy decision-making," in *Renewable Energy and the Public: from NIMBY to Participation,* ed P. Devine-Wright (London: Earthscan).

Haggett, C. (2009b). "Public engagement in planning for renewable energy," in *Planning for Climate Change: Strategies for Mitigation and Adaptation for Spatial Planners*, eds S. Davoudi, J. Crawford, and A. Mehmood (London: Taylor and Francis), 297–307.

Häme, L., Norppa, J., Salovaara, P., and Pylvänäinen, J. (2016). *Power line monitoring using optical satellite data. CIRED Workshop*, Helsinki, Paper 0383. Available online at: www.semanticscholar.org

Hanger, S., Komendantova, N., Schinke, B., Zejli, D., Ihlal, A., and Patt, A. (2016). Community acceptance of large-scale solar energy installations in developing countries: Evidence from Morocco. *Energy Res. Social Sci.* 14, 80–89. doi: 10.1016/j.erss.2016.01.010

Hasager, C. B., Badger, M., Mouche, A., Astrup, P., Stoffelen, A., and Karagali, I. (2010). "Offshore wind resource estimation using satellite images: what are the challenges?" in *Geophysical Research Abstracts,* eds H. Charlotte Bay, B. Merete, M. Alexis, A. Poul, S. Ad, and K. Ioanna (Munich: EGU), 2010–4650.

Hindmarsh, R., and Matthews, C. (2008). Deliberative speak at the turbine face: community engagement, wind farms, and renewable energy transitions, in Australia. *J. Environ. Policy Planning* 10, 217–232. doi: 10.1080/15239080802242662

Hu, T., and Smith, R. B. (2018). The impact of hurricane maria on the vegetation of dominica and puerto rico using multispectral remote sensing. *Remote Sens.* 10, 827–847. doi: 10.3390/rs10060827

IDEAM (2018). *Strengthening Hydrometeorological Monitoring. Early Warning Systems and Climate Services in Colombia: The IDEAM, a Successful Case in Latin America*. Bogotá: IDEAM.

International Energy Agency (2014). *Africa Energy Outlook: A Focus on Energy Prospects in Sub-Saharan Africa*. Paris: OECD/IEA. doi: 10.1787/weo-2014-en

International Energy Agency (2017a). *Energy Access Outlook 2017: From Poverty to Prosperity*. Paris: OECD/IEA.

International Energy Agency (2017b). *Southeast Asia Energy Outlook 2017*. Paris: OECD/IEA.

International Hydropower Association (2018). *Hydropower Status Report: Sector Trends and Insights*. London: IHA.

IPCC (2012), "Summary for policymakers," in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, eds C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, et al. (Cambridge; New York, NY: A Special Report of Working Groups, I., and II of the Intergovernmental Panel on Climate Change; Cambridge University Press), 1–19.

Isagen Productive Energy (2015). *Environmental Management Plan: Miel I Hydroelectric Power Plant*. Medellin: Isagen.

Jang, H. S., Bae, K. Y., Park, H., and Sung, D. K. (2016). Solar power prediction based on satellite images and support vector machine. *IEEE Trans. Sustain. Energy*, 7, 1255–1263. doi: 10.1109/TSTE.2016.2535466

Jao, I., Kombe, F., Mqalukore, S., Bull, S., Parker, M., Kamuya, D., et al. (2015). Research Stakeholders' views on benefits and challenges for public health research data sharing in kenya: the importance of trust and social relations. *PLoS ONE* 10:e0135545. doi: 10.1371/journal.pone.0135545

Jimenez, J. (2016). *Informe de Gestión*. Panama City: ESTESA.

KenGen (2018). *Request for Proposals (RFP) for Development of an Integrated Spatial Resources Management for KenGen*. Nariobi: KenGen.

Klein, L., Wamburu, J. M., and Hamann, H. (2018). *A Deep Learning Framework for Vegetation Management for Electrical Utilities Using Multispectral High-Resolution Satellite Imagery*. Washington, DC: American Geophysical Union, EP51E−1873.

Korup, O., Seidemann, J., and Mohr, C. H. (2019). Increased landslide activity on forested hillslopes following two recent volcanic eruptions in Chile. *Nat. Geosci.* 12, 284–289. doi: 10.1038/s41561-019-0315-9

Kosmopoulos, P. G., Kazadzis, S., El-Askary, H., Taylor, M., Gkikas, A., Prestakis, E., et al. (2018). Earth-observation-based estimation and forecasting of particulate matter impact on solar energy in Egypt. *Remote Sens.* 12, 1870–1893. doi: 10.3390/rs10121870

Kosowatz, J. (2018). *Using AI to Manage the Grid*. New York, NY: ASME.

Lahoz, W. A., and Schneider, P. (2014). Data assimilation: making sense of Earth Observation. *Front. Environ. Sci.* 28, 1–28. doi: 10.3389/fenvs.2014.00016

Langer, K., Decker, T., and Menrad, K. (2017). Public participation in wind energy projects located in Germany: which form of participation is the key to acceptance? *Renew. Energy* 112, 63–73. doi: 10.1016/j.renene.2017.05.021

Lantz, E., Roberts, O., and Dykes, K. (2017). *Trends, Opportunities, and Challenges for Tall Wind Turbine and Tower Technologies, Presented at the American Wind Energy Association Wind Power 2017 Conference, Anaheim, California*. Golden, CO: NREL.

Larson, E. C., and Krannich, R. S. (2016). "A Great Idea, Just Not Near Me!" understanding public attitudes about renewable energy facilities. *Soc. Nat. Resources* 29, 1436–1451. doi: 10.1080/08941920.2016.1150536

Lautenbacher, C. C. (2006). The global earth observation system of systems: Science serving society. *Space Policy* 22, 8–11. doi: 10.1016/j.spacepol.2005.12.004

Li, G., Tan, Q., Xie, C., et al. (2018). The transmission channel tower identification and landslide disaster monitoring based on INSAR, Intl. *Arch. Photogrammetry Remote Sens. Spat. Info Sci.* 42, 807–813. doi: 10.5194/isprs-archives-XLII-3-807-2018

Liu, L., Bouman, T., Perlaviciute, G., and Steg, L. (2019). Effects of trust and public participation on acceptability of renewable energy projects in the Netherlands and China. *Energy Res. Soc. Sci.* 53, 137–144. doi: 10.1016/j.erss.2019.03.006

Long, J., Giri, C., Primavera, J., and Trivedi, M. (2016). Damage and recovery assessment of the Philippines' mangroves following Super Typhoon Haiyan. *Mar. Pollut. Bull.* 109, 734–743. doi: 10.1016/j.marpolbul.2016.06.080

Longwei, L., and Dengsheng, L. (2016). Mapping population density distribution at multiple scales in Zhejiang Province using Landsat Thematic Mapper and census data. *Int. J. Remote Sens.* 37, 4243–4260. doi: 10.1080/01431161.2016.1212422

Maclean, I. M. D., Inger, R., Benson, D., Booth, C. G., Embling, C. B., Grecian, W. J., et al. (2014). Resolving issues with environmental impact assessment of marine renewable energy installations. *Front. Mar. Sci.* 75, 1–5. doi: 10.3389/fmars.2014.00075

Mann, M. L., Melass, E. K., and Malik, A. (2016). Using VIIRS Day/Night band to measure electricity supply reliability: preliminary results from Maharashtra, India. *Remote Sens.* 8, 711–723. doi: 10.3390/rs8090711

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., and Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories. *J. Informetr.* 12, 1160–1177. doi: 10.1016/j.joi.2018.09.002

McNeish, R., Rigg, K. K., Tran, Q., and Hodges, S. (2019). Community-based behavioral health interventions: developing strong community partnerships. *Eval. Program Plann.* 73, 111–115. doi: 10.1016/j.evalprogplan.2018.12.005

Merson, L., Phong, T. V., Nhan, L. N. T., Dung, N. T., Ngan, T. T. D., Kinh, V. N., et al. (2015). Trust, respect, and reciprocity: informing culturally appropriate data-sharing practice in Vietnam. *J. Empirical Res. Hum. Res. Ethics* 10, 251–263. doi: 10.1177/1556264615592387

Mojaddadi, H., Pradhan, B., Nampak, H., Ahmad, N., and bin Ghazali, A. H. (2017). Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS. *Geomat. Nat. Hazards Risk* 8, 1080–1102. doi: 10.1080/19475705.2017.1294113

Natarajan, L., Rydin, Y., Lock, S. J., and Lee, M. (2018). Navigating the participatory processes of renewable energy infrastructure regulation: a 'local participant perspective' on the NSIPs regime in England and Wales. *Energy Policy* 114, 201–210. doi: 10.1016/j.enpol.2017.12.006

National Academies of Sciences, Engineering and Medicine (2015). *Continuity of NASA Earth Observations from Space: A Value Framework*. Washington, DC: The National Academies Press.

National Academies of Sciences, Engineering and Medicine. (2018). *Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space*. Washington, DC: The National Academies Press.

National Research Council (2003). *Using Remote Sensing in State and Local Government: Information for Management and Decision Making*. Washington, DC: NAP.

National Research Council (2007). *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. Washington, DC: NAP.

Novellino, A., Jordan, C., Ager, G., Bateson, L., Fleming, C., and Confuorto, P., (2018). "Remote sensing for natural or man-made disasters and environmental changes" in *Geological Disaster Monitoring Based on Sensor Networks*, eds. T. Durrani, W. Wang, and S. Forbes (Singapore: Springer), 23–31. doi: 10.1007/978-981-13-0992-2_3

Oshri, B., Hu, A., Adelson, P., Chen, X., Dupas, P., Weinstein, J., et al. (2018). "Infrastructure Quality Assessment in Africa using Satellite Imagery and Deep Learning" in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: ACM), 616–625. doi: 10.1145/3219819.3219924

Parkins, J. R., Rollins, C., Anders, S., and Comeau, L. (2018). Predicting intention to adopt solar technology in Canada: The role of knowledge, public engagement, and visibility. *Energy Policy* 114, 114–122. doi: 10.1016/j.enpol.2017.11.050

Pellizzone, A., Allansdottir, A., De Franco, R., Muttoni, G., and Manzella, A. (2015). Exploring public engagement with geothermal energy in southern Italy: a case study. *Energy Policy* 85, 1–11. doi: 10.1016/j.enpol.2015.05.002

Quansah, D. A., Adaramola, M. S., and Mensah, L. D. (2016). Solar Photovoltaics in Sub-Saharan Africa – Addressing Barriers, Unlocking Potential. *Energy Procedia* 106, 97–110. doi: 10.1016/j.egypro.2016.12.108

Ramlal, B., Davis, D., and De Bellot, K. (2018). A rapid post-hurricane building damage assessment methodology using satellite imagery. *West Indian J. Eng.* 41, 74–83.

Rinaldo, R., Coppola, D., Walawalkar, R., and Thacker, H. (2017). *Integrated Applications for Microgrids in Developing Economies: Webinar*. Paris: ESA.

Robinson, C., Hohman, F., and Dilkina, B. (2017). "A deep learning approach for population estimation from satellite imagery" in *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities* (New York, NY: ACM), 47–54. doi: 10.1145/3149858.3149863

Román, M., Wang, R., Shrestha, R., Yao, T., and Kalb, V. (2019). *Black Marble User Guide Version 1.0*. Washington, DC: NASA.

Said, N., Pogorelov, K., Ahmad, K., Riegler, M., Ahmad, N., Ostroukhova, O., et al. (2018). "Deep learning approaches for flood classification and flood aftermath detection" *in MediaEval 18* (Sophia Antipolis).

Schöpfel, J. (2010). "Towards a Prague Definition of Grey Literature" in *Twelfth International Conference on Grey Literature: Transparency in Grey Literature* (Prague: Národní technická knihovna).

Schwarz, B., Pestre, G., Tellman, B., Sullivan, J., Kuhn, C., Mahtta, R., et al. (2018). "Mapping floods and assessing flood vulnerability for disaster decision-making: a case study remote sensing application in senegal" in *Advanced Remote Sensing Technology for Tsunami Modelling and Forecasting*, eds P. P. Mathieu and C. Aubrecht (Boca Raton: CRC Press), 293–300. doi: 10.1007/978-3-319-65633-5_16

SIEPAC (2004). *Geotechnical Study and Classification of Soils in the Line of Transmission*. SIEPAC Final Review, Revision 01: Tomo V – Nicaragua. San José: SIEPAC.

Stevens, F. R., Gaughan, A. E., Linard, C., and Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS ONE* 10:e0107042. doi: 10.1371/journal.pone.0107042

Stiles, G., and Murove, C. (2018). *SADC Renewable Energy and Energy Efficiency Status Report*. Paris: REN21.

Urmee, T., Harries, D., and Schlapfer, A. (2009). Issues related to rural electrification using renewable energy in developing countries of Asia and Pacific. *Renew. Energy* 34, 354–357. doi: 10.1016/j.renene.2008.05.004

Van Nguyen, M., Arason, S., Gissurarson, M., and Pálsson, P. G. (2015). *Uses of Geothermal Energy in Food and Agriculture: Opportunities for Developing Countries*. Rome: FAO.

Van Westen, C. J. (2000). Remote sensing for natural disaster management. *Int. Arch. Photogrammet. Remote Sens.* 33, 1609–1617.

Varshney, K. R., Chen, G. H., Abelson, B., Nowocin, K., Sakhrani, V., Xu, L., et al. (2015). Targeting villages for rural development using satellite image analysis. *Big Data* 3, 41–53. doi: 10.1089/big.2014.0061

Walker, G., Cass, N., Burningham, K., and Barnett, J. (2010). Renewable energy and sociotechnical change: imagined subjectivities of 'the public' and their implications. *Environ. Plann. A* 42, 931–947. doi: 10.1068/a41400

Wan, C., Zhao, J., Song, Y., Xu, Z., Lin,., J., et al. (2015). Photovoltaic and solar power forecasting for smart grid energy management. *CSEE J. Power Energy Syst.* 1, 38–46. doi: 10.17775/CSEEJPES.2015.00046

Wang, X., and Xie, H. (2018). A review on applications of remote sensing and Geographic Information Systems (GIS) in water resources and flood risk management. *Water* 10, 608–619. doi: 10.3390/w10050608

Washaya, P., Balz, T., and Mohamadi, B. (2018). Coherence change-detection with sentinel-1 for natural and anthropogenic disaster monitoring in urban areas. *Remote Sens.* 10:1026. doi: 10.3390/rs10071026

Wekesa, F., and Gichini, B. (2016). *The Use of GIS in Geothermal Resource Management—A Case Study of Olkaria Geothermal Project*. Nairobi: KenGen.

World Energy Council (2016). *World Energy Resources 2016*. London: WEC.

Young, O. R., and Onoda, M. (2017). "Chapter 1.4 Taxonomy of Roles," in *Satellite Earth Observations and Their Impact on Society and Policy*, eds M. Onoda and O. Young (Singapore: Springer), 13–17.

Yu, M., Yang, C., and Li, Y. (2018). Big data in natural disaster management: a review. *Geosciences* 8:165. doi: 10.3390/geosciences8050165

Conflict of Interest Statement: All authors are employed by Battelle Memorial Institute, a non-profit research institute. All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 03 September 2019
doi: 10.3389/fenvs.2019.00127

The Value of Near Real-Time Earth Observations for Improved Flood Disaster Response

Perry C. Oddo [1,2] and John D. Bolten [2]*

[1] Universities Space Research Association, Columbia, MD, United States, [2] Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, United States

Information is a critical resource in disaster response scenarios. Data regarding the geographic extent, severity, and socioeconomic impacts of a disaster event can help guide emergency responders and relief operations, particularly when delivered within hours of data acquisition. Information from remote observations provides a valuable tool for assessing conditions "on the ground" more quickly and efficiently. Here, we evaluate the social value of a near real-time flood impact system using a disaster response case study, and quantify the Value of Information (VOI) of satellite-based observations for rapid response using a hypothetical flooding disaster in Bangkok, Thailand. MODIS imagery from NASA's Land, Atmosphere Near real-time Capability for EOS (LANCE) system is used to produce operational estimates of inundation depths and economic damages. These rapid Earth observations are coupled with a decision-analytical model to inform decisions on emergency vehicle routing. Emergency response times from vehicles routed using flood damage data are compared with baseline routes without the benefit of advance information on road conditions. Our results illustrate how the application of near real-time Earth observations can improve the response time and reduce potential encounters with flood hazards when compared with baseline routing strategies. Results indicate a potential significant economic benefit (i.e., millions of dollars) from applying near real-time Earth observations for improved flood disaster response and management.

Keywords: value of Information, near real-time, emergency response, applied Earth observations, socioeconomic

OPEN ACCESS

Edited by:
Daniel Eric Irwin, Marshall Space Flight Center (NASA), United States

Reviewed by:
Guy Jean-Pierre Schumann, University of Bristol, United Kingdom
Daniel Lapidus, RTI International, United States

****Correspondence:***
Perry C. Oddo
perry.oddo@nasa.gov

Specialty section:
This article was submitted to Interdisciplinary Climate Studies, a section of the journal Frontiers in Environmental Science

Received: *01 June 2019*
Accepted: *16 August 2019*
Published: *03 September 2019*

Citation:
Oddo PC and Bolten JD (2019) The Value of Near Real-Time Earth Observations for Improved Flood Disaster Response. Front. Environ. Sci. 7:127. doi: 10.3389/fenvs.2019.00127

INTRODUCTION

Natural disasters like floods can have devastating societal impacts. Direct damages from flooding, such as the loss of human life or the destruction of infrastructure have immediate social ramifications, while indirect impacts like reduced business production or loss of income can lead to more protracted socioeconomic effects (Haraguchi and Lall, 2015). This is especially true in regions like the Lower Mekong River Basin (LMRB), where its combination of high population density, seasonal monsoons, and low-lying topography make it particularly susceptible to flooding (Gale and Saunders, 2013) (**Figure 1a**).

Managing flood hazards in such a flood-prone region requires an acute understanding of the risk of future events. Flood risks are defined by the function of the probability of occurrence, the exposure (e.g., population and assets subject to flooding), and vulnerability, which is a measure of the society's ability to cope with an event (Koks et al., 2015). Emergency management operations

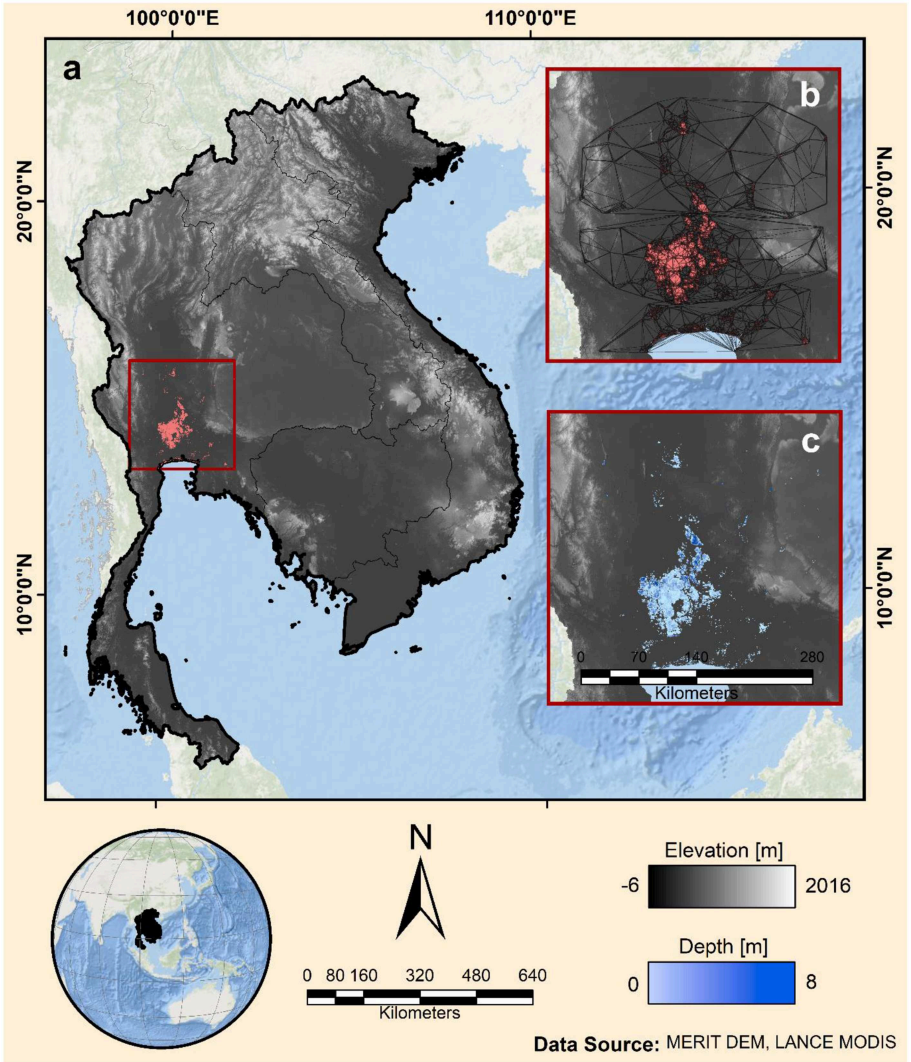


FIGURE 1 | (a) Map of Lower Mekong River Basin countries with flood extent from 2011 event (pink shaded region). **(b)** Study extent showing results of the triangular interpolated network (TIN) produced by extracting land elevations from around the perimeter of the flood extent. **(c)** Resulting inundation depth raster produced by flood impact analysis.

are typically divided into four phases: Preparedness, Response, Recovery, and Mitigation (Altay and Green, 2006). Preparedness refers to the capacity-building period before a disaster strikes. After the onset of an event, the Response phase focuses on saving lives and averting additional damages. Post-event activities include Recovery, in which communities seek to return to pre-flood capabilities, and Mitigation, which refers to resiliency initiatives meant to strengthen the community from future events (Howden, 2009).

Each stage in the disaster management cycle presents its own set of challenges, with the highest uncertainties occurring in the Response phase immediately after a disaster strikes (Okuyama, 2003; Ortuño et al., 2013). A single flood event can impact multiple interdependent systems, such as telecommunications, transportation, and power infrastructure, thereby hindering coordination between decision makers, emergency responders, and populations in need (Comfort et al., 2004).

In such situations, information about the extent, severity, and impacts of flooding becomes a time-critical resource. The 2005 World Disasters Report notes that in disaster response scenarios, information can be considered as vital a form of aid as "…water, food, medicine, or shelter," with respect to its ability to save lives and extend resources (Walter, 2005). Due to the high uncertainties and dynamic nature of the disaster response phase, information is most valuable when obtained as soon after the flood event as possible. In a survey of 52 Emergency Management Agencies, 82% expressed a need for flood impact information within 24 h of the event (Hodgson et al., 2009).

Geospatial technologies like remote sensing, aerial footage, or volunteered geographic information (VGI) provide a valuable

way of obtaining useful intelligence at broader scales and with shorter latency than traditional methods (Hodgson et al., 2009; Goodchild and Glennon, 2010; Haworth and Bruce, 2015). These Earth-observing technologies have been widely incorporated into the emergency management cycle, from providing early warning systems for flood preparation (Koriche and Rientjes, 2016) to improving hazard mapping to inform long-term planning and mitigation strategies (Shivaprasad Sharma et al., 2017). Yet despite the obvious benefits of Earth observations, comparatively little research has been done to quantify the value that these data provide, particularly in the Response phase (Hodgson et al., 2009).

To this end, there remain several questions regarding the extent to which satellite-based information can aid in disaster management and planning, rapid response cases, and in improving the allocation of resources. To investigate the utility of operational Earth observations during a flood, we present a simple decision-analytical model based on a hypothetical flood response scenario. We use a Value of Information (VOI) framework to identify management objectives that could benefit from applied satellite data, and determine the potential time and cost savings achieved when incorporating geospatial information into emergency vehicle routing. In this context, routing is defined as the turn-by-turn navigation from a predefined dispatch center to a location of potential need. We accomplish this by coupling flood inundation estimates from an operational near real-time (NRT) satellite-based flood monitoring system with an open-source routing platform to evaluate how management objectives perform both in the presence and absence of advance flood information. In doing so we attempt to address the following main questions:

(1) How does the presence of NRT flood impact information affect the response times of emergency vehicles when compared to baseline routes?
(2) What model parameters are most important in determining vehicle response times?
(3) What is the potential social value of rapid earth observations when applied to a disaster response scenario?

Thus, the objective of this study is not to assess the performance of the flood monitoring system against other flood monitoring methods or products. Rather, we aim to get a better understanding of the value of these or similar satellite-based NRT observations and quantify to what degree they can potentially improve and support disaster risk management and response decisions. The following sections introduce the concept of VOI and review its application to geospatial data. We then describe the model design using the 2011 Southeast Asia Floods as an illustrative example. Finally, we present the results and discuss the broader impacts from the analysis.

BACKGROUND AND PREVIOUS WORK

Quantifying the Value of Information

In a decision making context it is important to understand how the introduction of new information can improve a given strategy. VOI methodologies were originally presented in economics literature as a way to quantify the marginal benefits produced by reducing model uncertainties (Howard, 1966, 1968). In economic terms, the VOI describes the amount of money a rational agent would be willing to pay for new information before making a decision (Alfonso et al., 2016). While these concepts have been widely implemented in the field of decision analysis, they have become increasingly common in Earth science applications. Some recent examples include investigations into learning about potential climate thresholds (Keller et al., 2007), petroleum engineering (Bratvold et al., 2009), or drought monitoring (Bernknopf et al., 2017). Here, we extend the VOI concept to a flood-related disaster management scenario.

Since the creation of the first weather satellites, global Earth-observing technologies have raised new questions about the benefits of geospatial data (Mjelde et al., 1989; Obersteiner et al., 2017). Macauley (2006) provides a broad overview of the concept of VOI in decision making using a simple crop harvesting scenario (**Figure 2**). In this example, a farmer has a choice whether to harvest the crop immediately or to harvest over the course 2 days. Harvesting over 2 days would net a higher payout, yet there is a chance that if it rains, part of the harvest could be ruined. Improved information about the probability of rain would be valuable to the farmer, and the information is of the most value when the farmer's subjective uncertainty is highest. In the diagram describing this decision

Nature: Decision	Heavy Rain Tomorrow	No Rain Tomorrow
Harvest all today	$40,000	$40,000
Harvest over two days	$22,500	$45,000

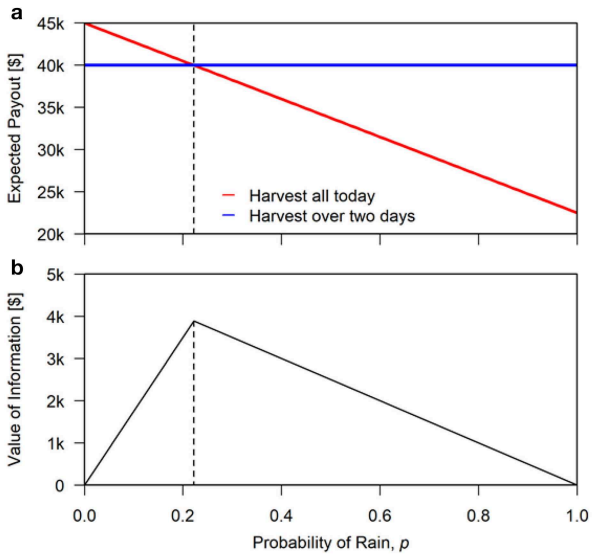


FIGURE 2 | Illustration of Value of Information based on the decision whether or not to harvest crops. Table shows the expected payout matrix for the decision of when to harvest. **(a)** Expected payout based on the subjective probability of heavy rain in the future. **(b)** Expected Value of Information for improved rain forecast. The dashed vertical line represents the point at which the farmer's uncertainty is highest, representing the highest value of information. Figure adapted from Macauley (2006).

problem, the vertical dashed line represents this threshold where the farmer's subjective belief about the probability of rain is the most uncertain (in this case, $p = 5/22.5$, or ∼22%). At this point, the most the farmer would be willing to pay is $3,888; above that, the expected costs of improved weather information outweigh its benefits. This simple approach can be applied in numerous ways to help quantify and convey the VOI for informing decisions. In this analysis, access to satellite-based maps of inundated regions serve as a proxy for information on the current state of the world.

The Value of Applied Earth Observations in Emergency Management

In the field of Emergency Response operations, improvements to satellite spatial resolution and latency have led to a wide adoption of geospatial informatics. Some data, such as those provided through NASA's Land, Atmosphere Near real-time Capability for EOS (LANCE) system, can commonly be made available within 3 h of overpass, making it suitable for NRT applications (Davies et al., 2015). Examples of such initiatives include NASA's NRT Flood Mapping (Ahamed et al., 2017; Fayne et al., 2017; Policelli et al., 2017), UMD's Global Flood Monitoring System (GFMS, Wu et al., 2014), and the Dartmouth Flood Observatory (DFO, Brakenridge and Anderson, 2006).

In the LMRB, several recent studies have demonstrated how rapid Earth observations can be used operationally to inform flood management. Ahamed and Bolten (2017) produced automatic flood extents by applying a dynamic surface water classifier to data from the Moderate-resolution Imaging Spectroradiometer (MODIS) sensors on the Aqua and Terra satellites. Oddo et al. (2018) then demonstrated how the resulting flood extents can be combined with socioeconomic data to produce rapid estimates of flood impacts using depth-damage curves for different types of land cover and infrastructure. This study attempts to determine how the flood detection and impact assessment metrics produced by the preceding analyses can be further applied to emergency response (**Figure 3**).

METHODS

To investigate how applied Earth observations can be used to improve flood response operations, we consider a hypothetical emergency response scenario based on the 2011 Southeast Asia Floods. In this scenario, we use the output of the previously described flood damage assessment to identify a number of population sites that may have been impacted and require emergency attention. Emergency response vehicles stationed at dispatch centers are routed to each of the population sites and their response times and navigation details are evaluated in the presence and absence of information on flood conditions. The following sections outline these steps in more detail.

Flood Impact Assessment: 2011 Southeast Asia Floods

The 2011 Southeast Asia Floods resulted in the highest-ever insured losses of any freshwater flood disaster (Gale and Saunders, 2013). In parts of Thailand, rainfall increase of up to 143% combined with land subsidence to produce widespread flooding in the region around Bangkok (Haraguchi and Lall, 2015). A surface flood extent raster for this 2011 flood was obtained through the DFO, which maintains an archive of historical events. Imagery collected by NASA's MODIS sensor shows the extent of surface inundation between December 24, 2011 and January 2, 2012 at a spatial resolution of 250-m (Brakenridge et al., 2011) (**Figure 1a**). The flood extent was vectorized using QGIS software and a triangular interpolated network (TIN) was generated by sampling land surface elevations around the perimeter (**Figure 1b**). The resulting TIN serves as an estimate of flood surface elevation across the detected extent. Subtracting the underlying digital elevation model ("Multi-Error-Removed Improved-Terrain"—MERIT DEM) produces a raster of estimated flood depths (**Figure 1c**) (Yamazaki et al., 2017). For a more detailed discussion of the methodology and limitations of this approach, see sections 3 and 5 of Oddo et al. (2018).

Estimated flood depths were intersected with a land use/land cover map produced by NASA SERVIR's Regional Land Cover Monitoring System (https://rlcms-servir.adpc.net/en/landcover/). Land cover classifications were derived from atmospherically-corrected imagery from Landsat 4, 5, 7, and 8 to produce a map of the entire LMRB at ∼30-m ground resolution. Damages to specific land cover types were assessed using regionally-derived depth-damage functions (Oddo et al., 2018). Additionally, damages to populations and infrastructure were estimated by intersecting flood depths with population data from NASA's Socioeconomic Data and Applications Center (SEDAC) and open-source infrastructure data from OpenStreetMap (OSM), respectively (CIESIN, 2016; OpenStreetMap Contributors, 2019).

The resulting socioeconomic damage map (**Supplementary Figure 1**) was used to identify a total of 75 potential population sites that may have been most highly impacted by the flooding. Emergency response dispatch sites were chosen as ambulance and fire station locations in OSM ($n = 10$). Finally, estimated flood depths were used to delineate areas that were considered Highly Flooded to average vehicles. The threshold for Highly Flooded areas was identified as 300 mm of inundation, according to the modeled relationship between flood depth and vehicle speed in Pregnolato et al. (2017) (**Figure 4**). This was found to be the average depth at which a passenger vehicle would begin to float, signifying areas that would cause the most significant delays to emergency vehicles. Regions within the flood extent that exceed this threshold were exported as a GeoJSON object using QGIS for use in the routing model. Those areas that impacted yet were below the 300-mm threshold were identified as simply Flooded.

Routing Model

The coordinate locations of the dispatch and populations sites were used as endpoints for the vehicle routing model, which was built using the open-source OpenRouteService (ORS) navigation service. ORS provides free location based-services generated from user-defined geographic data from OSM. In

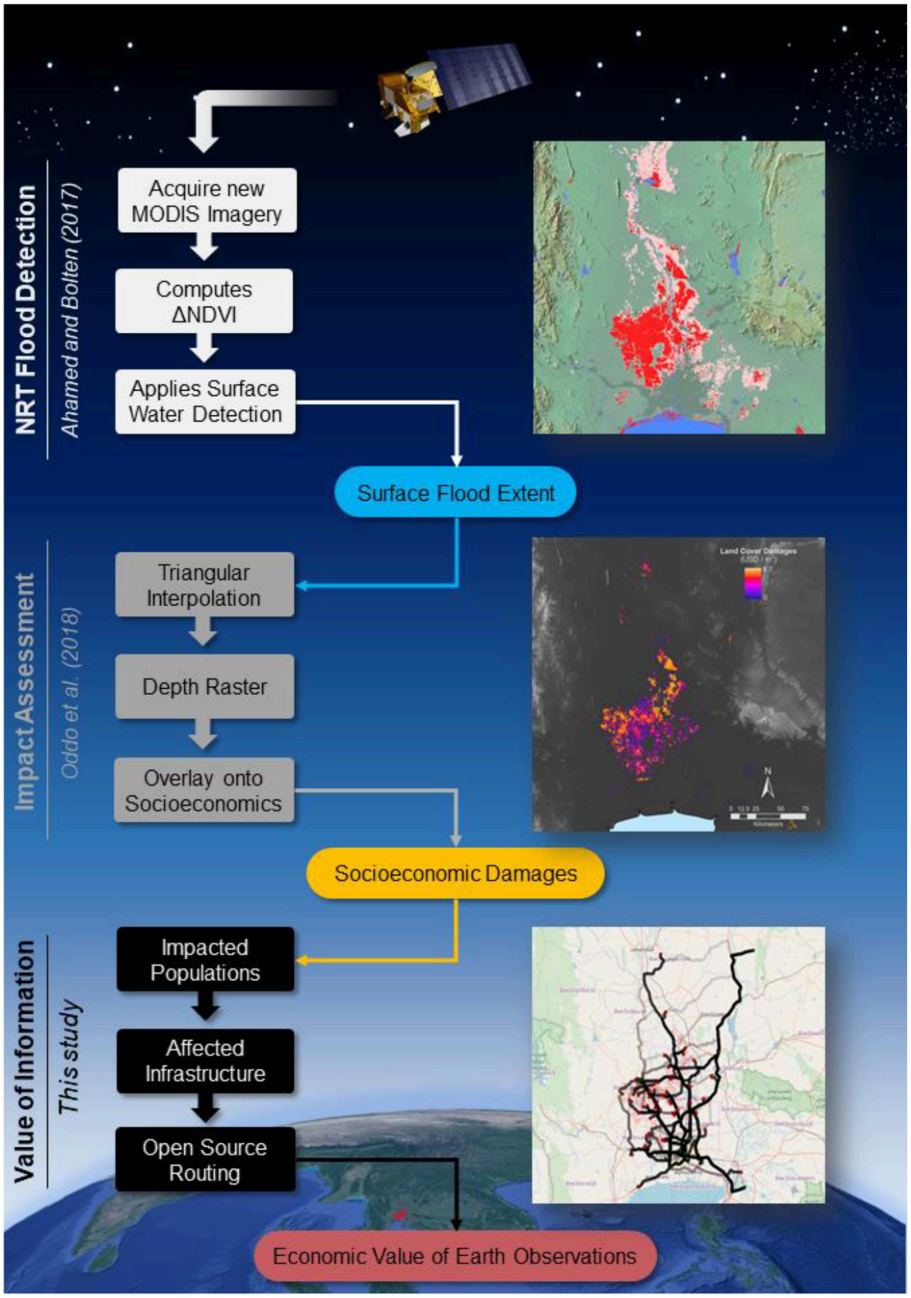


FIGURE 3 | Workflow for generating near real-time (NRT) routing information from Earth observations. MODIS imagery is ingested from LANCE server to produce surface water extents. Extents are digitized and used to estimate flood depths and damage estimates. Finally, these damages can be incorporated into the value of information analysis for emergency response.

addition to providing travel routes and navigation information via a graphical front-end, ORS also provides an application program interface (API) for the directions service with a variety of customizable parameters. Among these parameters is an "avoid_polygon" option, which allows users to identify the coordinates of polygon vertices to avoid when calculating the optimal routes (**Figure 5**). Route information is returned as a GeoJSON object, which includes information on distance (meters), travel time (seconds), and velocity (meters/second) for each segment along the route. Average velocities across an entire route were calculated for use in the decision analytical model by averaging the travel time and distance across each individual segment (see section Decision Analytical Model).

Vehicle routes were generated under both baseline conditions, and in the presence of flood information. Baseline routing simply used the coordinate locations from each of the 10 dispatch

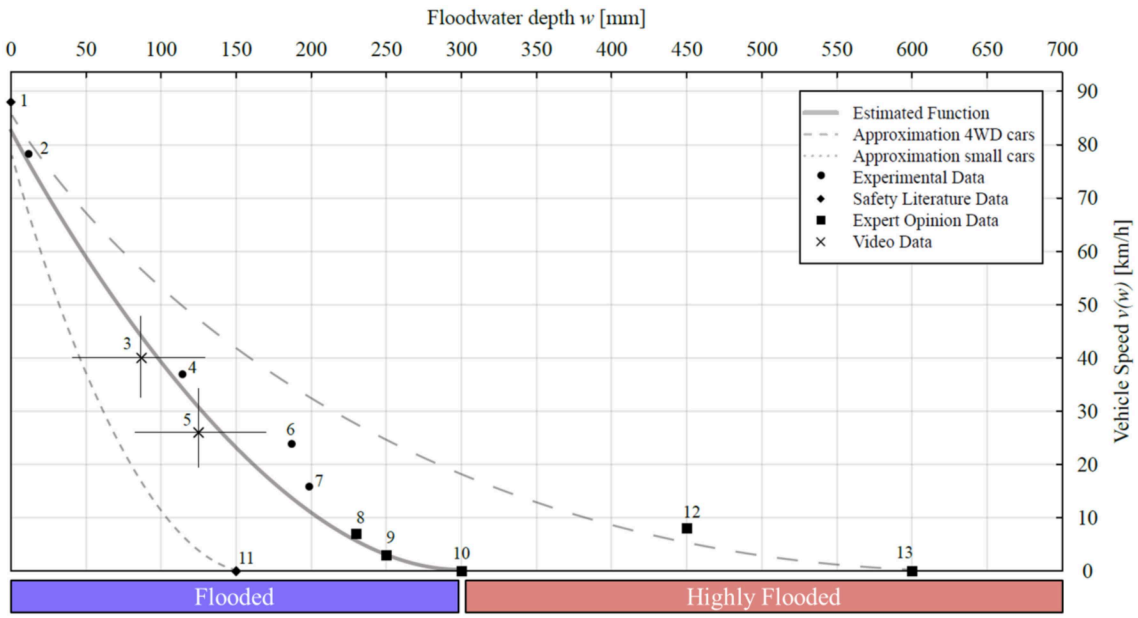


FIGURE 4 | Modeled relationship between flood depth and vehicle speed. Estimated function was used to determine extent of Highly Flooded zones for routing. Figure adapted from Pregnolato et al. (2017).

centers and found the optimal route to each of the 75 population sites for a total of 750 routes. Of these, a number of the routes resulted in an unsuccessful API response (i.e., no viable route found) and did not return an accompanying GeoJSON. Possibilities for undetermined routes could be due to incomplete road segments not attached to the full network or the lack of a geolocated address at either the start or end points. These were removed from the analysis, resulting in a total of 518 valid routes.

We contrast the baseline routes against those generated using the NRT flood information. In this scenario, we assume the vehicles have been given some advance warning of potentially adverse road conditions and choose to circumvent the most highly impacted regions. Here, we assign the coordinates of the Highly Flooded polygon to the "avoid_polygon" parameter when calculating routes (**Figure 5c**). Routes which intersect flooded regions are assumed to experience a decrease in velocity for the duration of the impacted road segment. A velocity reduction coefficient, derived from the empirical relationship between flood depth and vehicle velocity, is used to impose slower speeds on impacted routes (**Figure 6**). Routes which intersect Highly Flooded regions are assumed to experience more severe reductions, modeled here as reductions greater or equal to the 75th percentile of the coefficient distribution.

Decision Analytical Model

For each of the 518 valid route combinations, we evaluate how emergency response vehicles perform both under baseline conditions and with the benefit of NRT Earth observations. To do this we identify the following management objectives:

1. Minimize the length of impacted roads relative to the total route length (O_1). The objective function is:

$$\frac{1}{r}\sum_{n=1}^{r}\frac{D_F+D_{HF}}{D_T} \tag{1}$$

where D_F is the "flooded" length, D_{HF} is the Highly Flooded length, and D_T is the total route length for each route, r.

1. Minimize the expected value of emergency response time (O_2). The objective function is:

$$E\left[\frac{1}{r}\sum_{n=1}^{r}\frac{D_T}{\mu_r}+\frac{D_F}{(\mu_r*\alpha)}+\frac{D_{HF}}{(\mu_r*\beta)}\right]_N \tag{2}$$

In this formulation, μ represents the average velocity for each route, r, while α and β are stochastic rate reduction coefficients which reduce the average velocity by a specified percentage according to the degree of flooding. The $E[\ ]_N$ notation refers to the expected value for each of r routes over 10,000 uncertain states-of-the-world. We focus on the expected (average) outcome due to its emphasis on classic decision theory, which states that a rational agent will seek to optimize expected utility (Von Neumann and Morgenstern, 1945).

Sensitivity Analysis

Finally, we perform a one-at-a-time sensitivity analysis to determine how variations in the model parameter inputs affect the objective outcomes (Hamby, 1994). Sensitivity analyses can serve as useful diagnostic tools to identify the parameters that

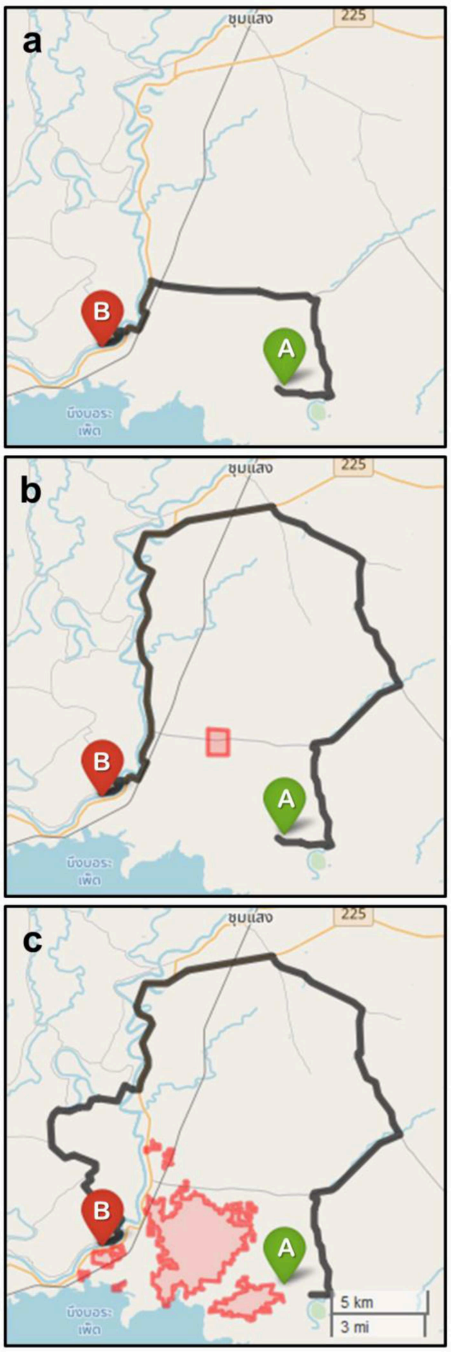


FIGURE 5 | (a) Demonstration of OpenRouteService routing API from origin point A to destination point B. **(b)** Example of "avoid_polygon" parameter in which the system will route around a user-specified roadblock. **(c)** Coordinates of the flood vertices used as "avoid_polygon" feature to simulate Highly Flooded areas.

may require additional calibration and to identify potential knowledge gaps. The one-at-a-time method is known as a local sensitivity analysis because it quantifies the extent to which individual parameters influence model output. We perform this analysis by varying individual parameters in isolation while holding all others constant. The isolated parameters are sampled from the 1st to 99th percentile of its prior distribution (**Supplementary Figure 2**). The results of this analysis then serve to rank the parameters in order of their impact to model variance (Saltelli, 2002).

RESULTS

The geographic distribution of baseline routes differed significantly from the routes that avoided the Highly Flooded areas (**Figure 7a**). Without any forewarning of possible flood conditions, baseline routes optimized the route for fastest arrival, often taking the most geographically direct path. Doing so, however caused many of the baseline routes to unknowingly intersect Flooded and Highly Flooded road segments. Routes that avoided Highly Flooded areas were an average of 11.1 km longer across all 518 route combinations. That said, by avoiding the Highly Flooded areas, the routes utilizing the satellite-based maps of flood inundation also incidentally avoided much of the less severe Flooded regions. Baseline routes contained an average of 16.6 km of impacted roadway ($D_F + D_{HF}$), while the avoidance routes by definition encountered no Highly Flooded areas (**Figure 7b**), and contained an average of 8.7 km of Flooded roadway.

The response times for the baseline and avoidance routes showed similar distributions. Baseline routes results had lower minimum and maximum response times than the corresponding avoidance routes, yet the median, mean (expected), and third quartile response times of the avoidance routes were lower than the baseline counterparts (**Figure 8**). When evaluating for the expected response time over each uncertain SOW, routes that avoided the Highly Flooded regions (and thereby much of the Flooded region as well), were on average ∼9 min faster.

The results of the one-at-a-time sensitivity analysis demonstrate which parameters exert the largest influence on the response time objective (**Figure 9**). The left pane shows the percent of the total model variance attributable to each of the individual parameters, with the width of the colored bars representing the magnitude of the influence. Similarly, the vertical displacement of the curves in the right pane indicate the degree of sensitivity as parameters are varied from the 1st to the 99th percentile of their prior distribution. We see that the rate reduction coefficient for Flooded road segments, α, has the highest degree of influence over the response time objective. The comparatively smaller influence of the Highly Flooded coefficient, β, may be explained by the fact that, on average, Highly Flooded segments only comprised about 2.3% of the total route length. Flooded segments comprised an average of 14% of the total route length, indicating that vehicles were roughly six times more likely to encounter roads that were Flooded but not necessarily ones that were Highly Flooded.

DISCUSSION AND CONCLUSIONS

The disaster response scenario described in this study represents a theoretical example of how NRT Earth observations can

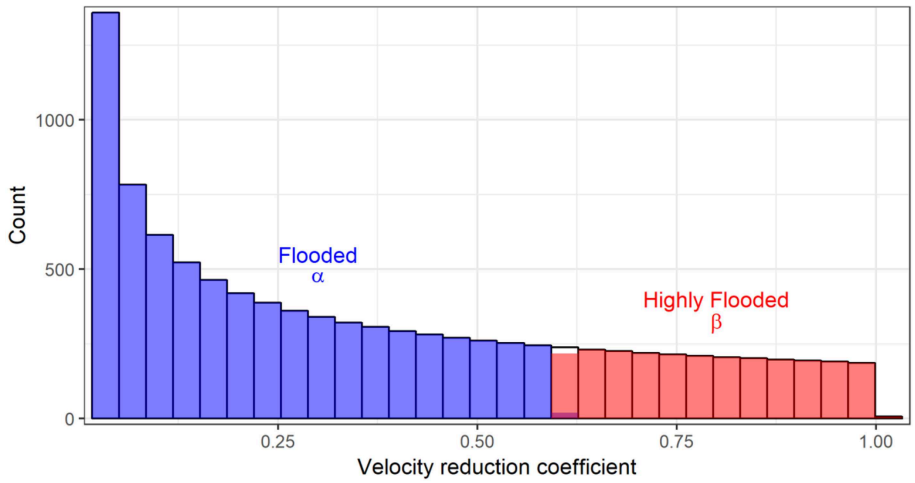


FIGURE 6 | Rate reduction coefficient used to adjust velocity of flooded road segments. Distribution derived from empirical relationship between water depth and vehicle velocity outlined in **Figure 4**.

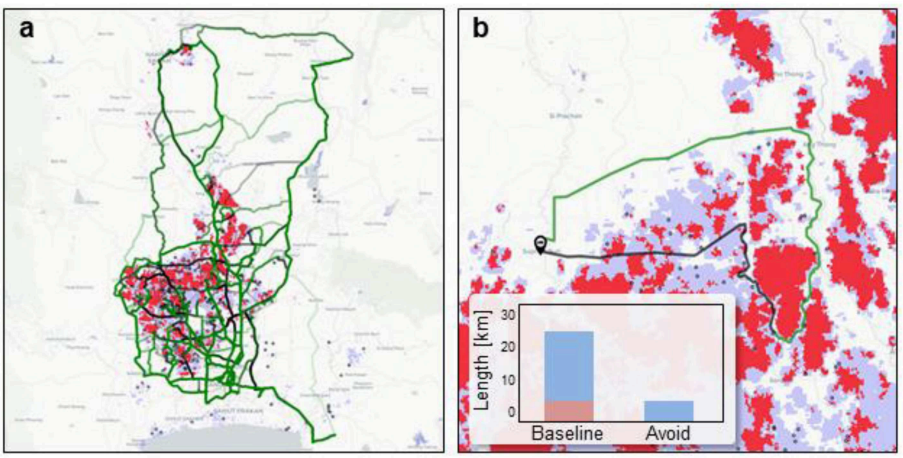


FIGURE 7 | (a) Geographic distribution of routes under baseline conditions (black) and with advance Earth observations (green) with opacity indicating density of routes. **(b)** Example of single route combination. Inset graph shows relative lengths of Flooded and Highly Flooded roads (blue and red shading, respectively).

potentially be used to inform a disaster management decision. When evaluating the social value of that decision, we can relate our management objectives to an established economic metric. In a review of Thai emergency services, Jaldell et al. (2014) investigated the relationship between ambulance response time and mortality to quantify the social benefits of improved emergency operations. They found the value of a 1-min decrease in response time for each dispatch over the course of a year totaled 1.6 billion Thai baht (∼$50,500,000 US). While Jaldell et al. (2014) doesn't explicitly consider explore flood disaster scenarios, they found that the greatest monetary savings occurred during medical emergency and traffic accident calls—both scenarios being likely results of extreme flooding. The potential time savings demonstrated in this analysis are highly dependent on the fidelity of the inundation estimates and the choice of velocity reduction coefficient. Yet while the apparent improvements may appear small initially (∼9 min, on average), when viewed in aggregate, the economic value of the improved information can result in substantial cost reductions and potential lives saved.

An important component of a NRT flood mapping system, as demonstrated here, is how readily it can be deployed to produce potentially useful information. While geospatial data is increasingly commonplace in disaster response and humanitarian logistics, only a small fraction of studies operationally integrate real- or near-real time data (Özdamar and Ertem, 2015; Yagci Sokat et al., 2016). Some of the variables discussed in this analysis have non-trivial uncertainties (e.g., the depth raster and resulting damage map (see section Caveats and Future Research Needs for a discussion of these limitations). That said, imperfect information—delivered operationally and at a latency determined by regional service

providers and disaster responders—may often outweigh more reliable information that comes too late (Eidsvik et al., 2015).

We find that emergency response times are highly dependent on the detected flood inputs, as well as the relationship between estimated inundation depths and vehicle speeds. When accounting for parametric uncertainties and evaluating for model objectives over different route combinations, we find that routes which circumvent the most impacted regions (i.e., Highly Flooded) also avoid additional Flooded regions, resulting in shorter expected response times. The social value of these Earth observations, particularly when evaluated in the context of potential lives saved, can be on the order of millions of dollars annually.

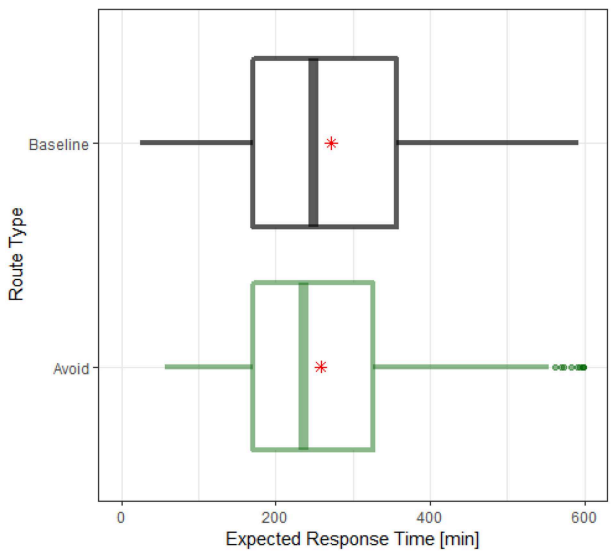


FIGURE 8 | Boxplots of expected (mean) response times across 10,000 states-of-the-world for 518 routes. Red star shows mean values.

CAVEATS AND FUTURE RESEARCH NEEDS

While the results presented here demonstrate the potentially high value of applied satellite information, the analysis has some notable caveats and limitations. One such limitation is the resolution and availability of the data used to generate the flood detection and impact analysis. Because time is a critical factor in any emergency response situation, MODIS imagery was selected to produce the flood detection due to its fast revisit time (twice daily) and low latency (~3 h). An important tradeoff is that the imagery is only of moderate resolution (250-m) and is unable to penetrate cloud cover, making it not as suited for applications requiring high-resolution flood maps. Future implementations of the analysis could feasibly utilize any flood extent—regardless of the sensor used to generate it—providing that imagery is available immediately following the event.

The impact assessment portion of the analysis also includes non-trivial uncertainties. Depth estimates were generated using the MERIT DEM, which improves on many of the sources of error in other global elevation datasets (e.g., vegetation biases, striping, and speckling), yet still has significant errors in vertical accuracy (Yamazaki et al., 2017). The use of open-source materials, including the OpenRouteService routing platform and infrastructure data from OpenStreetMap provide important benefits for accessibility and scalability, yet they often present the challenge of being incomplete and potentially inconsistent. Therefore, the socioeconomic damage map used to identify the potential population sites (**Supplementary Figure 1**) also has associated uncertainties. Despite this, we find that the provisional damage estimates can still be instructive for identifying areas that may be most highly impacted [for more detail, see the discussion section 5.2—Damage Estimate Validation in Oddo et al. (2018)].

Another broad limitation of this analysis is how it models the complex geophysical and behavioral dynamics inherent in a real-world disaster response scenario. The flood extent used here is currently treated as static. In reality, floods are highly dynamic phenomena, causing inundated areas to change over time. Furthermore, in a real-life scenario there could be

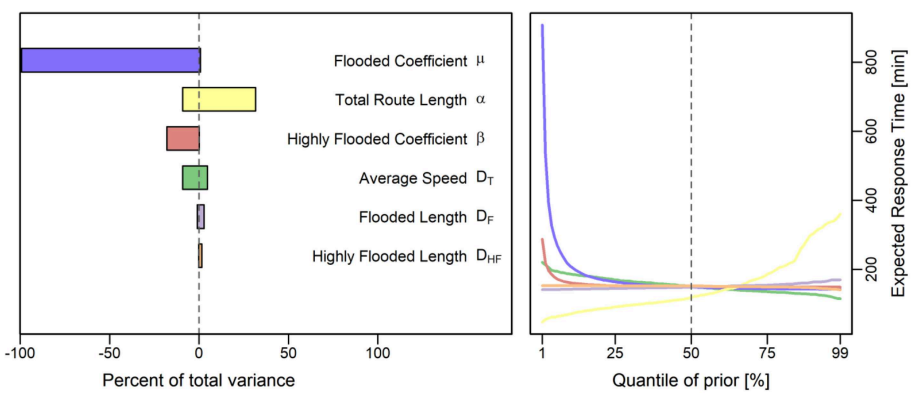


FIGURE 9 | One-at-a-time (OAT) sensitivity analysis for the emergency response time objective. Width of the bars and steepness of curve inclines indicate the degree of sensitivity to each model parameter.

a number of compounding factors that could affect vehicle response times during flood conditions. Congestion from other cars due to evacuations, for instance, would likely be a significant factor, particularly in a densely populated area like Bangkok. This analysis only considers road impediments due to flood inundation, whereas other factors (e.g., fallen trees, construction, or downed power lines) would likely also occur.

On the behavioral side, the model contains several simplifying assumptions for how individuals respond to the potentially chaotic conditions during a flood event. For one, it is assumed that the satellite data is the only source of information on regional road conditions, and that this information is able to be communicated readily to emergency responders. In reality, there may be other sources of intelligence available (i.e., radios, satellite phones, or aerial imagery) and the ability to process and transmit flood maps may be impaired. We also make assumptions for how emergency responders utilize the satellite data. Currently, once a driver decides on a specified route (either the most direct route under baseline conditions or when avoiding the Highly Flooded regions), they do not deviate over the course of the trip. A more realistic portrayal would allow a rational driver to continue until encountering an impediment before subsequently choosing to re-route. Finally, we only consider ground transportation as a mechanism for emergency response, whereas a coordinated flood operation would likely involve support through other means (e.g., helicopters or boats).

Due to the assumptions described here, this analysis is not intended to be prescriptive in how to route emergency vehicles in a real-world flood situation. Rather, the results are intended to serve as a didactic example of how applied Earth observations can be operationally combined with econometric data to produce insights for decision-making. Future refinements to this system could include routes constrained by actual emergency vehicle observations, more advanced behavior on the part of the drivers, and a more dynamic treatment of how flooded areas change through time.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

PO and JB conceived and designed the broader research objectives. PO performed the analysis, analyzed the data, and wrote the manuscript. JB provided the funding support, project administration, reviewing/editing, and supervision.

FUNDING

Funding for this work was graciously provided by the NASA Applied Sciences Program.

ACKNOWLEDGMENTS

The authors would like to sincerely thank Aakash Ahamed, Joseph Spruce, and members of the Bolten Research Group for exploratory research and thoughtful discussion of this work. The authors would also like to gratefully thank G. Robert Brackenridge for providing a clean version of the flood extent used to produce the study location maps. Support for this study provided by the NASA Applied Sciences Program and SERVIR-Mekong, with data and cooperation from the Asian Disaster Preparedness Center (ADPC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2019.00127/full#supplementary-material

REFERENCES

Ahamed, A., Bolten, J., Doyle, C., and Fayne, J. (2017). "Near real-time flood monitoring and impact assessment systems," in *Remote Sensing of Hydrological Extremes Springer Remote Sensing/Photogrammetry*, ed V. Lakshmi (Cham: Springer International Publishing), 105–118. doi: 10.1007/978-3-319-43744-6_6

Ahamed, A., and Bolten, J. D. (2017). A MODIS-based automated flood monitoring system for southeast asia. *Int. J. Appl. Earth Obs.* 61, 104–117. doi: 10.1016/j.jag.2017.05.006

Alfonso, L., Mukolwe, M. M., and Di Baldassarre, G. (2016). Probabilistic flood maps to support decision-making: mapping the value of information. *Water Resour. Res.* 52, 1026–1043. doi: 10.1002/2015WR017378

Altay, N., and Green, W. G. (2006). OR/MS research in disaster operations management. *Eur. J. Oper. Res.* 175, 475–493. doi: 10.1016/j.ejor.2005.05.016

Bernknopf, R., Brookshire, D., Kuwayama, Y., Macauley, M., Rodell, M., Thompson, A., et al. (2017). The value of remotely sensed information: the case of a GRACE-enhanced drought severity index. *Weather Clim. Soc.* 10, 187–203. doi: 10.1175/WCAS-D-16-0044.1

Brakenridge, G. R., Kettner, A., Policelli, F., and Slayback, D. (2011). *NASA-MODIS Rapid Response Record of 2011 Flooding in Thailand*. Dartmouth Flood Observatory. Available online at: http://floodobservatory.colorado.edu/RapidResponse/2011ThailandRR.html (accessed January 23, 2019).

Brakenridge, R., and Anderson, E. (2006). "MODIS-based flood detection, mapping and measurement: the potential for operational hydrological applications," in *Transboundary Floods: Reducing Risks Through Flood Management*, eds J. Marsalek, G. Stancalie, and G. Balint (Dordrecht: Springer), 1–12. doi: 10.1007/1-4020-4902-1_1

Bratvold, R. B., Bickel, J. E., and Lohne, H. P. (2009). "Value of information in the oil and gas industry: past, present, and future," in *SPE Annual Technical Conference and Exhibition* (Anaheim, CA: Society of Petroleum Engineers).

CIESIN (2016). *Gridded Population of the World, Version 4 (GPWv4): Population Density*. Available online at: http://dx.doi.org/10.7927/H4NP22DQ (accessed November 7, 2018).

Comfort, L. K., Ko, K., and Zagorecki, A. (2004). Coordination in rapidly evolving disaster response systems: the role of information. *Am. Behav. Sci.* 48, 295–313. doi: 10.1177/0002764204268987

Davies, D. K., Murphy, K. J., Michael, K., Becker-Reshef, I., Justice, C. O., Boller, R., et al. (2015). "The use of NASA LANCE imagery and data for near real-time applications," in *Time-Sensitive Remote Sensing*, eds C. D. Lippitt, D.

A. Stow, and L. L. Coulter (New York, NY: Springer New York), 165–182. doi: 10.1007/978-1-4939-2602-2_11

Eidsvik, J., Mukerji, T., and Bhattacharjya, D. (2015). *Value of Information in the Earth Sciences: Integrating Spatial Modeling and Decision Analysis*. Cambridge, UK: Cambridge University Press.

Fayne, J. V., Bolten, J. D., Doyle, C. S., Fuhrmann, S., Rice, M. T., Houser, P. R., et al. (2017). Flood mapping in the lower Mekong River Basin using daily MODIS observations. *Int. J. Remote Sens.* 38, 1737–1757. doi: 10.1080/01431161.2017.1285503

Gale, E. L., and Saunders, M. A. (2013). The 2011 Thailand flood: climate causes and return periods. *Weather* 68, 233–237. doi: 10.1002/wea.2133

Goodchild, M. F., and Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *Int. J. Digit. Earth* 3, 231–241. doi: 10.1080/17538941003759255

Hamby, D. M. (1994). A review of techniques for parameter sensitivity analysis of environmental models. *Environ. Monit. Assess.* 32, 135–154. doi: 10.1007/BF00547132

Haraguchi, M., and Lall, U. (2015). Flood risks and impacts: a case study of Thailand's floods in 2011 and research questions for supply chain decision making. *Int. J. Disaster Risk Reduct.* 14, 256–272. doi: 10.1016/j.ijdrr.2014.09.005

Haworth, B., and Bruce, E. (2015). A review of volunteered geographic information for disaster management: a review of VGI for disaster management. *Geogr. Compass* 9, 237–250. doi: 10.1111/gec3.12213

Hodgson, M. E., Davis, B. A., and Kotelenska, J. (2009). "Remote sensing and GIS data/information in the emergency response/recovery phase," in *Geospatial Techniques in Urban Hazard and Disaster Analysis Geotechnologies and the Environment*, eds P. S. Showalter and Y. Lu (Dordrecht: Springer), 327–354. doi: 10.1007/978-90-481-2238-7_16

Howard, R. A. (1966). Information value theory. *IEEE Trans. Syst. Sci. Cybern.* 2, 22–26. doi: 10.1109/TSSC.1966.300074

Howard, R. A. (1968). The foundations of decision analysis. *IEEE Trans. Syst. Sci. Cybern.* 4, 211–219. doi: 10.1109/TSSC.1968.300115

Howden, M. (2009). "How humanitarian logistics information systems can improve humanitarian supply chains: a view from the field," in *Proceedings of the 6th International ISCRAM Conference* (Gothenburg).

Jaldell, H., Lebnak, P., and Amornpetchsathaporn, A. (2014). Time is money, but how much? the monetary value of response time for thai ambulance emergency services. *Value Health* 17, 555–560. doi: 10.1016/j.jval.2014.05.006

Keller, K., Kim, S.-R., Baehr, J., Bradford, D. F., and Oppenheimer, M. (2007). "What is the economic value of information about climate thresholds?," in *Human-Induced Climate Change: An Interdisciplinary Assessment*, eds M. Schlesinger, H. Kheshgi, J. Smith, F. de la Chesnaye, J. M. Reilly, T. Wilson, et al. (Cambridge, UK: Cambridge University Press), 343–354. doi: 10.1017/CBO9780511619472.033

Koks, E. E., Jongman, B., Husby, T. G., and Botzen, W. J. W. (2015). Combining hazard, exposure and social vulnerability to provide lessons for flood risk management. *Environ. Sci. Policy* 47, 42–52. doi: 10.1016/j.envsci.2014.10.013

Koriche, S. A., and Rientjes, T. H. M. (2016). Application of satellite products and hydrological modelling for flood early warning. *Phys. Chem. Earth Parts ABC* 93, 12–23. doi: 10.1016/j.pce.2016.03.007

Macauley, M. K. (2006). The value of information: Measuring the contribution of space-derived earth science data to resource management. *Space Policy* 22, 274–282. doi: 10.1016/j.spacepol.2006.08.003

Mjelde, J. W., Sonka, S. T., and Peel, D. S. (1989). *The Socioeconomic Value of Climate and Weather Forecasting: A Review*. Champaign, IL: Illinois State Water Survey.

Obersteiner, M., Balkovič, J., Böttcher, H., Bouma, J. A., Fritz, S., Fuss, S., et al. (2017). "The value of global earth observations," in *Satellite EarthObservations and Their Impact on Society and Policy*, eds M. Onoda and O. R. Young (Singapore: Springer Singapore), 137–142.

Oddo, P. C., Ahamed, A., and Bolten, J. D. (2018). Socioeconomic impact evaluation for near real-time flood detection in the lower mekong river basin. *Hydrology* 5:23. doi: 10.3390/hydrology5020023

Okuyama, Y. (2003). Economics of natural disasters: A critical review. *Res. Pap.* 12, 20–22.

OpenStreetMap Contributors (2019). *OpenStreetMap*. Available online at: https://www.openstreetmap.org/ (accessed November 7, 2017).

Ortuño, M. T., Cristóbal, P., Ferrer, J. M., Martín-Campo, F. J., Muñoz, S., Tirado, G., et al. (2013). "Decision aid models and systems for humanitarian logistics. a survey," in *Decision Aid Models for Disaster Management and Emergencies*, eds B. Vitoriano, J. Montero, and D. Ruan (Paris: Atlantis Press), 17–44. doi: 10.2991/978-94-91216-74-9_2

Özdamar, L., and Ertem, M. A. (2015). Models, solutions and enabling technologies in humanitarian logistics. *Eur. J. Oper. Res.* 244, 55–65. doi: 10.1016/j.ejor.2014.11.030

Policelli, F., Slayback, D., Brakenridge, B., Nigro, J., Hubbard, A., Zaitchik, B., et al. (2017). "The NASA global flood mapping system," in *Remote Sensing of Hydrological Extremes Springer Remote Sensing/Photogrammetry*, ed. V. Lakshmi (Cham: Springer International Publishing), 47–63. doi: 10.1007/978-3-319-43744-6_3

Pregnolato, M., Ford, A., Wilkinson, S. M., and Dawson, R. J. (2017). The impact of flooding on road transport: a depth-disruption function. *Transp. Res. Part Transp. Environ.* 55, 67–81. doi: 10.1016/j.trd.2017.06.020

Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk Anal.* 22, 579–590. doi: 10.1111/0272-4332.00040

Shivaprasad Sharma, S. V., Roy, P. S., Chakravarthi, C., Srinivasarao, G., and Bhanumurthy, V. (2017). Extraction of detailed level flood hazard zones using multi-temporal historical satellite data-sets – a case study of Kopili River Basin, Assam, India. *Geomat. Nat. Hazards Risk* 8, 792–802. doi: 10.1080/19475705.2016.1265014

Von Neumann, J., and Morgenstern, O. (1945). Theory of games and economic behavior. *Bull. Amer. Math. Soc.* 51, 498–504. doi: 10.1090/S0002-9904-1945-08391-8

Walter, J. (Ed). (2005). *World Disasters Report 2005: Focus on Information in Disasters, 1st Edn.* Dordrecht; Norwell, MA: Kumarian Press.

Wu, H., Adler, R. F., Tian, Y., Huffman, G. J., Li, H., and Wang, J. (2014). Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.* 50, 2693–2717. doi: 10.1002/2013WR014710

Yagci Sokat, K., Zhou, R., Dolinskaya, I. S., Smilowitz, K., and Chan, J. (2016). Capturing real-time data in disaster response logistics. *J. Oper. Supply Chain Manag.* 9:23. doi: 10.12660/joscmv9n1p23-54

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., et al. (2017). A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.* 44:2017GL072874. doi: 10.1002/2017GL072874

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

TECHNOLOGY AND CODE
published: 09 October 2019
doi: 10.3389/fenvs.2019.00148

Enabling Stakeholder Decision-Making With Earth Observation and Modeling Data Using Tethys Platform

E. James Nelson[1], Sarva T. Pulla[2], Mir A. Matin[3], Kiran Shakya[3], Norm Jones[1], Daniel P. Ames[1], W. Lee Ellenburg[2,4], Kel N. Markert[2,4], Cédric H. David[5], Benjamin F. Zaitchik[6], Patrick Gatlin[7] and Riley Hales[1]*

[1] Department of Civil and Environmental Engineering, Brigham Young University, Provo, UT, United States, [2] National Aeronautics and Space Administration Marshall Space Flight Center, SERVIR, Science Coordination Office, Huntsville, AL, United States, [3] International Centre for Integrated Mountain Development (ICIMOD), SERVIR Team, Kathmandu, Nepal, [4] Earth System Science Center, The University of Alabama in Huntsville, Huntsville, AL, United States, [5] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States, [6] Department of Earth and Planetary Sciences, John Hopkins University, Baltimore, MD, United States, [7] National Aeronautics and Space Administration Marshall Space Flight Center, Earth Science Branch, Huntsville, AL, United States

OPEN ACCESS

Edited by:
Niall Patrick Hanan, New Mexico State University, United States

Reviewed by:
Nidhi Nagabhatla, United Nations University Institute for Water Environment and Health, Canada
Yang Hong, University of Oklahoma, United States

***Correspondence:**
E. James Nelson
jimn@byu.edu

Specialty section:
This article was submitted to Freshwater Science, a section of the journal Frontiers in Environmental Science

Received: *06 May 2019*
Accepted: *17 September 2019*
Published: *09 October 2019*

Citation:
Nelson EJ, Pulla ST, Matin MA, Shakya K, Jones N, Ames DP, Ellenburg WL, Markert KN, David CH, Zaitchik BF, Gatlin P and Hales R (2019) Enabling Stakeholder Decision-Making With Earth Observation and Modeling Data Using Tethys Platform. Front. Environ. Sci. 7:148. doi: 10.3389/fenvs.2019.00148

Tethys Platform is an open source framework for developing web-based applications for Earth Observation data. Our experience shows that Tethys significantly lowers the barrier for cloud-based app development, simplifies the process of accessing scalable distributed cloud computing resources and leverages additional software for data and computationally intensive modeling. The Tethys software development kit allows users to create web apps for visualizing, analyzing, and modeling Earth Observation data. Tethys platform provides a collaborative environment for scientists to develop and deploy several Earth Observation web applications across multiple Tethys portals. We work in partnership with leading regional organizations world-wide to help developing countries use information provided by earth-observing satellites and geospatial technologies for managing climate risks and land use. This paper highlights the several Tethys portals and web applications that were developed as part of this effort. Implementation of the Tethys framework has significantly improved the Application Readiness Level metric for several NASA projects and the potential impact of Tethys to replicate and scale other applied science programs.

Keywords: Tethys platform, earth observations, decision-making, hydroinformatics, SERVIR

INTRODUCTION

SERVIR is a joint initiative of the National Aeronautics and Space Administration (NASA) and the United States Agency for International Development (USAID) that seeks to build the capacity of local decision-makers dealing with a wide range of climate-related problems by making global earth observation data and associated tools available. The SERVIR model brings together regional hubs supported by USAID grants with scientific experts in the areas of Agriculture and Food Security, Water Resources and Hydroclimatic Disasters, Land Cover and Land Use Change and Ecosystems, and Weather and Climate.

For example, the NASA Applied Sciences Program funds scientists from United States (US) institutions with the expectation that technology transfer to the hubs and local stakeholders in their regions occurs by the end of the 3-year grant period. The NASA Applied Sciences Program measures research becoming integrated into stakeholder and end-user decision-making using a 9-point Application Readiness Level (ARL) scale where ARL 1 represents basic research and ideas in their infancy to ARL 9 where data, models, and tools are approved, and fully integrated by the stakeholders and have sustained use in making decisions (see **Figure 1**).

The hydrologic cycle and other earth systems including the atmosphere and land surface processes are extremely complex integrated systems. Understanding and simulating the hydrologic cycle has been a scientific challenge for many decades, which is further complicated by the vast amount of data, computational horsepower, and human resources it requires (Sood and Smakhtin, 2015). In too many ways and places, our understanding of the hydrologic system is incomplete, and humanity suffers from a lack of information that leads to uninformed decisions. To address this challenge, the NASA/USAID SERVIR program is leveraging advances in earth observations, numerical weather prediction, supercomputing, hydrologic modeling, cloud services, big data visualization, and the collaboration of the scientists that make up the SERVIR Applied Science Team (AST).

The term "Big Data" has been used variously to describe the massive and growing quantity of data available for scientific research, as well as advanced machine learning methods for interpreting and deriving meaning from extremely large databases. When considering the massive datasets themselves—before attempting to automatically extract relationships and predictions from the data—it is critical but also challenging to achieve basic data management and visualization. This problem is exacerbated by the variety of often domain-specific file formats and software tools that have evolved to support these data.

For example, within the hydrologic modeling community, time-series oriented data is often stored in simple text files or Microsoft Excel files. The Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) has developed networked database systems using web services to advance data storage and sharing beyond these simple methods (Tarboton et al., 2009; Ames et al., 2012; Kadlec et al., 2015; Horsburgh et al., 2016). Additionally, the Open Geospatial Consortium (OGC) has promulgated data structure and storage standards such as WaterML and TimeseriesML (OGC, 2012). However, even with these developments over the past decade, much time series hydrology data continues to be stored and shared using less efficient and outdated methods. In the atmospheric sciences, multivariate flexible data and file format standards such as NetCDF (Rew and Davis, 1990) are commonly used. And, in the case of earth observations data, the Hierarchical Data Format (HDF) has been widely adopted (Folk et al., 1999). This distinction between time series and spatially-oriented data has been termed a "digital divide" and various efforts have been undertaken to integrate spatial and temporal data standards, thereby "crossing" this digital divide (Maidment et al., 2010; Salas et al., 2012).

While tight integration of spatial (e.g., earth observation) and temporal (e.g., hydrologic observations) data in a unified data model and associated file format would be a potentially optimal solution to this earth systems science big data problem, another point-of-view on the challenge is to maintain original datasets in their "native" file formats and rather, develop

FIGURE 1 | NASA application readiness level scale. Available online at: https://www.nasa.gov/sites/default/files/files/ExpandedARLDefinitions4813.pdf (accessed September 20, 2016).

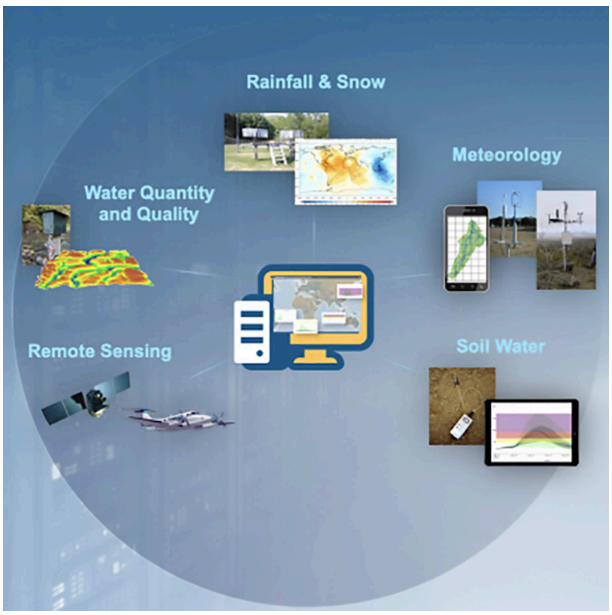


FIGURE 2 | Cyberinfrastructure requirements for hydrologic information development.

visualization and analysis tools that can work with these formats directly. This is the approach that has largely been adopted by SERVIR Hubs working to make earth science data actionable for decision-makers and stakeholders. Indeed, various web application technologies exist to support such decision-support tools that operate on spatial and temporal data. For example, Esri Story Maps technology has been used for community outreach in the SERVIR program (Adams et al., 2016) and Esri ArcGIS Online technology has been used to develop interactive flood inundation forecast and mapping tools (Souffront Alcantara, 2018). In addition to these well-established commercial tools for web application development, a number of custom tools built on open source web technologies have recently been presented in the literature for the analysis and visualization of large spatial and temporal data sets (Brendel et al., 2019; da Costa et al., 2019; González et al., 2019; Markert et al., 2019; Strömbäck et al., 2019). The remainder of this paper explores the use of a relatively new big data visualization web application development toolkit called Tethys Platform and its adoption and use within the SERVIR community. Tethys Platform was developed to help lower barriers to web application development in water and environmental modeling (Swain et al., 2016) and has been used successfully as an open source data portal in several applications (Ames et al., 2018; Jackson et al., 2018; Markert et al., 2018; Souffront Alcantara et al., 2018).

TETHYS PLATFORM

In the era of big data, desktop computing is not efficient or sustainable, which has the benefit of leveling the computational and decision-making landscape for stakeholders, especially in developing countries where the resources to maintain hardware, software, and technical capacity is limited. International development and relief agencies pour millions into strengthening capacity of water resources managers with often only a small return on investment, because while initial investment in cyberinfrastructure and training may yield working systems, the ability of the receiving countries/agencies to maintain the systems is limited. **Figure 2** illustrates the cyberinfrastructure requirements for developing good hydrologic information from earth observation and modeling resources. At the center are computational resources that must have internet access and bandwidth to continually update both *in situ* and remotely sensed data along with numerical weather predictions that are used to drive hydrologic and hydraulic models.

The challenge is that there remains a heavy burden to both stand up such systems as well as operate and maintain them in such a way that actionable information is sufficiently reliable to enable informed decision making for national hydrometeorological services. Even after making the investment of such systems, a large amount of the effort and resources are expended just to produce the information necessary to make good decisions. Cloud computing has become a powerful and affordable solution that can aid in shifting the decision-making responsibilities of local agencies from one of developing the reliable hydrologic information needed for informed decision-making to one of focusing on the application and decision-support tools that are populated with established hydrologic information services from reliable sources (see **Figure 3**).

Tethys Platform is a web-based app development framework for rapid deployment of end-user-focused tools that follow modern, consistent, scalable, cross-platform, reusable, web programming paradigms. Tethys is a relatively new software system built on commonly used web programming frameworks (e.g., Django, GeoServer, PostGIS, and OpenLayers). It is stable and supported by a growing user and developer community. Tethys Platform leverages recent advances in cloud computing to facilitate better use of large earth observation data sets and water resource models as decision-making tools. These modeling and visualization tools can be hosted on a server and used by multiple remote users via a web interface, which eliminates the need to procure and maintain high performance hardware typically required by models. Further, it deals with issues related to software installation and platform incompatibilities (Mac vs. PC vs. Linux, etc.), monitor and install software updates, or download large data sets; problems that are exacerbated in regions where financial and technical capacity can be limited. An internet connection and a web browser are all that is required to access the models and associated data, which means challenges associated with downloading data and updating software are not a barrier to sustainable use.

The Tethys Platform software architecture is illustrated in **Figure 4**. Tethys is built on the Django framework and Python programming language—significantly lowering the barrier for app development (Swain et al., 2016). Tethys Platform apps are hosted in a Tethys Portal and are intended to ease the burden of science information access by enabling web based interaction with spatial resources stored in repositories such

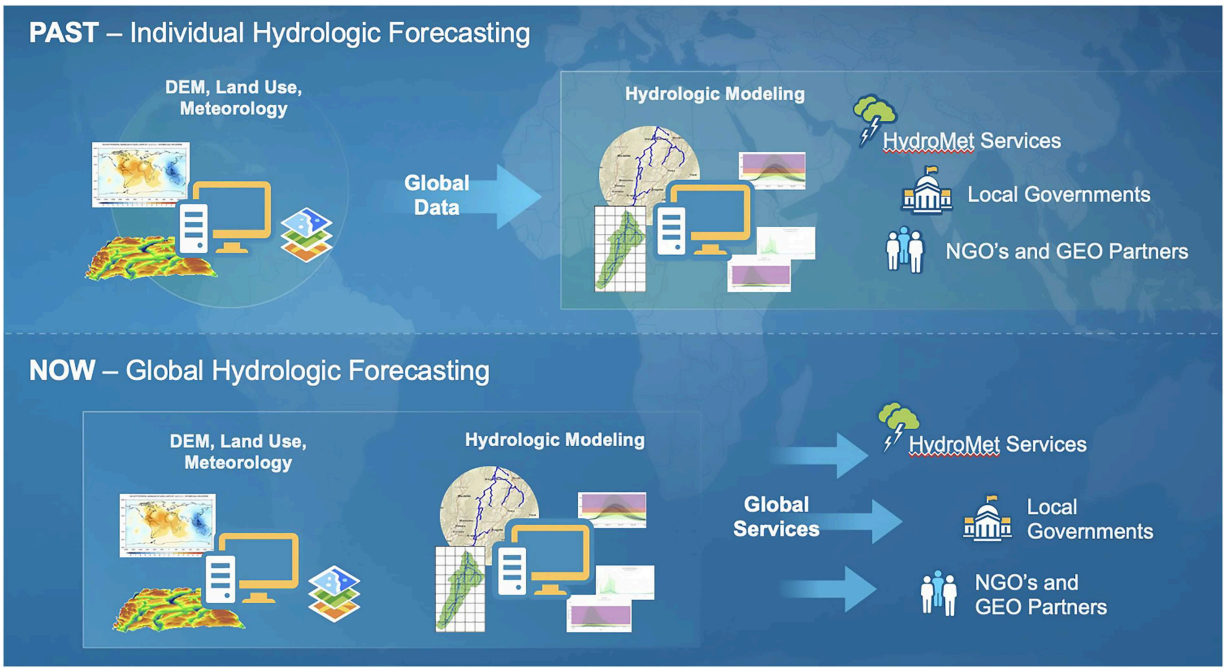


FIGURE 3 | Disruptive technologies for hydrologic information delivery.

as the commercial ArcGIS Online or open source approaches such as GeoServer and Openlayers (Swain et al., 2016). The base of the system is the Tethys Software Development Kit (SDK) which integrates a broad suite of open source tools for rapid development of web-based water resource data and modeling applications. The Tethys software suite includes components for distributed computing, spatial publishing, geoprocessing, spatial data management, and visualization. It includes external connections for cloud computing and dataset storage and supports integration of both open source components (GeoServer, OpenLayers, etc.) and proprietary systems (ArcGIS Online, ArcGIS Server, and ArcGIS JavaScript Mapping API) so that applications for stakeholders within collaborative organizations such as SERVIR can be custom-tailored to address a variety of needs and circumstances.

App views, or web pages, use the Django templating language, but the Tethys framework provides a base template that includes a standard layout for app pages with areas for a header, navigation links, action buttons, and primary content rather than requiring developers to start from scratch with each template using Django. This reduces the amount of repetitive coding required for developing web apps and also leads to a familiar user interface experience.

Tethys Portal is the Tethys Platform component that provides the primary runtime environment for Tethys web apps. It is implemented as a Django website project and it extends Django capabilities to provide the core website functionality that is often taken for granted in modern web applications. It includes a user account system complete with user profiles and a password reset mechanism for forgotten passwords. It also provides a landing page that presents the associated Tethys Platform instance and an app library page that provides an access point for installed apps. It includes an administrator backend that can be used to manage user accounts, permissions, link to elements of the software suite, and customize the instance. This architecture and the associated apps have been developed at BYU and also now being used by the SERVIR Global science coordination office (**Figure 5**).

SERVIR—HKH APPLIED SCIENCE TEAM

Tethys Platform, described above, has been used extensively to support SERVIR Hub data visualization and dissemination, and as the basis for several regional decision support tool. In particular, we have applied this technology to support the Hindu Kush Himalayan (HKH) SERVIR region. This region is often referred as a disaster hotspot due to increasing geophysical and hydro-meteorological events resulting in growing loss of lives and livelihood support systems. The region accounts for 36% (940 out of 2,564) of disaster events in Asia between 1900 and 2015 (Guha-Sapir, 2009). Flooding is the most prominent disaster that affects all the countries in the region (Shrestha and Pradhan, 2015). International Center for Integrated Mountain Development (ICIMOD) is a regional intergovernmental organization that aims to support sustainable development in the HKH region through information and knowledge generation and sharing and evidence-based decision making to reduce the impact of climate change. ICIMOD hosts the SERVIR HKH hub serving five countries including Afghanistan, Bangladesh, Myanmar, Nepal, and Pakistan. The aim is to enhance the capacity of relevant institutes and people to use earth observation and

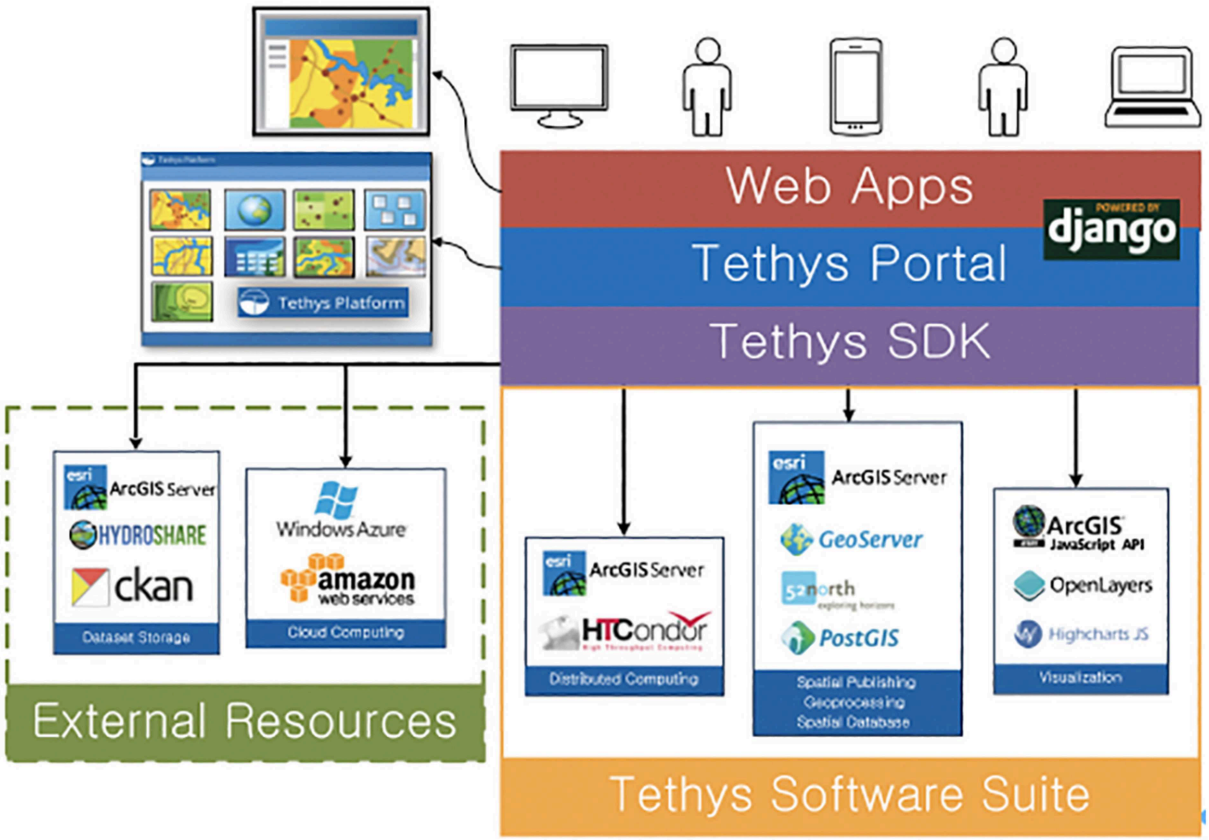


FIGURE 4 | Tethys platform cloud-based app architecture for delivering SERVIR decision support tools.

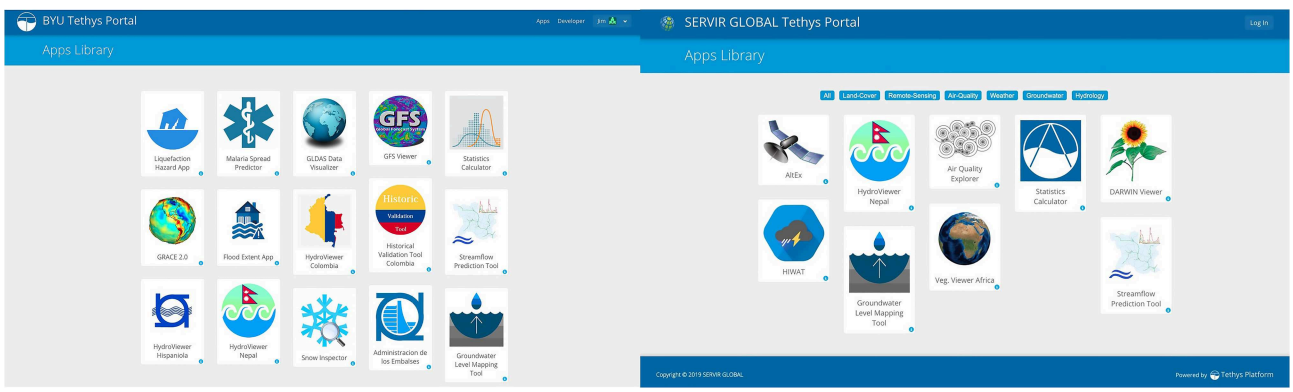


FIGURE 5 | BYU Tethys portal **(Left)** and SERVIR global Tethys portal **(Right)**.

geospatial technology-based solutions in four service areas, including agriculture and food security, water resources and hydroclimatic disasters, land use landcover and ecosystem, and weather and climate.

To better serve the HKH countries, a series of consultation and assessments were conducted to understand their needs and capacities. Flood early warning was identified as one of the key priorities by all of the countries. However, the existing capacity and type of information needs varied by government agencies, community organizations, and countries. Early warning and transboundary flow were identified as key needs in Bangladesh for both riverine and flash flood. In Nepal, both riverine flood and flash flood warnings were identified as priorities. At present, all the countries are familiar with the Global Flood Awareness

System (GLOFAS) to support their flood warning service (Alfieri et al., 2013). However, because of the coarse resolution of the GLOFAS-based forecasts, they are applicable to only a few major river systems. SERVIR promotes the use of earth observations and builds capacity in the region by linking projects of the NASA AST to the region's needs.

The SERVIR AST projects in the HKH region aim to enhance the capacity of ICIMOD to better serve the stakeholders flood early warning needs. In this context, the AST team brings the tools and models to address the modeling and information gaps and build capacity of ICIMOD to address the needs of its stakeholders. The ICIMOD team engages with users and stakeholders to develop customized dissemination tools, build capacity of stakeholders and validate the model results. The following sections describe the four primary AST projects and how each was able to leverage the Tethys Platform to provide a consistent and easy to use system for water management and forecasting of both floods and long-term storage and drought.

South Asia Streamflow Prediction From ECMWF Ensemble Forecasts

A high-density large-scale streamflow prediction system covering South Asia was developed using the GloFAS runoff, ERA Interim data, and the RAPID routing model with sufficient detail to provide local forecasts where decisions are made. The workflow to generate these forecasts was deployed completely on the cloud. Two principal Tethys web applications exist to interact with the results, while a REST application programmer interface (API) has also been developed to easily retrieve data without the need of a web interface. GloFAS is an ensemble hydrologic model that generates 51 different runoff forecasts for the major rivers of the world on a global grid with a resolution of 16 km^2 on a continuous basis. A 52nd forecast is generated at a resolution of 8 km^2. GloFAS was released in 2011 by ECMWF and the European Commission's Joint Research Center (JRC) as part of the Copernicus Emergency Management Services and has been operational since July 2011. The GloFAS system is composed of an integrated hydrometeorological forecasting chain and of a monitoring system that analyzes daily results and shows forecast flood events on a dedicated web platform (Alfieri et al., 2013). This model uses real-time and historical observations in combination with a Data Assimilation System (DAS) and a Global Circulation Model (GCM). The underlying framework used to create GloFAS is ECMWF's Integrated Forecasting System (IFS). GloFAS uses the HTESSEL for its land surface scheme. HTESSEL is a hydrologically revised version of the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) model (Balsamo et al., 2009). This new land surface scheme corrected the absence of a surface runoff component in its predecessor, among other minor improvements.

The ERA-Interim data is the result of a global atmospheric reanalysis produced also by ECMWF. This data covers from January 1980 to December 2014 (35 years) for the entire globe. One of the advantages of using reanalysis is that the data provides a global view that encompasses many essential climate variables in a physically consistent framework, with only a short time delay (Dee et al., 2011). This type of data becomes invaluable in areas where no actual observed data are available and provides a surrogate in regions where no historical observations are available. A runoff derivative of this atmospheric reanalysis was produced on a 40 km^2 global grid using a land surface model simulation in HTESSEL.

Streamflow Prediction Tool Tethys App

The Streamflow Prediction Tool was created as part of the initial experiment to produce higher resolution results on stream networks in the United States (Snow, 2015). The application now provides support for global watersheds and is the primary web application used for visualization and dissemination through an Application Programming Interface (API) to develop customized solutions (see **Figure 6**).

Nepal Streamflow Prediction System Tethys App

As an alternate to accessing the entire South Asia stream networks, or any of the larger modeled regions, a custom viewer for a specific country or region can be developed through the API. In this case a national model for Nepal was created as a separate Tethys web application so that it could be customized for use by the Department of Hydrology and Meteorology (DHM). The streamflow network is derived from the subset of the larger South Asia network and then using the unique ID of each river the API to retrieve a streamflow forecast or retrospective historical simulation can be retrieved and displayed. In this instance in order to make the application more responsive a workflow updates a local database of all streams through the API each day as they are computed and then in real time the app accesses forecasts from the local database rather than through the API. Additional layers which indicate important geographic references to provinces and districts have been added as could any other layer's functionality as needed by the specific stakeholders (see **Figure 7**).

Bangladesh Transboundary Flow Tethys App

Bangladesh is a country that sits at the downstream of the Ganges and Brahmaputra rivers, while the country itself represents only about 8% of the entire drainage area belonging to those two major networks. Because of this there are repeated floods whose flows originate upstream and out of the country such that good information is difficult to obtain. Because of this the most important need was to have an application that provided the streamflow forecast at these two major transboundary locations along with other minor river networks that also originate outside the country. The ECMWF ensemble forecasts enabled through the streamflow prediction tool API along with a Tethys application that displays the forecasts at each of the transboundary rivers allows the Bangladesh Flood Forecast and Warning Center (FFWC) access to a probabilistic forecast with 15 days of lead time that can be used in downstream hydraulic and flood mapping applications (see **Figure 8**).

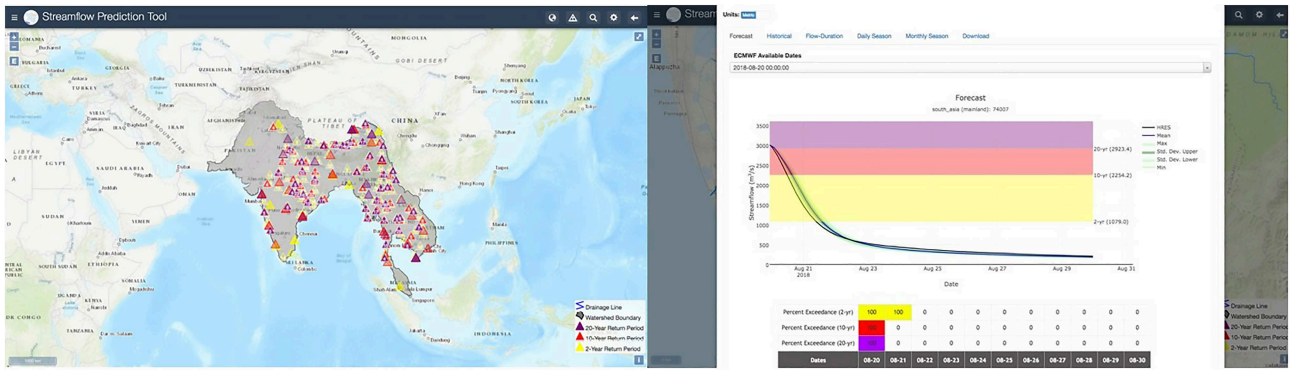


FIGURE 6 | Tethys streamflow prediction tool for visualization and dissemination of the 15-day ECMWF ensemble global forecasts.

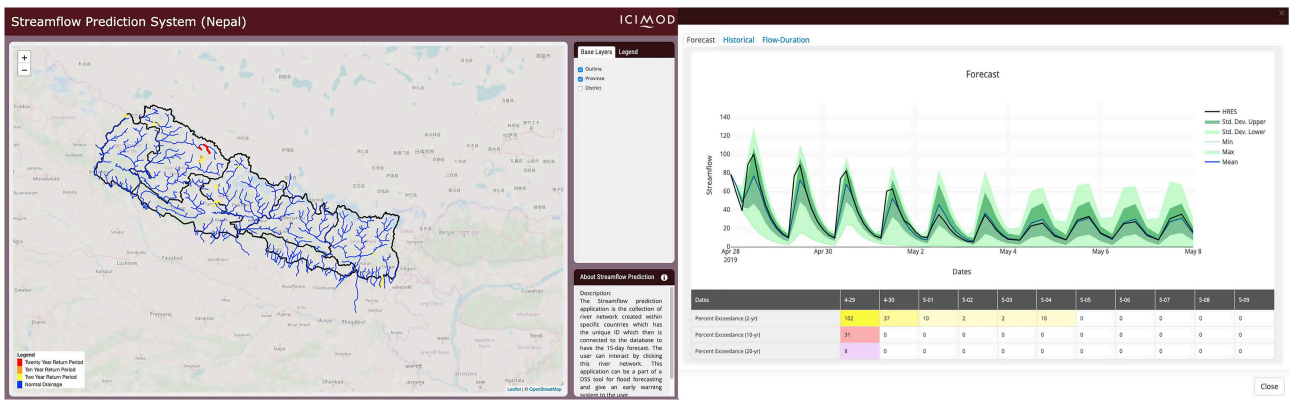


FIGURE 7 | Nepal streamflow prediction system derived from the global streamflow prediction tool.

The HKH Sub-seasonal-to-seasonal Hydrological Forecast System (HKH-S2S)

The HKH and downstream areas in South Asia experience some of the most significant sub-seasonal to seasonal (S2S) variability on earth. The South Asian monsoon is both the lifeblood of the region and a source of significant hazard. Its precipitation sustains agriculture, feeds mountain snowpack and glaciers, and replenishes surface and groundwater resources. But torrential monsoon rains triggers floods and landslides, sub-seasonal monsoon dry spells can cause crops to fail, and years of weak monsoon can endanger food and water security across the region. S2S prediction of HKH meteorology is a grand challenge for atmospheric research and modeling, but significant progress has been made in recent years. Forecasts like those produced by the NASA Global Modeling and Assimilation Office (GMAO) and the NOAA Climate Prediction Center (CPC) now offer meaningful skill that can be applied to disaster preparedness and water and agriculture management. This is particularly true when these atmospheric forecasts are applied to predict hydrological conditions, including drought outlooks and estimates of S2S flood risk. Hydrological forecasts derive skill both from the S2S meteorological prediction and from accurate estimates of initial hydrological conditions, including water storage in snow and soil moisture, so an S2S hydrological forecast can be more skillful than the meteorological forecast that drives it (see **Figure 9**).

Recognizing this potential, HKH-S2S makes use of a land data assimilation system (LDAS)—in this case, the South Asia Land Data Assimilation System (Ghatak et al., 2018)—to provide near-real time hydrological monitoring. The LDAS consists of a suite of gridded land surface models, parameterized using satellite-derived estimates of vegetation, land use, and other parameters, that draw meteorological forcing from a combination of downscaled Global Data Assimilation System (GLDAS) atmospheric analysis fields and precipitation estimates from the Climate Hazards Group InfraRed Precipitation with Stations (Funk et al., 2014) product. The LDAS also has the capability to assimilate terrestrial water storage anomaly estimates from GRACE and snow-covered area estimates from MODIS. The LDAS provides automated near real-time monitoring, and these monitoring products are disseminated through a Tethys app (**Figure 10**). The LDAS is also used to provide initial conditions for ensemble S2S forecasts. These ensemble forecasts draw meteorological forcing fields from the NASA GEOS forecast system (other forecasts may be added in the future), downscaled using the Generalized Analog and

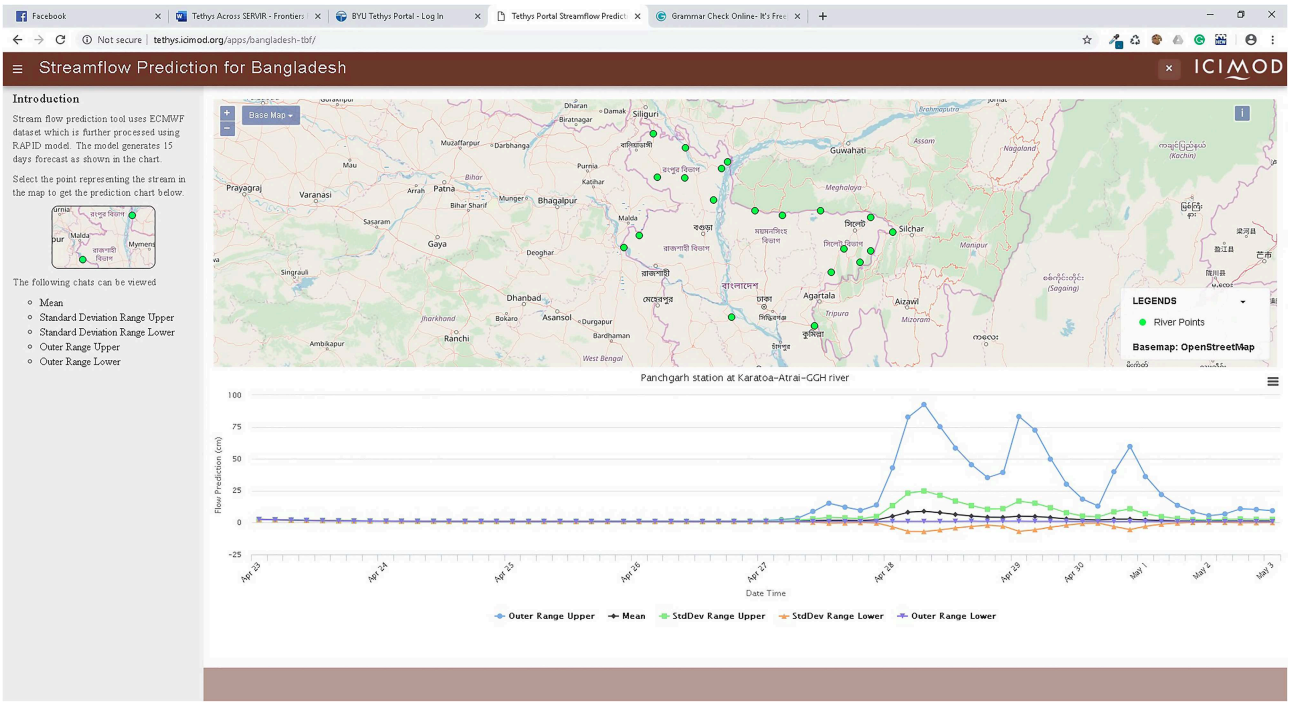


FIGURE 8 | Bangladesh transboundary flows Tethys application.

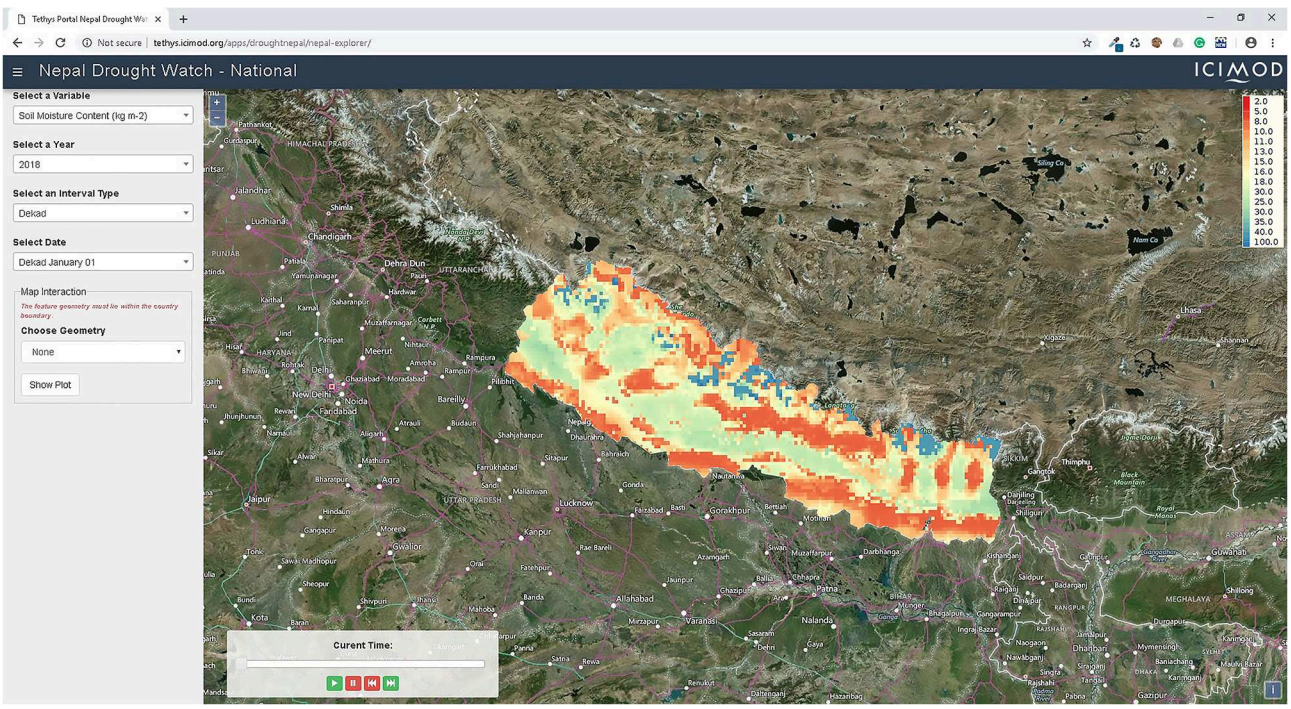


FIGURE 9 | Nepal drought watch—national Tethys application.

Regression Downscaling (GARD) tool (Gutmann et al., 2018). The resulting forecasts are used to generate probabilistic forecasts of drought indicators defined in consultation with end users throughout the region. These drought indicators can be accessed as maps and timeseries products via two different Tethys apps that have been created. The national drought monitoring

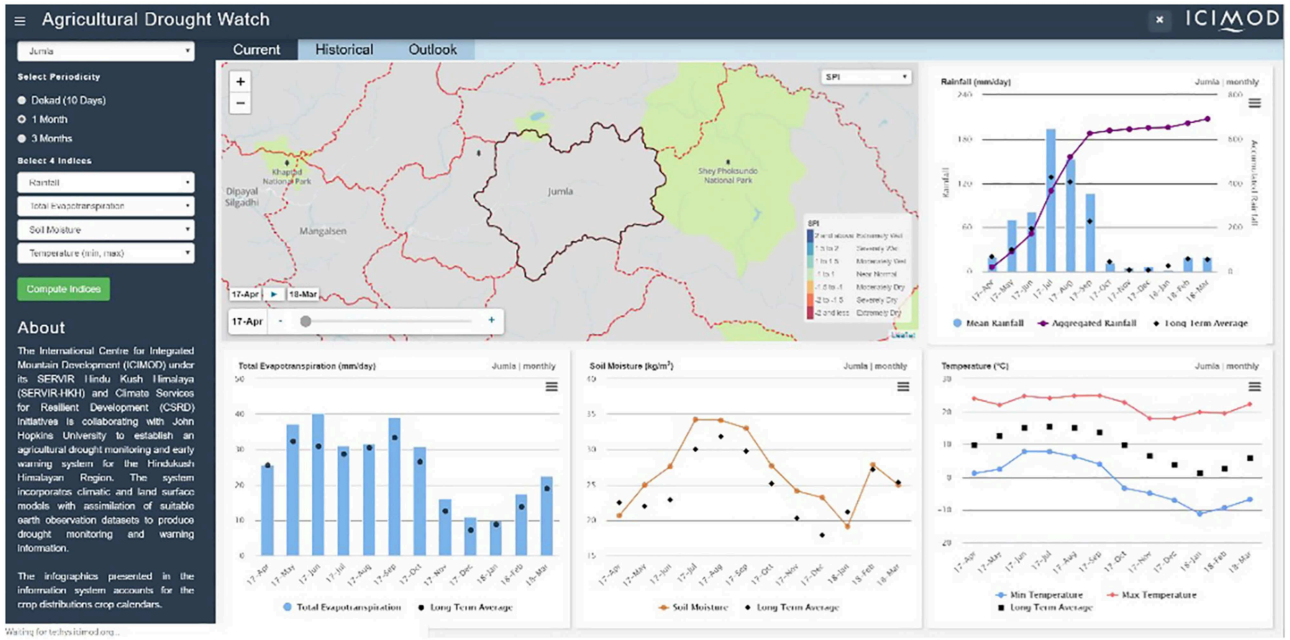


FIGURE 10 | The SERVIR-HKH agricultural drought watch Tethys app.

system uses five parameters (soil moisture, precipitation, air temperature, evapotranspiration, and standard precipitation index) from the SALDAS dataset as the drought indicators, which is further grouped in dekad (10 days), monthly and quarterly intervals, making it more relevant as a drought application. The system is highly interactive and can view a historical dataset in time-series, draw points or polygon and view the charts of the interested drought indices. During the SERVIR program, large amounts of climate data has been generated by scientific communities but the use of that data/information by decision makers at local and management levels remain low. In-order to fill the information gap, the drought monitoring system provides a good visualization platform for these drought parameters to meet the diverse needs of decision maker.

A second high-level agricultural drought watch Tethys app was created to aggregate data over Nepal Districts (i.e., states) and provide a series of plots with statistics that can be helpful in assessing risk and making plans for agricultural productivity.

Grace

The Gravity Recovery and Climate Experiment (GRACE) mission (Tapley et al., 2004), a collaborative effort between NASA and the German Aerospace Center (DLR), measured monthly anomalies in the Earth's gravity field from 2002 to 2017. Because water has a high density—1,000 kg m^{-3}–changes in terrestrial water storage can be inferred from the gravity anomalies measured by the two twin satellites of GRACE. Note that the total quantity of water integrated vertically at any given place of the Earth is not provided by GRACE, but rather its variations compared to a long-term mean. In other words, GRACE can tell us if there is more or less water than usual (i.e., Terrestrial Water Storage Anomalies: TWSA), but not how much water there actually is (i.e., Terrestrial Water Storage: TWS). GRACE has also proven to be a useful asset in estimating the otherwise unknown changes in groundwater storage. To do so, estimates of water storage anomalies—for snow, soil moisture, canopy interception, and surface water—are obtained from numerical models (e.g., GLDAS: Rodell et al., 2004 and removed from the TWSA estimates to extract Groundwater Storage Anomalies (GWSA). The value of such a methodology has been demonstrated in numerous studies (e.g., Rodell et al., 2009; Famiglietti et al., 2011; Castle et al., 2014) and is therefore mature for water management applications. Yet, the analysis involved in the combination of hydrological models and satellite observations is complex and has been a barrier for the broader use of GRACE-based estimates of Terrestrial Water Storage or Groundwater Storage Anomalies.

An open-source software specifically tailored for the joint analysis of GRACE and GLDAS data was therefore developed to ease the access to satellite-based water availability assessment. The software allows for the generation of on-the-fly estimates of storage anomalies for the various components of the hydrologic cycle and only requires the description of the study domain—in the form of a shapefile—as input. A specific Tethys Application was then developed to allow web-based servicing and exploration of the data (**Figure 11**). The app displays aggregate times series and raster animations of GRACE-derived water storage anomalies for selected subregions. It displays results derived using the JPL, CSR, and GFZ signal processing methods (Frappart et al., 2010, 2011). It also displays the surface water and soil moisture storage components obtained from the GLDAS model (Syed et al., 2008). The groundwater storage component

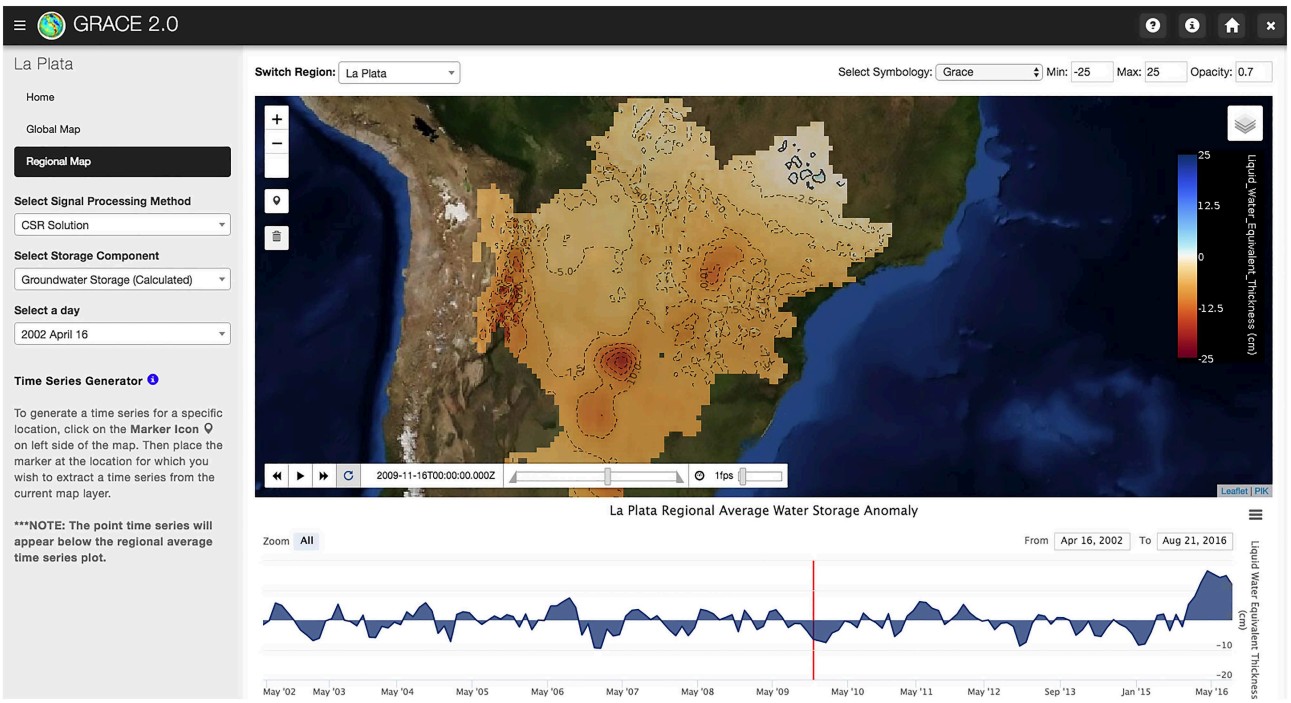


FIGURE 11 | GRACE Tethys app illustrating groundwater storage anomaly for the La Plata Basin in South America.

is found by subtracting the surface water and soil moisture components from the total water storage obtained from the GRACE data.

High Impact Weather Assessment Toolkit (HIWAT)

The High Impact Weather Assessment Toolkit (HIWAT) was designed for filling a gap in weather forecasting services across the HKH region. In particular, HIWAT provides high resolution, probabilistic-based forecasts of extreme weather hazards (e.g., damaging winds/tornadoes, hail, lightning and flash flooding). To do this, HIWAT utilizes the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008) in the framework of the Unified Environmental Modeling System (UEMS) (NWS, 2018), which is a collection of scripts that facilitate a relatively simple means for managing the workflow of numerical weather prediction (NWP). HIWAT consists of a 12-member ensemble WRF configuration that includes several combinations of four microphysical and three planetary boundary layer parameterizations relevant to thunderstorm initiation and evolution. To further capture the range of possible forecast solutions, each ensemble member uses a different initial and boundary condition provided by the NCEP Global Ensemble Forecast System (Zhou et al., 2017). Each HIWAT HKH member uses a 12-km resolution outer domain over South Asia to downscale the GEFS initial/boundary conditions to a 4-km resolution convection-allowing model domain over Nepal and Bangladesh.

A demonstration of HIWAT over the HKH region was conducted during the pre-monsoon and monsoon seasons of 2018. The objective was to assess the capability and utility of HIWAT for enhancing early warning services in the region. A Tethys application was developed to facilitate efficient forecast product interpretation (**Figure 12**). One of the most useful features of this application is its data interrogation capabilities, thereby enabling decision-makers to readily assess where, what time and the probability of thunderstorm-related hazards during the forecast period. This interactive app enables the user to select a point (or draw a polygon) anywhere on the map to interrogate the hourly forecasts data. Hence a decision-maker (e.g., operational weather forecaster) can quickly assess where strong storms can be expected during the first 24-h in the region from eastern Nepal, across northern Bangladesh to northeast India and central Bangladesh. Going a step further, the user can also quickly ascertain frequent lightning, hail and damaging winds can be expected in northern Bangladesh.

SERVIR SCIENCE COORDINATION

Given the low barrier to development as well as the replicability of current applications and the success amongst principal investigators assigned to the SERVIR-HKH hub, the SERVIR Science Coordination Office (SCO) has leveraged the Tethys platform to support existing applications, deployment of regional Tethys portals beyond the HKH, and application development for hubs and global services. The demand for web applications from other regions supporting SERVIR services has grown

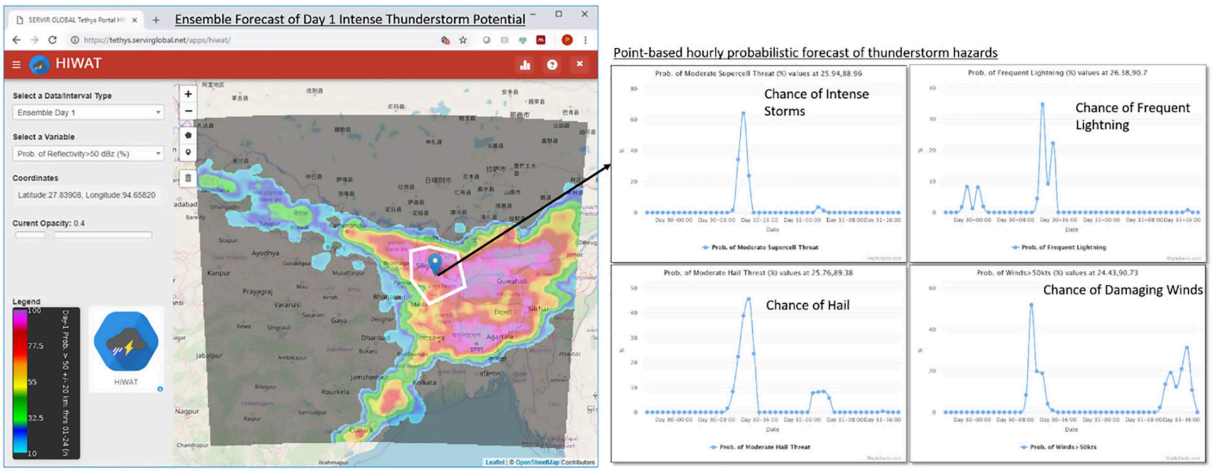


FIGURE 12 | Tethys-based application for HIWAT ensemble forecast weather hazards.

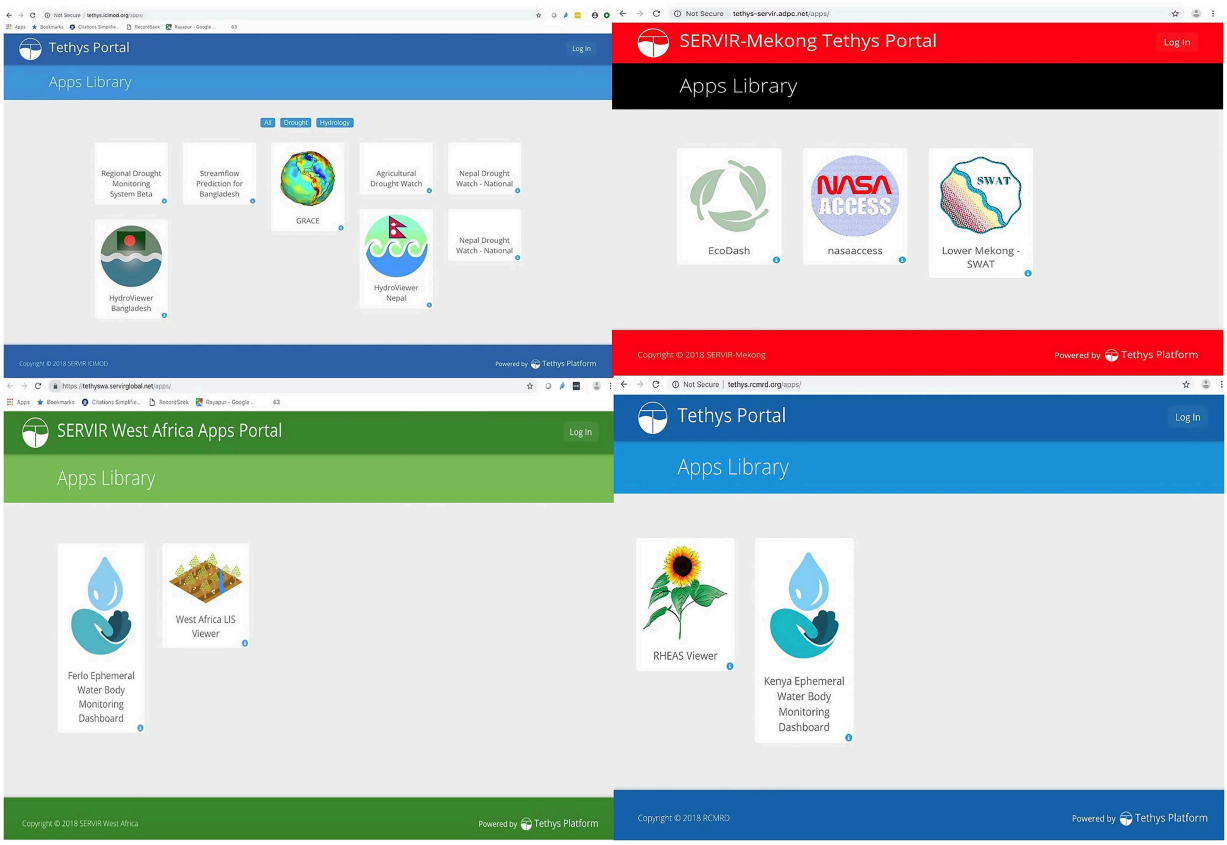


FIGURE 13 | Tethys portals for SERVIR regional hubs. (top left) ICIMOD http://tethys.icimod.org/apps (top right) ADPC http://tethys-servir.adpc.net (bottom left) AGRHYMET (bottom right) http://tethys.rcmrd.org.

and the Tethys application enables quick prototyping and deployment of the applications within the portal framework leading to additional Tethys portals in all of SERVIR's hubs (**Figure 13**).

In the case of East and Southern Africa, RCMRD is running a complex coupled land surface crop model within an assimilation framework call the Regional Hydrologic Extremes Assessment System (RHEAS) (Andreadis et al., 2017). RHEAS was developed

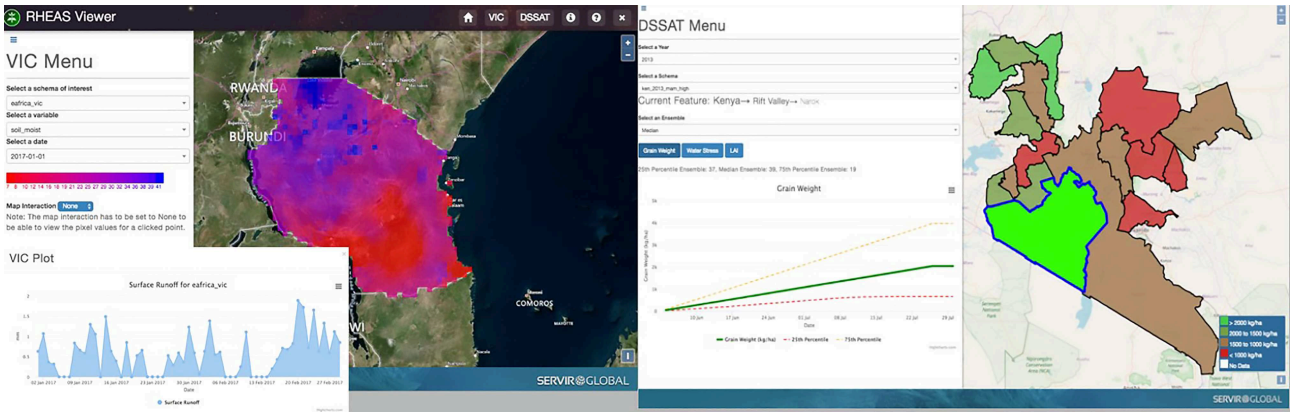


FIGURE 14 | Tethys-based application for the visualization of RHEAS outputs. The portal allows for the display and inspection of land surface outputs such as drought onset, standardized precipitation, or soil moisture **(Left)** and crop yield trends and water stress **(Right)**.

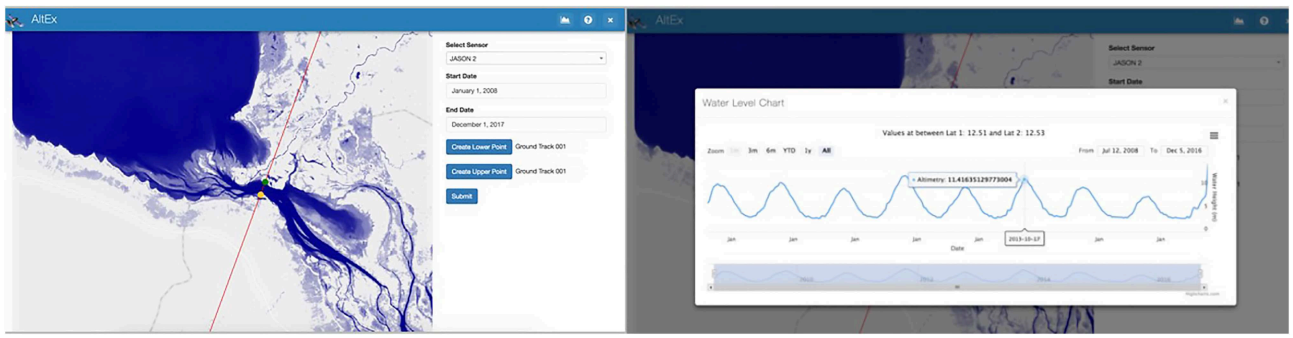


FIGURE 15 | AltEx web application user interface over Tonle Sap, Cambodia **(Left)** and water level chart as rendered in AltEx web application **(Right)**. Figure taken from Markert et al. (2019).

as part of an AST project and is used to support national ministries in East and Southern Africa in monitoring and forecasting drought and its impact on agriculture. The RHEAS system is run in a Linux environment in a command line interface, making it not ideal for end-users in the region. Tethys Platform is being used to efficiently access the RHEAS database (PostGIS) and visualize both the land surface components in a user-friendly geographical interface. Though still in development, **Figure 14** shows both components of the RHEAS output visualization. The portal allows the user to inspect the land surface outputs, generate, plot and download time series data. On the crop model page, the app generates the ensemble spread of crop yield by county for the season at hand. Another case that the SERVIR SCO has supported with the hubs is the reimplementation of the EcoDash (Poortinga et al., 2018) and the Mekong Suspended Sediment application into Tethys Platform.

As a global application provided by the SERVIR SCO to the hubs and the Earth science community, the Altimetry Explorer (AltEx) (Markert et al., 2019) was developed using Tethys to make global water levels data derived from altimetry data available to stakeholders. The AltEx application automatically extracts Jason-2 and Jason-3 data from data sources and allows users to dynamically select overpass segments to display water level data. **Figure 15** displays the AltEx user interface where a water layer is provided, users select a start and end point on the altimetry ground track that intersect with water, and the extracted water level time series is provided. This application is also API enabled allowing users programmatically query the application and use the returned water level data for other use. As future needs for global services and regional applications arise the SERVIR SCO will continue to support the development and use of Tethys within the SERVIR network.

CONCLUSIONS AND FUTURE DIRECTIONS

This paper presented and discussed a number of web applications developed by SERVIR AST members primarily using the open source Tethys Platform software development toolkit for accessing, visualizing, analyzing spatial and temporal earth observation data and providing stakeholder decision support. While many tools and technologies exist to support these types of applications, we found that the Tethys Platform did indeed lower many barriers to web app development allowing for team members with a variety of technical backgrounds to actively contribute to creating these apps. The developed tools meet our

definition of a "web app" as essentially single function, self-describing, "easy" to use, web-based applications presented in a unified portal. As such, these web apps can be considered significantly distinct from "web pages" inasmuch as these apps are extremely modular and can be installed and uninstalled in different Tethys Portals as needed to support the needs of individual user communities. Some of the apps developed and presented here could be readily deployed and made useful in any of the SERVIR Hubs, however, regionally specific apps, or apps that are focused problems not present in certain hubs, can be installed or installed in specific portals. In all more than 20 Tethys apps were created in support of the applied science developed by all four of the principal investigators of the HKH hub and deployed across five different SERVIR portals as well as the BYU portal.

Because much of this work presented here has only recently been completed, we recognize the lack of extensive user statistics, usability analysis, and other metrics indicating exactly how successful these apps are in supporting on the ground decision making. We anticipate addressing a part of this deficiency in the near future by deploying new versions of the Tethys Portals that integrate web-based analytics code to gather usage statistics and user information. We also anticipate interesting use and usability results to naturally emerge in the form of white papers, and technical reports from the SERVIR Hubs. The wide-ranging use of Tethys within SERVIR in such a short time is evidence of its promise within SERVIR and perhaps across similar applied sciences groups, however it is still early and feedback is needed to further improve its usability for SERVIR and other organizations that which to join the growing community of this open-source project.

In terms of future technical work in this area, we are actively pursuing the development of an "App Warehouse" which will function much like an app store for a mobile device, allowing user communities to readily launch an empty decision support data portal and, through a few clicks and server commands, identify and install apps from a library of Tethys Apps. This work is currently underway as part of the NASA Water Science Team GeoGLOWS project and is expected to yield operational technology within the timeframe of this project. We expect that this new effort will open more doors for both stakeholder communities and also app developers in much the same way that the availability of an app store for mobile devices has resulted in an explosion of mobile apps and app developers. The Tethys App Warehouse will likely host a relatively small number of apps (in comparison to the Apple App Store, for example) but it is expected that this approach can help motivate both the development, deployment, and adoption of earth observation data visualization and decision support tools throughout the developing and developed world.

SOFTWARE AVAILABILITY

Tethys Platform:

- Purpose: Integrated package of open source web app development tools and app portal environment targeted toward water and environmental science and engineering.
- Source Code: https://github.com/tethysplatform/tethys.
- Developers: Brigham Young University, independent code contributors, and the Tethys Platform Steering Committee.
- License: Open Source under the BSD 2-Clause "Simplified" License.
- Initially Release Date: 2015.
- Current Release Version: 2.1.

Additional Software Tools:

- All of the apps presented in this paper are provided through open source licenses on the Brigham Young University Hydroinformatics Research Lab github.com repository located at: https://github.com/BYU-Hydroinformatics.

AUTHOR CONTRIBUTIONS

EN led this project and coordinated the research and writing of this article. SP, MM, KS, NJ, DA, WE, KM, CD, BZ, PG, and RH developed the web applications presented herein, provided intellectual contributions, and contributed to the writing of the article.

FUNDING

This work was supported by the NASA ROSES SERVIR Applied Research Grant NNX16AN45G and from USAID under SERVIR-HKH. CD was supported by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA, including a grant from the SERVIR Applied Sciences Team. The views and interpretations in this paper are those of the authors and are not necessarily attributable to ICIMOD, USAID, NASA, or BYU.

REFERENCES

Adams, E., Flores, A., Muench, R., Coulter, D., Limaye, A., and Irwin, D. (2016). "Using Esri story map technology to demonstrate SERVIR global success stories," in *AGU Fall Meeting Abstracts* (San Francisco, CA).

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., et al. (2013). GloFAS–global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.* 17, 1161–1175. doi: 10.5194/hess-17-1161-2013

Ames, D. P., Horsburgh, J. S., Cao, Y., Kadlec, J., Whiteaker, T., and Valentine, D. (2012). HydroDesktop: web services-based software for hydrologic data discovery, download, visualization, and analysis. *Environ. Model. Softw.* 37, 146–156. doi: 10.1016/j.envsoft.2012.03.013

Ames, D. P., Jones, N., and Nelson, J. (2018). "Rapid web app development to enable earth science data visualization and analysis using Python and Tethys platform," in *AGU Fall Meeting Abstracts* (Washington, DC).

Andreadis, K. M., Das, N., Stampoulis, D., Ines, A., Fisher, J. B., and Granger, S. (2017). The regional hydrologic extremes assessment system: a software framework for hydrologic modeling and data assimilation. *PLoS ONE* 12:e0176506. doi: 10.1371/journal.pone.0176506

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., Van den Hurk, B., Hirschi, M., et al. (2009). A revised hydrology for the ECMWF model: Verification

from field site to terrestrial water storage and impact in the Integrated Forecast System. *J. Hydrometeorol.* 10, 623–643. doi: 10.1175/2008JHM1068.1

Brendel, C. E., Dymond, R. L., and Aguilar, M. F. (2019). An interactive web app for retrieval, visualization, and analysis of hydrologic and meteorological time series data. *Environ. Model. Softw.* 117, 14–28. doi: 10.1016/j.envsoft.2019.03.003

Castle, S. L., Thomas, B. F., Reager, J. T., Rodell, M., Swenson, S. C., and Famiglietti, J. S. (2014). Groundwater depletion during drought threatens future water security of the Colorado River Basin. *Geophys. Res. Lett.* 41, 5904–5911. doi: 10.1002/2014GL061055

da Costa, R. T., Manfreda, S., Luzzi, V., Samela, C., Mazzoli, P., Castellarin, A., et al. (2019). A web application for hydrogeomorphic flood hazard mapping. *Environ Model Softw.* 118, 172–186. doi: 10.1016/j.envsoft.2019.04.010

Dee, D. P., Uppalla, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 593–597. doi: 10.1002/qj.828

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K. J., and Syed, T. H. (2011). Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophys. Res. Lett.* 38, 1–4. doi: 10.1029/2010GL046442

Folk, M., Mcgrath, R. E., and Yeager, N. (1999). "HDF: an update and future directions," in *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293)* (Hamburg: IEEE), 273–275.

Frappart, F., Ramillien, G., Leblanc, M., Tweed, S. O., Bonnet, M.-P., and Maisongrande, P. (2011). An independent component analysis filtering approach for estimating continental hydrology in the GRACE gravity data. *Remote Sens. Environ.* 115, 187–204. doi: 10.1016/j.rse.2010.08.017

Frappart, F., Ramillien, G., Maisongrande, P., and Bonnet, M.-P. (2010). Denoising satellite gravity signals by independent component analysis. *IEEE Geosci. Remote Sens. Lett.* 7, 421–425. doi: 10.1109/LGRS.2009.2037837

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., et al. (2014). A quasi-global precipitation time series for drought monitoring. *US Geol. Surv. Data Series* 832, 1–12. doi: 10.3133/ds832

Ghatak, D., Zaitchik, B., Kumar, S., Matin, M., Bajracharya, B., Hain, C., et al. (2018). Influence of precipitation forcing uncertainty on hydrological simulations with the NASA South Asia land data assimilation system. *Hydrology* 5, 1–22. doi: 10.3390/hydrology5040057

González, A., Gleeson, J., and Mccarthy, E. (2019). Designing and developing a web tool to support strategic environmental assessment. *Environ. Model. Softw.* 111, 472–482. doi: 10.1016/j.envsoft.2018.10.014

Guha-Sapir, D. (2009). *EM-DAT: The Emergency Events Database [Online].* Available online at: https://www.emdat.be/, https://www.emdat.be/database. (accessed March 2019).

Gutmann, E., Hamman, J., Eidhammer, T., and Clark, M. (2018). "Evaluating downscaling methods," in *EGU General Assembly Conference Abstracts* (Vienna), 20.

Horsburgh, J. S., Aufdenkampe, A. K., Mayorga, E., Lehnert, K. A., Hsu, L., Song, L., et al. (2016). Observations data model 2: a community information model for spatially discrete earth observations. *Environ. Model. Softw.* 79, 55–74. doi: 10.1016/j.envsoft.2016.01.010

Jackson, E. K., Estep, E., Nelson, J., and Perez, F. (2018). An open source hydropower feasibility web application for the dominican republic. *Open Water J.* 5:2. Available online at: https://scholarsarchive.byu.edu/openwater/vol5/iss1/2

Kadlec, J., Stclair, B., Ames, D. P., and Gill, R. A. (2015). WaterML R package for managing ecological experiment data on a CUAHSI HydroServer. *Ecol. Informatics.* 28, 19–28. doi: 10.1016/j.ecoinf.2015.05.002

Maidment, D., Salas, F., Domenico, B., and Nativi, S. (2010). "Crossing the digital divide: connecting GIS, time series and space-time arrays," *AGU Fall Meeting Abstracts* (San Francisco, CA).

Markert, K. N., Pulla, S. T., Lee, H., Markert, A. M., Anderson, E. R., Okeowo, M. A., et al. (2019). AltEx: an open source web application and toolkit for accessing and exploring altimetry datasets. *Environ. Model. Softw.* 117, 164–175. doi: 10.1016/j.envsoft.2019.03.021

Markert, K. N., Pulla, S. T., Limaye, A. S., Lee, H., Saah, D. S., Ate, P., et al. (2018). "Towards democratization of water related data in developing regions using cloud computing and open source geospatial web applications," *AGU Fall Meeting Abstracts* (Washington, DC).

NWS (2018). *Unified Environmental Modeling System*. National Weather Service Science and Training Resource Center.

OGC (2012). *OGC® WaterML 2.0: Part 1- Timeseries*. in ed P. Taylor.

Poortinga, A., Clinton, N., Saah, D., Cutter, P., Chishtie, F., Markert, K. N., et al. (2018). An operational before-after-control-impact (BACI) designed platform for vegetation monitoring at planetary scale. *Remote Sens.* 10:760. doi: 10.3390/rs10050760

Rew, R., and Davis, G. (1990). NetCDF: an interface for scientific data access. *IEEE Comput. Graph. Appl.* 10, 76–82. doi: 10.1109/38.56302

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., et al. (2004). The global land data assimilation system. *Bull. Am. Meteorol. Soc.* 85, 381–394. doi: 10.1175/BAMS-85-3-381

Rodell, M., Velicogna, I., and Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature* 460, 999–1002. doi: 10.1038/nature08238

Salas, F., Boldrini, E., Maidment, D., Nativi, S., and Domenico, B. (2012). Crossing the digital divide: an interoperable solution for sharing time series and coverages in earth sciences. *Nat. Hazards Earth Syst. Sci.* 12, 3013–3029. doi: 10.5194/nhess-12-3013-2012

Shrestha, A. B., and Pradhan, N. S. (2015). "Strengthening flash flood risk management in the Hindu Kush Himalayas," in *Building Resilience to Water-Related Disasters in the Asia-Pacific Region of the 7th World Water Forum* (Kathmandu: ICIMOD).

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D., Duda, M. G., et al. (2008). *A Description of the Advanced Research WRF Version 3*. NCAR Technical Note NCAR/TN-475+STR.

Snow, A. (2015). *Downscaling ECMWF Global Runoff Datasets Using Geoprocessing Tools and RAPID to Produce Highresolution Stream Forecasts*. Provo, UT: Master of Science, Brigham Young University.

Sood, A., and Smakhtin, V. (2015). Global hydrological models: a review. *Hydrol. Sci. J.* 60, 549–565. doi: 10.1080/02626667.2014.950580

Souffront Alcantara, M. A. (2018). *A Flexible Service-Oriented Approach to Address Hydroinformatic Challenges in Large-Scale Hydrologic Predictions*. Ph.D. Dissertation, Brigham Young University, Provo, Utah.

Souffront Alcantara, M. A., Kesler, C., Stealey, M. J., Nelson, E. J., Ames, D. P., and Jones, N. L. (2018). Cyberinfrastructure and web apps for managing and disseminating the national water model. *J. Am. Water Resour. Assoc.* 54, 859–871. doi: 10.1111/1752-1688.12608

Strömbäck, L., Pers, C., Strömqvist, J., Lindström, G., and Gustavsson, J. (2019). A web based analysis and scenario tool for eutrophication of inland waters for Sweden and Europe. *Environ. Model. Softw.* 111, 259–267. doi: 10.1016/j.envsoft.2018.07.012

Swain, N. R., Christensen, S. D., Snow, A. D., Dolder, H., Espinoza-Dávalos, G., Goharian, E., et al. (2016). A new open source platform for lowering the barrier for environmental web app development. *Environ. Model. Soft.* 85, 11–26. doi: 10.1016/j.envsoft.2016.08.003

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., and Wilson, C. R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resour. Res.* 44, 1–15. doi: 10.1029/2006WR005779

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M. M. (2004). GRACE measurements of mass variability in the earth system. *Science* 305, 503–505. doi: 10.1126/science.1099192

Tarboton, D. G., Horsburgh, J. S., Maidment, D. R., Whiteaker, T., Zaslavsky, I., Piasecki, M., et al. (2009). "Development of a community hydrologic information system," in *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation* (Cairns, QLD), 988–994.

Zhou, X., Zhu, Y., Hou, D., Luo, Y., Peng, J., and Wobus, R. (2017). Performance of the New NCEP global ensemble forecast system in a parallel experiment. *Weather Forecast* 32, 1989–2004. doi: 10.1175/WAF-D-17-0023.1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 22 October 2019
doi: 10.3389/fenvs.2019.00165

Capacity Building Approach and Application: Utilization of Earth Observation Data and Geospatial Information Technology in the Hindu Kush Himalaya

Rajesh Bahadur Thapa, Mir A. Matin and Birendra Bajracharya*

International Centre for Integrated Mountain Development, Kathmandu, Nepal

While the Earth observation (EO) data and geospatial information technology (GIT) are getting more open and accessible, lack of skilled human resources and institutional capacities are limiting effective applications in the Hindu Kush Himalayan (HKH) region. This paper aims to present the capacity building approach and applications designed to fill these gaps and empower decision makers and practitioners in using EO data and GIT through information education and training. The capacity building approach consists of four components: assessment, design, implementation, and monitoring (ADIM). The assessment component focuses on identifying the needs and priorities of capacity building for targeted groups or institutions. The design component develops training content in order to execute the plan in coordination with subject matter experts (SME). The implementation component executes the capacity building activity in any of these four formats—standard training, training of trainers, on-the-job training, and exposure learning. The monitoring component helps to identify the participants' expectations, learning achievements, and feedback so as to improve future capacity building events. In the application of ADIM, we conducted needs assessment in four countries, designed 26 types of capacity building contents and implemented 39 capacity building events. A range of thematic topics—from agriculture and food security, water resources and hydro-climatic disasters, land use, land cover and ecosystem, weather and climate services, to crosscutting issues—were covered in the events. Altogether, the activities reached out to over 1,000 individuals (35% of them women) from over 200 unique institutions in 30 countries. Institutional capacity was built for universities in Afghanistan and Bangladesh to design and deliver courses independently. The capacity of partner agencies were built to co-design and co-develop data and applications. The approach also experienced challenges in the nomination process and in identifying women participants due to the lack of women professionals in the field and in the respective agencies. The ADIM approach and its workflow focused on bridging the gap between the current trend and progression in EO and GIT fields and the existing state of capacity of the agencies involved in the decision-making process. It promoted gender equity, adopted frontier technologies, engaged SMEs and provided sustainable solutions, which are starting to bring success stories in the region.

Keywords: capacity building, informal education, SME, SERVIR, GIT, HKH region

OPEN ACCESS

Edited by:
Daniel Eric Irwin,
Marshall Space Flight Center (NASA), United States

Reviewed by:
Ana Prados,
Community College of Baltimore County, United States
Lauren Michelle Childs-Gleason,
Science Systems and Applications, Inc., United States

****Correspondence:***
Rajesh Bahadur Thapa
rajesh.thapa@icimod.org;
thaparb@gmail.com

Specialty section:
This article was submitted to Land Use Dynamics, a section of the journal Frontiers in Environmental Science

Received: *14 June 2019*
Accepted: *03 October 2019*
Published: *22 October 2019*

Citation:
Thapa RB, Matin MA and Bajracharya B (2019) Capacity Building Approach and Application: Utilization of Earth Observation Data and Geospatial Information Technology in the Hindu Kush Himalaya. Front. Environ. Sci. 7:165. doi: 10.3389/fenvs.2019.00165

INTRODUCTION

The increasing trend of extreme weather events and climate uncertainty poses new challenges for governments, communities, and individuals around the world. The mountain ecosystems of the Hindu Kush Himalaya (HKH) region is more vulnerable to climate change (Sharma et al., 2019). The Himalayan ecosystem houses a population of around 250 million people in the mountains; it is geologically fragile and experiences climate change and rapid social transformations driven by globalization, human conflicts, migration, infrastructural development, urbanization, and tourism. The complex interaction of these drivers is imposing major environmental and social consequences on the region. To mitigate these consequences, regional governments, communities, and individuals must work together whereby they can design and implement measures to adapt to change, deal with loss and damage, set up credible measures to monitor various parameters, adopt new technologies and methods, and raise awareness on climate change. Like other developing countries, the countries in the HKH region also lack the human, technical, and institutional capacities to deal with many of these challenges (World Bank, 2005; GEO, 2006; Merino and Carmenado, 2012; ECBI, 2018). Thus, they require customized capacity building approach and implementation programs which enhance organizational and technical abilities, as well as relationships and values that enable countries, organizations, groups, and individuals to carry out relevant functions and achieve their development objectives (Morgan, 2006; Balcazar et al., 2008; Whittle et al., 2012; Chandler and Kennedy, 2015). The challenges are multifaceted as they influence actions at local, subnational, national, and regional scales. The need for capacity building actions is highlighted in Article 6 of United Nations Framework Convention on Climate Change (UNFCCC), Article 10 of the Kyoto Protocol and Article 11 of the Paris Agreement. As the importance of information derived from Earth observing satellites and geospatial technologies is being widely recognized for supporting decisions on climate change mitigation worldwide, in this study, we focused on capacity building in the area of data utilization from Earth observation and geospatial information technology (EO&GIT). This paper aims to present the capacity building approach and applications designed to empower decision makers and relevant practitioners by providing them with geospatial data, tools, products, and services in order to act locally on climate-sensitive issues such as disasters, agriculture, water, land use, and ecosystems in the HKH region.

The HKH region covers the highest mountain range in the world where difficult terrain and high degree of inaccessibility present a formidable challenge to collect and manage data and information (Bajracharya et al., 2015). EO data, in combination with emerging GIT, helps to acquire information from remote regions and provides viable solutions to bridge the important data and knowledge gaps in the HKH region. The SERVIR-HKH Initiative (http://servir.icimod.org) of the International Centre for Integrated Mountain Development (ICIMOD), one of the five hubs (implementing agencies) of SERVIR Global (https://www.servirglobal.net/), is working on bridging the capacity and data gaps in the region. SERVIR is a joint development initiative of the National Aeronautics and Space Administration (NASA) and the United States Agency for International Development (USAID). It provides support to build the capacity of governmental institutions and other development stakeholders whereby EO data and GIT can be incorporated into the decision-making process; whereby free and open information sharing can be promoted through national and regional platforms and collaborations; whereby innovative, user-tailored analyses, decision-support services, and trainings that advance scientific understanding can be developed; and thus information can be delivered to those who need it (SERVIR Global, 2018).

Capacity building is an integral part of the SERVIR program globally as well as of the SERVIR-HKH Initiative for the HKH region. Capacity building for SERVIR-HKH refers to a process wherein the necessary skills and abilities are developed for using EO data and GIT in order to make and execute decisions that enable effective, efficient, and sustainable results. The initiative organizes training programs in the development and use of EO and GIT applications for partner institutions and end users to maximize the benefit of GIT in the region. The initiative provides technical support and conducts customized trainings, and also shares opportunities as per the needs of its regional partners. SERVIR-HKH strengthens the capacity of ICIMOD to forge better partnerships with the national institutions in the HKH. Capacity building efforts under SERVIR-HKH focus primarily on four thematic areas: Agriculture and Food Security; Water Resources and Hydro-climatic Disasters; Land Cover, Land Use and Ecosystems; and Weather and Climate Services. Capacity building is at the center of this whole initiative that supports institutions, communities and individuals in Afghanistan, Bangladesh, Myanmar, Nepal, and Pakistan (SERVIR-HKH, 2016). In an effort to meet the capacity building demands from these countries, we have designed and implemented an efficient capacity building approach. Details about the approach, implementation results, and the lessons learnt are discussed in the following sections.

OVERVIEW OF CAPACITY BUILDING GAPS AND NEEDS IN THE REGION

While developing successful capacity building in EO&GIT applications, we need to focus on the requirements of the national institutions and the end users of the system. The essence of implementing successful EO applications lies in the capacities of the implementing agencies and end users to maintain and utilize these applications to support the intended decision-making process. In the HKH region, a number of country consultation workshops were carried out in Afghanistan, Bangladesh, Nepal, and Pakistan to identify the national priorities, requirements, and gaps in terms of data, information products and the capacity of the institutions to utilize EO&GIT in supporting decision-making processes in the respective countries (ICIMOD, 2015, 2016a,b,c,d).

Data and Knowledge Gaps

The assessment of the institutions across the countries in the region showed varying degrees of geospatial capacities in

terms of GIT infrastructure and human resources. Some gaps in geospatial capacities, including in the areas of data, and physical and human resources, were observed. For instance, most of the institutions in Bangladesh indicated that they needed hydro-climatic data on a daily and weekly basis in order to develop early warning forecasting products (e.g., for riverine floods, flash floods, and abnormal water surge in the coastal areas) and also to predict related vulnerabilities such as riverbank erosion, crop loss, and landslide hazards (ICIMOD, 2016a). Institutions in Nepal mentioned that they needed data based on elevation as there are big variations in terms of land and climate variables (ICIMOD, 2016b,c). Afghanistan's institutions mentioned that their agencies relied on annual data to produce geospatial products for irrigation planning and also for assessing the impacts of hazards such as drought, floods and landslides (ICIMOD, 2015). In some places, data was required on a daily basis, in some others it was seasonal, while in yet others, it was annual—so as to perform various kinds of spatial analysis and to produce final map products and services. There were also knowledge gaps in terms of EO&GIT applications at decision- and policy making levels. In the case of some countries, prominently Afghanistan, there was insufficient coordination between the universities and the government agencies.

Lack of Human Resources and Gender Equality

Many institutions in the region are lacking in strategy to sustain GIT projects as the majority of the geospatial applications are project based, thereby creating challenges in maintaining and upgrading GIT infrastructures, as well as in updating geospatial data and strengthening human resources after the completion of the project. Government agencies across the region have been finding it difficult to retain qualified professionals for a longer term in the field of EO&GIT. Besides, there is a large gender gap in capacity building as indicated by the educational attainment indicator of the region, where Bangladesh, Nepal and Pakistan are ranked 116th, 123rd, and 139th, respectively, among 149 countries (World Economic Forum, 2018). This is also reflected in the gender imbalance in the GIT workforce in the HKH region. Despite the vital role that women play in environmental management and decision-making, their involvement and participation in GIT is rather low. While the crucial role played by female professionals in geospatial applications is recognized, their numbers are still less in the institutions. Thus, capacity building and gender equality have been acknowledged as twin challenges in building pathways to sustainable development (ICIMOD, 2017). Capacity building that promotes regional cooperation and knowledge exchange via organizing trainings and workshops with the collaboration of the regional member countries (RMCs) will pave the way for the next generation to see more women in the sphere of science and technology, thereby bridging the gender gap.

Lack of Institutional and Technical Capacity

The provision for sharing geospatial data among the institutions in the region is poor. For instance, while Pakistan has fairly good capacities and infrastructure for using geospatial technologies, there are no platforms available for sharing data and products beyond single intuitional use (ICIMOD, 2016d). There is also a need to build geospatial capacities (infrastructural as well as in terms of human resources) in assessing the impacts of hydro-climatic disasters (e.g., flood-affected area mapping in Bangladesh, landslide susceptibility in Nepal, change detection in irrigated agriculture in Afghanistan) and in setting up early warning and forecasting systems in Afghanistan, Bangladesh, Nepal, and Pakistan. Furthermore, there is a lack of institutional capacity in the educational arena to institute a formal degree in geospatial technology. Presently, geospatial technology is being taught in the region through modules as part of other degree programs like in geography, water resources, and urban planning. Moreover, the curriculum for these modules are not up to date, and hence not in tune with the advancement in the field; the lack of trained faculty is also a limitation in this area.

In order to fulfill all the gaps and needs systematically and effectively, operational capacity building approach with verified applications on the use EO&GIT is necessary to strengthen the capacity of government institutions and other stakeholders in the region.

MATERIALS AND METHODS

Keeping in view the diverse capacities of the institutions in the region, the gaps need to be bridged in a coordinated manner through customized training programs, institutional strengthening, coordination, and regional cooperation. The capacity building process must develop the skills and ability of the users of EO&GIT so that they can make and execute decisions that ensure effective and efficient results. There are many ways in which the capacity of the users can be built and they have been evolving over the years (Potter and Brough, 2004; World Bank, 2005; GEO, 2006; Balcazar et al., 2008; Wignaraja, 2009; Merino and Carmenado, 2012; Whittle et al., 2012; Chandler and Kennedy, 2015; Bergeron et al., 2017; Raynor et al., 2018). Here, we have opted for a strategy based on the lessons that have been learnt; so we have developed a sequential capacity building workflow (**Figure 1**) with four major tasks: Assessment, Design, Implementation and Monitoring—ADIM for short.

Assessment

The capacity assessment of the targeted institutions is very important to identify gaps, requirements, and priority needs for delivering effective capacity building programs. In this regard, four country consultation workshops (**Table 1**) were organized in 2015–2016 where gaps, needs, priorities, and capabilities of organizations, communities and individuals were systematically assessed. The need assessments were focused on the four thematic areas of SERVIR: Agriculture and Food

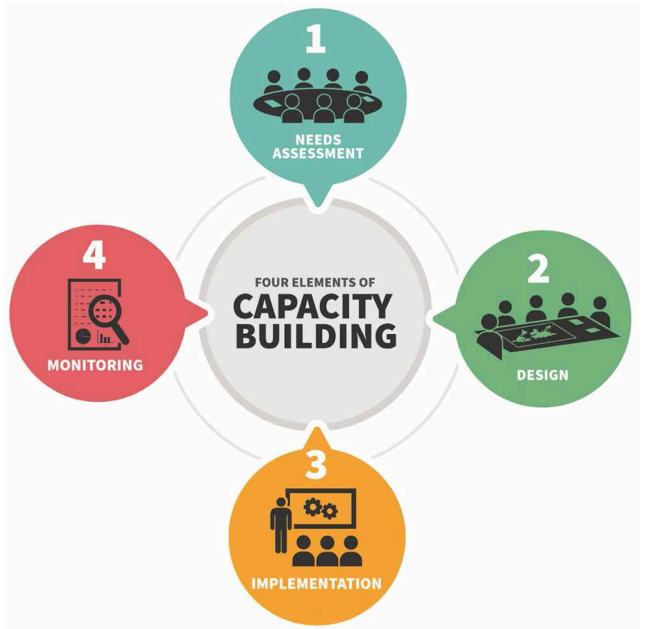


FIGURE 1 | ADIM—capacity building workflow.

TABLE 1 | Country consultation workshops in four countries in 2015–2016.

Country	Location	No. of institutions	No. of participants	Date
Afghanistan	Kabul	13	27	14–15 December 2015
Bangladesh	Dhaka	20	67	26 January 2016
Nepal	Kathmandu	12	39	10 February 2016 29 March 2016
Pakistan	Islamabad	23	37	23 February 2016

Participants from the relevant government organizations, universities, and research institutes, bilateral and multilateral organizations attended these workshops.

Security (AFS); Water Resources and Hydro-climatic Disasters (WRHD); Land Use, Land Cover and Ecosystems (LULC&E); and Weather and Climate Services (WCS). The assessment process helped us to gain an understanding about the necessary EO&GIT products and services, physical capabilities (hardware and software), and the available human resources in the targeted organizations of the selected countries. The following specific activities were completed during the consultation process in each country:

- Introduced the overall framework and objectives of SERVIR-HKH to the relevant partner agencies;
- Assessed the existing context in the country in terms of the institutional capacity and expertise in the application of EO&GIT in various climate services; and studied the status of data availability and data access and sharing mechanisms;
- Identified the priorities for specific applications, products and decision support tools in the relevant thematic areas; and
- Explored potential collaboration with institutions to engage in the implementation of relevant capacity building activities.

The country consultation workshops helped to promote active user engagement in strengthening the on-going partnerships, building new ones and enhancing more engagements within the SERVIR-HKH team and with the users and other stakeholders. The partnership approach framework and principles were used in the workshops as one of the tools for fostering more engagements internally and externally, and for building awareness about capacity building among the SERVIR-HKH team and other stakeholders. With the feedback mechanism in place to know about the users'/partners' requirements in terms of specific applications, capacity building needs/demands were identified. This crucial information served as inputs to start on the Design workflow.

Design

After understanding the requirements from the assessment process, we, along with the SMEs, started designing the curriculum based on background notes, learning objectives, expected outcomes, targeted audience, and the daily agenda. The necessary materials (theoretical and hands-on exercises), including PPTs, manual, and exercise data, were prepared in modular approach with real-world examples for each activity. The training program devoted most of its time (60–80%) to hands-on activities. Here, four types of capacity building activities were designed: on-the-job training (OJT), standard training, training of trainers (ToT), and exposure learning.

OJT activities are semi-structured, focusing on building the capacity of partner institutes to enhance their capacity to develop, operate and maintain specific applications and services. They are designed keeping in view the background and skills of the participants. In OJT, the participants worked with the SMEs at ICIMOD on a rotational basis and learnt specific tasks objectively. After a certain period of guided training (1 or 2 weeks), independent assignments related to specific applications were handed out which were to be finished within a given period of time, along with the submission of a report. As for "standard trainings," they are curriculum based designed for specific purposes; some are general remote sensing and GIS trainings while some are application oriented toward a specific subject or cross-thematic areas. Professionals from various backgrounds participate in these types of training. In addition, internships are also accommodated. In the case of ToTs, they are aimed to build the capacity of institutes to deliver training at their respective institutes and countries so as to reach out to more participants. These are designed for teaching staff from academic institutions and focus not only on content but also on skills related to the delivery of training (e.g., communication and presentation). After receiving such training, it is expected that the participants will be able to transfer knowledge to wider audiences in their respective institutions. In ToT, the participants are also engaged

in refining training materials, for instance, in the development of hands-on exercises with local data and case studies. A ToT program is different from a standard training program as it includes additional contents in the form of adult training, scientific communication, and monitoring and evaluation; also, in this specific project, back-end support were rendered to the participants of ToT while they were conducting trainings in their home institutions. As regards "exposure learning," it is of a short term and designed to make the participants aware of current development, applications, benefits, and future prospects of EO&GIT. And these activities were often organized alongside professional conferences, workshops, competitions and technical exchanges. In exposure learning, the competitions were meant for the youth; the conferences and workshops for decision- and policymaking professionals; and the technical exchanges for high-tech professionals.

While OJT and ToT were considered as institutional capacity building activities, the other two were considered as individual capacity building activities. The institutional capacity building activities were very specific and focused on the requirements of particular institutions; while the individual capacity building activities were more general in nature, linking thematic applications with common topics that cater to a wider audience. In this regard, we worked closely with the thematic leads and developed a training definition document (TDD) for each capacity building activity. The TDD provides a comprehensive capacity building activity plan that covers baseline information, requirements, participants, training schedule, gender considerations, activity inputs and the expected outputs from particular products, and services. In addition, in the case of some activities, we opted for a co-development module where multiple stakeholders such as in-house professionals, SMEs (Flores-Anderson et al., 2019) and user partners developed curriculum and training materials, and shared the resources. The use of open source GIS/RS software/platforms such as Quantum GIS, SNAP, and Google Earth Engine were encouraged in these activities.

Implementation

This refers to the execution of a capacity building activity. To implement the activities, we categorized participants into three groups: policy/decision makers; technical professionals; and youth. The makers of policies and decisions form a very crucial set of target audience who can greatly influence the successful implementation of any EO&GIT-related services in partner institutions. Therefore, to bring in the participation of the national institutions in utilizing the applications, it is important to make the leadership aware and convinced about these technologies and their potential applications. So, exposure learning events were organized for this category of participants. The technical professionals are those who are responsible for the development of databases and EO applications in different thematic areas; they run the applications and derive the products which can be used by the decision makers and general users. So, the OJT, ToT and standard training activities were offered to this category. As for the youth, their role as agents of change and effective communicators has been especially recognized in view of the emergence of numerous youth networks working on climate change. And as EO&GIT is a vital tool that improves our understanding about the climate change phenomenon, we organized exposure learning and standard training activities in this area, specifically targeting youth or early career professionals. Such as NASA SpaceApps challenge, miss technology, internship, selected training for young women (ICIMOD, 2018a) were organized for this group. Importantly, maintaining gender balance among the participants was given high priority. Strategically, we tried to ensure that at least 30% women took part in each capacity building activity.

Furthermore, the capacity building activities were organized at both national and regional levels. The national-level training courses were conducted jointly by ICIMOD and a local institute in the partner country. The selection of participants was based on the nomination process wherein we had invited nominations from ICIMOD's partner institutes in the region relevant to the training theme; but when the trainings had broader goals and required wider participation, the field was opened out. OJT, ToT, and exposure learning activities were based on nomination; while either nomination or open call or both were applied in the case of standard training activity. Sometimes, individual requests for internship were also considered in standard trainings. Based on the activity theme, basic eligibility criteria for participation were also set. And by using the basic principles of communication, we were able to ensure that the relevant training opportunities reached the targeted group of people; gender-sensitive language was used to encourage the maximum participation of women; we reached out to the disadvantaged groups too; and also conveyed our objectives to communities of practice and social media channels in order to gain a larger audience.

Monitoring

Monitoring and evaluation is an important part of any capacity building event. It helps to improve performance and achieve better results. Monitoring and evaluation has to do with the quality and relevance of capacity building efforts. Structured pre- and post-assessment tools were developed and used for each type of capacity building activity, except for the exposure learning type, in order to monitor a participant's expectations, his or her progress in understanding the contents, and to get feedback for continuous improvement. We conducted pre- and post-training evaluation surveys for each capacity building event so that the cognitive growth of the participants could be monitored and documented. A representative pre- and post-assessment results after the monitoring and evaluation practice from one of the capacity building activities is reported.

In this research, we also compiled and used data for results analysis from the country consultation reports (ICIMOD, 2015, 2016a,b,c,d) and the capacity building events data during October 2015–September 2018 (the first 3 years of SERVIR-HKH Phase II).

RESULTS AND DISCUSSIONS

Capacity Building Needs and Priorities

During the country consultation meetings, several capacity building needs and priorities were identified and grouped into

thematic areas (**Table 2**). Most of the needs and priorities—such as climate data analysis for drought monitoring in the AFS thematic area; strengthening capacity on regional flood outlook in the WRHD thematic area (SERVIR-HKH, 2016); and the South Asia Land Data Assimilation System (SALDAS) in the WCS thematic area—were common for all countries. In the case of three countries—Afghanistan, Bangladesh and Nepal—capacity building in land cover mapping and monitoring system, and hydrological modeling were in the priority list. We could also observe common needs and priorities in the case of Bangladesh and Nepal. However, many country-specific unique needs and priorities—such as WebGIS, water resource management, and wheat area mapping for Afghanistan; and REDD and food security information analysis for Nepal—were also identified.

Furthermore, the institutions in Bangladesh and Afghanistan insisted that more OJT programs for government institutions were required for the effective implementation of EO&GIT tools in their planned activities. Afghanistan stated that universities and training institutes needed to include EO&GIT in their curriculum in order to ensure the sustainability of capacity building activities; while Bangladesh required more trainings for university faculty so that they could replicate and disseminate capacity building activities in the longer run. In addition, during our assessments, we identified common needs and priorities of all countries in crosscutting areas such as, cloud computing, empowering women and youth and policymakers in the EO&GIT field, and enhancing the capacity to adopt emerging technologies.

Capacity Building Activity Design

Based on the results from the country consultation exercises, we designed 26 capacity building modules covering various topics from the thematic areas (**Table 3**). These modules were used to deliver OJT, standard training, ToT and exposure learning type of activities multiple times in the last 3 years. The AFS theme has five modules—two in OJT and the rest in standard training type of capacity building formats. Most of the modules consist of a daily training activity plan and associated materials in the form of PPT, alongside theoretical discussions and hands-on tutorials. In addition, the training module for climate data analysis for drought monitoring consists of a manual. Similarly, we designed six training activity modules for the WRHD theme. All these modules were used in the standard training program. Since the expected outcomes from OJT and ToT were different from the standard training program, the modules on "mapping and monitoring of glaciers" and "Remote Sensing and GIS for water resources management" were slightly modified to fit, respectively, into the OJT and ToT activities. All modules in this theme consist of PPT and hands-on tutorials to cover both theory and practice in the field. Also, detailed training manuals are available for three of the modules—hydrological modeling; mapping and monitoring of glaciers; and remote sensing and GIS for water resource management. Like the WRHD theme, six training modules were designed for the LULC&E theme, where five of them were delivered in the standard training format. A course on tree cover estimation was also developed for OJT activity. One of the five standard training modules, i.e., on land cover and land use mapping, was slightly modified to be used for

TABLE 2 | Capacity building needs and priorities in the four countries.

Capacity building needs by thematic areas	Countries
Agriculture and food security (AFS)	
Food security information analysis	Nepal
Climate data analysis for drought monitoring	Afghanistan, Bangladesh, Nepal, Pakistan
WebGIS application development; irrigation information management; wheat area mapping	Afghanistan
Application of synthetic aperture radar data and cloud computing technology	Afghanistan, Pakistan
Water resources and hydro-climatic disasters (WRHD)	
Strengthening capacity on regional flood outlook	Afghanistan, Bangladesh, Nepal, Pakistan
Hydrological modeling	Afghanistan, Bangladesh, Nepal
Water resources management; mapping, and monitoring of glaciers, glacial lakes and snow cover	Afghanistan
Enhancing capacity in stream flow forecasting tools and flood early warning system	Bangladesh, Nepal
Land use, land cover, and ecosystems (LULC&E)	
Land cover mapping and monitoring system	Afghanistan, Bangladesh, Nepal
Enhancing capacity to understand ecosystem services; application of synthetic aperture radar for forest monitoring; tree cover estimation	Bangladesh, Nepal
Reducing emissions from deforestation and forest degradation (REDD)	Nepal
Weather and climate services (WCS)	
Implementation of South Asia Land Data Assimilation System (SALDAS)	Afghanistan, Bangladesh, Nepal, Pakistan
Crosscutting*	
Cloud computing; empowerment of women, youth, decision-/policymaking professionals; GIS and remote sensing applications development; topographic applications; and adaptation of emerging EO and GIT	Afghanistan, Bangladesh, Nepal, Pakistan

**Indicates inclusion of multiple thematic areas.*

ToT activity. All modules consist of PPT and tutorials. Besides, a manual was developed for land cover and land use mapping. The Synthetic Aperture Radar (SAR) for the forest monitoring module was covered by a part of the SAR Handbook (Flores-Anderson et al., 2019), a collaborative effort of SERVIR networks and partners worldwide.

The WCS thematic area consisted of only one course module as this theme has only one service (SALDAS). The module was in the standard training format and supported by PPT and hands-on tutorial materials. Among the total number of modules, eight are in the crosscutting areas. These modules deal with cross-disciplinary topics and are linked to more than one thematic area. Five modules were designed for the standard training format, while the Google Earth Engine module was extended to ToT activity with slight modification. The GIS application development module was designed for OJT activity. Under the crosscutting areas, the exposure learning visit type of capacity building activity was divided into two parts: capacity building conferences, i.e., Geospatial World Forum and Global Forest Observations Initiative; and competition events, i.e., NASA SpaceApp and "Miss Technology." We closely collaborated with the event organizers for developing an agenda and supporting the activities wherein senior policymakers and technical professionals would get exposure learning on EO&GIT in the conferences and the youth would get exposure learning during the competitions. In the exposure learning type of activities, only PPT materials of corresponding conferences were available while the R&D prototypes developed through competitions were also available.

TABLE 3 | Designing capacity building content, type, materials during SERVIR years (10/2015–09/2018).

Thematic areas	Modules	CB type	Materials
AFS	5	OJT, ST	PPT, manual, hands-on tutorials
WRHD	6	OJT, ST, ToT	PPT, manual, hands-on tutorials
LULC&E	6	OJT, ST, ToT	PPT, book, manual, hands-on tutorials
WCS	1	ST	PPT, hands-on tutorials
Crosscutting	8	OJT, ST, ToT, EL	PPT, hands-on tutorials, R&D prototype

AFS, Agriculture and Food Security; WRHD, Water Resources and Hydro-climatic Disasters; LULC&E, Land Use, Land Cover and Ecosystems; WCS, Weather and Climate Services; Crosscutting indicates inclusion of multiple themes; ST, Standard Training; OJT, On-the-Job Training; ToT, Training of Trainers; EL, Exposure Learning.

TABLE 4 | Capacity building activities by thematic area during October 2015–September 2018.

Capacity building			Participants		
Thematic areas	Event	Institution	Total	Male (%)	Female (%)
AFS	5	30	68	83.82	16.18
WRHD	12	32	264	80.30	19.70
LULC&E	6	49	179	75.42	24.58
WCS	2	4	14	64.29	35.71
Crosscutting	14	158	476	48.95	51.05
Total	39	213	1,001	64.54	35.46

Capacity Building Activity Implementation

Using the course modules and materials, we implemented 39 capacity building events in the last 3 years (**Table 4**). These events served 1,001 people—a combination of technical professionals, policy/decision makers and youth—from 213 unique institutions in the HKH region and beyond. The participating individuals were from government ministries, departments and local offices, as well as from INGOs, NGOs, academic institutions, research organizations, and the private sector.

Most of the events (14) were crosscutting in nature and served half of the total institutions. Twelve events were organized within the WRHD theme where 264 participants from 32 partner institutions were engaged in various types of capacity building activities. The second major thematic area was LULC&E under which six events were organized to serve 49 institutions. Five capacity building events were organized under the AFS thematic area to build the capacity of 68 people from 30 institutions. Two capacity building events were organized under the WCS thematic area to train 14 professionals from four partner institutions.

In terms of gender, 35% of those who participated in the capacity building activities were women. This shows progress toward the goal of the Fourth Mid-Term Action Plan (MTAP) of ICIMOD which proposed to ensure 30% participation of women in event organizations (ICIMOD, 2017). While gender balance was observed in the crosscutting areas where female participants exceeded the number of male ones, more than 75% of the participants in the other thematic areas, such as AFS, WRHD, and LULC&E, were male. In this regard, we shall continue to strive toward ensuring gender balance in all the events by requesting partner institutions to make gender-balanced nominations, and by providing extra seats for women, implementing women-only programs and by establishing childcare facilities on campus. The participants in these capacity building events under the four thematic areas had been chosen based on the institutional nomination process where selection decisions were completely made by partner institutions.

As many as 28 out of the 39 capacity building events were in the standard training format (**Table 5**). This type of training alone served more than 778 participants, including 34% women, from 188 partner institutions. Standard training type of capacity building activities are often more general in content and serve various purposes of the EO&GIT applications. Therefore, it can accommodate greater number of participants and institutions. Exposure learning was the second major activity in the capacity building program. Through its capacity building-related competitions and conferences, it engaged nearly 200 people from 42 institutions. Here, as compared to other types of activities, a smaller gender gap was observed.

There were only two ToTs and four OJTs type of capacity building activities in the last 3 years. Since these two types of activities were very specific to particular institutions on

TABLE 5 | Different types of capacity building activities during October 2015–September 2018.

Capacity building			Participants		
Type	**Event**	**Institution**	**Total**	**Male (%)**	**Female (%)**
Standard training	28	188	778	65.94	34.06
On-the-job training	4	6	16	87.50	12.50
Training of trainers	2	4	12	83.33	16.67
Exposure learning	5	42	195	55.90	44.10

TABLE 6 | Country participation in capacity building events during October 2015–September 2018.

Capacity building			Participants		
Country	**Event**	**Institution**	**Total**	**Male (%)**	**Female (%)**
Afghanistan	18	22	274	87.23	12.77
Bangladesh	12	31	91	73.63	26.37
Bhutan	8	6	14	78.57	21.43
China	2	3	5	60.00	40.00
India	8	10	17	64.71	35.29
Myanmar	7	18	42	57.14	42.86
Nepal	26	94	473	49.68	50.32
Pakistan	10	13	17	70.59	29.41
Others	11	33	68	64.71	35.29

Others include: Australia, Brazil, Canada, Germany, Finland, France, Gabon, United Kingdom, Guatemala, Italy, Kenya, Cambodia, Mongolia, Malawi, Mozambique, Netherlands, Papua New Guinea, Thailand, United States and Vietnam.

EO&GIT applications, they served a lesser number of people and institutions. However, these activities were found to have an effective impact on the institutional capacity building of partner organizations. To ensure sustainability, we collaborated with our partner agencies for targeted capacity building on specific applications in order to set up self-managing information systems that could meet organizational needs. This was achieved through a combination of ToT and OJT. The ToT program helped to attain multiplicity in our training courses by developing trainers who could conduct courses independently and transfer the knowledge to a wider audience such as university students. We also collaborated with universities in the HKH region for the ToT program. We have successfully trained the faculty of Kabul University and Jahangirnagar University, respectively, in EO&GIT applications for water resource management and big data analysis using GEE. Thus, ToT and OJT types of capacity building activities are playing instrumental roles in institutional capacity building in the region. Through this process, Kabul University has independently organized training courses for its students and updated its curriculum on EO and GIS courses. Meanwhile, Jahangirnagar University has established an institute for remote sensing and GIS, while independent training programs were conducted in Bangladesh by trainers who had received ToT from SERVIR-HKH.

Although the SERVIR-HKH programs focused on Afghanistan, Bangladesh, Myanmar, Nepal, and Pakistan, they also provided capacity building services to other regional member countries of ICIMOD (Bhutan, China, and India) and beyond (**Table 6**). Among these countries, Nepal and Afghanistan were on the top of the list in terms of participation and events. Along with specific events organized for Nepali women and youth, many capacity building events were held in Nepal which eventually provided more opportunities to the professionals working in that country; thus, Nepal found itself on the top of the list. In the 26 out of the 39 events that were organized, we served 473 participants from 94 institutions in Nepal. As part of our special focus on EO&GIT capacity building of our Afghanistan partners, 274 participants from 22 institutions participated in 18 capacity building events. In the case of Bangladesh, the capacity building activities provided opportunities to a relatively larger number of institutions (31) where 91 professionals participated in 12 different types of activities. Among the countries within the ambit of SERVIR, Pakistan registered the lowest number of participants—only 17 professionals from 13 institutions attended the 10 capacity building events. This may be due to the lesser number of Pakistan-focused services in SERVIR and delay in product development. However, there is a push toward implementing more capacity building programs in Pakistan.

A significant number of institutions from Bhutan (6), India (10), Myanmar (18), and China (3) participated in various types of capacity building events. We also served some 33 institutions from 20 countries outside the HKH region on EO&GIT capacity building—some of them were Thailand, Cambodia, Vietnam, Mongolia, Malawi, and Kenya. In terms of gender balance in participation, Nepal was the most impressive, registering 50% women participants. Such a balance has been achieved through women-focused programs in Nepal such as "Miss Technology" and "Empowering Women in GIT" (ICIMOD, 2018a). The institutions from Myanmar and China maintained the minimum participation level of 40% women, while the number of women were very less from Afghanistan. This may be due to the fact that very few women are engaged in EO&GIT-related professions in that country; thus, more effort has to be put in to bring in more women from Afghanistan. Similarly, the participants from Bangladesh and Pakistan were mostly male—73.63 and 70.56%, respectively; here too, efforts are required to increase the participation of women.

Monitoring of Capacity Building Activities

Monitoring and evaluation is an integral part of capacity building activities. We conduct pre- and post-assessments of each type of capacity building activity, except for exposure learning visits. The pre-assessment provides guidelines about the participants' awareness and expectations about the capacity building content, while the post-assessment provides the level of knowledge acquisition by the participants and feedback for the purpose of improvement. In this paper, we present pre- and post-assessment monitoring and evaluation results from one of the 39 capacity building activities. It is a representative application of monitoring activity as it will be difficult to accommodate the results from each activity due to unique characteristics, response, and feedback.

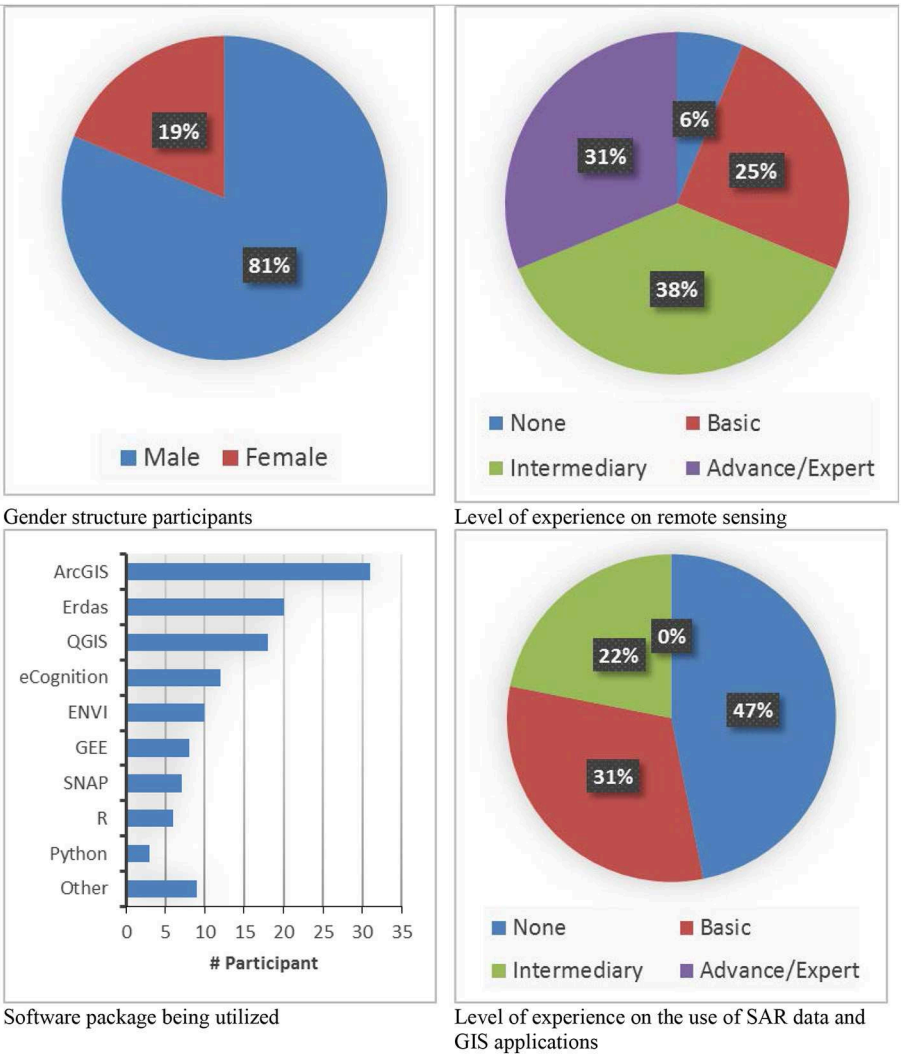


FIGURE 2 | Structural composition of the participants and their technical skills.

A case from training on "SAR for Mapping of Forest Degradation and Deforestation" is presented here with details about its targeted audience, course contents and the results from the monitoring exercise. The training aimed to provide theoretical and practical knowledge so that SAR data and its applications could be effectively used to map forest degradation and deforestation. This 5-day training program covered theoretical background while providing hands-on knowledge about basic processing procedures using relevant data sets from active radar systems; it also outlined the limitations and error sources of each processing technique. Further, it covered various steps on SAR data processing and data analysis (ICIMOD, 2018b). The training was attended by 32 participants from SERVIR networks and partner institutes from ICIMOD member countries, including Afghanistan, Bangladesh, Bhutan, India, Myanmar, Nepal, and Pakistan. Two SMEs from US institutions were also brought in as resource persons.

We conducted two pre- and post-training evaluation surveys. These surveys focused on assessing the participants' level of knowledge and skills in the subject matter. This helped in understanding the overall experience of the participants. **Figure 2** provides an overview of the composition of the participants and their technical skills before the training. Each participant responded to all questionnaires in both surveys. Due to various external reasons and the nomination process in selecting the participants, the gender structure of the training was largely imbalanced, as only about 19% women professionals took part in the program. In terms of remote sensing knowledge, most of the participants reported that they had either intermediary or expert level of knowledge. Almost all the participants were using ArcGIS for geospatial analysis in their works. Very few of them were using Python. About 53% of the participants were familiar with SAR data and GIS applications; the remaining participants reported that they had never used SAR data. From the survey, we

TABLE 7 | Participants' response to the technical capacity building support (in %).

QN	Description	Strongly agree	Agree	No comment	Disagree	Strongly disagree
1	The presentation was clear and to the point	70.0	30.0	0.0	0.0	0.0
2	The training was interactive	73.3	23.3	3.3	0.0	0.0
3	The presenter(s)/facilitator(s) were experts in their subject domain	96.7	3.3	0.0	0.0	0.0
4	The training achieved its goals and objectives	50.0	46.7	3.3	0.0	0.0
5	The materials/handouts were useful	70.0	23.3	6.7	0.0	0.0
6	The presentations were interesting and practical	66.7	33.3	0.0	0.0	0.0
7	Adequate time was provided for attendee questions	73.3	26.7	0.0	0.0	0.0
8	The content was well organized and easy to follow	56.7	40.0	3.3	0.0	0.0
9	The training met my expectations	53.3	43.3	3.3	0.0	0.0
10	Appreciation of coffee break and lunch	70.0	26.7	3.3	0.0	0.0
11	Appreciation of hotel and other services provided	40.0	20.0	40.0	0.0	0.0
12	Appreciation of training facilities (equipment, space, internet, etc.)	46.7	33.3	20.0	0.0	0.0
13	Appreciation of logistics (airport pickup, transport, travel arrangements, etc.)	53.3	10.0	36.7	0.0	0.0

found that most of the participants expected to gain theoretical as well practical knowledge about SAR data analysis for forest monitoring applications.

In the post-training assessment, 13 questions were related to the participants' response to the technical capacity building support (**Table 7**). The results show that the participants were satisfied with the training as all of them either responded with "Strongly Agree" or "Agree;" no one ticked on "Disagree" or "Strongly Disagree." However, some of the questions, such as QN# 11–13, received "No Comment," which was mainly due to responses from the residential participants from Nepal as they did not use the accompanying facilities such as hotel and transport.

Table 8 provides the results on whether the participants' knowledge and skill levels increased after the training. All the participants responded that their capacity had increased and assured us that they would apply the knowledge that they had gained from the training. Most of the participants agreed that they have learnt the basic principles of SAR and processing steps such as radiometric terrain correction, SAR data visualization and image interpretation, time series data analysis, forest change detection, Python scripting and automation, and the use of Jupiter notebook and Hype3 application. They also stated that they have acquired brief theoretical knowledge about interferometric SAR and flood applications. Some of the participants also wanted to learn about time-series data creation method, cloud processing system for bulk data analysis, use of artificial intelligence in mapping, methods of forest degradation mapping, and way to measure confidence intervals of change detection; besides, they wanted more tutorials on Python API for SAR processing and data cube. They also sought to learn about the procedures in data fusion of optical and radar, polarimetric and interferometric SAR for forest applications, biomass estimation and uncertainty reporting. They noted that SAR applications on forest reference emission level (FREL) and measurement, reporting and verification (MRV), biomass change detection, classification and mapping are very important, but had not been covered in the training. Thus, for the to-do list of future trainings, the following have been identified: advanced SAR data analysis; hands-on exercise with complete case study; operational SAR for forest applications; coding to access voluminous data; more scripting for large-scale data processing; automated system for forest degradation identification; forest health monitoring using SAR; and more discussions on cloud computing implementation. One of the shortcomings some of the participants pointed out was that the time for discussions was too short.

TABLE 8 | Learning–knowledge/skills abilities gained (in %).

QN	Description	Yes	No
14	Has your capacity increased because of the training?	100	0
15	Will you be able to apply the knowledge that you have gained?	100	0

All these primary pieces of information derived from the pre-assessment survey were very helpful in refining the training module during the training period itself to meet the expectations of the participants. The results from the post-assessment survey were helpful in improving the next training module. In addition, we are planning to conduct tracer surveys to understand the status of capacity building in the past 3 years. The relevant questionnaire of this survey will be sent to all the participants who attended any of the 39 training programs. Furthermore, institutional capacity assessment using OCAT (Organizational Capacity Assessment Tool) is also planned to examine the long-term impact of the capacity building activities on select partner institutions in Afghanistan, Bangladesh, Myanmar, Nepal, and Pakistan.

Challenges and Opportunities

We have presented the approach adopted for building the capacity of individuals and organizations on EO&GIT in the region where 39 events were conducted successfully in the last 3 years. Some institutional success stories, from establishing a GIS lab to preparing a glacier data inventory, and many more are

documented in http://servir.icimod.org/stories. However, there are many challenges and opportunities that we learnt about while implementing the capacity building activities.

Challenges

Bringing the key stakeholders to the meeting table for consultations and engaging them in prioritizing the capacity needs of their country was a big challenge. In some cases, it was also difficult to get policy-level people who could provide wider inputs on capacity gaps and the EO&GIT needs of the country.

Influx of variety of EO data and emerging GIT are making challenges to keep frequent updating capacity of individuals and institutions. When new technologies emerge, new curriculum, materials and programs on capacity building become necessary, but they are ultimately resource intensive and the appropriate SMEs may not be available immediately.

OJT, ToT, exposure learning and most of the standard trainings were nomination based, which made it difficult to ensure gender balance in certain capacity building events. Furthermore, sometimes, we received nominations of less relevant people for highly technical training modules. Geopolitical tension between some of the countries in the region was also a major challenge—it made difficult to ensure the participation of particular countries within the stipulated time frame. In the HKH region, while there are lesser number of women professionals in the EO&GIT field, there's also the added factor of women having more social and religious obligations than men; thus, bringing more women to the capacity building events was also a key challenge.

Moreover, the region has diverse sociocultural systems, so we had to pay special attention to this context while organizing a capacity building event. Sometimes, language was also a barrier. We provide best efforts to engage resource persons with multiple language skills, however, some time it becomes either difficult or cost intensive for language translation services.

We also received the feedback that after investing much resources in capacity building, many organizations in the region were facing challenges in retaining technical professionals for a longer tenure. Retaining technical professionals is a key challenge in the region because of reshuffle, promotion, and administrative engagement of professionals within the organization, transfer of professionals into other organization, and lucrative opportunities outside the organization. Challenges were also faced in the monitoring and evaluation aspect of capacity building activities, as some participants were either reluctant or less motivated to provide true responses to the survey questionnaires.

Opportunities

The country consultation meetings provided opportunities to bring the stakeholders closer and engage them in the assessment of capacity building gaps and needs in their respective countries. Their priority list helped in the designing of activity and in the smooth implementation of plans. They ensured demand and ownership which eventually helped them to include capacity building experiences in their workflows.

The capacity building activities also provided a platform for people from various segments and contexts—from young men and women, experienced professionals and technicians, to policymakers—to work together in the larger interests of the region; this helped in achieving great benefits from a small investment. The influx of a wide variety of EO data, tools and literature provided cost-effective choices on developing training materials and implementing capacity building activities in some resource-starved areas. The co-organization of capacity building activities had multiplier effects in terms of learning, value-added knowledge and a greater feeling of ownership among the stakeholders.

The EO&GIT communities are now moving toward open access tools and data which reduces the cost of capacity building of individuals as well as institutions. As for the nomination-based selection process, it reduced the liability of selecting the right participants from the partner institutions. Meanwhile, employing the services of SMEs brought cutting-edge expert knowledge to the region.

As the number of women engaged in this field is rather less, there is a vast opportunity to design women-focused capacity building programs which can empower them in the field of EO&GIT and nurture a gender-balanced workforce in the future. In terms of multicultural participation in the capacity building events, it provided a great opportunity to interact with people from different social backgrounds and countries, which ultimately strengthened partnerships and professional networks. The monitoring and evaluation process provided the opportunity to understand the expectations and learning achievements, trace the impacts of the activities, and consider the necessary measures that must be taken in future capacity building plans.

CONCLUSION AND WAY FORWARD

The simple but robust ADIM (assessment, design, implementation, and monitoring) approach presented in this paper was able to cater to the needs as is evident from the results. This innovative approach identified gaps and needs, designed efficient capacity building activities, implemented the plans to achieve a lasting impact and monitored the delivery results so that a feedback mechanism could be established. This approach is cost-efficient, sustainable and gender balanced, and so should go a long way in securing effective workforce development. The complete workflow focused on bridging the gap between current trends/progressions in the EO&GIT field and the existing state of capacity of the agencies involved in the decision-making process. The OJT and ToT type of capacity building activities played a very important role in institutional capacity building. The standard training type of activities were powerful in building individual capacity and improving the gender balance in the technical professional workforce. The exposure learning visit brought young people together and provided a competitive environment of learning wherein they gained and shared knowledge among themselves and others. In addition, this type of capacity building program brought together high-level policy/decision makers, and they were exposed to and became aware of emerging geospatial technologies—thus, they are now capable of integrating such

technology in their institutional workflows. Our approach was successful in delivering capacity building activities based on the priorities of our partners, and this should strengthen the ability of governments and development stakeholders in the region to sustain the program. By engaging global experts as SMEs and building the capacity on the use of EO&GIT in the region, we have laid a strong foundation for positive technological interventions. Despite the challenges, the approach was able to provide capacity building services to a large number of unique organizations and communities worldwide; it will also pave way toward a more gender-balanced workforce in the future. The paper reveals the several opportunities that arise from capacity building activities—from sociocultural experiential learning to the use of frontier technology. Our approach promoted gender equity, adopted frontier technology, engaged SMEs and provided sustainable solutions, all of which will have a positive bearing on the region. Apart from this effective capacity building approach, the detailed sharing of unique experiences will be of great help to capacity building practitioners in the HKH region and beyond.

As for the way forward, there is still much that needs to be fixed improvement in the coming years. Regular organizational capacity assessment (OCA) and tracer surveys, at least once in 5 years, are mandatory to monitor the periodic impact of the capacity building efforts in the region. In fact, we have already started conducting OCA and tracer surveys. There's also the avenue to set up an interactive web-based training portal to host existing and upcoming training materials for wider distribution. Such a portal would enable video-based online tutorials whereby the users can learn by themselves about all the aspects of capacity building. And that will surely have a path breaking impact on building capacities in the region and beyond.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

RT, MM, and BB design the conceptual framework of the manuscript. RT collected the data. All authors analyzed the data and wrote the manuscript.

ACKNOWLEDGMENTS

We, the authors, wish to thank our reviewers for their creative comments and suggestions that helped us to improve this manuscript. We are also thankful to the entire SERVIR-HKH team, the SERVIR Science Coordination Office and all the subject matter experts for their support in the capacity building activities. ICIMOD gratefully acknowledges the assistance of its core donors: the governments of Afghanistan, Australia, Austria, Bangladesh, Bhutan, China, India, Myanmar, Nepal, Norway, Pakistan, Sweden, and Switzerland. The views and interpretations in this paper are those of the authors alone and so are not necessarily attributable to ICIMOD.

REFERENCES

Bajracharya, S. R., Maharjan, S. B., Shrestha, F., Guo, W., Liu, S., Immerzeel,W., et al. (2015). The glaciers of the Hindu Kush Himalayas: current status and observed changes from the 1980s to 2010. *Int. J. Water Res. Dev.* 31, 161–173. doi: 10.1080/07900627.2015.1005731

Balcazar, Y. S., Balcazar, F. E., Ritzler, T. T., and Iriarte, E. G. (2008). Capacity building and empowerment: a panacea and challenge for agency-university engagement. *Int. J. Commun. Res. Engage.* 1, 179–196. doi: 10.5130/ijcre.v1i0.626

Bergeron, K., Abdi, S., DeCorby, K., Mensah, G., Rempel, B., Manson, H. (2017). Theories, models and frameworks used in capacity building interventions relevant to public health: a systematic review. *BMC Public Health* 17:914. doi: 10.1186/s12889-017-4919-y

Chandler, J., and Kennedy, K. S. (2015). *A Network Approach to Capacity Building*. Washington, DC: National Council of Nonprofits. Available online at: www.councilofnonprofits.org

ECBI (2018). *Pocket Guide to Capacity Building for Climate Change*. European Capacity Building Initiative.

Flores-Anderson, A., Herndon, K., Cherrington, E., and Thapa, R. (2019). *The SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation, 1*. Huntsville, AL: NASA.

GEO (2006). *Capacity Building Strategy*. Document No. 13, GEO-III.

ICIMOD (2015). *Needs Assessment for SERVIR-HKH in Afghanistan*. National Consultation Workshop Report, Ministry of Agriculture, Irrigation and Livestock (MAIL), Kabul, 14–15 December 2015.

ICIMOD (2016a). *Needs Assessment for SERVIR-HKH in Bangladesh*. National Consultation Workshop Report, Local Government Engineering Department (LGED), Dhaka, 26 January 2016.

ICIMOD (2016b). *Needs Assessment for SERVIR-HKH in Nepal – Development of Agriculture Advisory System to Support Food Security and Livelihood Improvement in Nepal*. National Consultation Workshop Report, ICIMOD, Kathmandu, 10 February 2016.

ICIMOD (2016c). *Needs Assessment for SERVIR-HKH in Nepal – Climate Services and Hydro-Climatic Disaster Risk Management in Nepal*. National Consultation Workshop Report, ICIMOD, Kathmandu, 29 March 2016.

ICIMOD (2016d). *Needs Assessment for SERVIR-HKH in Pakistan*. National Consultation Workshop Report, Pakistan Council of Research in Water Resources (PCRWR), Islamabad, 23 February 2016.

ICIMOD (2017). *Fourth Medium-Term Action Plan 2018–2022*. Available online at: https://www.icimod.org/resource/30287 (accessed May 30, 2019).

ICIMOD (2018a). *Empowering Women in GIT*. Training Report, Kathmandu, Nepal, 25–28 June 2018.

ICIMOD (2018b). *Regional Training on Synthetic Aperture Radar for Monitoring of Forest Carbon Stocks*. Training Report, Kathmandu, Nepal, 30 April—4 May 2018.

Merino, S. S., and Carmenado, I. R. (2012). Capacity building in development projects. *Procedia* 46, 960–67. doi: 10.1016/j.sbspro.2012.05.231

Morgan, P. (2006). *The Concept of Capacity*. European Centre for Development Policy Management (ECDPM), The Netherlands. Available online at: www.ecdpm.org

Potter, C., and Brough, R. (2004). Systemic capacity building: a hierarchy of needs. *Health Policy Plan.* 19, 336–45. doi: 10.1093/heapol/czh038

Raynor, J., Cardona, C., Knowlton, T., Mittenthal, R., and Simpson, J. (2018). *Capacity Building 3.0 – How to Strengthen the Social Ecosystem*. Brief Paper, TCC Group. Available online at: www.tccgrp.com

SERVIR Global (2018). *A Retrospective Report: 2014–2018*. Available online at: www.servirglobal.net

SERVIR-HKH (2016). *Service Area Planning*. Planning Document, 48.

Sharma, E., Molden, D., Rahman, A., Khatiwada, Y. R., Zhang, L., Singh, S. P., et al. (2019). "Introduction to the hindu kush himalaya assessment," in *The Hindu Kush Himalaya Assessment – Mountains, Climate Change, Sustainability and People*, eds P. Wester, A. Mishra, A. Mukherji, and A. B. Shrestha (Cham: Springer Nature Switzerland AG), 1–16.

Whittle, S., Colgan, A., and Rafferty, M. (2012). *Capacity Building What the Literature Tells Us*. Dublin: Centre for Effective Services.

Wignaraja, K. (2009). *Capacity Development: A UNDP Primer*. United Nations Development Programme, New York, NY.

World Bank (2005). *Capacity Building in Africa – An OED Evaluation of World Bank Support*. Operation Evaluation Department, Washington, DC. (No. 34351).

World Economic Forum (2018). *The Global Gender Gap Report*. Available online at: www.weforum.org

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

TECHNOLOGY AND CODE
published: 23 October 2019
doi: 10.3389/fenvs.2019.00158

Hydrologic Modeling as a Service (HMaaS): A New Approach to Address Hydroinformatic Challenges in Developing Countries

Michael A. Souffront Alcantara[1], E. James Nelson[2], Kiran Shakya[3], Christopher Edwards[2], Wade Roberts[2], Corey Krewson[1], Daniel P. Ames[2], Norman L. Jones[2] and Angelica Gutierrez[4]*

[1] Aquaveo, LLC, Provo, UT, United States, [2] Department of Civil and Environmental Engineering, Brigham Young University, Provo, UT, United States, [3] International Centre for Integrated Mountain Development (ICIMOD), Lalitpur, Nepal, [4] National Oceanic and Atmospheric Administration (NOAA), Silver Springs, MD, United States

OPEN ACCESS

Edited by:
Ashutosh S. Limaye, National Aeronautics and Space Administration (NASA), United States

Reviewed by:
Teresa Ferreira, University of Lisbon, Portugal
Sharad Kumar Jain, National Institute of Hydrology (Roorkee), India
Kel Markert, University of Alabama in Huntsville, United States

***Correspondence:**
Michael A. Souffront Alcantara
msouffront@aquaveo.com

Specialty section:
This article was submitted to Freshwater Science, a section of the journal Frontiers in Environmental Science

Received: *31 May 2019*
Accepted: *30 September 2019*
Published: *23 October 2019*

Citation:
Souffront Alcantara MA, Nelson EJ, Shakya K, Edwards C, Roberts W, Krewson C, Ames DP, Jones NL and Gutierrez A (2019) Hydrologic Modeling as a Service (HMaaS): A New Approach to Address Hydroinformatic Challenges in Developing Countries. Front. Environ. Sci. 7:158. doi: 10.3389/fenvs.2019.00158

Hydrologic modeling can be used to aid in decision-making at the local scale. Developed countries usually have their own hydrologic models; however, developing countries often have limited hydrologic modeling capabilities due to factors such as the maintenance, computational costs, and technical capacity needed to run models. A global streamflow prediction system (GSPS) would help decrease vulnerabilities in developing countries and fill gaps in areas where no local models exist by providing extensive results that can be filtered for specific locations. However, large-scale forecasting systems come with their own challenges. These New hydroinformatic challenges can prevent these models from reaching their full potential of becoming useful in the decision making process. This article discusses these challenges along with the background leading to the development of a large-scale streamflow prediction system. In addition, we present a large-scale streamflow prediction system developed using the GloFAS-RAPID model. The developed model covers Africa, North America, South America, and South Asia. The results from this model are made available using a Hydrologic Modeling as a Service approach (HMaaS) as an answer to some of the discussed challenges. In contrast to the traditional modeling approach, which makes results available only to those with the resources necessary to run hydrologic models, the HMaaS approach makes results available using web services that can be accessed by anyone with an internet connection. Web applications and services for providing improved data accessibility, and addressing the discussed hydroinformamtic challenges are also presented. The HydroViewer app, a custom application to display model results and facilitate data consumption and integration at the local level is presented. We also conducted validation tests to ensure that model results are acceptable. Some of the countries where the presented services and applications have been tested include Argentina, Bangladesh, Colombia, Peru, Nepal, and the Dominican Republic. Overall, a HMaaS approach to operationalize a GSPS and provide meaningful and easily accessible results at the local level is provided with the potential to allow decision makers to focus on solving some of the most pressing water-related issues we face as a society.

Keywords: cyberinfrastructure, data visualization, hydroinformatics, hydrologic modeling, XaaS

1. INTRODUCTION

The creation of a global high-resolution streamflow prediction system fills a critical need for many water-related application areas, including food security, climate change, and risk reduction. The United Nations (UN) has adopted a set of goals that aspire to greater prosperity for our society while maintaining a sustainable approach. The list, known as the Sustainable Development Goals (SDGs), includes seventeen different goals aimed at areas of need such as poverty and hunger. This set of goals highlights how important water is for the success of humankind as more than half of the seventeen goals are directly related to water, and one can argue that many other goals if not all are indirectly and positively affected by a greater understanding and use of water resources. Complementary to the UN's SDGs, the SENDAI Framework for Disaster Risk Reduction constitutes an agreement endorsed by the UN to reduce disaster risk, and subsequently the losses of lives, livelihoods, and environmental assets at the individual, community, and country scale due to natural disasters.

Early warning systems have been identified a one of the main strategies to help reduce environmental risks, especially those due to hydrological events (Hallegatte, 2012; Alfieri et al., 2013; Wilhite et al., 2014; Cools et al., 2016). The main concept behind any disaster risk reduction or mitigation is to lower the costs of such events. The effectiveness of flood preparedness has been proven by various general and localized estimates that compare the initial cost of the initiative with the potential cost of a given flood event or a number of them (Godschalk et al., 2009; Kelman, 2013; Kull et al., 2013). Developed countries usually possess the resources required to develop and operate models that provide the necessary information to drive their own flood warning system. The US National Water Model and the European Flood Awareness System are prime examples of such models. While most of the developed world has adequate data, models, tools, and experience, developing countries often lack the capacity to produce and maintain their own modeling infrastructure, which in turn increases their vulnerability. Organizations like the World Bank have recognized that international assistance is essential for developing countries to overcome vulnerability. With floods being one of the most recurrent and costly natural disaster around the world, the development of a global streamflow prediction system (GSPS) as a source to feed local early warning systems also has the potential to markedly improve risk reduction, especially in areas lacking the resources to develop their own models. A GSPS that supplements and fills gaps in local information can be used to help us understand how to better respond to extreme events such as floods and droughts, and prepare accordingly.

The development of a functional global high-resolution hydrologic model was deemed one of the "grand challenges" within hydrology (Wood et al., 2011). A functional global model must have sufficient resolution to be relevant at local scales. The development of large-scale high-resolution models has become a focus for many hydro-meteorological scientists in response to this challenge.

In recent years, a number of large-scale models have emerged (Rodell et al., 2004; Lindström et al., 2010; Alfieri et al., 2013; NOAA, 2016). The development of such models has been possible due to the evolution of hydrologic modeling, which includes a number of internal scientific advances, but also a vertical expansion where elements from other sciences such as meteorology have been integrated. As a result, we have increased our ability to predict hydrologic events by linking atmospheric and land surface models so they can work as one integrated hydrometeorological model. Advances in other disciplines, such as information technology and computer science, have also made the development of larger-scale models possible, by providing local access to large datasets that cannot be downloaded and explored on a desktop environment. In addition, probabilistic forecasts offer an alternative to incorporate the uncertainty introduced by the inputs used to run a hydro-meteorological model through ensemble forecasting (Demeritt et al., 2013). This expansion of hydrologic modeling opens the door for greater application in all the earth sciences and provides valuable support to solving the wider set of interdisciplinary problems articulated in the SDG's. **Figure 1** shows a concept example of how hydrometeorological models can provide water intelligence in a multidisciplinary environment that aims to solve complex problems.

While many advancements and improvements to hydrometeorological models have been and are being made, there are major challenges remaining to make these large-scale models relevant at the local scale where decisions are made. For example. the inherent uncertainty introduced by models themselves can be significant and should not be overlooked (Butts et al., 2004). In addition, while traditional discharge calibration from observed discharge can improve model performance in a specific area, it is difficult to find a single parameterization that works well for a large-scale model given its inherent (Sperna Weiland et al., 2015).

On the other hand, the amount of data produced by large-scale models presents yet a new hydroinformatic challenge. Furthermore, integrating and communicating model results has historically been a major challenge due to the evolving nature of hydrology and hydrologic models (Beran and Piasecki, 2009).

In general, communicating water data to different groups (e.g., scientists, emergency responders, decision makers, and the general public) has also been a major challenge due to their distinct contexts and needs (Souffront Alcantara et al., 2017). The answer to this challenge is being answered by the adoption of standards, a push to create Earth Observation Systems (EOS) and model results that can be accessed as services, and the creation of derivative tools that facilitate the interpretation and application of data.

Replicating a hydrologic model, with the same or different inputs and coverage, requires technical skill, and computational resources. The overall cost of deploying, running, and maintaining the model are also limiting factors. Since decision makers and stakeholders are not expected to have the skills necessary to provide sustainable hydrologic modeling predictions, capacity building and specific training at the technical level is usually the solution. However, this is often

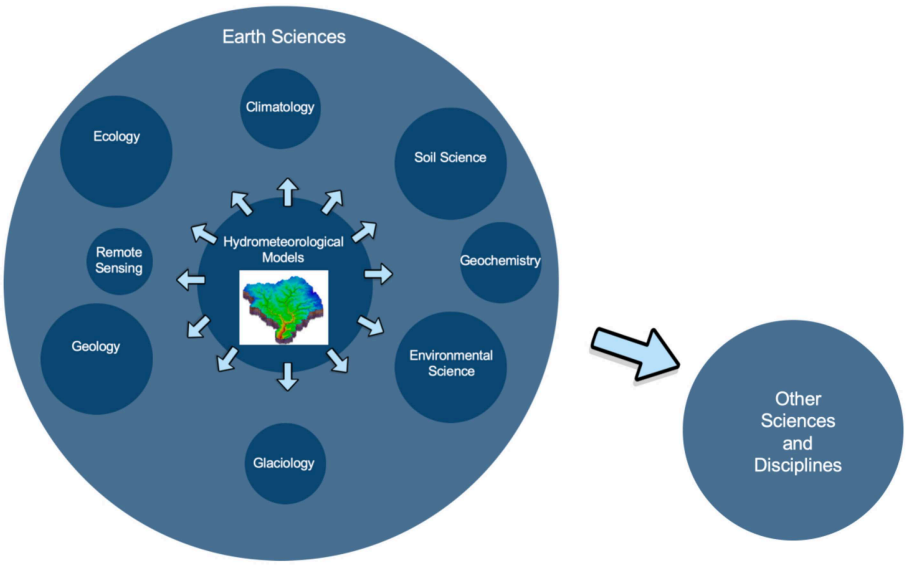


FIGURE 1 | The expansion of hydrology.

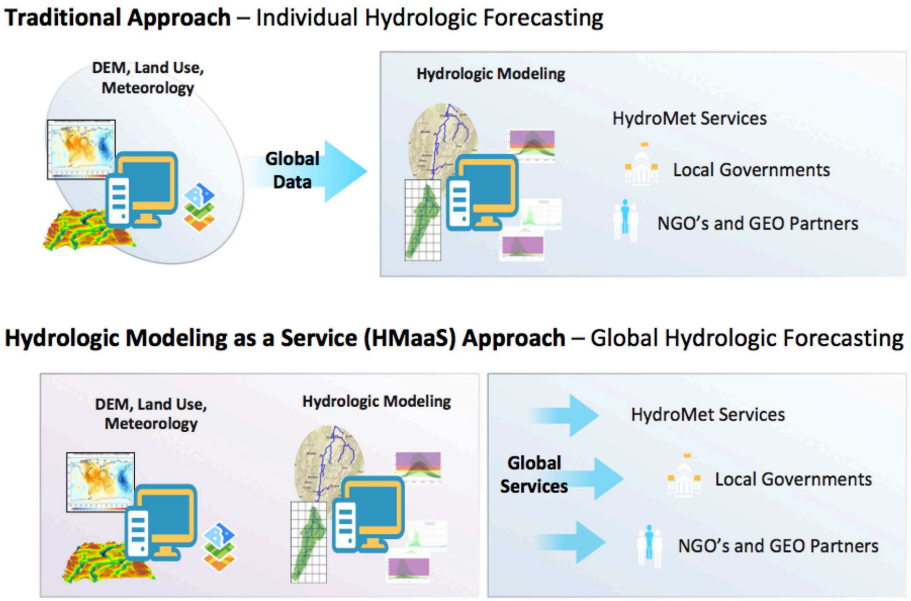


FIGURE 2 | Traditional vs. HMaaS approach.

a short-term solution mainly due to maintenance costs after the end of the project/funds, and to the loss of the original trained staff over time. While a number of large-scale models that provide hydrologic information useful for areas lacking a local model already exist, the available resolution for these models is usually not adequate at the local scale. A HMaaS approach solves these issues by taking advantage of the latest cloud computing and information communication technologies to provide model results as a service at a meaningful resolution, thus alleviating local maintenance costs, reducing the necessary technical training, and allowing investors to focus on providing training and funds for the actual problems that hydrologic modeling is needed for such as water distribution issues, and early warning systems (**Figure 2**).

This paper summarizes our effort to create a streamflow prediction system covering most of the globe with sufficient detail to be useful locally while emphasizing the need to make results readily available to different user groups using a state-of-the-art service-oriented technology. Implementation and validation results are presented. Additionally, the extended challenges

resulting from the creation of such a model are discussed in detail.

2. HYDROLOGIC MODELING AND HYDROINFORMATIC CHALLENGES

Communicating model results has historically been a major barrier between engineers and scientists, and decision makers. A successful model needs to provide clear and actionable information to meet the demand of its user community. However, in the case of hydrologic models, there is a range of distinct users with very specific, but totally different needs. These groups range from scientists to the general public. The nature of scientific research makes data discovery and retrieval a need that requires constant attention. This is not the case for other user groups. In other words, finding model data is not a priority in decision making. Therefore, most models often fail to be relevant to other groups due to the difficulty of obtaining model results in a relatively straightforward way.

Modeling as a Service (MaaS) is a distribution mechanism in which a provider makes a model, or modeling results available to stakeholders through the use of web services. This concept, which evolved from the Software as a Service (SaaS) (Choudhary, 2007), and the Anything as a Service (XaaS) (Duan et al., 2015) principles has gained speed as an answer to challenges in the deployment of environmental models in general. Roman et al. (2009) discussed the challenge of migrating stand-alone applications to services on the web. Furthermore, Li et al. (2017) proposed a MaaS as a solution to the many challenges of deploying models in the geospatial sciences.

The realization that even if a robust model that provides clear and accurate results won't be useful unless results are readily available and presented in context has opened the doors to addressing some of these extended challenges in the field of Hydroinformatics. These challenges do not only cover communication issues like data accessibility, relevancy, and clarity; but also include big data issues like storage, maintenance, and metrics tracking, and adoption issues like ownership, partnering, branding, and overall implementation alternatives at the local level. Adding to these issues is data validation, which has traditionally been a model challenge, but more so in the case of large-scale models. We have divided these hydroinformatic challenges in four main areas in order to better discuss them.

- Big Data
- Communication
- Adoption
- Validation

2.1. Big Data

A GSPS requires a solid cyberinfrastructure where results can be computed, stored, visualized, and retrieved. Moreover, a continuous operational forecast system requires a workflow that can be run automatically. This would include the download and organization of model inputs, which would add to the already large amount of data produced by the model. Therefore, the cyberinfrastructure for a global model is bound to include organizational tasks to download, archive, and delete data. Traditionally, hydrologic models have been run on local servers, however with the latest advances in Information and Communication Technologies (ICT), and in accordance with the MaaS concept, cloud storage and computing has become an indispensable resource.

Cloud computing offers a number of advantages for the development of an operational global forecast prediction system using a MaaS approach. Some of the most obvious advantages include: the removal of expensive computing hardware and storage for every local agency, cloud cyberinfrastructures are scalable, maintenance time and costs are removed with machines being maintained by the cloud provider. In addition, the entire system can be managed from one place (usually a dashboard). A task manager can handle the entire workflow from data input collection to model results storing. This is not unique to cloud computing environments, but it becomes a must when dealing with High Performance Computing (HPC) as would be the case with a global high resolution streamflow prediction system.

2.2. Communication

In the last few decades, the emergence of standards for the sharing and distributing hydrologic data has made communicating and disseminating water data much easier. Some of these standards include WaterML, which offers a simple structure for working with time series data (Almoradie et al., 2013); netCDF, which offers a more solid structure for working with multi-dimensional data (Rew and Davis, 1990); and GIS open web service standards like Web Mapping Service (WMS), Web Feature Service (WFS), and Web Processing Service (WPS), which offer a common denominator for exposing geospatially enabled water data in a dynamic way that is compatible with most available web-based visualization tools.

The adoption of the standards mentioned above has helped reduce the existing gap between data producers and data users in the hydrologic community. However, most of the focus on data communication is usually placed on scientific/research users. Furthermore, water data needs to be effectively communicated not only to the scientific community, but also to decision makers, emergency responders, and the general public. Water data needs to be presented as actionable information that is accessible and understandable for all user levels (Souffront Alcantara et al., 2017).

A solution to communicating results to the broad set of groups needing access to results is to develop intuitive web applications and services that allow users to interact with the data according to their specific needs. HMaaS through the use of a web app has many benefits. Results can be displayed using open standards, while other functionality can be added to satisfy user needs from a simple web browser. Web apps can successfully link the back-end cyberinfrastructure needed to generate forecast results with state-of-the art web development technologies to create a dynamic environment where users from different levels can access information that is relevant to them by taking advantage of open standards like WaterML, and OGC's WMS, WFS, and WPS.

2.3. Adoption

Adopting a new technology usually depends on the estimated benefits and costs of implementation. In the case of a large-scale streamflow prediction system, there are a number of general and specific factors that will determine such benefits and costs, and therefore influence implementation at the local level. Some of the general factors include the existence of a local system, and the disposition of the local community to incorporate or integrate a global system. In such a case, the global system's value would most likely be in serving as a secondary tool to trigger action, to corroborate when an extreme event is forecasted by the local system, or fill gaps from the limitations of local models in space or application. Obviously, the greatest value of a global system comes when there is no local system available.

More specific factors regarding the adoption of a global forecasting system include the time it takes to adopt new technologies, and who would take responsibility for the success/failure of the model in predicting events accurately. Principles like the Technology Acceptance Model (TAM) suggest that the adoption of a new technology depends on the perceived ease of use and usefulness of the technology (Davis, 1986). In theory, a HMaaS system offers a relatively ease of use by eliminating the costs of producing the model in favor of offering forecast results as web services that can be consumed by anyone and through programmatic means to develop derivative applications as needed. However, it is important to notice that while a forecast is provided, model results still need to be interpreted by able professionals, and decision support systems that enable responses to forecasted events remain the responsibility of the local community. Therefore, an understanding of model assumptions, limitations, and application is required at the local level. In addition, each country/region that decides to implement a global prediction system will have a vested interest in the good performance of the model. To this end, a mechanism to provide feedback and keep track of model performance is necessary.

The success or failure of the model to predict imposes certain responsibility on the owner of the model. But with a global system, ownership may not be initially clear. While the developer of the model provides results, interpretation, and response to the model fall at the local level. In practical terms, the weight of the decision support system developed from the model is of far more importance than the generation of a model. As a result, it is advised that a multi-criteria approach be used to support decisions whenever possible. Examples of such systems usually include multiple models, or observation data integration (Niswonger et al., 2014; Wan et al., 2014; Horita et al., 2015; Svoboda et al., 2015; Ahmadisharaf et al., 2016). Based on these factors, users may welcome or reject ownership and therefore responsibility over certain aspects of a global model. To this end, there are a number of implementation levels that would depend on what is determined to work better at the local level by the local agency itself.

1. External model consumption through a web app: The model is accessed from a generic web app developed to display the complete global model. Additional functionality in the app would allow for extraction and visualization of data for a specific area. This generic app could be hosted by an international organization working with different countries/regions.
2. Internal model consumption through a web app: The model is generated on-premise and displayed and accessed the generic web app. Internal generation would allow for computation of areas of interest only.
3. External model consumption through web services: The model is accessed through open standards and a REST API, and displayed using a customizable web app or integrated into an existing visualization tool.

2.4. Validation

The accuracy and uncertainty of a model need to be quantified before forecasts can be trusted for any decision-making. Traditionally, models are tested and calibrated for specific areas. This poses an additional challenge for a large-scale forecast system. Given the global extent, validation and calibration would be a very arduous task. To this end, many large-scale models have instead carried over the uncertainty of their inputs by presenting an ensemble result that accounts for input uncertainty.

Another way the accuracy of the forecast can be evaluated is by comparing results to observed data. Assuming a global model has been adopted at a regional or local scale, the model could be easily compared to regional or local observed data. Moreover, a global forecast that uses open standards improves the ability to compare with any other existing dataset. However, a mechanism to facilitate data comparison would be needed to ensure that comparisons could be made in any specific area following similar criteria.

3. MATERIALS AND METHODS

A high-density large-scale streamflow prediction system covering most of the world has been developed using GloFAS runoff, ERA Interim data, and the RAPID routing model. The workflow to generate these forecasts was deployed completely on the cloud. Two main web applications exist to interact with the results, while a REST API has also been developed to easily retrieve data without the need of a web interface, or for which custom views and subareas can be created in a separate web interface. A number of validation tests have also been performed to assert that: (1) the high-density routed forecasts yield, in essence, the same result as the original GloFAS and ERA Interim result; (2) variability on the chosen resolution to route the runoff does not alter the results at a given location; (3) model results are close to observed data at different locations around the world.

GloFAS is an ensemble hydrologic model that generates 51 different runoff forecasts for the major rivers of the world on a global grid with a resolution of 16 km^2 on a continuous basis. A 52nd forecast is generated at a resolution of 8 km^2. GloFAS was released in 2011 by ECMWF and the European Commission's Joint Research Centre (JRC) as part of the Copernicus Emergency Management System (CEMS), and has been quasi-operational since July 2011, and fully

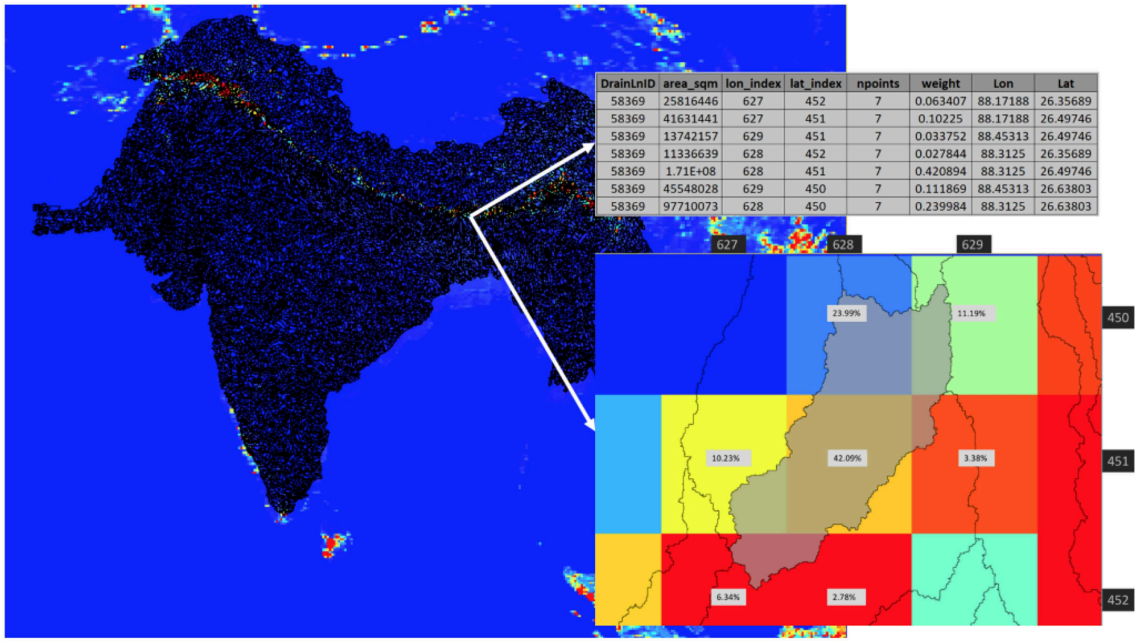


FIGURE 3 | River network and subbasin generation example for the South Asia region.

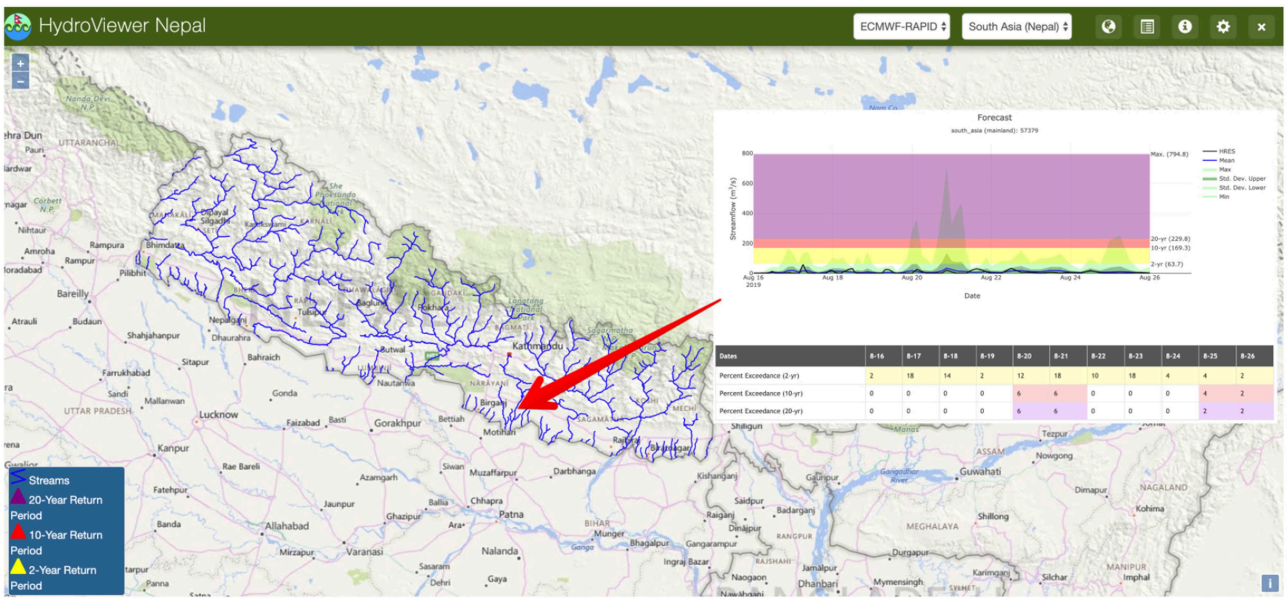


FIGURE 4 | HydroViewer app design.

operational since April 2018. The GloFAS system is composed of an integrated hydrometeorological forecasting chain and of a monitoring system that analyzes daily results and shows forecast flood events on a dedicated web platform (Alfieri et al., 2013). This model uses real-time and historical observations in combination with a Data Assimilation System (DAS) and a Global Circulation Model (GCM). The underlying framework used to create GloFAS is ECMWF's Integrated Forecasting System (IFS). GloFAS uses HTESSEL for its land surface scheme. HTESSEL is a hydrologically revised version of the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) model (Balsamo et al., 2008). This new land surface scheme corrected the absence of a surface runoff component in its predecessor, among other minor improvements.

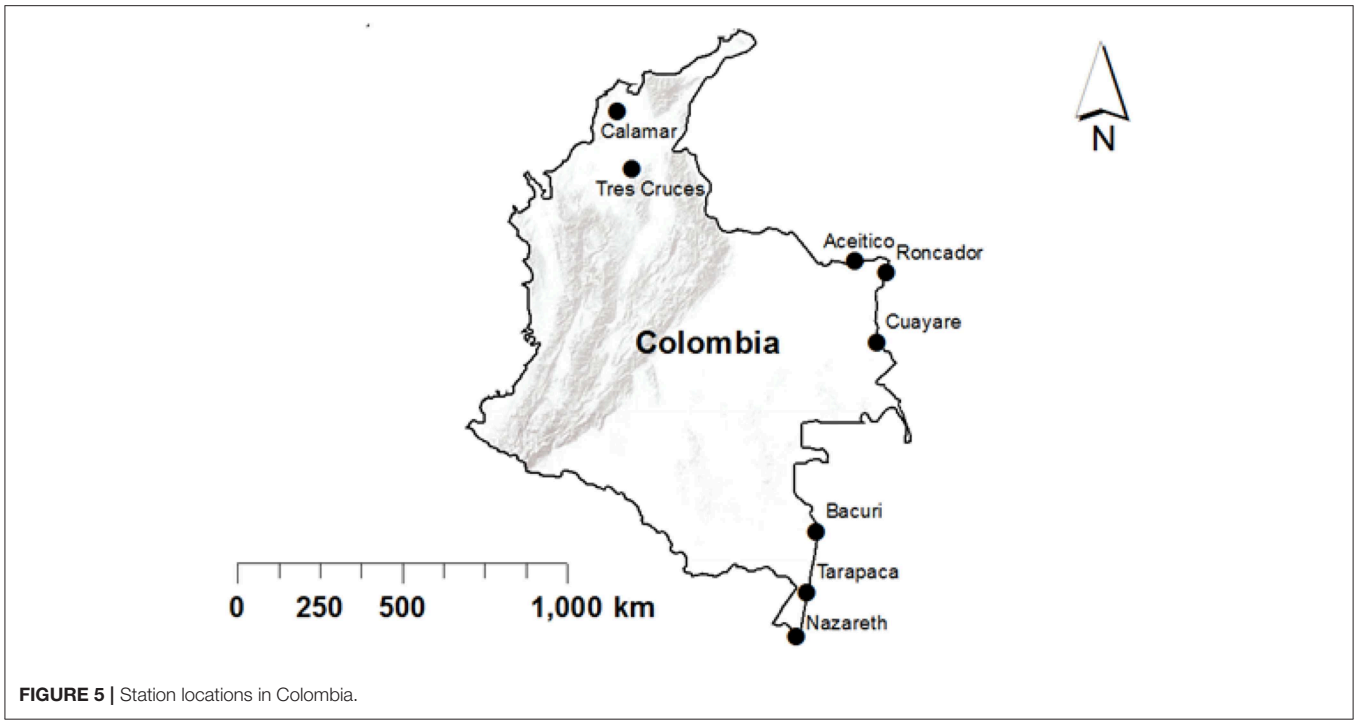


FIGURE 5 | Station locations in Colombia.

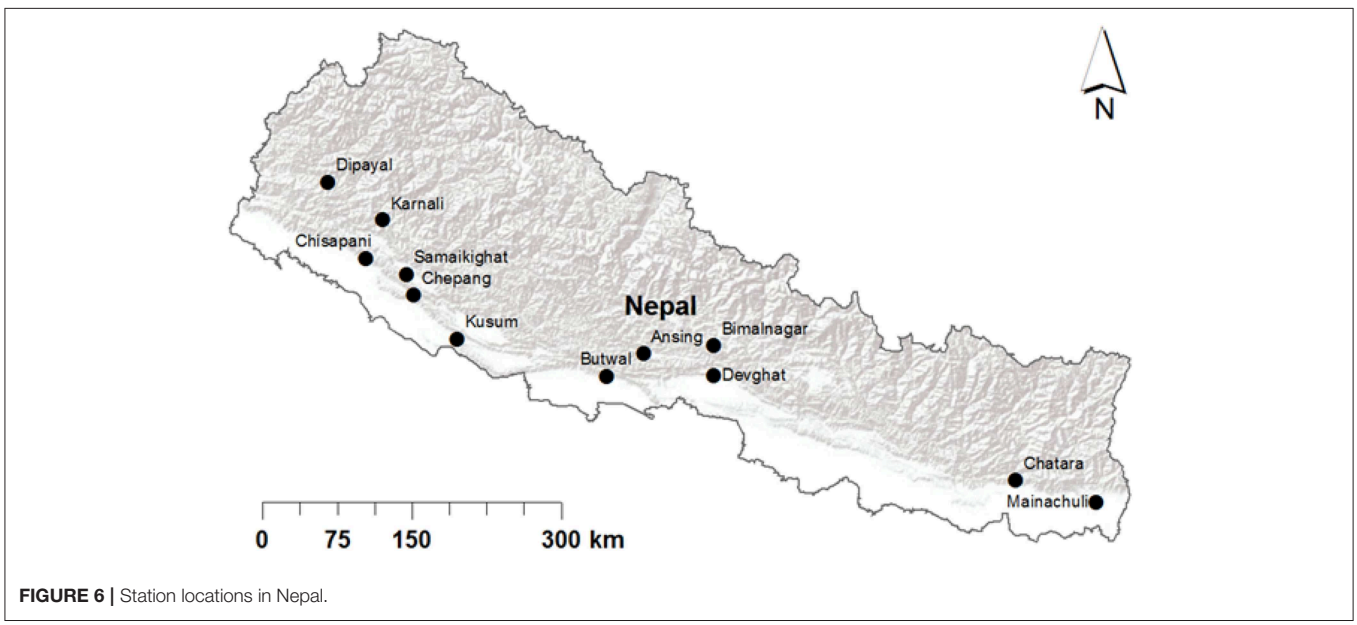


FIGURE 6 | Station locations in Nepal.

The ERA-Interim data is the result of a global atmospheric reanalysis produced also by ECMWF. This data covers from January 1980 through December 2014 (35 years) for the entire globe. One of the advantages of using reanalysis is that the data provides a global view that encompasses many essential climate variables in a physically consistent framework, with only a short time delay (Dee et al., 2011). This type of data becomes invaluable in areas where no actual observed data are available. A runoff derivative of this atmospheric reanalysis was produced on a 40 km^2 global grid using a land surface model simulation in HTESSEL.

GloFAS forecasts can be visualized from their main website (http://www.globalfloods.eu/glofas-forecasting/), which combines the forecasts from GloFAS and the simulated historic run from the ERA Interim to provide an awareness system that displays warning points and the probability of an event occurring based on the ensemble forecasts and return periods extracted from the ERA data.

TABLE 1 | Watershed attributes.

Basin	Area [km^2]	USGS Stream Gauge ID
Arizona	1013.89	09494000, 09492400
Idaho	2085.71	13340600
Montana	2391.49	07014500, 07013000, 07014000
New York	262.32	01413500, 01413408, 01413398
Oregon	532.69	14306500, 14306400, 14306100
Colombia	1767	N/A

TABLE 2 | Catchment resolution description.

Resolution	Number of catchments	Average area [km^2]
Low	3	506.94
Medium	7	217.26
High	20	75.64

RAPID is a numerical model that simulates the propagation of water flow waves in networks of rivers composed of tens to hundreds of thousands of river reaches (David et al., 2016). The RAPID model is based on the Muskingum method, which has a time and a dimensionless parameter as its main variables. RAPID successfully created a way to efficiently adapt the Muskingum method to any river network.

In an effort to create a higher density version of GloFAS that would include smaller, but important streams Snow et al. (2016) combined GloFAS with the River Application for Parallel Computation of Discharge (RAPID) routing model covering the main hydrologic regions within the United States. This work addresses GloFAS' density challenge by routing model results through a predefined river network that provides results not only for major rivers but for any potential river in the world. The Streamflow Prediction Tool (SPT), a web app similar to the main GloFAS application, was also originally developed as part of this work. The SPT provides an intuitive user interface that allows for the easy lookup and visualization of results. Other advances of this app include the capability to present dynamic hydrographs as opposed to static images. We have improved the SPT by incorporating a REST API, and improving the visualization of results.

We created a river network and weight tables for Africa, North America, South America, and South Asia following the methodology presented by Snow et al. (2016) as shown in **Figure 3**. A river network for a specific area is created using the HydroSHEDS dataset, which is a hydrographic dataset based on elevation data from the Shuttle Radar Topography Mission (SRTM) that provides data at a global scale (Lehner et al., 2008). In addition to generating hydrography, this preprocessing also generates weight tables, and Muskingum/RAPID parameters for converting the gridded results from GloFAS to a vector-based forecast using the river network.

3.1. Implementation and Visualization

We have deployed two web applications to display results using the Tethys Platform framework. Tethys is a web framework for facilitating the development of water resources web applications. It includes specific open-source software components that address the unique development needs of water resources web apps with the main goal to lower the barrier of web app development for water resource scientists and engineers (Swain et al., 2016). The first web app, the SPT, was originally developed by Snow et al. (2016). The SPT provides an interactive map where users can select a specific river reach and display a hydrograph for that reach with a 10-day forecast and the 2, 10, and 20 year return periods corresponding to that reach. Some of the improvements to the SPT include the visual design of the app, especially the graph area, but more significantly the incorporation of a REST API.

A REST API is a web service that can be used to access data without the need of a web interface. REST APIs use the http protocol to request data where parameters are passed through a Uniform Resource Locator (URL) string using a predetermined organization. This development facilitates integration of our forecast results with third-party web apps, or any other workflow; the automation of forecast retrievals using any programing language; and the development of derived applications that consume these results through the API and further process them as opposed to incurring on the same computational costs of generating their own forecast results. This last use, allows for the development of lightweight applications that provide complex results by relying on APIs from other apps.

The HydroViewer app is an example of such a lightweight web application. It was designed to visualize streamflow forecasts for specific regions using different model alternatives, which can be added to the app in a relatively easy way. So far the app includes the aforementioned GloFAS-RAPID model, the South Asia Land Assimilation System (SALDAS), and the High Intensity Weather Assessment Toolkit (HIWAT) model for monitoring intense thunderstorms. This app relies on the use of REST APIs to retrieve and visualize water data as opposed to incurring into computational costs. The HydroViewer app was also designed to allow customizations for the specific region it is deployed to. This allows users to rebrand the web app and integrate it into their system. **Figure 4** shows the HydroViewer app design.

A cloud-computing environment approach was used to deploy our workflow and make it accessible on the Internet. Two Virtual Machines were deployed on the cloud, one for performing the main computations necessary to generate the forecasts, and the other for hosting spatial web services for data visualization purposes.

3.2. Validation

Modeled data validation is essential for determining the value and limitations of the data. Jackson et al. (2019) compiled a number of commonly used error metrics that can be used to compare hydrologic modeled data to observed data. Some of these metrics include the Root Mean Square Error (RMSE) and derivatives, Coefficient of Determination, Coefficient of Correlation, Anomaly Correlation Coefficient, Nash-Sutcliffe Efficiency (NSE), and the Spectral Angle. Most of these error metrics have been compiled in a Python package called HydroStats (https://github.com/BYU-Hydroinformatics/Hydrostats).

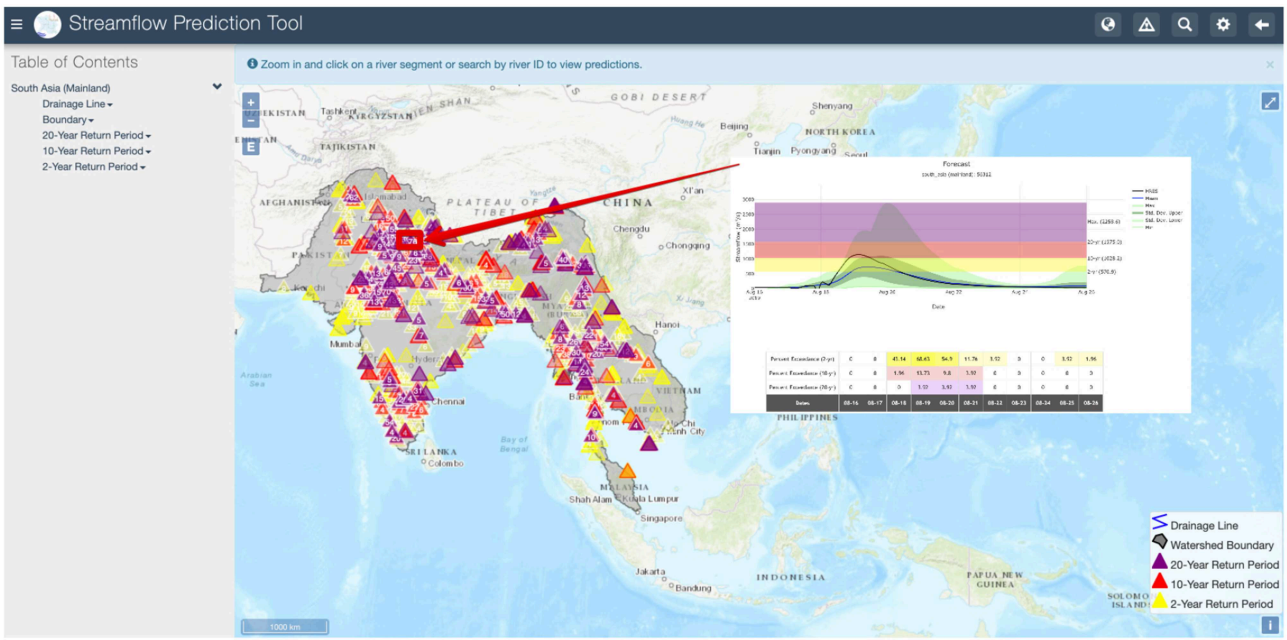


FIGURE 7 | Updated streamflow prediction tool.

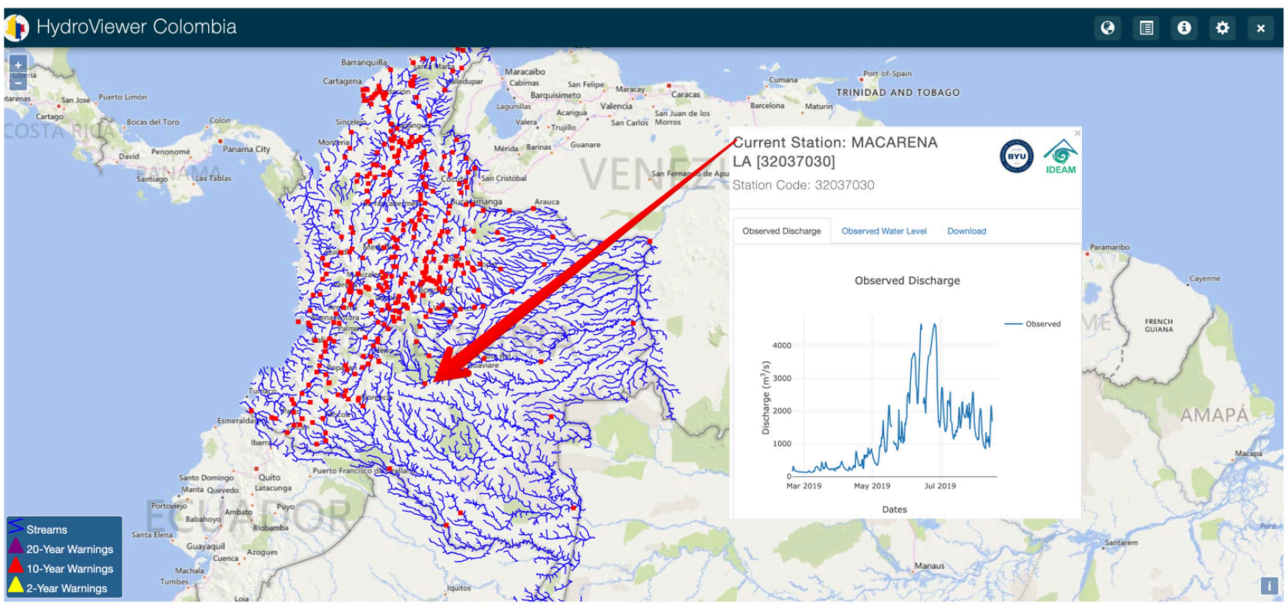


FIGURE 8 | Hydroviewer Colombia displaying observed data customization.

Using HydroStats, we compared our modeled results to observed data from Colombia, and Nepal. We analyzed eight stations for the former, and 12 stations for the latter (**Figures 5**, **6**). In our analysis, we used a number of different metrics. We used the anomaly correlation coefficient, the root mean square error, the interquartile range normalized root mean square error, the Nash-Sutcliffe Efficiency metric, the Pearson correlation coefficient, the Spearman correlation coefficient, the spectral angle metric, the improved Kling-Gupta efficiency, and the refined index of agreement. We chose to use this suite of metrics to give a more complete picture of how well the simulated data correlates to the observed data (Krause et al., 2005).

We performed a comparison between our high-density routed results with the gridded result from GloFAS at selected locations.

Data was collected from six GloFAS locations found in Nepal including Chatara, Chepang, Chisapani, Devghat, Kusum, and Parigaun. Our assumption was that if our result had similar trends and values to those of the original GloFAS runoff then it meant that our RAPID processing did not introduce any significant bias by converting the gridded GloFAS results to a higher density vector result based on a river network. In addition, we also assumed that the results of this comparison could be applied to other areas outside of the locations used for the comparison.

Data was collected every day for 9 weeks and summarized weekly. We used the mean flow of both datasets to perform the comparison as the best representation from all the ensembles. The flows from the GSPS were easily accessed through the use of the developed REST API. Because the flows from GloFAS came exclusively in a hydrograph image, values had to be digitized from the hydrograph images.

Multiple watersheds from distinct regions in the United States were tested to determine the effect of varying the catchment area resolution of the sub-basins within the watershed. The following criteria were used to select the watersheds.

- Watershed size of several hundred square kilometers.
- United States Geological Survey (USGS) gage station proximity to mouth.
- Relatively pristine area with no reservoirs.

Potential watersheds were selected from the USGS Hydro-Climatic Data Network, a collection of roughly 700 watersheds with relatively unimpaired flows.

The selected sites included: the Meramec River near Sullivan, MO; the East Branch Delaware River at Margaretville, NY; the Alsea River near Tidewater, OR; the White River near Fort Apache, AZ; and North Fork Clearwater River near Canyon Ranger Station, ID. Another similar site, the Negro River in Colombia, was also tested (see **Table 1**).

The GloFAS-RAPID historical simulation was run for each watershed at three different resolutions (**Table 2**). The streamflow at the basins' mouths were compared using HydroStats. The resulting streamflows were also compared to observed data from USGS stream gage stations.

4. RESULTS

A GSPS covering Africa, North America, South America, and South Asia at a resolution of 350 m^2 was developed and deployed using cloud services and following a MaaS approach. The cloud cyberinfrastructure and workflows develop provide an alternative to the storing and data management side of the big data challenges described in section 2.1. Two web applications as well as a REST API were developed to communicate forecast results and provide alternatives that users can choose depending on their needs. These web applications and services directly address the communication challenges described in section 2.2. A series of validation tests were also performed on the results to determine that (1) our downscaling process did not alter results compared to the original GloFAS forecasts, (2) changing the catchment area of a river reach did not alter results downstream; that is, streamflow volume remained the same for downstream reaches, and (3) modeled results were close to observed results at different locations around the world.

The new SPT provides visualization of our GloFAS-RAPID results as well as data retrieval in CSV and WaterML formats. Forecast results are available in the app for 1 week, after which they are removed and archived. Forecasts for a specific reach can be accessed by clicking on the reach. A pop-up window displays the dynamic hydrograph, which includes common interactions like zoom in or out, and data download as an image or CSV file. The hydrograph includes the 2, 10, and 20-year return periods to provide context of how much water is too much for a specific reach. The 51 ensembles are displayed using statistics that include the mean, min, max, and standard deviation. A percent exceedance table also displays the probability of a specific flow value surpassing a return period based on the prediction in each individual ensemble (**Figure 7**).

The SPT REST API was developed to facilitate data access. It includes methods to programmatically retrieve forecast statistics, as well as individual forecast ensembles. It also provides methods to retrieve the computed 35 year historic simulation, and derivatives such as return periods of each river reach within the regions. The REST API includes the following methods:

- **GetForecasts**: a method to extract forecast statistics from the 51 different ensembles available from the GloFAS-RAPID results. The available statistics are mean, max, min, and standard deviation. A high-resolution 52nd ensemble result is also available.
- **GetEnsemble**: a method to extract individual ensembles. Each ensemble can be retrieved separately, or a range of ensemble can be selected.
- **GetHistoricData**: a method to extract the 35 years of historic simulated data for a specific river reach.
- **GetReturnPeriods**: a method to extract the 2, 10, and 20 year return periods for a specific river reach calculated using the historic simulation.
- **GetAvailableDates**: a method for extracting the available forecasted dates.
- **GetWarningPoints**: a method that returns the center of a river reach along with information about the forecasted flow and if it is greater than any of the calculated return periods for that reach.

The REST API is the key functionality behind the HMaaS approach. It allows for programmatic data retrieval, and in turn, for the development of lightweight applications that provide results by relying on the API as opposed to local computational resources.

The HydroViewer app is a lightweight web application that allows users to display relevant data and customize the web app according to stakeholder needs. This app makes use of web services to display results as opposed to replicating the hardware, software, and modeling cyberinfrastructure to generate its own hydrological forecasts. The app uses the REST API to access forecast results and publicly available geospatial web services to display hydrographic data. The interface of the app can be customized to display the colors and logo of the organization it

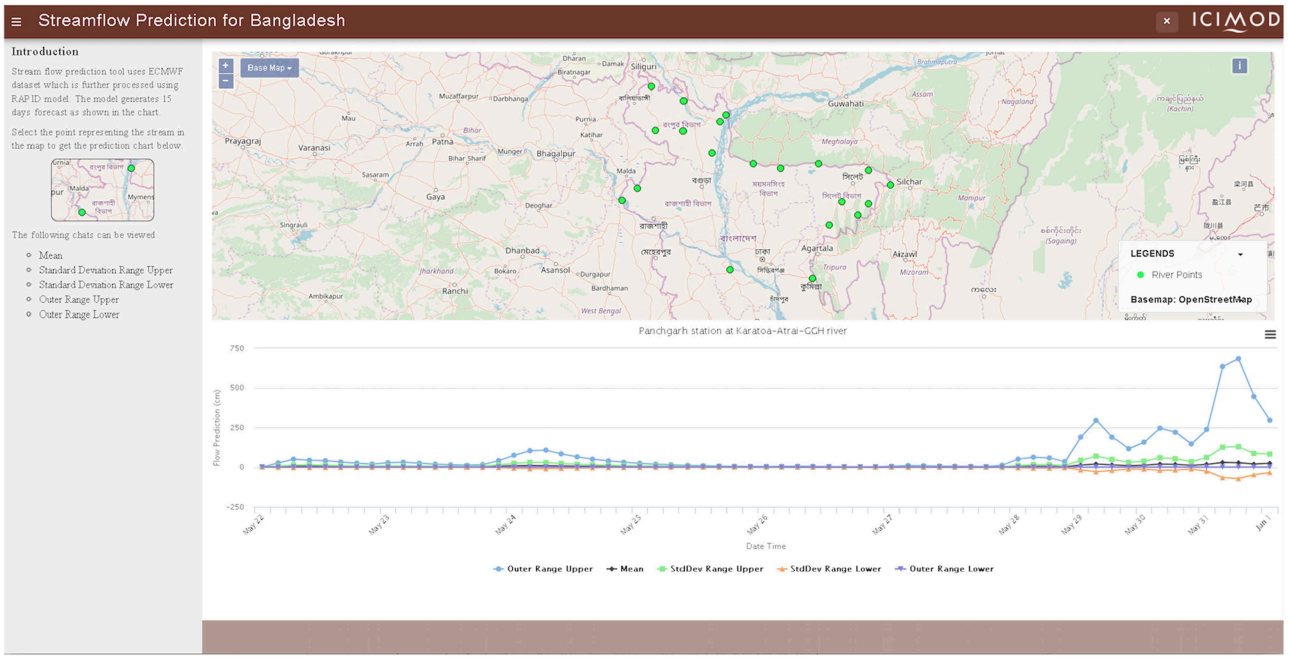


FIGURE 9 | Bangladesh's Transboundary Streamflow Prediction tool.

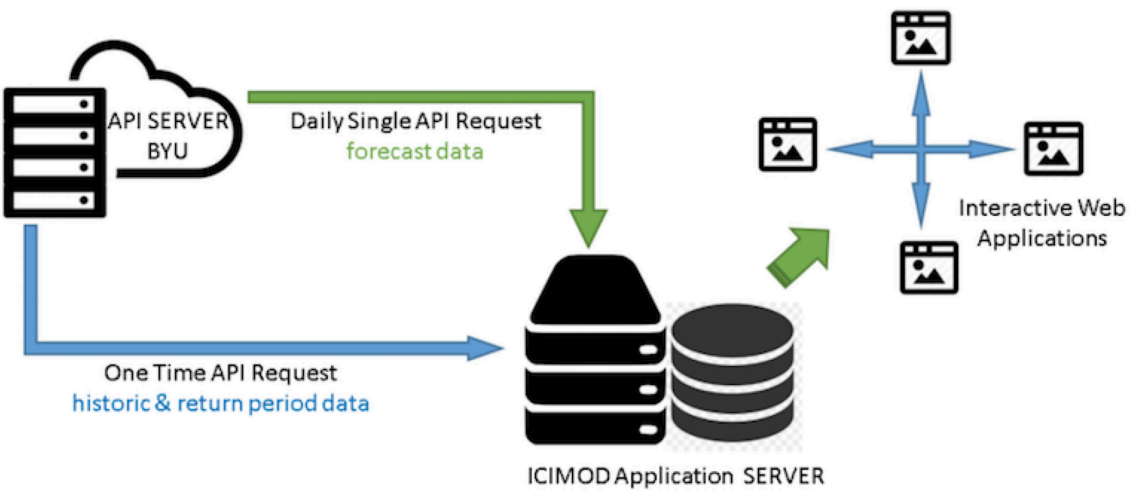


FIGURE 10 | ICIMOD workflow.

is deployed for, thus allowing users at the local scale to rebrand it as their own and market it as their own product. In addition, the HydroViewer app was designed with the principle of visualizing hydrologic results from different models, not only the GloFAS-RAPID model.

Customizations for different organizations also include the addition of hydrographs displaying observed data, data comparison displays, or the inclusion of other important geospatial data such as districts or country boundaries. Instances of the HydroViewer have been deployed for the following countries: Argentina, Bangladesh, Brazil, Colombia, La Hispaniola (The Dominican Republic, and Haiti), Nepal, and Peru. **Figure 8** shows the customized HydroViewer for Colombia.

The incorporation of a REST API has enabled the development of more complex web applications that use forecast results retrieved using the REST API. Some of these web apps include flood mapping, reservoir monitoring, and statistical analysis applications. These apps benefit from a REST API by consuming the forecast results made available through the

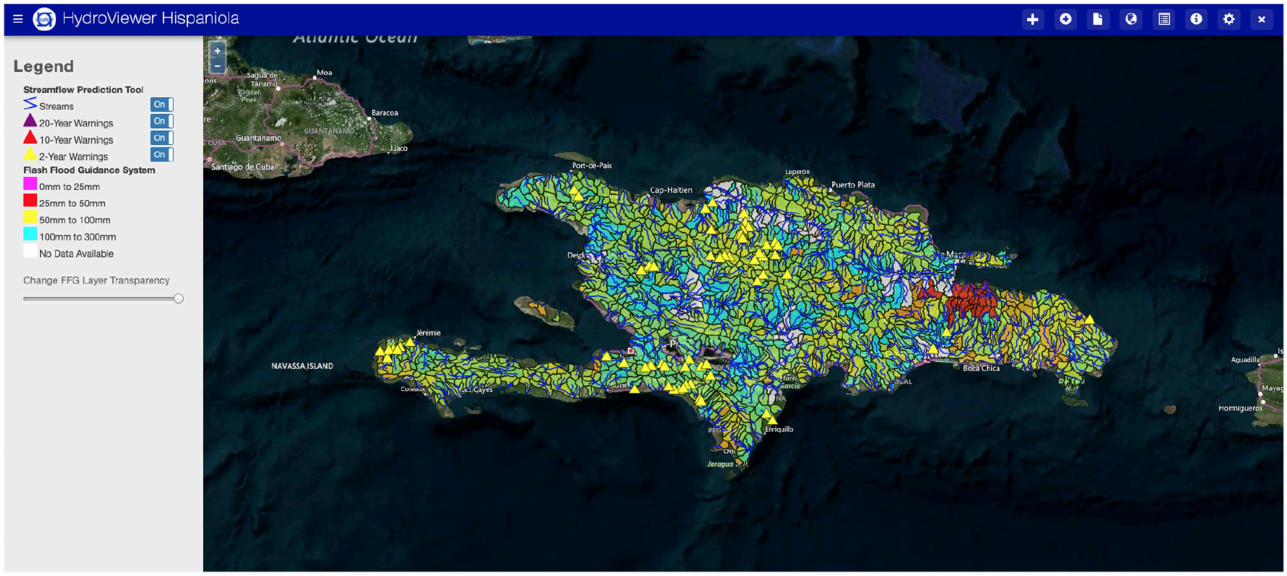


FIGURE 11 | Hispaniola HydroViewer displaying both the Flash Flood Guidance and the Global Streamflow Prediction systems.

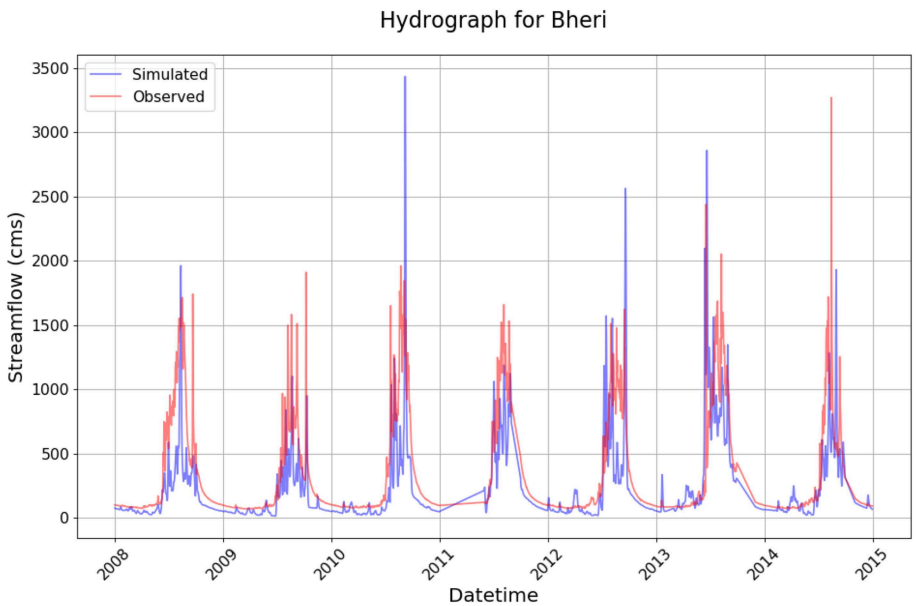


FIGURE 12 | Simulated vs. observed data for Bheri station in Nepal.

REST API endpoints. This allows for the creation of specialized apps that do not have the need to spend computational resources on recalculating essential input data such as streamflow.

The developed REST API was used to develop custom applications at the International Centre for Integrated Mountain Development (ICIMOD) and also during trainings at the national and regional level to retrieve hydrologic information. One of the applications developed is the Bangladesh Transboundary Streamflow Prediction Tool (**Figure 9**). This app provides streamflow predictions for Bangladesh's Flood Forecasting and Warning Center (FFWC), in combination with observed data from twenty stations near the international border areas of Bangladesh. The data produced is mainly used as an input to feed internal hydraulic models in an effort to improve lag-time estimations.

TABLE 3 | Error metrics for Colombia and Nepal (**Data sheet 1** has the charts for Colombia and **Data Sheet 2** the charts for Nepal).

Location	ACC	RMSE	RMSLE	NRMSE (IQR)	NSE	R (Pearson)	R (Spearman)	SA	KGE (2012)	dr
COLOMBIA										
Aceitico	0.827	2286.773	0.608	0.502	0.478	0.827	0.902	0.366	0.594	0.660
Bacuri	0.620	4761.954	0.455	0.890	−0.293	0.621	0.717	0.370	0.513	0.435
Calamar	0.666	4451.048	0.436	0.636	0.276	0.666	0.711	0.372	0.481	0.627
Cuayare	0.693	3465.712	0.449	0.563	0.441	0.693	0.804	0.425	0.677	0.680
Nazareth	0.854	10722.792	0.421	0.502	0.416	0.854	0.888	0.250	0.561	0.610
Roncador	0.802	8468.616	0.414	0.396	0.604	0.802	0.882	0.368	0.719	0.760
Tarapaca	0.530	2324.521	0.348	0.826	−0.096	0.530	0.588	0.298	0.414	0.457
Tres Cruces	0.579	1223.259	0.460	0.650	0.306	0.579	0.623	0.405	0.480	0.617
NEPAL										
Asaraghat	0.679	455.667	0.621	0.746	0.112	0.679	0.837	0.614	0.560	0.693
Babai	0.320	137.209	0.926	2.778	0.048	0.320	0.880	1.109	0.090	0.658
Bheri	0.694	315.363	0.727	0.889	0.393	0.695	0.811	0.637	0.507	0.730
Kaligandaki	0.690	427.634	0.773	0.953	0.393	0.690	0.849	0.654	0.493	0.727
Kamali	0.722	1132.257	0.624	0.772	0.446	0.722	0.858	0.584	0.605	0.739
Kankai	0.447	168.701	0.694	3.269	0.192	0.447	0.852	1.010	0.294	0.669
Marsyangdi	0.600	228.728	0.966	0.823	0.156	0.600	0.774	0.754	0.238	0.663
Narayani	0.702	1383.591	0.789	0.700	0.351	0.702	0.855	0.627	0.411	0.707
Rapti	0.411	235.208	0.710	2.277	0.085	0.411	0.906	1.032	0.317	0.717
Saptakosi	0.774	1103.025	0.432	0.536	0.527	0.774	0.912	0.511	0.696	0.789
Seti	0.691	234.586	1.173	1.212	0.258	0.691	0.653	0.674	0.271	0.651
Tinaukhola	0.135	631.093	1.840	17.618	−0.055	0.135	0.452	1.331	−0.279	0.695

In addition, ICIMOD has fully integrated the HMaaS services into their cyberinfrastructure (**Figure 10**). ICIMOD has improved the performance of their applications and data center by implementing a workflow that downloads daily forecast data during low demand times using the REST API. The stored data is then used for different applications during high demand.

The Dominican Republic provides another example where the web services and visualization tools have helped strengthen vulnerabilities. An array of derivative applications that take advantage of the REST API have been subsequently developed. These applications range from reservoir storage monitoring to flood mapping and risk management. In particular, the custom version of the HydroViewer app for the Dominican Republic provides another layer of information by combining the previously existing Flash Flood Guidance system (Georgakakos, 2006) with the developed GSPS (**Figure 11**).

In general, the development of the HydroViewer app and the REST API facilitate the adoption and integration of the developed streamflow prediction system by providing a lightweight application that can be easily deployed and customized to visualize and interpret results, and providing a way for results to be integrated and combined with existing resources through the use of the REST API.

4.1. Validation Results

We compared our historic simulation results to observed data from 20 different locations in Nepal and Colombia. **Figure 12** shows that the routed historic simulation successfully follows the same pattern as the observed data and captures most events with a tendency to under-predict. **Data sheet 1** has the charts for Colombia and **Data Sheet 2** the charts for Nepal. **Table 3** shows a summary of the error metrics when comparing forecasted results with observed data at the selected locations.

We performed an analysis to determine if our GloFAS-RAPID routed results were similar to the coarser GloFAS results. Data was collected for 9 weeks during the summer of 2017 and summarized weekly.

We found that GloFAS-RAPID provides a very similar result to the original GloFAS and follows trends with very similar shapes. This information demonstrates that even though GloFAS-RAPID is routing results over smaller watersheds, results from the same locations are still very similar in volume, with the main differences being the initialization methods used with each model, and the differences in the terrain and hydrography used for the routing. **Table S1** corresponds to the validation exercise.

Finally, we performed an analysis to determine if our selected watershed size for routing results had any effect or introduced any variability on forecasted results. This was done by comparing forecasted results at the mouth of a watershed using three different spatial decompositions of the watershed upstream.

As expected, the results from varying resolutions at the mouth of all the tested watersheds did not yield any significant differences in the results. These results are consistent with the fact that the RAPID preprocessing methodology assigns a percentage of the total runoff volume to each sub-basin. The sum of these volumes at the mouth of a watershed should always be about the same. Aside from initial validation, data validation for a large-scale forecast prediction system at specific locations is a complicated task. This is in part due to the extent covered by the model. Local involvement is necessary to validate results and to provide feedback about the model. The collaboration efforts described above, as well as the development of validation tools, and accessibility tools such as REST APIs that facilitate forecasted and observed data analyses, provide a long term approach to validating and improving overall model results at the local level.

5. CONCLUSION

The traditional hydrologic modeling approach presents a major barrier for areas that lack the necessary resources to run a model. A HMaaS was developed to answer the need for water information in areas lacking the resources to run their own models. A large-scale streamflow prediction system based on the ECMWF ensemble global runoff forecast. However, this new model presents a series of challenges to run in an operational environment and to make the resulting streamflow information useful at the local scale. These "hydroinformatic" challenges were divided into four categories: big data, data communication, adoption, and validation. The developed model provides a high-density result by routing runoff volume from ECMWF using the RAPID routing model. A HMaaS approach was used to provide an answer to the communication challenges faced by a model covering such a large area. A cloud cyberinfrastructure was developed to host model workflows, inputs, and outputs. Web applications were deployed to expose results over the Internet. Web services such as a REST API and geospatial services were created to provide accessibility to forecasted results. Additional web applications were created with the main goal to allow customizations and provide flexibility for local agencies to use results according to specific needs. These projects were demonstrated in different countries around the world. Some of these countries include: Argentina, Bangladesh, Brazil, Colombia, Haiti, Peru, Nepal, Tanzania, the Dominican Republic, and the United States. We tested our results by comparing our forecasts to observed data. We determined that our model results are in essence the same as the GloFAS results, but in a higher density. We also determined that the our forecasted results are usually close to observed values and are able to capture most extreme events. Finally, we analyzed the effect of density variations on our model, and determined that sub-basin sizes do not significantly affect results at the mouth of the watershed.

DATA AVAILABILITY STATEMENT

The high-density results from our GloFAS-RAPID model runs can be accessed through the SPT or for a specific area using the HydroViewer app. These apps are currently available online at two different portals: the NASA SERVIR app portal (https://tethys.servirglobal.net/apps/), and the BYU app portal (https://tethys.byu.edu/apps/). The source code

for the latest version of the SPT can be found at https://github.com/BYU-Hydroinformatics/tethysapp-streamflow_prediction_tool, while detailed documentation including installation and use can be found at https://byu-streamflow-prediction-tool.readthedocs.io/en/latest/. The source code and documentation for the HydroViewer app can be found at https://github.com/BYU-Hydroinformatics/hydroviewer.

AUTHOR CONTRIBUTIONS

EN, DA, and NJ helped conceive and guide the research for this work. MS helped conceive the original idea of Hydrologic Modeling as a Service as well as carry the bulk of the research and implementation of the global model presented. KS developed a derivative tool that makes use of the model service through its REST API and helped integrate the services in ICIMOD. CE carried the multiple resolution comparison analysis presented. WR developed the HydroStats package and performed the comparison analysis using observed data. CK compared the presented model to the original GloFAS model. AG guided efforts in comparing the traditional and hydrologic modeling as a service approaches through activities in the Group on Earth Observations (GEO).

FUNDING

This work was supported by the NASA ROSES SERVIR Applied Research Grant NNX16AN45G.

ACKNOWLEDGMENTS

The authors wish to recognize collaborators: Cedric David and Alan Snow who developed key software; Christel Prudhomme, Florian Pappenberger, and Ervin Zsoter from ECMWF and Peter Salamon of JRC; the Copernicus Emergency Management System that made GloFAS forecasts available; NASA SERVIR for opportunity to test and validate model results; the Esri team who helped develop the global services visualization; and Microsoft's AI for Earth who provided cloud services necessary to run the model.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2019.00158/full#supplementary-material

REFERENCES

Ahmadisharaf, E., Kalyanapu, A. J., and Chung, E.-S. (2016). Spatial probabilistic multi-criteria decision making for assessment of flood management alternatives. *J. Hydrol.* 533, 365–378. doi: 10.1016/j.jhydrol.2015.12.031

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., et al. (2013). GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.* 17, 1161–1175. doi: 10.5194/hess-17-1161-2013

Almoradie, A., Jonoski, A., Popescu, I., and Solomatime, D. (2013). Web based access to water related data using OGCWaterML 2.0. *Int. J. Adv. Comput. Sci. Appl.* 3, 83–89. doi: 10.14569/SpecialIssue.2013.030310

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., et al. (2008). A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the integrated forecast system. *J. Hydrometeorol.* 10, 623–643. doi: 10.1175/2008JHM1068.1

Beran, B., and Piasecki, M. (2009). Engineering new paths to water data. *Comput. Geosci.* 35, 753–760. doi: 10.1016/j.cageo.2008.02.017

Butts, M. B., Payne, J. T., Kristensen, M., and Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *J. Hydrol.* 298, 242–266. doi: 10.1016/j.jhydrol.2004.03.042

Choudhary, V. (2007). "Software as a service: implications for investment in software development," in *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)* (Waikola, HI: IEEE), 209a. doi: 10.1109/HICSS.2007.493

Cools, J., Innocenti, D., and O'Brien, S. (2016). Lessons from flood early warning systems. *Environ. Sci. Policy* 58, 117–122. doi: 10.1016/j.envsci.2016.01.006

David, C. H., Famiglietti, J. S., Yang, Z. L., Habets, F., and Maidment, D. R. (2016). A decade of RAPID–Reflections on the development of an open source geoscience code. *Earth Space Sci.* 3, 226–244. doi: 10.1002/2015EA000142

Davis, F. D. Jr. (1986). *A technology acceptance model for empirically testing new end-user information systems: theory and results (dissertation).* Ph.D. thesis, Massachusetts Institute of technology, Cambridge, MA, United States.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 553–597. doi: 10.1002/qj.828

Demeritt, D., Nobert, S., Cloke, H. L., and Pappenberger, F. (2013). The European Flood Alert System and the communication, perception, and use of ensemble predictions for operational flood risk management. *Hydrol. Process.* 27, 147–157. doi: 10.1002/hyp.9419

Duan, Y., Fu, G., Zhou, N., Sun, X., Narendra, N. C., and Hu, B. (2015). "Everything as a service (XaaS) on the cloud: origins, current and future trends," in *2015 IEEE 8th International Conference on Cloud Computing* (Waikola, HI: IEEE), 621–628. doi: 10.1109/CLOUD.2015.88

Georgakakos, K. P. (2006). Analytical results for operational flash flood guidance. *J. Hydrol.* 317, 81–103. doi: 10.1016/j.jhydrol.2005.05.009

Godschalk, D. R., Rose, A., Mittler, E., Porter, K., and West, C. T. (2009). Estimating the value of foresight: aggregate analysis of natural hazard mitigation benefits and costs. *J. Environ. Plan. Manage.* 52, 739–756. doi: 10.1080/09640560903083715

Hallegatte, S. (2012). *A Cost Effective Solution to Reduce Disaster Losses in Developing Countries: Hydro-Meteorological Services, Early Warning, and Evacuation*. Waikola, HI: The World Bank. doi: 10.1596/1813-9450-6058

Horita, F. E. A., de Albuquerque, J. P., Degrossi, L. C., Mendiondo, E. M., and Ueyama, J. (2015). Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks. *Comput. Geosci.* 80, 84–94. doi: 10.1016/j.cageo.2015.04.001

Jackson, E. K., Roberts, W., Nelsen, B., Williams, G. P., Nelson, E. J., and Ames, D. P. (2019). Introductory overview: error metrics for hydrologic modelling–A review of common practices and an open source library to facilitate use and adoption. *Environ. Modell. Softw.* 119, 32–48. doi: 10.1016/j.envsoft.2019.05.001

Kelman, I. (2013). *Disaster Mitigation Is Cost Effective*. Washington, DC: World Bank.

Krause, P., Boyle, D. P., and Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* 5, 89–97. doi: 10.5194/adgeo-5-89-2005

Kull, D., Mechler, R., and Hochrainer-Stigler, S. (2013). Probabilistic cost-benefit analysis of disaster risk management in a development context. *Disasters* 37, 374–400. doi: 10.1111/disa.12002

Lehner, B., Verdin, K., and Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *EOS Trans. Am. Geophys. Union* 89, 93–94. doi: 10.1029/2008EO100001

Li, Z., Yang, C., Huang, Q., Liu, K., Sun, M., and Xia, J. (2017). Building model as a service to support geosciences. *Comput. Environ. Urban Syst.* 61, 141–152. doi: 10.1016/j.compenvurbsys.2014.06.004

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., and Arheimer, B. (2010). Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrol. Res.* 41, 295–319. doi: 10.2166/nh.2010.007

Niswonger, R. G., Allander, K. K., and Jeton, A. E. (2014). Collaborative modelling and integrated decision support system analysis of a developed terminal lake basin. *J. Hydrol.* 517, 521–537. doi: 10.1016/j.jhydrol.2014.05.043

NOAA (2016). *National Water Model: Improving NOAA's Water Prediction Services*. Waikola, HI.

Rew, R., and Davis, G. (1990). NetCDF: an interface for scientific data access. *IEEE Comput. Graph. Appl.* 10, 76–82. doi: 10.1109/38.56302

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., et al. (2004). The global land data assimilation system. *Bull. Am. Meteorol. Soc.* 85, 381–394. doi: 10.1175/BAMS-85-3-381

Roman, D., Schade, S., Berre, A. J., Bodsberg, N. R., and Langlois, J. (2009). "Model as a service (MaaS)," in *AGILE Workshop: Grid Technologies for Geospatial Applications* (Hanover).

Snow, A. D., Christensen, S. D., Swain, N. R., Nelson, E. J., Ames, D. P., Jones, N. L., et al. (2016). A high-resolution national-scale hydrologic forecast system from a global ensemble land surface model. *J. Am. Water Resour. Assoc.* 52, 950–964. doi: 10.1111/1752-1688.12434

Souffront Alcantara, M. A., Crawley, S., Stealey, M. J., Nelson, E. J., Ames, D. P., and Jones, N. L. (2017). Open water data solutions for accessing the national water model. *Open Water J.* 4:3. Available online at: https://scholarsarchive.byu.edu/openwater/vol4/iss1/3

Sperna Weiland, F. C., Vrugt, J. A., van Beek, R. L., Weerts, A. H., and Bierkens, M. F. (2015). Significant uncertainty in global scale hydrological modeling from precipitation data errors. *J. Hydrol.* 529, 1095–1115. doi: 10.1016/j.jhydrol.2015.08.061

Svoboda, M. D., Fuchs, B. A., Poulsen, C. C., and Nothwehr, J. R. (2015). The drought risk atlas: Enhancing decision support for drought risk management in the United States. *J. Hydrol.* 526, 274–286. doi: 10.1016/j.jhydrol.2015.01.006

Swain, N. R., Christensen, S. D., Snow, A. D., Dolder, H., Espinoza-Dávalos, G., Goharian, E., et al. (2016). A new open source platform for lowering the barrier for environmental web app development. *Environ. Modell. Softw.* 85, 11–26. doi: 10.1016/j.envsoft.2016.08.003

Wan, Z., Hong, Y., Khan, S., Gourley, J., Flamig, Z., Kirschbaum, D., and Tang, G. (2014). A cloud-based global flood disaster community cyber-infrastructure: development and demonstration. *Environ. Modell. Softw.* 58, 86–94. doi: 10.1016/j.envsoft.2014.04.007

Wilhite, D. A., Sivakumar, M. V. K., and Pulwarty, R. (2014). Managing drought risk in a changing climate: the role of national drought policy. *Weather Climate Extremes* 3, 4–13. doi: 10.1016/j.wace.2014.01.002

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., et al. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resour. Res.* 47, 1–10. doi: 10.1029/2010WR010090

Conflict of Interest: MS and CK were employed by Aquaveo LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 23 October 2019
doi: 10.3389/fenvs.2019.00155

An Open-Source Tool to Facilitate the Processing of GRACE Observations and GLDAS Outputs: An Evaluation in Bangladesh

Adam J. Purdy[1,2], *Cédric H. David*[1]*, *Md. Safat Sikder*[1], *John T. Reager*[1], *Hrishikesh A. Chandanpurkar*[1], *Norman L. Jones*[3] and *Mir A. Matin*[4]

[1] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States, [2] Department of Environmental Science, University of San Francisco, San Francisco, CA, United States, [3] Department of Civil and Environmental Engineering, Brigham Young University, Provo, UT, United States, [4] International Centre for Integrated Mountain Development, Kathmandu, Nepal

Bangladesh lies at the intersection of the Ganges, Brahmaputra, and Meghna rivers with a combined average discharge of 38,000 m^3s^{-1} ranking fourth globally. Despite the volume of water flowing through and seasonally inundating parts of the landscape, groundwater reliance is necessary to support an intensive agricultural industry. Here we use newly-developed open-source software to combine observations from the Gravity Recovery and Climate Experiment (GRACE) satellites with hydrologic estimates of land water storage from the Global Land Assimilation Data System (GLDAS) to isolate basin-scale groundwater anomalies in Northwest Bangladesh from 2002 to 2016. We place our estimates in the context of previously-published water management estimates and our results suggest the largest losses in water storage are due to groundwater abstractions with groundwater storage decreasing at a rate of 0.88 cm yr^{-1}. We estimate basin-averaged total water storage loss from 2002 to 2016 at 27.92 cm with groundwater and surface water storage loss accounting for 12.46 cm or 44.6%. For Bangladesh, a region where 80% of landcover is dedicated for agricultural use and over half of the country's population is employed in the agricultural sector, the estimated declines in water storage hold long-term implications for the livelihood and food supply of the region.

Keywords: water resources, groundwater, sustainable management, food security, open-source software

OPEN ACCESS

Edited by:
Ashutosh S. Limaye, National Aeronautics and Space Administration (NASA), United States

Reviewed by:
Hyongki Lee, University of Houston, United States
Augusto Getirana, National Aeronautics and Space Administration, United States

***Correspondence:**
Cédric H. David
cedric.david@jpl.nasa.gov

Specialty section:
This article was submitted to Freshwater Science, a section of the journal Frontiers in Environmental Science

Received: *31 May 2019*
Accepted: *25 September 2019*
Published: *23 October 2019*

Citation:
Purdy AJ, David CH, Sikder MS, Reager JT, Chandanpurkar HA, Jones NL and Matin MA (2019) An Open-Source Tool to Facilitate the Processing of GRACE Observations and GLDAS Outputs: An Evaluation in Bangladesh. Front. Environ. Sci. 7:155. doi: 10.3389/fenvs.2019.00155

INTRODUCTION

Having a holistic understanding of water availability is necessary to combat the challenges of managing water resources under a changing climate (Famiglietti and Rodell, 2013; Rodell et al., 2018). With evidence of more intense rainfall (Durack et al., 2012), an increased probability of flooding (Milly et al., 2002), decreasing glacier melt (Gardner et al., 2013), changes in seasonal snowpack (Smith and Bookhagen, 2018), and shifts in the intensity of drought (Trenberth et al., 2014), water managers require as much relevant information as possible to meet water needs of a growing populations to support food and energy production (Vörösmarty et al., 2000; Milly et al., 2005). Traditionally humans have met agricultural and urban water demands by routing surface water or pumping groundwater. While *in situ* surface water observations have provided rich datasets to support management (Fekete et al., 2012), limited observations of groundwater

have prevented a complete understanding of water availability and use to support sustainable management (Taylor et al., 2013; Famiglietti, 2014).

Globally, groundwater provides a drinking water source for half of the world (IGRAC, 2018) and supports over 40% of irrigated lands worldwide (Siebert et al., 2010). While groundwater use varies greatly region to region depending on infrastructure and climate, all regions increase reliance under drought (Famiglietti et al., 2011; Castle et al., 2014). With 5 billion people expected to feel the impacts on freshwater availability from a changing climate, regions that are already over-reliant on groundwater will be more susceptible to devastating consequences of degraded quality and reduced availability (Vörösmarty et al., 2000). Unfortunately, many regions still lack the infrastructure to monitor changes in availability across basins scales.

In 2002, with the launch of the Gravity Recovery And Climate Experiment (GRACE) mission (Tapley et al., 2004), the global extent of this shared problem of over-reliance on groundwater began to emerge (Famiglietti et al., 2011; Voss et al., 2013; Richey et al., 2015). Numerous studies have utilized GRACE observations of terrestrial water storage anomalies (TWSa), the combined anomalies of snow, soil moisture, canopy water, surface water, and groundwater, to study the global water cycle and impacts on water resources (Equation 1).

$$TWSa = SMa + SWEa + CANa + SWa + GWa \quad (1)$$

where, *SMa* is soil moisture anomaly, *SWEa* is the snow water equivalent anomaly, *CANa* is canopy intercepted water anomaly, and *SWa* is surface water storage anomaly. GRACE observations have supported quantifying: flood potential (Reager and Famiglietti, 2009), drought (Thomas et al., 2014), basin evapotranspiration (Rodell et al., 2011), global discharge (Chandanpurkar et al., 2017), and the fingerprint of human management on the water cycle at basin scales (Anderson et al., 2012; Castle et al., 2016; Massoud et al., 2018). GRACE's unique ability to quantify total terrestrial water storage anomalies facilitates tracking rates of groundwater depletion across the globe (Famiglietti, 2014). From India and the Middle East to the California Central Valley and Colorado River Basin, GRACE observations have revealed how each region relies on groundwater to meet freshwater demands from irrigated farmland, industry, and growing populations (Rodell et al., 2009; Famiglietti et al., 2011; Voss et al., 2013; Castle et al., 2014). Many of the regions with evidence of water loss have limited surface water availability, other regions, such as Bangladesh, show evidence of persistent storage declines despite high rates of mean annual rainfall and river discharge (Shamsudduha et al., 2012; Burgess et al., 2017).

Bangladesh lies at the confluence of the Ganges, Brahmaputra, and Meghna Rivers with a mean discharge topping 38,000 $m^3 s^{-1}$ (Gain et al., 2011). The volume of water flowing in these rivers makes up large fraction of the total water within the country (Getirana et al., 2017). Furthermore, Bangladesh experiences seasonally intense rainfall during the monsoon season from June to September (Dash et al., 2012). In fact, some Eastern parts of the country experience more than 2 m yr^{-1} of annual rainfall, among the highest rates in the world. Despite seasonal inundations on much of the landscape, a large agriculture industry relies on groundwater to irrigate farmland (**Figure 1**). In Bangladesh, where over 160 million people live, more than half the population relies on agriculture to support their livelihood (Food Agriculture Organization, 2011). The importance of agriculture to sustain food production and local workforces places a strain on regional water resources. Previous *in situ* measurements of documented declines in groundwater spurred legislation in 1999 and 2001 to curtail over-use (Food Agriculture Organization, 2011). The Northwest Region is home to the highest fractions of irrigated farmland in the country and relies heavily on groundwater during the dry season to maintain crop production. Unfortunately, this happens to be the area in Bangladesh most susceptible to drought and its impacts (Dey et al., 2012; Alamgir et al., 2015). Water management challenges in Bangladesh are compounded by water contamination (e.g., arsenic) and salinity intrusion (Nickson et al., 2000; Mahmuduzzaman et al., 2014). During prolonged dry years the arsenic-rich groundwater being pumped to the surface for irrigation accumulates in soils used to grow rice (Roberts et al., 2010). In addition, projected increases in peak flow intensity may limit this region's ability to store surface water (Gain et al., 2011). Like many similar regions around the world

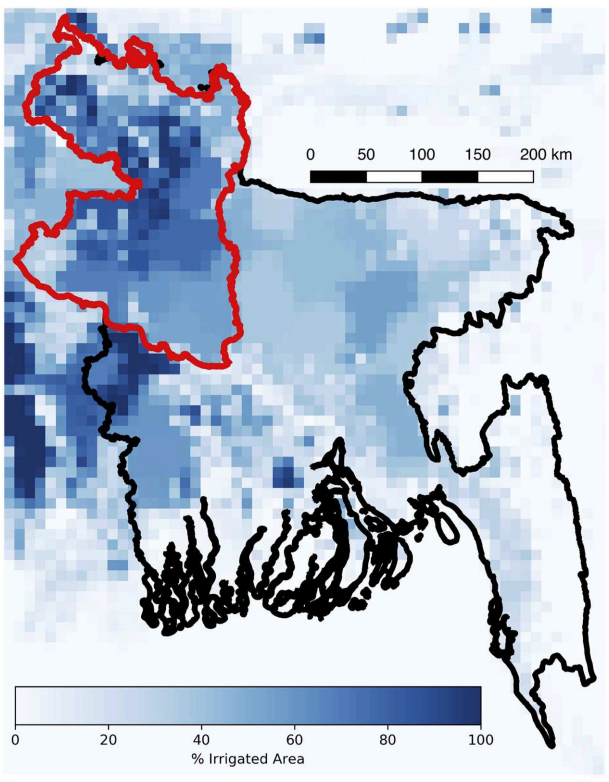


FIGURE 1 | Fraction of irrigated area in Bangladesh (Siebert et al., 2013). Country boundary in black, Northwest Bangladesh region (i.e., Rangpur and Rajshahi divisions) boundary in red. The irrigated area data were accessed from FAO AQUASTAT http://www.fao.org/nr/water/aquastat/irrigationmap/.

facing complex water management challenges, regional scale observations of water storage changes hold potential to support long-term planning (Adhikary et al., 2013).

Here we use 159 months of GRACE data, from April 2002 through January 2017, to evaluate water storage changes in Bangladesh. To isolate changes in groundwater storage we supplement GRACE observations with model estimates of soil moisture, snow, and canopy storage. While GRACE observations (2002–2007) have been demonstrated to accurately capture the seasonal water cycle including the monsoon intensity over the entire Bengal Basin (Steckler et al., 2010; Shamsudduha et al., 2012), regions with the most intense groundwater use have yet to be analyzed in isolation and the shorter records used in previous analysis limited the computation of regional trends. In addition to a national scale analysis over Bangladesh, we focus additional analysis on Northwest Bangladesh where irrigated farmland and groundwater use is among the highest in the country (**Figure 1**). Furthermore, we here present the Satellite Hydrology Bits Analysis and Mapping (SHBAAM), an entirely open-source python-based tool, to perform regional-scale water budget analysis using GRACE observations and model data for a given region of interest.

DATA AND METHODS

Study Region

In Rangpur and Rajshahi, the two Divisions that make up Northwest Bangladesh (**Figure 1**), groundwater is the predominant source of water for irrigation. Compared to the rest of the country, this region experiences substantially less rainfall (Dash et al., 2012). Northwest Bangladesh is bordered to the South by the Padma River, the main distributary of the Ganges River, to the West and North by India, and to the East by the Jamuna River, the lower course of the Brahmaputra River. While seasonal inundation occurs within the floodplains of these major rivers, farmers still pump groundwater during the dry season to support crops such as rice, wheat, potato and fruits including lychees and mangos, among others (Food Agriculture Organization, 2011). On top of the large presence of agriculture, Bangladesh is one of the most densely populated countries in the world with more than 1,100 people per km^2. These two factors exacerbate stress on water resources and create a need to bring the regional water back into balance.

GRACE Terrestrial Water Storage Observations

In this analysis, we use the Jet Propulsion Laboratory Release 5 GRACE mascon solutions from April 2002 through January 2017 (Watkins et al., 2015). This dataset provides mass equivalent terrestrial water storage (TWS) anomalies on a 0.5° grid. Gain factors, generated with CLM4 model outputs, were used as multiplying factors along with the GRACE TWS anomalies to minimize leakage errors and distribute mass changes within each 3° mascon solution (Wiese et al., 2016). Additionally, we remove the temporal mean for the entire time period and present TWS anomalies relative to the complete record from 2002 to 2017. To evaluate changes in all the components of the water budget, we supplement GRACE observations with model estimates of snow water equivalent, canopy storage, soil moisture storage, and surface water storage; as discussed below.

GLDAS Land Surface Model Elements

Outputs from the Global Land Data Assimilation System (GLDAS, Rodell et al., 2004) provide multiple land surface model (LSM) estimates globally. GLDAS utilizes observation-based forcing datasets of precipitation and radiation to run land surface models at 3-hourly timesteps to globally resolve fluxes and storage of water and energy. We use monthly outputs from four models within GLDAS including: Noah (NOAH) (Chen et al., 1996, 1997; Ek, 2003), Variable Infiltration Capacity (VIC) (Liang et al., 1994), Common Land Model (CLM) (Dai et al., 2003), and Mosaic (MOS) (Koster et al., 2000). Each LSM provides an independent estimate of snow water equivalent, soil moisture, and canopy water storage. While each LSM uses the a common set of soil types and atmospheric forcing, the maximum depth of soil water storage varies from 1.9 m (VIC), 2.0 m (NOAH), 3.0 m (MOS), to 3.43 meters (CLM), and the total number of layers ranges from 3 (VIC and MOS), 4 (NOAH), to 10 (CLM). Surface water storage, such as lakes, reservoirs, and rivers, is not directly resolved by any of the GLDAS models. Instead, we apply the Routing Application for Parallel computation of Discharge (RAPID) river routing model (David et al., 2011) to estimate mean surface water storage for each GLDAS model across every reach in the Ganges, Brahmaputra, and Meghna Basins (Sikder et al., in press). Sikder et al. (in press) simulated 3-hourly discharge using all GLDAS models for the Ganges-Brahmaputra-Meghna Basin on a 15-arc second river network from HydroSHEDS (Lehner et al., 2008). These discharge simulations result in standard errors $\leq$2% with correlation coefficients ranging from 0.6 for the Meghna River to 0.8 for the Brahmaputra River. We sum the volume from all the river reaches within Bangladesh and compute monthly mean surface water storage and associated uncertainty using multi-model statistics (Richey et al., 2015; Thomas et al., 2017). To compare LSM and river routing outputs with GRACE TWS anomalies, each storage component was converted to an equivalent water height anomaly by removing the temporal mean from April 2002 to January 2017 and dividing by the surface area.

Computing Groundwater Anomalies With SHBAAM

The Satellite Hydrology Bits Analysis and Mapping (SHBAAM) software developed here completes the pre and post-processing tasks on GRACE TWS anomalies that are necessary to study the terrestrial water cycle and compute changes in groundwater. SHBAAM is an open-source Python and bash shell toolbox that is available online via GitHub (https://github.com/c-h-david/shbaam) or Docker (https://hub.docker.com/r/chdavid/shbaam). This toolbox automates the downloading and processing of the GRACE and GLDAS datasets to evaluate changes in water storage for any region around the globe. Groundwater anomalies for a given region are computed by rearranging (Equation 1):

$$GWa = TWSa - SMa - SWEa - CANa - SWa \quad (2)$$

SHBAAM simplifies this computation by preparing and subsetting the necessary prerequisite data to the same temporal resolution and spatial domain. Scripts retrieve GRACE terrestrial water storage anomalies, land-ocean masks, and gain factors from the Physical Oceanography Distributed Active Archive Center (PO.DAAC), and individual monthly GLDAS snow water equivalent, soil moisture, and canopy storage from the Goddard Earth Science Data and Information Service Center (GES-DISC). Missing GRACE observations are automatically filled according to Hamlington et al. (2019). First, GRACE *TWSa* data are multiplied by spatially distributed gain factors. Then the scaled *TWSa* data are detrended and the annual climatology is removed. Gaps in the remaining data, the interannual variation, are filled using cubic interpolation. Lastly, the climatology and trend are reintroduced to provide a continuous *TWSa* record. These gap-filling steps provide more accurate estimates of missing observations during winter maximums and summer minimums. Additional processing steps include concatenation of the individual monthly GLDAS LSM outputs and LSM anomaly computation to compare with GRACE TWS anomaly observations. Any shapefile in a geographic coordinate reference system can be used in SHBAAM for the joint regional analysis of GRACE and GLDAS. SHBAAM finds the intersecting grid coordinates for both GRACE and GLDAS and computes area-weighted averages for each water storage component to determine groundwater storage anomalies within the domain following (Equation 2). Groundwater anomalies are computed using the multi-model mean values. Outputs from SHBAAM include tables of monthly water storage anomalies and figures of the time-series changes in each component (**Figure 2**). These output tables support supplemental analyses on basin water storage changes, such as trend analysis or quantifying the total volume of water lost.

Error variances in *SWEa*, *SMa*, and *CANa* are computed using the multi-model standard deviation. The errors in *TWSa* are on average 7 cm for the study region. Errors in monthly groundwater assume the absence of error covariances and are computed as:

$$\sigma GWa = \sqrt{(\sigma TWSa)^2 - (\sigma SMa)^2 - (\sigma SWEa)^2 - (\sigma CANa)^2 - (\sigma SWa)^2} \quad (3)$$

where σXa is the standard deviation or error variance in each hydrologic component *X*. SHBAAM can be used to reproduce the results from this study or to complete similar analyses on any basin globally granted that the basin size is large-enough to be used along with GRACE data.

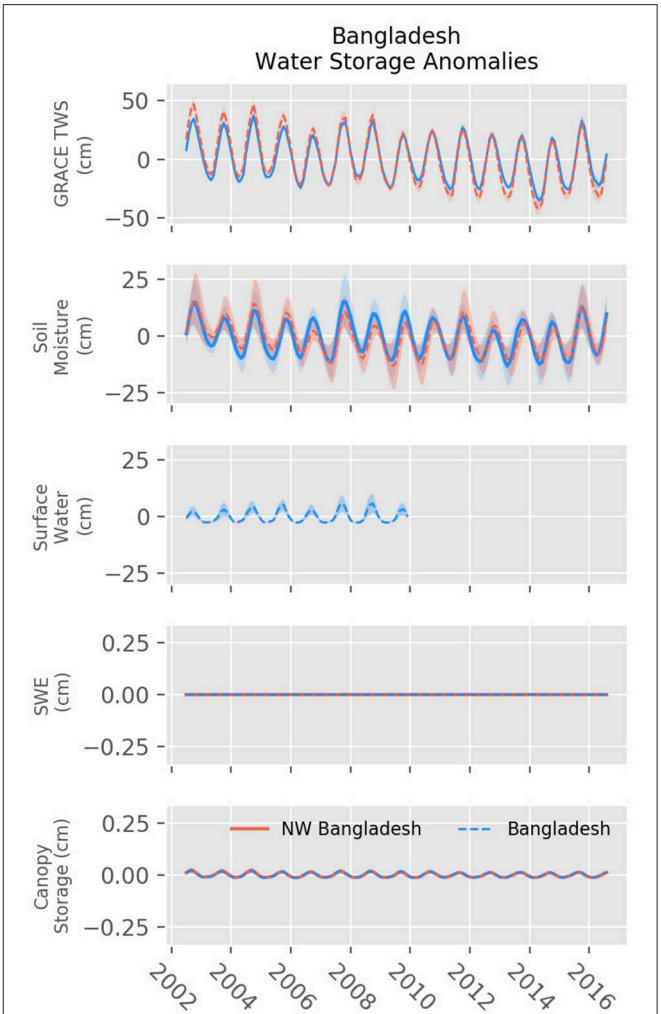


FIGURE 2 | Changes in water storage anomalies for GRACE terrestrial water storage (Top), soil moisture (2nd from Top), surface water (3rd from Top), snow water equivalent (4th from Top), and canopy water storage (Bottom). Red lines show changes in storage for Bangladesh. Blue dashed lines show changes in storage for Northwest Bangladesh. GRACE observations show a decline in storage from 2003 to 2016.

RESULTS

Hydrographs for each water storage component from GRACE and GLDAS reveal variations in water storage and similar annual amplitudes for terrestrial water storage and soil moisture (**Figure 2**). In both Bangladesh and Northwest Bangladesh snow water equivalent and canopy water storage anomalies are negligible. The SHBAAM outputs show declines in GRACE terrestrial water storage anomalies from 2002 to 2017 for both Bangladesh and Northwest Bangladesh. Trends in *TWSa* reveal average storage declines at a rate of −0.85 cm yr^{-1} (−1.16 km^3 yr^{-1}) and −1.99 cm yr^{-1} (−0.60 km^3 yr^{-1}), respectively. Soil moisture anomalies show similar declines from 2002 to 2015, but a recovery in 2016. Non-zero trends for each water storage component are listed in **Table 1**. We estimate total storage lost for each region using the trend of decline and the area. From 2002 to 2017 GRACE observations reveal Bangladesh lost a total of 16.26 km^3 freshwater. During the same time period we find the Northwest Region lost 8.42 km^3 of water. Outputs from SHBAAM reveal two hot-spots of water decline in the Northwest region (**Figure 3**).

Groundwater anomalies uncover large declines in storage across the time period (**Figure 4**). Groundwater storage decreases across Bangladesh at a rate of −0.75 cm yr^{-1} (−1.02 km^3

yr^{-1}) and in Northwest Bangladesh at a rate of -0.88 cm yr^{-1} (-0.27 km^3 yr^{-1}). These rates are consistent with previously published ranges of *in situ* estimates (-0.85 to -1.61 km^3 yr^{-1}) (Shamsudduha et al., 2012). During the study period, Bangladesh lost approximately 14.3 km^3 with the Northwest region accounting for roughly 3.74 km^3. Interestingly, we find changes in surface water storage anomalies, as computed from routed GLDAS runoff do not change the trajectory of Bangladesh-wide changes in groundwater anomalies (**Supplementary Figure 1**). The Rangpur and Rajshahi Divisions lie in regions more susceptible to pre-monsoon and post-monsoon droughts and are known to rely more heavily on groundwater to irrigate farmland.

TABLE 1 | Trends for water storage components in Bangladesh and Northwest Bangladesh.

Component	Bangladesh	Northwest Bangladesh
Total Water Storage	-0.890 cm yr^{-1}	-2.24 cm yr^{-1}
Soil Moisture	-0.134 cm yr^{-1}	-1.355 cm yr^{-1}
Groundwater	-0.755 cm yr^{-1}	-0.886 cm yr^{-1}

Trends are computed for 2003–2016. Trends of 0 cm yr^{-1} are not displayed for canopy water storage or snow water equivalent.

We compare changes in water storage and groundwater storage for all of Bangladesh to published values for 2008 (Food Agriculture Organization, 2011). Previously reported total water withdrawal for 2008 is estimated at 35.87 km^3, with groundwater extraction accounting for approximately 80% or 28.42 km^3. The same report estimates groundwater recharge to be 21 km^3 yr^{-1}. Here we estimate the amplitude of groundwater storage for 2007, 2008, and the mean climatology for the GRACE record to be 30.7, 36.0, and 27.6 km^3, respectively (**Figure 5**). Our estimates are therefore of the same order of magnitude as study previous, although they notably differ by over 30%. This disagreement can be attributed to numerous factors, including: (1) a potential underestimation in water withdrawal estimates used previously for the region or (2) known limitations in the

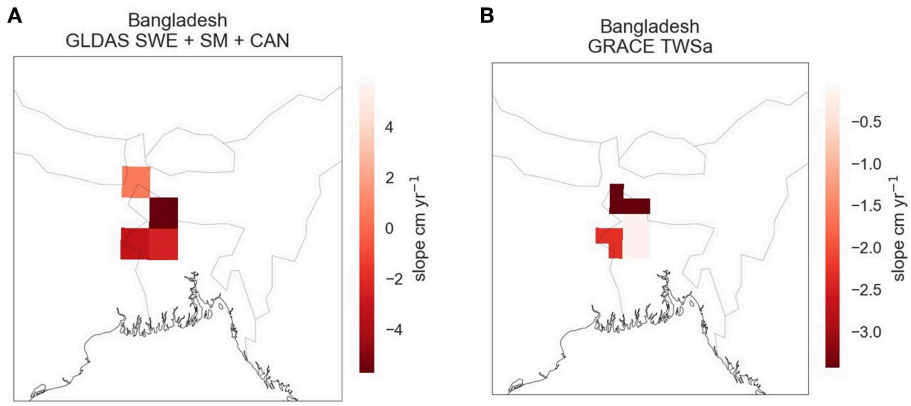


FIGURE 3 | Changes in water storage represented as the slope from 2002 to 2017 for the sum of GLDAS water storage components **(A)** and GRACE total water storage **(B)**. The plots demonstrate the difference in spatial resolution of the two datasets (GLDAS 1° and GRACE 0.5°).

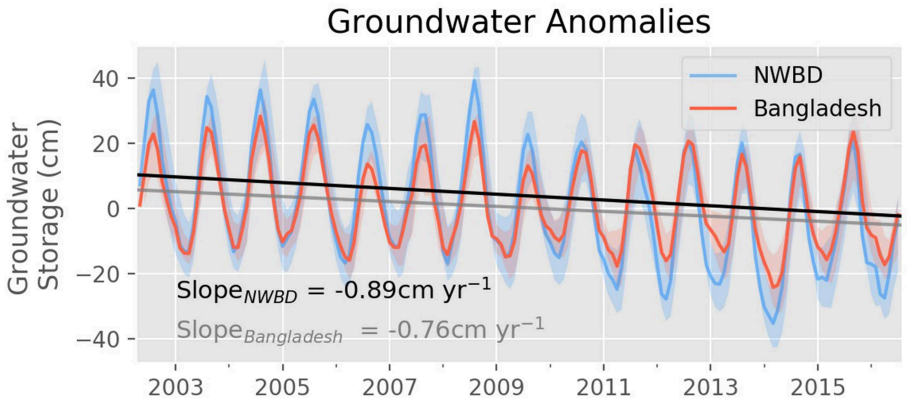


FIGURE 4 | Changes in groundwater anomalies for Bangladesh (red line) and Northwest Bangladesh (blue line). Slopes for Bangladesh (gray) and Northwest Bangladesh (black) show a decline in groundwater from 2002 to 2017. Shaded regions represent monthly groundwater uncertainty for each study region as computed by Equation (3).

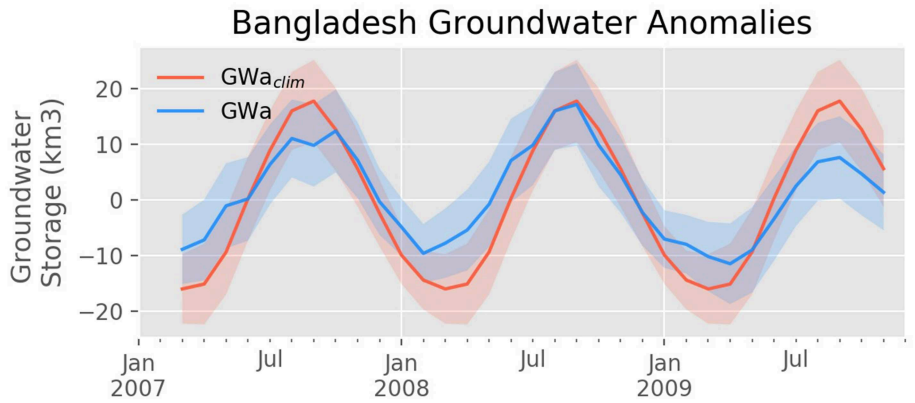


FIGURE 5 | Changes in groundwater storage with uncertainty for Bangladesh. The blue line indicates the monthly changes in groundwater storage and the red line indicates the corresponding monthly climatology. Shaded regions represent monthly groundwater uncertainty as computed by Equation (3).

combination of GRACE and GLDAS to estimate changes in groundwater storage (Scanlon et al., 2018), (3) high uncertainty in surface water storage anomalies, or (4) natural aquifer recharge and discharge processes make up for the difference. Nonetheless, our methodology provides an alternative means to estimating regional water storage changes from remotely-sensed observations combined with global hydrologic simulations, and is here indicative of over-consumptive use.

DISCUSSION

Despite large amounts of annual discharge from the Ganges, Brahmaputra, and Meghna, GRACE observations suggest that intense agricultural practices sustained water storage declines from 2002 to 2017 in Bangladesh. Previous exploration of this regional signal found changes in precipitation being the driving cause for total water storage declines (Rodell et al., 2018). Here, we find evidence that groundwater depletion contributes to this regional signal, especially in Northwest Bangladesh. Despite observations and model estimates indicating potentially dire circumstances, further on-ground validation is required to validate the magnitude of potential groundwater over-reliance. Recent work revealed GRACE observations and GLDAS modeled total water storage disagree on trend direction across the majority of global river basins (Scanlon et al., 2018). We found consistent trends across models in Northwest Bangladesh but contrasting trends across the Bangladesh study region (**Supplementary Figure 2**). Multi-model statistics help quantify the uncertainty in the monthly GWa estimates but relating the SHBAAAM outputs to previous studies can reveal opportunities to further improve these estimates. For example, the ratio of SWSa to TWSa estimated by GLDAS is lower than previous studies using the Noah-MP model with HyMAP river routing (Getirana et al., 2017). Further efforts to quantify seasonal and inter-annual surface water storage anomalies in the region may shed a clearer light on the degree of severity for particular regions.

While the uncertainty of groundwater anomalies computed with GRACE increases for smaller scale analyses, our results still capture trends and spatial patterns of groundwater reliance similar to previous *in situ* analysis for the region (Shamsudduha et al., 2012). The findings from this study suggest groundwater use exceeds natural recharge rates in Bangladesh extending to as recently as 2017, especially in the Rangpur and Rajshah Divisions. For this region, projected changes in surface discharge pose to only further complicate regional water management by increasing seasonal groundwater reliance and increasing the difficulty of storing surface water with larger peak flows. In addition to these potential challenges in managing freshwater quantity, Bangladesh also faces rising sea levels that intrude on coastal aquifers, and arsenic accumulation in farm soils during dry spells (Roberts et al., 2010; Mahmuduzzaman et al., 2014). These local issues mirror similar challenges faced around the world that require as much information as possible to understand and plan for actionable solutions.

GRACE and GLDAS-based hydrologic estimates provide a potential tool for water managers to support long-term management strategies by evaluating regional changes with relatively low latency, constraining basin-scale modeled estimates, and identifying regions facing or moving toward severe water shortages. The outputs from SHBAAM are not meant to replace *in situ* monitoring as many water management decisions are made at scales finer than the GRACE satellites can resolve. Instead, these datasets hold potential to support regional and seasonal management by complementing efforts focused on short-term forecasting of transboundary water flows (e.g., Biancamaria et al., 2011). SHBAAM lowers the barriers to access and use of these NASA datasets to support water management and scientific research. The increased exposure to these datasets may lead to new, creative avenues for water managers to harness the value of this tool and relevant associated datasets.

Since its generation, successful SHBAAM tutorials have been given to trainers at international agency partners through the NASA SERVIR program. However, creating and maintaining open-source software such as SHBAAM does not come without challenges. Hosting the toolbox and documentation on Github and Docker reduces many of the challenges such as installation requirements for different operating systems. Recent effort has

moved toward simplifying the coding interface through the introduction of Python-backed notebooks. These steps can be expected to further broaden the user base, and overcome a general lack of familiarity around command line interfaces outside of the research community. Additionally, with the recently launched GRACE-Follow On and improvements to GLDAS, updates to SHBAAM processing will be required to leverage near-real time observations and guide water management into the next decade (Tapley et al., 2019).

DATA AVAILABILITY STATEMENT

The software and datasets generated for this study can be found in the https://github.com/c-h-david/shbaam. Data downloaded by this software are freely available with a NASA EarthData Login. GRACE data are available at https://podaac-opendap.jpl.nasa.gov. GLDAS model output are available at https://disc.gsfc.nasa.gov.

AUTHOR CONTRIBUTIONS

CD and MM conceived this study. AP, CD, and MS completed this analysis. All authors contributed to writing the manuscript.

FUNDING

AP, CD, MS, JR, and HC were supported by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA, including a grant from the SERVIR Applied Sciences Team. MM was supported by USAID under SERVIR HKH at ICIMOD.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2019.00155/full#supplementary-material

REFERENCES

Adhikary, S. K., Das, S. K., Saha, G. C., and Chaki, T. (2013). "Groundwater drought assessment for barind irrigation project in Northwestern Bangladesh," in *20th International Congress on Modelling and Simulation, At Modelling and Simulation Society of Australia and New Zealand* (Adelaide, SA).

Alamgir, M., Shahid, S., Hazarika, M. K., Nashrrullah, S., Harun, S. B., et al. (2015). Analysis of meteorological drought pattern during different climatic and cropping seasons in bangladesh. *J. Am. Water Resour. Assoc*. 51, 794–806 doi: 10.1111/jawr.12276

Anderson, R. G., Lo, M.-H., and Famiglietti, J. S. (2012). Assessing surface water consumption using remotely-sensed groundwater, evapotranspiration, and precipitation. *Geophys. Res. Lett*. 39, 1–6. doi: 10.1029/2012GL052400

Biancamaria, S., Hossain, F., and Lettenmaier, D. P. (2011). Forecasting transboundary river water elevations from space. *Geophys. Res. Lett*. 38:L11401. doi: 10.1029/2011GL047290

Burgess, W. G., Shamsudduha, M., Taylor, R. G., Zahid, A., Ahmed, K. M., Mukherjee, A., et al. (2017). Terrestrial water load and groundwater fluctuation in the Bengal Basin. *Sci. Rep*. 7:3872. doi: 10.1038/s41598-017-04159-w

Castle, S. L., Reager, J. T., Thomas, B. F., Purdy, A. J., Lo, M. H., Famiglietti, J. S., et al. (2016). Remote detection of water management impacts on evapotranspiration in the Colorado River Basin. *Geophys. Res. Lett*. 43, 5089–5097. doi: 10.1002/2016GL068675

Castle, S. L., Thomas, B. F., Reager, J. T., Rodell, M., Swenson, S. C., and Famiglietti, J. S. (2014). Groundwater depletion during drought threatens future water security of the Colorado River Basin. *Geophys. Res. Lett.* 41, 5904–5911. doi: 10.1002/2014GL061055

Chandanpurkar, H. A., Reager, J. T., Famiglietti, J. S., and Syed, T. H. (2017). Satellite- and reanalysis-based mass balance estimates of global continental discharge (1993-2015). *J. Clim*. 30, 8481–8495. doi: 10.1175/JCLI-D-16-0708.1

Chen, F., Janji,ć, Z., and Mitchell, K. (1997). Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model. *Boundary Layer Meteorol.* 85, 391–421. doi: 10.1023/A:1000531001463

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H. L., Koren, V., et al. (1996). Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res. Atmos*. 101, 7251–7268. doi: 10.1029/95JD02165

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., et al. (2003). The common land model. *Bull. Am. Meteorol. Soc*. 8:1013–1024. doi: 10.1175/BAMS-84-8-1013

Dash, B. K., Rafiuddin, M., Khanam, F., and Islam, M. N. (2012). Characteristics of meteorological drought in Bangladesh. *Nat. Hazards*. 64, 1461–1474. doi: 10.1007/s11069-012-0307-1

David, C. H., Maidment, D. R., Niu, G.-Y., Yang, Z.-L., Habets, F., and Eijkhout, V. (2011). River network routing on the NHDPlus dataset. *J. Hydrometeorol*. 12, 913–934. doi: 10.1175/2011JHM1345.1

Dey, N., Alam, M., Sajjan, A., Bhuiyan, M., Ghose, L., Ibaraki, Y., et al. (2012). Assessing environmental and health impact of drought in the Northwest Bangladesh. *J. Environ. Sci. Nat. Resour*. 4, 89–97. doi: 10.3329/jesnr.v4i2.10141

Durack, P. J., Wijffels, S. E., and Matear, R. J. (2012). Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. *Science* 27, 455–458. doi: 10.1126/science.1212222

Ek, M. B. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res*. 108:8851. doi: 10.1029/2002JD003296

Famiglietti, J. S. (2014). The global groundwater crisis. *Nat. Clim. Chang*. 4, 945–948. doi: 10.1038/nclimate2425

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K. J., Syed, T. H., et al. (2011). Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophys. Res. Lett*. 38:L03403. doi: 10.1029/2010GL046442

Famiglietti, J. S., and Rodell, M. (2013). Water in the balance. *Science* 340, 1300–1301. doi: 10.1126/science.1236460

Fekete, B. M., Looser, U., Pietroniro, A., and Robarts, R. D. (2012). Rationale for monitoring discharge on the ground. *J. Hydrometeorol*. 12, 1977–1986. doi: 10.1175/JHM-D-11-0126.1

Food and Agriculture Organization (2011). *Irrigation in Southern and Eastern Asia in Figures*: *AQUASTAT Survey*.

Gain, A. K., Immerzeel, W. W., Sperna Weiland, F. C., and Bierkens, M. F. P. (2011). Impact of climate change on the stream flow of the lower Brahmaputra: trends in high and low flows based on discharge-weighted ensemble modelling. *Hydrol. Earth Syst. Sci*. 15, 1537–1545. doi: 10.5194/hess-15-1537-2011

Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., et al. (2013). A reconciled estimate of Glacier 2003 to 2009. *Science* 340, 852–857. doi: 10.1126/science.1234532

Getirana, A., Kumar, S., Girotto, M., and Rodell, M. (2017). Rivers and floodplains as key components of global terrestrial water storage variability. *Geophys. Res. Lett*. 44, 359–368. doi: 10.1002/2017GL074684

Hamlington, B. D., Reager, J. T., Chandanpurkar, H., and Kim, K. Y. (2019). Amplitude modulation of season variability in terrestrial water storage. *Geophys. Res. Lett.* 46, 4404–4412. doi: 10.1029/2019GL082272

IGRAC (2018). *Groundwater Overview*. UN Water 60.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., and Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general

circulation model 1. Model structure. *J. Geophys. Res. Atmos*. 105, 24809–24822. doi: 10.1029/2000JD900327

Lehner, B., Verdin, K., and Jarvis, A., (2008). New Global Hydrography Derived From Spaceborne Elevation Data. *EOS* 89, 93–94. doi: 10.1029/2008EO100001

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res*. 99 14415–14428.

Mahmuduzzaman, M., Ahmed, Z. U., Nuruzzaman, A. K. M., and Ahmed, F. R. S. (2014). Causes of salinity intrusion in coastal belt of Bangladesh. *Int. J. Plant Res*. 4, 8–13. doi: 10.5923/s.plant.201401.02

Massoud, E. C., Purdy, A. J., Miro, M. E., and Famiglietti, J. S. (2018). Projecting groundwater storage changes in California's Central Valley. *Sci. Rep*. 8:12917. doi: 10.1038/s41598-018-31210-1

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., et al. (2005). Stationarity is dead: whither water management? *Science* 319, 573–574. doi: 10.1126/science.1151915

Milly, P. C. D., Wetherald, R. T., Dunne, K. A., and Delworth, T. L. (2002). Increasing risk of great floods in a changing climate. *Nature*. 415, 514–517. doi: 10.1038/415514a

Nickson, R. T., Mcarthur, J. M., Ravenscroft, P., Burgess, W. G., and Ahmed, K. M. (2000). Mechanism of arsenic release to groundwater, Bangladesh and West Bengal. *Appl. Geochem*. 15, 403–413. doi: 10.1016/S0883-2927(99)00086-4

Reager, J. T., and Famiglietti, J. S. (2009). Global terrestrial water storage capacity and flood potential using GRACE. *Geophys. Res. Lett*. 36:L23402. doi: 10.1029/2009GL040826

Richey, A. S., Thomas, B. F., Lo, M. H., Reager, J. T., Famiglietti, J. S., Voss, K., et al. (2015). Quantifying renewable groundwater stress with GRACE. *Water Resour. Res*. 51, 5217–5238. doi: 10.1002/2015WR017349

Roberts, L. C., Hug, S. J., Dittmar, J., Voegelin, A., Kretzschmar, R., Wehrli, B., et al. (2010). Arsenic release from paddy soils during monsoonflooding. *Nat. Geosci*. 3, 53–59. doi: 10.1038/ngeo723

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., et al. (2018). Emerging trends in global freshwater availability. *Nature* 557, 651–659. doi: 10.1038/s41586-018-0123-1

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., et al. (2004). The global land data assimilation system. *Bull. Am. Meteorol. Soc*. 2004, 381–394. doi: 10.1175/BAMS-85-3-381

Rodell, M., McWilliams, E. B., Famiglietti, J. S., Beaudoing, H. K., and Nigro, J. (2011). Estimating evapotranspiration using an observation based terrestrial water budget. *Hydrol. Process*. 25, 4082–4092. doi: 10.1002/hyp.8369

Rodell, M., Velicogna, I., and Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature* 460, 999–1002. doi: 10.1038/nature08238

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller Schmied, H., van Beek, L. P. H., et al. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proc. Natl. Acad. Sci*. 115, E1080–E1089. doi: 10.1073/pnas.1704665115

Shamsudduha, M., Taylor, R. G., and Longuevergne, L. (2012). Monitoring groundwater storage changes in the highly seasonal humid tropics: validation of GRACE measurements in the Bengal Basin. *Water Resour. Res*. 48:W02508. doi: 10.1029/2011WR010993

Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P., et al. (2010). Groundwater use for irrigation - a global inventory. *Hydrol. Earth Syst. Sci*. 14, 1863–1880. doi: 10.5194/hess-14-1863-2010

Siebert, S., Henrich, V., Frenken, K., and Burke, J. (2013). *Update of the Digital Global Map of Irrigation Areas to Version 5*. Institute of Crop Science and Resource Conservation, Land and Water Division, Food and Agriculture Organization of the United Nations, Rome. doi: 10.13140/2.1.2660.6728

Smith, T., and Bookhagen, B. (2018). Changes in seasonal snow water equivalent distribution in high mountain Asia (1987 to 2009). *Sci. Adv*. 4:e1701550. doi: 10.1126/sciadv.1701550

Steckler, M. S., Nooner, S. L., Akhter, S. H., Chowdhury, S. K., Bettadpur, S., Seeber, L., et al. (2010). Modeling earth deformation from monsoonal flooding in Bangladesh using hydrographic, GPS, and gravity recovery and climate experiment (GRACE) data. *J. Geophys. Res. Solid Earth*. 115:B08407. doi: 10.1029/2009JB007018

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science* 305, 504–505. doi: 10.1126/science.1099192

Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., et al. (2019). Contributions of GRACE to understanding climate change. *Nat. Clim. Chang*. 9, 358–369. doi: 10.1038/s41558-019-0456-2

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., Van Beek, R., Wada, Y., et al. (2013). Ground water and climate change. *Nat. Clim. Chang*. 3, 322–329. doi: 10.1038/nclimate1744

Thomas, A. C., Reager, J. T., Famiglietti, J. S., and Rodell, M. (2014). A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophys. Res. Lett*. 41, 1537–1545. doi: 10.1002/2014GL059323

Thomas, B. F., Caineta, J., and Nanteza, J. (2017). Global assessment of groundwater sustainability based on storage anomalies. *Geophys. Res. Lett*. 44, 11445–11455. doi: 10.1002/2017GL076005

Trenberth, K. E., Dai, A., Van Der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., et al. (2014). Global warming and changes in drought. *Nat. Clim. Chang*. 4, 17–22. doi: 10.1038/nclimate2067

Vörösmarty, C. J., Green, P., Salisbury, J., and Lammers, R. B. (2000). Global water resources: vulnerability from climate change and population growth. *Science* 289, 284–288. doi: 10.1126/science.289.5477.284

Voss, K. A., Famiglietti, J. S., Lo, M., De Linage, C., Rodell, M., and Swenson, S. C. (2013). Groundwater depletion in the middle east from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water Resour. Res*. 49, 904–914. doi: 10.1002/wrcr.20078

Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., and Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *J. Geophys. Res. Solid Earth* 120, 2648–2671. doi: 10.1002/2014JB011547

Wiese, D. N., Landerer, F. W., and Watkins, M. M. (2016). Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resour. Res*. 52, 7490–7502. doi: 10.1002/2016WR019344

Disclaimer: The views and interpretations in this paper are those of the authors and are not necessarily attributable to ICIMOD, USAID, NASA, or BYU.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 31 October 2019
doi: 10.3389/fenvs.2019.00171

Evaluation of Available Global Runoff Datasets Through a River Model in Support of Transboundary Water Management in South and Southeast Asia

Md. Safat Sikder[1], Cédric H. David[1], George H. Allen[2], Xiaohui Qiao[3], E. James Nelson[3] and Mir A. Matin[4]*

[1] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States, [2] Department of Geography, Texas A&M University, College Station, TX, United States, [3] Department of Civil and Environmental Engineering, Brigham Young University, Provo, UT, United States, [4] International Centre for Integrated Mountain Development, Kathmandu, Nepal

Numerical models have become essential tools for simulating and forecasting hydro-meteorological variability, and to help better understand the Earth's water cycle across temporal and spatial scales. Hydrologic outputs from these numerical models are widely available and represent valuable alternatives for supporting water management in regions where observations are scarce, including in transboundary river basins where data sharing is limited. Yet, the wide range of existing Land Surface Model (LSM) outputs makes the choice of datasets challenging in the absence of detailed analysis of the hydrological variability and quantification of associated physical processes. Here we focus on two of the world's most populated transboundary river basins—the combined Ganges-Brahmaputra-Meghna (GBM) in South Asia and the Mekong in Southeast Asia—where downstream countries are particularly vulnerable to water related disasters in the absence of upstream hydro-meteorological information. In this study, several freely-available global LSM outputs are obtained from NASA's Global Land Data Assimilation System (GLDAS) and from the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis-interim/Land (ERA-interim/Land) and used to compute river discharge across these transboundary basins using a river network routing model. Simulations are then compared to historical discharge to assess runoff data quality and identify best-performing models with implications for the terrestrial water balance. This analysis examines the effects of meteorological inputs, land surface models, and their spatio-temporal resolution, as well as river network fineness and routing model parameters on hydrologic modeling performance. Our results indicate that the most recent runoff datasets yield the most accurate simulations in most cases, and suggest that meteorological inputs and the selection of the LSM may together be the most influential factors affecting discharge simulations. Conversely, the spatial and temporal resolution of the LSM and river model might have the least impact on the quality of simulated discharge, although the routing model parameters affect the timing of hydrographs.

Keywords: water balance, discharge, global LSM, GLDAS, ERA-interim, GBM, Mekong

OPEN ACCESS

Edited by:
Ashutosh S. Limaye, National Aeronautics and Space Administration (NASA), United States

Reviewed by:
Sujay V. Kumar, National Aeronautics and Space Administration, United States
Olaf Schroth, Weihenstephan-Triesdorf University of Applied Sciences, Germany

****Correspondence:***
Md. Safat Sikder
md.safat.sikder@jpl.nasa.gov

Specialty section:
This article was submitted to Land Use Dynamics, a section of the journal Frontiers in Environmental Science

Received: *30 May 2019*
Accepted: *11 October 2019*
Published: *31 October 2019*

Citation:
Sikder MS, David CH, Allen GH, Qiao X, Nelson EJ and Matin MA (2019) Evaluation of Available Global Runoff Datasets Through a River Model in Support of Transboundary Water Management in South and Southeast Asia. Front. Environ. Sci. 7:171. doi: 10.3389/fenvs.2019.00171

INTRODUCTION

South and Southeast Asia are currently home to the world's most densely populated areas (FAO, 2016). This region is also characterized by extreme hydrologic variability with ∼70–80% of the total annual rainfall occurring during the short June to September window of the summer monsoon, causing flooding to be a regular annual event (Mirza, 2011; Hoang et al., 2019). In addition, the combination of climate change and anthropogenic water diversions from rivers affect the region through increased drought frequency (Khandu et al., 2016). Such demographic and hydrologic extremes together make the two principal rivers of South and Southeast Asia—the combined Ganges-Brahmaputra-Meghna (GBM) and the Mekong—some of the world's largest rivers (e.g., Dai et al., 2009) and most populous transboundary basins (Webster et al., 2010; Lakshmi et al., 2018). Surface water from these rivers provides great benefits because it helps support critical agricultural and energy production needs for over 690 million people (FAO, 2016), i.e., a tenth of the human population. Yet, the benefits of surface water also come with challenges, most notably for the downstream parts in these basins, which were determined to have the world's highest risks of exposure to floods, but also to droughts (UNEP, 2016).

In Bangladesh, which is situated in the downstream portion of the GBM basin, losses due to flooding are severely hampering the economic growth of the country. Approximately 80% of the country consists of floodplains such that, in a typical year, about a third of Bangladesh is flooded during the monsoon (Brouwer et al., 2007). During extreme flood years, two thirds of the country can be inundated for ∼3 months, causing widespread devastation on the region (Mirza et al., 2003). The mean annual loss caused by normal flooding in Bangladesh is estimated about US$175 million (Mirza, 2011). The World Resources Institute ranks Bangladesh as first in the world in terms of percentage of country GDP regularly exposed to flooding and second in terms of population exposed (Priya et al., 2017).

While the Mekong River Basin is also characterized by the intensity of its wet season (Hoang et al., 2019), the most critical ongoing hydrologic challenge is the expected increase of large hydropower dams on the main stem of the Lower Mekong River (Kummu and Sarkkula, 2008; Bonnema and Hossain, 2017). The proposed construction of 11 hydropower dams on the now free-flowing river (Orr et al., 2012) is anticipated to have a significant impact on the region's ecohydrology (Pokhrel et al., 2018) and aquaculture—the Mekong contains the world's largest inland fisheries (Hecht et al., 2019). These degradations are expected to continue to grow in the future with a total of 90 and 136 dams to be built within the Lower Mekong basin by the year 2030 and 2060, respectively (Räsänen et al., 2012). Recent studies reporting a 10-fold increase from water storage capacity from 2 to 20% of annual flow by 2025 indicate that the drastic impacts of the proposed dam constructions on fisheries, agriculture, and the environment are likely to persist (Kummu et al., 2010; Hecht et al., 2019).

The high population density and hydrologic extremes in South and Southeast Asia therefore pose a variety of challenges to water management. The transboundary aspect of river basins in the area and associated geopolitical challenges in data sharing make water management endeavors rely heavily on simulations from computer models (Hossain et al., 2014). Much of the existing literature in the region has therefore focused on detailed hydrologic modeling endeavors with locally-tailored models (e.g., Nishat and Rahman, 2009; Hossain et al., 2017). In the case of Bangladesh, much effort has been dedicated to flood mitigation through forecasting to further the economic progress and food security with particular emphasis on increased lead time of flood forecasting systems (CEGIS, 2006; Webster et al., 2010), i.e., the latency between available forecasts and real-time events. Other notable applications include the flood forecasting system of Sikder and Hossain (2018), and antecedent studies of hydrologic and hydraulic model development (e.g., Siddique-E-Akbor et al., 2014; Maswood and Hossain, 2015).

Similarly, the planned and ongoing development of hydroelectric dams in the Lower Mekong Basin has been a subject of intense national and international argument for the stakeholder countries of the Mekong basin and numerical models have played an important role in such debate. Hanington et al. (2017) showed how hydrological models can be used for water resources planning and management in the Mekong delta to support the agricultural production. Räsänen et al. (2012) studied the hydrological impact of the proposed dams on the downstream portion of the Mekong basin. Haddeland et al. (2006) and Tatsumi and Yamashiki (2015) investigated the effect of water diversions on the water and energy balances of the Mekong basin using hydrological models. Hoang et al. (2019) used a hydrological model to study the impact of climate change, construction of dams and flow diversion on the future flow of the Mekong river. Similar efforts by Johnston and Kummu (2012) and Pokhrel et al. (2018) have studied the expected evolution of the Mekong water resources in in the context of a changing climate.

Besides the aforementioned water resources applications of hydrologic models in the GBM and the Mekong basin, Land Surface Models (LSMs) are also used as a core tool for understanding the spatio-temporal variation of hydro-meteorological variables and associated physical processes. The importance of understanding the water cycle and quantifying its various fluxes using LSMs is even more acute in the case of ungauged and transboundary regions, where such data can prevent large-scale disasters (Siddique-E-Akbor et al., 2014; Murshed and Kaluarachchi, 2018). At this time, several operational global LSMs are producing continuous estimates of different hydrological fluxes, which can be used as an easily-accessible alternative to locally-tailored hydrological models. Numerous studies have used these readily available LSM outputs to analyze different components of the water cycle. For example, Lakshmi et al. (2018) quantified water availability in the world's major river basins using the Global Land Data Assimilation System (GLDAS) model outputs (Rodell et al., 2004).

Perhaps the most widespread use of GLDAS or other global LSMs in South and Southeast Asia is along with the Gravity Recovery and Climate Experiment (GRACE) data to determine groundwater fluctuations and changes in water storage. Rodell et al. (2009) and Chinnasamy et al. (2015) used GLDAS soil

moisture along with GRACE to estimate the groundwater depletion rate in North India. Khandu et al. (2016) used GRACE along with soil moisture estimates from different LSMs to estimate the influence of precipitation extremes on the Total Water Storage (TWS) in the GBM basin. A recent study by Murshed and Kaluarachchi (2018) used GLDAS soil moisture data to estimate freshwater availability in the Ganges Delta and demonstrated how LSMs can be used to identify the complex issues of water security.

Global LSM outputs are also often applied throughout the world's largest basins for comparison and validation against satellite-derived data. Syed et al. (2008) used GLDAS and GRACE separately to characterize the Terrestrial Water Storage Change (TWSC) in major river basins and found a good agreement in satellite- and model-derived TWSC. Rodell et al. (2011) used GRACE and other observed and modeled data together to estimate evapotranspiration for a few major river basins and compared the results with models. Chen et al. (2013) compared the Advanced Microwave Scanning Radiometer—Earth Observing System (AMSR-E) and all four GLDAS LSM-derived soil moisture with the *in-situ* station averaged soil moisture data in the Tibetan Plateau and reported that three out of four GLDAS LSMs perform better than the satellite product in that region, though the LSMs generally underestimated soil moisture.

Global LSM data products therefore have strong potential for practical applications in ungauged and transboundary river basins, including in South and Southeast Asia, and the accuracy of these models is therefore critical for associated water resources management endeavors (Wang et al., 2011). For these LSMs to be most useful, they must be validated against *in-situ* data in anticipation for practical (i.e., "real-world") applications, as done previously in several published studies. Berg et al. (2005) used GLDAS and NLDAS models to prepare hydro-meteorological forcing data (i.e., the continuous input data used to run the model) for global soil moisture estimation and compared their simulations with *in-situ* soil moisture data to find good agreement between the anomaly of observed and modeled soil moisture. Similarly, Bi et al. (2016) used *in-situ* soil moisture data to validate the GLDAS LSMs in the Tibetan Plateau and found that models can accurately capture the temporal variations, but systematically underestimate soil moisture. However, validating LSMs using distributed hydrological fluxes is challenging, since the evaporation, soil moisture, and groundwater are difficult to measure *in-situ*, and *in-situ* measurements only capture the local state.

River flow does not face the same distributed challenge because it is the integration of all upstream hydrologic processes and many *in-situ* discharge datasets in the major rivers of the world have continuous record. Therefore, river discharge can also be used to validate global LSM simulations, granted that the discharge is routed correctly (Zaitchik et al., 2010), the effects of anthropogenic activities in hydrological processes are addressed properly, and noting that observations are often difficult to obtain (Hossain et al., 2014). Such a validation is critical because it allows for the verification of the LSM estimates in an integrated manner. Yet, most of the validations of existing operational LSMs using river flow have been conducted in data-rich regions of the world. For example, Mitchell et al. (2004) used the *in-situ* discharge of the Contiguous United States (CONUS) to validate the North American Land Data Assimilation System (NLDAS). Xia et al. (2012b) updated the same validation study using the LSMs of NLDAS-2. In contrast, very few studies were conducted in other parts of the world. To the extent of our knowledge, only Zaitchik et al. (2010) used discharge estimates for the world's major rivers to validate GLDAS, and identified that the four LSMs of GLDAS performed differently and with distinct geographic patterns to estimate the river flow. They also found that the choice of meteorological forcing has a notable impact in the simulated discharge.

Recently-updated global LSM products (e.g., GLDAS-2) are now available but have not yet been evaluated against observed river flow. Such a validation effort is particularly needed for the transboundary river basins of South and Southeast Asia where global LSMs can provide valuable estimates of water fluxes and states. In this context, the primary goal of this study is to identify optimal global LSMs for estimating river flow in South and Southeast Asia. Furthermore, the accuracy of the model-simulated river discharge can be expected to be sensitive to the choice of the LSM, meteorological forcing, spatio-temporal resolution of the models, and routing model parameters. A secondary goal of this study is hence to identify the most influential factors affecting the accuracy of the simulated flow.

This paper is organized as follows. Section Model and Data describes the data and models used in this study, including the global LSMs, the river routing model, and the *in-situ* discharge data. The Methodology follows in section Methodology and section Results and Discussions describes our results and discusses our analysis. Finally, section Conclusions presents our conclusions.

MODEL AND DATA

Land Surface Models

Several publicly available LSM outputs from different projects were considered for this study. Model outputs from 2001 to 2009 were used for river routing simulation and preliminary analysis. This time window was carefully chosen to ensure that the data from all global LSMs considered are available. A list and the details of these LSMs are provided in **Table 1**, and further discussed below.

The GLDASv1 (denoted by GLDAS hereafter) consists of four different LSMs. These models are the Mosaic model (Koster and Suarez, 1996), the Noah model (Chen et al., 1996; Koren et al., 1999), the Common Land Model (CLM) (Dai et al., 2003), and the Variable Infiltration Capacity (VIC) model (Liang et al., 1994). A combination of meteorological datasets were used as the input (i.e., "forcing") for these LSMs: the National Oceanic and Atmospheric Administration (NOAA) Global Data Assimilation System (GDAS) atmospheric analysis fields, the spatio-temporally disaggregated NOAA Climate Prediction Center Merged Analysis of Precipitation (CMAP) fields, and the Air Force Weather Agency's AGRicultural METeorological modeling system (AGRMET) method based *in-situ* downward

TABLE 1 | Details of the Land Surface Model (LSM) outputs used in this study.

Projects	Version	Forcing			LSM	Resolution	
		Precipitation	Radiation	Others		Temporal	Spatial
GLDAS	v1	CMAP	AGRMET	GDAS	CLM v2.0 Mosaic Noah v2.7.1 VIC	3 h	1°
	v2.0	Princeton meteorological forcing			Noah v3.3	3 h	0.25°, 1°
	v2.1	GPCP	AGRMET	GDAS	Noah v3.3	3 h	0.25°, 1°
ECMWF	ERA-interim/Land	GPCP v2.1	ECMWF Re-Analysis		HTESSEL	Daily	80 km

shortwave and longwave radiation fields. The GLDASv1 outputs are available from January 1979 to date. More details about these models can be found in Rodell et al. (2004) and Rui and Beaudoing (2017).

An updated GLDAS is now available and denoted GLDAS-2. Two different versions of GLDAS-2 are available: GLDASv2.0 and GLDASv2.1. Both GLDASv2.x used more recent meteorological forcing than GLDAS in a unique LSM which is a newer version of Noah. GLDASv2.0 used the Princeton meteorological forcing data (Sheffield et al., 2006) that is available from January 1948 to December 2010. GLDASv2.1 used a combination of NOAA/National Center for Environmental Prediction's GDAS atmospheric analysis fields, the spatio-temporally disaggregated Global Precipitation Climatology Project (GPCP) precipitation fields, and the updated AGRMET downward shortwave and longwave radiation fields as the forcing data. Other enhancements in GLDAS-2 include switching to MODIS based land surface parameter datasets, and initialization of soil moisture over desert (Rui and Beaudoing, 2019). The model outputs of GLDASv2.1 are available for the period of January 2000—present.

Another set of global LSM outputs is available from the European Centre for Medium-Range Weather Forecasts (ECMWF). These simulations used the ECMWF Re-Analysis (ERA)-Interim data and the GPCPv2.1 adjusted precipitation field as forcing for the latest version of the Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) LSM to produce the ERA-Interim/Land products. The outputs of this model are covering a period of January 1979 to December 2010. More details of this dataset can be found in Balsamo et al. (2012, 2015).

River Network Routing Model

The Routing Application for Parallel computatIon of Discharge (RAPID) (David et al., 2011) is used as the river routing model in this study to derive the daily flow throughout the river basins. RAPID uses the Muskingum method (McCarthy, 1938) to calculate flow at all nodes of a given river network using surface and subsurface runoff from an LSM as inputs. One of the key advantages of RAPID is that it can efficiently be executed in a parallel computing environment by using a matrix form of the Muskingum method. More detailed descriptions of the Muskingum method and its use to propagate runoff from LSMs are available in David et al. (2011). The river network data for this study were obtained from the Hydrological Data and Maps Based on Shuttle Elevation Derivatives at Multiple Scales (HydroSHEDS) (Lehner et al., 2008). The fine resolution river network was directly extracted from the 15 arc second HydroSHEDS river network. The coarse river network was derived from the upscaled 0.1° HydroSHEDS grid (Alfieri et al., 2013; Snow, 2015). Apart from the given surface and subsurface runoff from any LSM and a given river network, the basic model setup requires two sets of model parameters. These parameters are the Muskingum dimensionless diffusion coefficient (denoted by x) and flow wave propagation time (denoted by k); and the parameters can vary spatially on a reach-by-reach basis although they are temporally constant.

Observed Discharge Data

Observed river flow data for different locations of the GBM and Mekong river basins were obtained from the Bangladesh Water Development Board (BWDB) and the Mekong River Commission (MRC), respectively. The location of these *in-situ* discharge stations are shown in **Figure 1**. The GBM discharge data were available for the entire routing model simulation period (2001–2009). The Mekong river flow data were available up to 2007 in a few stations, while the data were available up to 2006 in most stations. Therefore, the *in-situ* flow data in between 2001 and 2006 were considered for the performance analysis of the daily flow in all river stations for consistency. All available *in-situ* flow stations downstream of the Tonlé Sap Lake were excluded from the analysis due to its complex hydrological behavior.

METHODOLOGY

The primary objective of this study is to evaluate the performance of each global LSM in South and Southeast Asia by comparing routed runoff from a river model to daily *in-situ* river flow observations. The secondary objective is to identify the factors that significantly affect the simulated flow. To get a preliminary perspective on the relative performance of each LSM, the basin-averaged mean annual precipitation, evapotranspiration, and runoff were computed. At first, the basin-scale mean annual water budget error analysis was carried out in a way similar to Xia et al. (2012a) who used 28 years of data to evaluate LSM

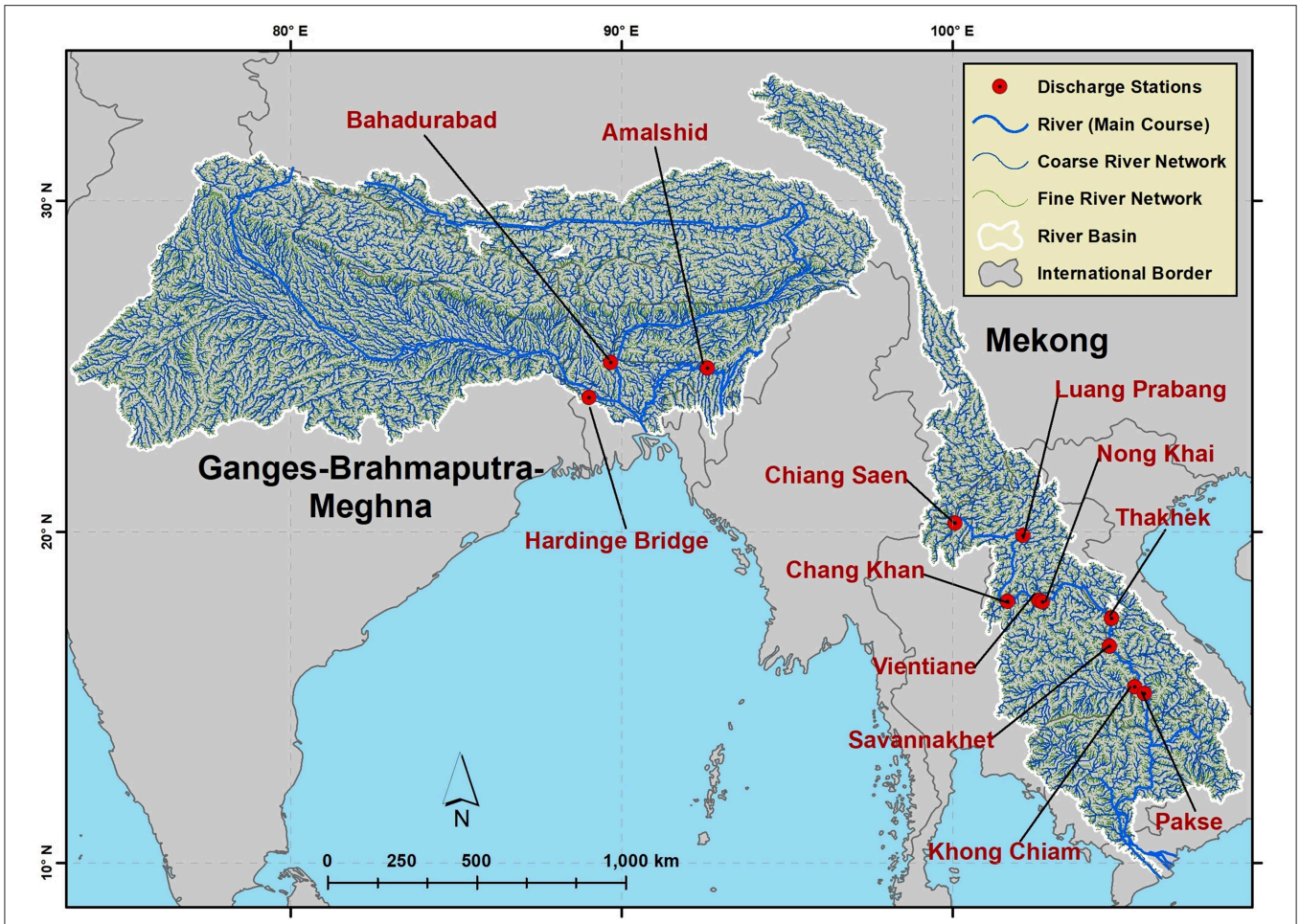


FIGURE 1 | Location of the Ganges-Brahmaputra-Meghan (GBM) and Mekong river basins along with the *in-situ* discharge stations. The fine and coarse river networks used for flow routing are shown in the background.

performance in the CONUS. In this study, the mean annual total model input (i.e., precipitation) and mean annual total model outputs (i.e., evapotranspiration, Runoff) were compared while assuming that the change in water storage for a long-term average should be negligible or close to zero. Here, 9 years of data (2001–2009) were used, and the difference between the mean-annual total input and output was reported as a percentage of the total input (i.e., precipitation). Note that the assumption of water balance closure over a 9-year period can be debated in light of the numerous past studies highlighting the effect of groundwater extraction on the terrestrial water balance (e.g., Rodell et al., 2009; Purdy et al., 2019) although it is justified in this study given our intended focus on readily-available LSM outputs. We define the mean annual percentage error in water budget as:

$$error_{ma} = \frac{(ET + Q_s + Q_{sb})_{ma} - (Rain + Snow)_{ma}}{(Rain + Snow)_{ma}} * 100\ \% \quad (1)$$

Where, *error*, *ET*, Q_s, Q_{sb}, *Rain*, *Snow* are for the area-averaged water budget error, the LSM simulated evapotranspiration, surface runoff, sub-surface runoff, input rainfall, and snowfall, respectively.

The evapotranspiration-precipitation and runoff-precipitation ratios were also derived from the area averaged mean annual variables. The spatio-temporal averaged runoff was then compared to the observed mean annual discharge, itself derived from the *in-situ* discharge of the most downstream station of each river basin. These stations are the Pakse, Hardinge Bridge, Brahmaputra, and Amalshid in the Mekong, Ganges, Brahmaputra, and Meghna basins, respectively (**Figure 1**). Note that all the area-averaged values discussed above were calculated with respect to these downstream *in-situ* stations. Therefore, the area covered by the spatio-temporal averaged variables is the same as the upstream basin area of each station (i.e., not the entire basin but a significant portion of it).

The river network routing model (i.e., RAPID) was then used to derive the daily flow at multiple locations in the river basins, and compared with available *in-situ* discharge data. To identify the factors that affect the simulated flow significantly, the simulated daily flows were compared to the observed discharge while varying the choice of the LSM, atmospheric forcing, model resolution, and routing parameters. Several experimental setups were therefore designed to reveal the relative impact of

these four different design factors on the simulated flow. The comparisons were performed using a series of five traditional metrics: correlation coefficient, Nash–Sutcliffe efficiency (NSE), root mean square error (RMSE), bias, and standard error.

The impact of the choice of the LSM was first assessed. Since GLDAS uses the same forcing in four different LSMs, data only from this version (i.e., GLDASv1) were used for this initial analysis. The ECMWF ERA-interim/LAND derived daily flow was also used here for comparison with the other four LSMs, although the forcing data and the LSM resolution of ECMWF are different. To be consistent, in all cases the routing model was executed using the fine resolution river network. The name of each test case contains the project name and version, LSM name, model spatial resolution (i.e., 10 for 1° LSM outputs), and the river network resolution. For example, the name for the GLDAS Mosaic derived flow from the 1° runoff using the fine river network is: GLDAS-MOS10-Fine.

Precipitation—the primary input of water to land surface models—and other meteorological inputs are also sources of uncertainty in simulated streamflow. Therefore, a performance analysis was conducted for different forcing inputs to the same LSM. The Noah LSM outputs are available which was forced by different meteorological datasets; GLDAS, GLDASv2.0, and GLDASv2.1. Although GLDAS-2 products are available in two different spatial resolutions (i.e., 0.25 and 1°), only 1° resolution LSM products were used here for consistency with GLDAS. Like the previous experiment, the daily flow was derived using only the fine resolution river network in all cases, and the ECMWF ERA interim/LAND derived daily flow was used here as reference. It is important to note that the Noahv2.7 was used in GLDAS, while both GLDAS-2 versions used Noahv3.3 as the LSM. Also, the LSM used with the ECMWF ERA interim data was different from Noah, and has a different model resolution (∼80 km). The experimental design nomenclature remains similar, but different GLDAS (i.e., v1, v2.x) versions were used here.

In a third experiment, models with different spatio-temporal resolutions were used. To control for LSM and forcing variabilities, only the GLDASv2.1 was used here with different spatial and temporal resolution, and discharge was simulated using fine (∼5 km) and coarse (∼20 km) river networks (**Figure 1**). The GLDASv2.1 (Noah LSM) model outputs are available in 0.25 and 1° spatial resolution and with 3 h temporal resolution. The 3 hourly data were temporally averaged to derive daily forcing for the routing model with different temporal resolutions (i.e., as daily LSM input). The name of each test case starts with the project name and version, followed by the LSM name, resolution (i.e., 10 and 025 for 1 and 0.25°, respectively), temporal resolution only for the daily scale (i.e., D) and the river network resolution (i.e., fine or coarse). For example, both GLDASv2.1-NOAH025-Coarse and GLDASv2.1-Noah025D-Coarse stand for the simulated river flow using the coarse river network with the 0.25 GLDASv2.1 Noah LSM output, but for the 3-hourly and daily temporal resolutions, respectively.

In a final experiment, river routing simulations were performed by changing the two parameters of the routing model (i.e., RAPID). These parameters are the diffusion coefficient (denoted by x) and propagation time (denoted by k) of the Muskingum method. Although, it is possible to vary sets of parameters independently for each reach in RAPID, it is not practical to do so for the large river basins with numerous river reaches used here. Therefore, both sets of parameters were first determined and subsequently changed through multiplication by a spatially-constant scale factor from experience. The initial values of the storage constants (k) of all cells of the model were determined from a spatiotemporally-constant wave celerity of 1 km/h while accounting for the variable length of river reaches. Similarly, the weighting coefficients (x) were initially set to the commonly accepted value of 0.1. Note that no specific parameter calibration is performed in this study. All the test simulations were then conducted using the ECMWF ERA interim/LAND outputs. The name of each test case starts with the model name and followed by the k scale factor (i.e., k02 and k035 for k scale factor 0.2 and 0.35, respectively), the x scale factor (i.e., x2, x3, and x4 for x scale factor 2, 3, and 4, respectively), river network resolution. For example, ERAi-Land-k02-x3-Coarse means the RAPID model was simulated using the ECMWF ERA interim/LAND with coarse resolution river network, while the scaling factor for k and x were 0.2 and 3, respectively. Note that the values of the scale factors are based on accepted ranges of values (e.g., Fread, 1993) and on past experience from previous RAPID studies (e.g., David et al., 2011).

The timing of hydrograph is also a concern for discharge simulations and motivate the use of an additional metric. To determine the accuracy of the hydrograph timing, a lagged cross-correlation was used (e.g., David et al., 2011; Allen et al., 2018). The lagged cross-correlation determines the correlation between two timeseries as a function of lag that is added between them. Optimal simulations flow should therefore show the maximum correlation with observations when the lag time is zero. The lagged cross-correlation between the simulations and observations is shown in Equation (2).

$$\rho = \frac{\sum_{t=1}^{n}\left[{Q_{ob}}^{(t)} - \overline{Q_{ob}}\right]\left[{Q_{sim}}^{(t+\tau_{lag})} - \overline{Q_{sim}}\right]}{\sqrt{\sum_{t=1}^{n}\left[{Q_{ob}}^{(t)} - \overline{Q_{ob}}\right]^2\left[{Q_{sim}}^{(t+\tau_{lag})} - \overline{Q_{sim}}\right]^2}} \quad (2)$$

Where, ρ, Q_{ob}, Q_{sim}, t, and τ_{lag} are the lagged cross correlation, observed flow, simulated flow, time step, and lag time, respectively.

TABLE 2 | Mean annual error in model water budget with respect to the input precipitation (%).

Model/Basin	Mekong	Ganges	Brahmaputra	Meghna
GLDAS-CLM	0.45	0.06	0.51	0.02
GLDAS-MOS	4.63	1.04	2.35	−0.36
GLDAS-NOAH	1.05	−0.18	0.5	−0.21
GLDAS-VIC	1.52	0.61	1.28	0.3
GLDASv2.0-NOAH	0.1	−0.06	0.05	0.17
GLDASv2.1-NOAH	0.15	−0.41	−0.01	0.11
ECMWF-ERAi/Land	0.15	−0.21	0.23	0.33

A maximum of 10-day lag time was considered in this study, and correlations between the observed and (–)10-day to (+)10-day temporal offset of the simulated flow were calculated.

RESULTS AND DISCUSSIONS

The analysis of the mean annual water budget following Equation (1) is summarized in **Table 2** where the errors are shown here as a percentage of the input (i.e., precipitation). Assuming steady state, the error should be close to zero for the most optimized LSM, hence demonstrating that the water balance is properly satisfied. Our results show that the overall performance of the newer versions of GLDAS (i.e., GLDAS-2) and ECMWF is better in terms of the accuracy in water balance. Among the four LSMs of GLDAS, CLM is maintaining the lowest water balance errors in all the river basins, while Mosaic shows the largest errors. Additionally, and although all the four GLDAS models are using the same input data, the variation in water balance errors is significant within these LSMs. This may be due to differences in the fluxes and storage within soil layers given that the number of soil layers and their associated depths differ among models. The Mosaic model has three soil layers up to 350 cm below the ground, while the Noah model is using four soil layers up to 200 cm. CLM and VIC are using ten layers up to 343.3 cm and three layers up to 190 cm depth, respectively (Bi et al., 2016; Rui and Beaudoing, 2017).

The evapotranspiration-precipitation ratio was then used to evaluate the models, as shown in **Figure 2**, and suggests that the newest projects (i.e., GLDAS-2 and ECMWF) have relatively similar behaviors throughout all river basins (with little ET/P ratio variability on a per-basin basis) while the earlier GLDAS LSMs (i.e., GLDASv1) have inconsistent behaviors (high ET/P variability in each basin). Although consistency among different models cannot be directly linked to simulation quality, it does add confidence in simulations. CLM may be showing the closest evapotranspiration-precipitation ratio to the newer version of the models. However, if the mean annual evapotranspiration (**Figure 2**, values are reported above the bars in mm/year) is considered in place of the ratio, then VIC and Noah appear to be the closest to the newer version of the models and CLM would underestimates the mean annual evapotranspiration. A similar pattern was found by Xia et al. (2012b), where simulated evapotranspiration was compared with observations for the CONUS using the NLDAS-2, and where VIC and Noah produce similar evapotranspiration. The GLDAS-2 and ECMWF were forced with similar amounts of input precipitation (**Figure 2**, values are reported below the bars in mm/year), while the GLDAS LSMs were forced with low-biased precipitation relative to GPCP (Rui and Beaudoing, 2019). The case of the Ganges basin may be of particular interest here, given that the amounts of input precipitation are all very close in GLDAS, GLDAS-2, and ECMWF. In this basin, CLM shows consistently lower evapotranspiration-precipitation ratios, while Noah is more consistent with newer simulations in GLDAS-2 and ECMWF. Our experiments are therefore generally inconclusive in recommending a specific LSM for evapotranspiration given that both evapotranspiration-precipitation ratio and precipitation itself both vary greatly in available experiments. However, the expected increase in the quality of precipitation dataset and the consistency of precipitation and evapotranspiration-precipitation ratios among latest projects provides some level of confidence in GLDAS-2 and ECMWF simulations.

Similarly, the runoff-precipitation ratio was derived (**Figure 3**) along with the model simulated mean annual runoff values (**Figure 3**, above the bars in mm/year) and the observed mean annual runoff (**Figure 3**, on top of the charts). Here again GLDAS-2 and ECMWF are showing similar results, although the comparison with observed runoff (computed from observed discharge) is here able to confirm better performance in terms of mean annual runoff for the most recent projects. All the four earlier GLDAS LSMs are consistently underestimating runoff which is perhaps due to underestimated precipitation for this earlier dataset. CLM appears to perform relatively better in terms of runoff compared to the other GLDAS LSMs. This may be related to CLM's underestimation of evapotranspiration combined with lower amounts of input precipitation (**Figure 2**), generating relatively larger runoff. However, CLM also overestimates runoff in the Ganges basin due to more accurate precipitation there, which is associated with the underestimated evapotranspiration by CLM in Ganges basin. This confirms that a more accurate precipitation as input in the earlier GLDAS LSMs could have changed their accuracy. In general, GLDAS-2 and the ECMWF provide more accurate runoff estimation, in addition to low errors in annual water balance (**Table 2**). However, all LSMs significantly underestimate runoff in the case of the Brahmaputra basin, which may be due to unrealistic runoff-precipitation ratios or flawed amounts of precipitation in that basin.

The RAPID model was then used to derive daily flows from available global LSMs runoff throughout river basins and simulated flows were compared with the *in-situ* measurements. The analysis was carried out for the four aforementioned experimental setups, since the simulated flow may be sensitive to the selection the LSM, meteorological forcing, model resolution, and model parameters. A sample output of RAPID simulation is shown in **Figure 4** for the Pakse station in the Mekong basin where the simulated and observed discharge are plotted as a function of: different LSMs (**Figure 4A**), different meteorological forcing (**Figure 4B**), different spatio-temporal resolutions of the models (**Figure 4C**), and different river network routing model parameters (**Figure 4D**). Note that the number of test cases considered in each experiment differs; and that this lack of consistency in the number of ensembles of the experimental setup may impede the fairness of comparisons. However, the ensemble sizes for the first two experiments (i.e., for different LSMs and different meteorological forcing) is dictated by data availability and beyond our control. Nevertheless, this preliminary analysis suggests that the simulated flow is mostly influenced by the selection of the LSM and by the meteorological forcing than it is influenced by the spatio-temporal resolution of the models and by the routing model parameters. Further analysis was conducted in all discharge stations of **Figure 1** for

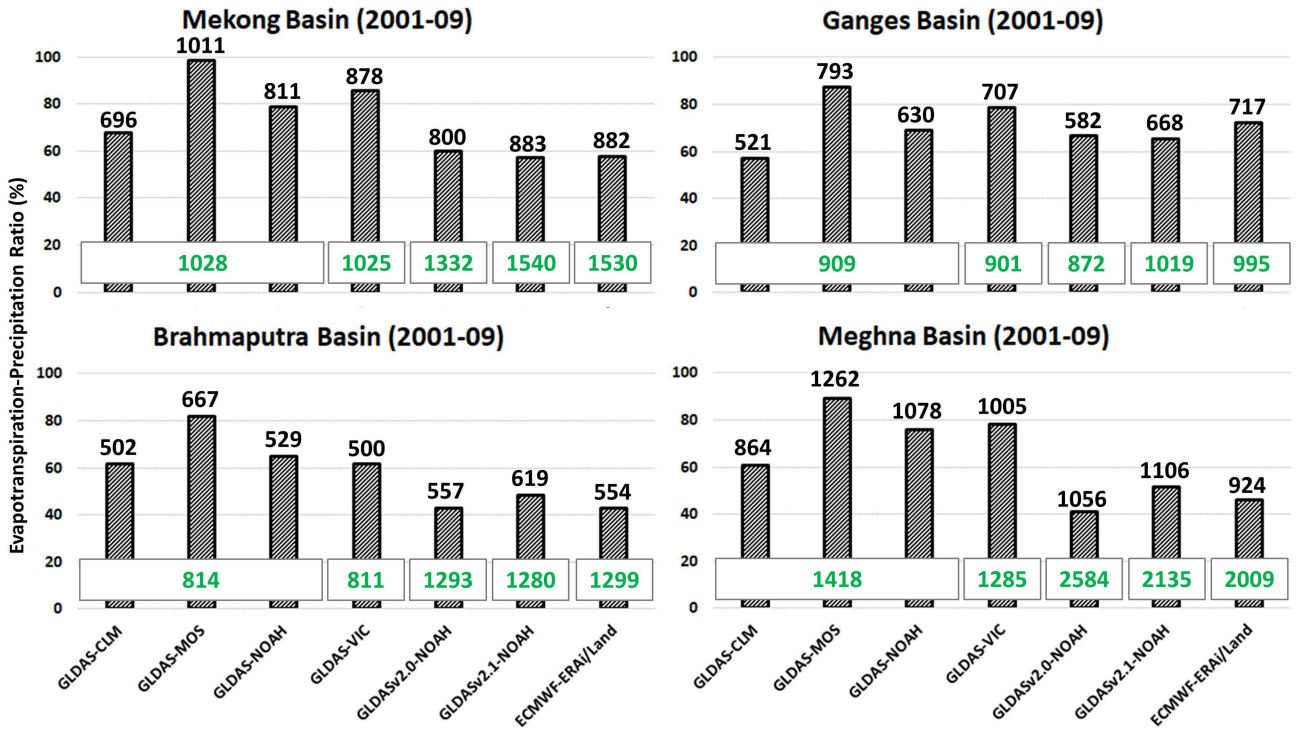


FIGURE 2 | Mean annual Evapotranspiration-Precipitation ratio the Mekong and GBM basins. All the values shown above the bars and in the boxes are the simulated mean annual evapotranspiration and input precipitation in mm, respectively.

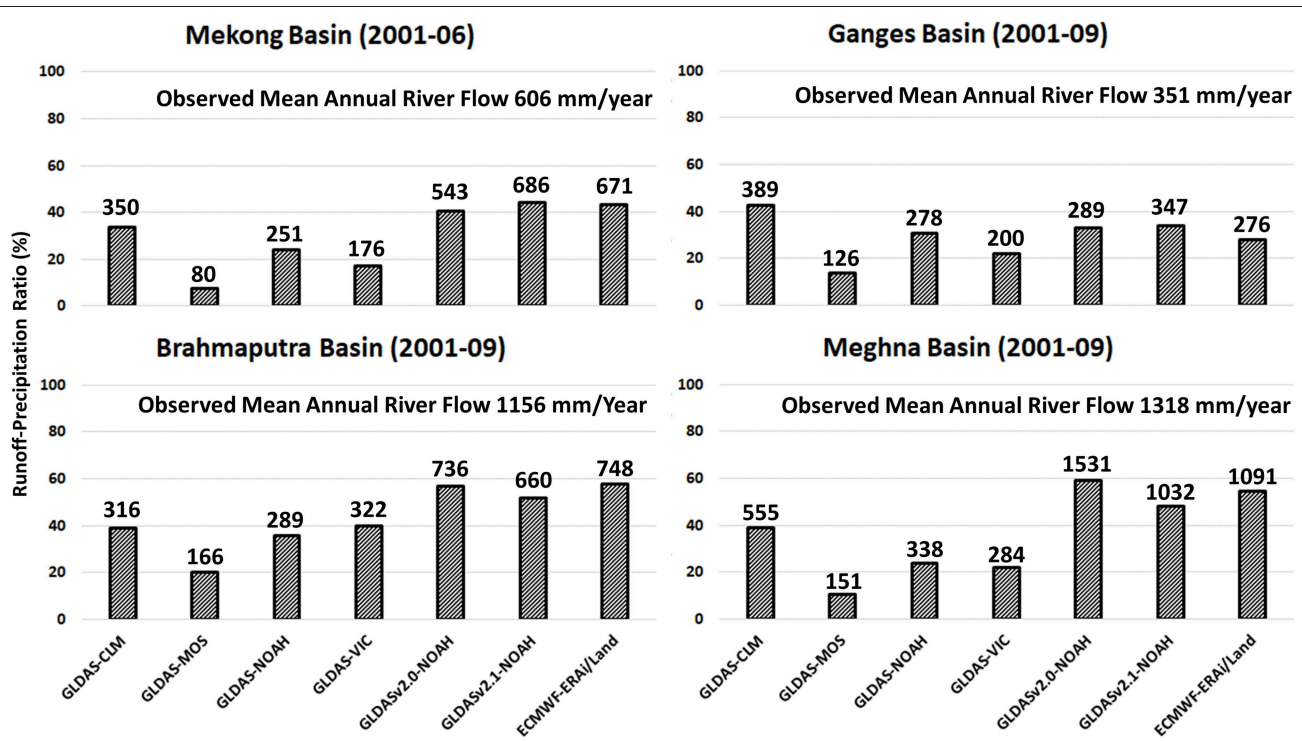


FIGURE 3 | Mean annual Runoff-Precipitation ratio of the Mekong and GBM basins, along with the observed mean annual river flow. All the values reported above the bars are the simulated mean annual runoff in mm. Analysis for Mekong basin was conducted up to 2006, constrained by the *in-situ* river flow data availability.

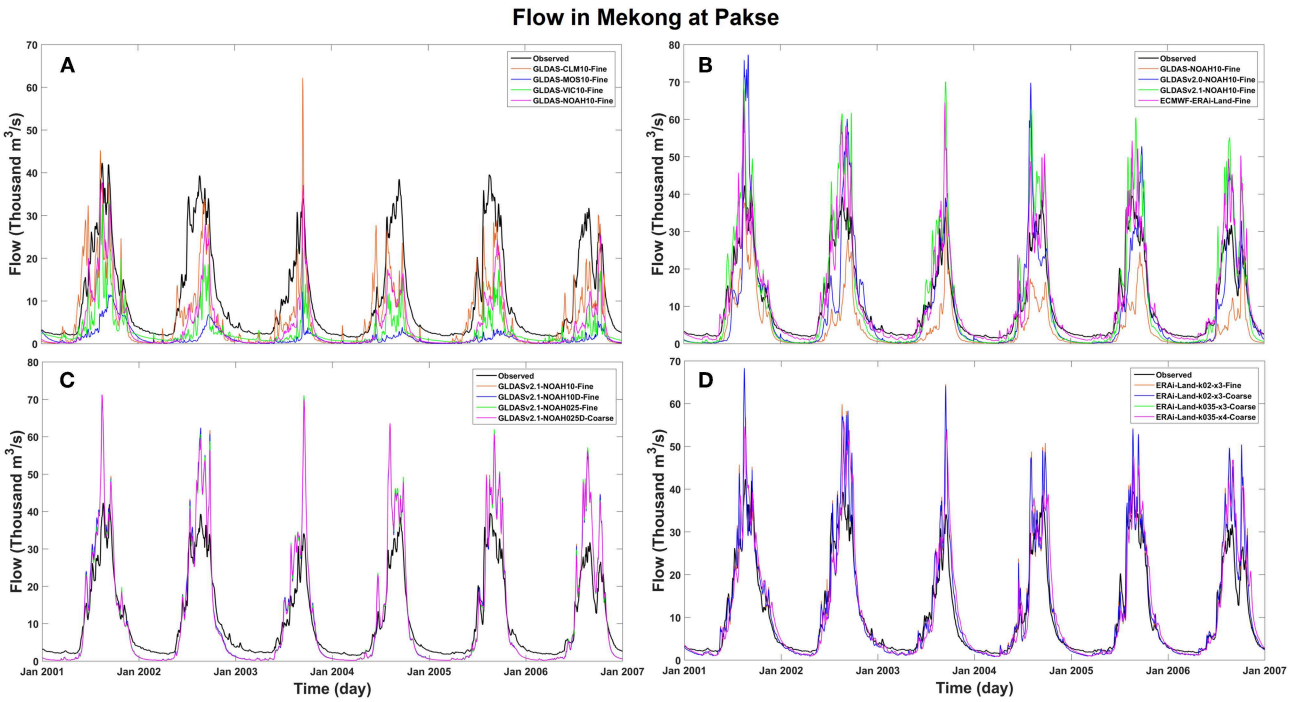


FIGURE 4 | Example of the simulated daily flow hydrograph in the Mekong basin as a function of different **(A)** LSM, **(B)** input forcing, **(C)** spatio-temporal resolution of LSM and routing model, and **(D)** routing model parameter. Observed river flow is shown in black line.

these four different design factors using the correlation, NSE, RMSE, bias, and standard error, as discussed below.

Figure 5 shows the evaluation of the global LSMs as a function of different land surface models used. In this experiment, all the four models of GLDAS were considered, which were forced by the same meteorological data. Outputs using the ECMWF—forced with a different set of data—are also shown as a reference for comparison. Based on the five metrics used in this analysis (**Figure 5**), the selection of the LSM is an influential factor in river discharge simulations, as expected because LSMs are responsible for determining the amount of runoff that is available as inputs to rivers. The NSEs show that CLM performs consistently better among the other GLDAS LSMs. This can be explained by the low estimation of evapotranspiration in CLM combined with the underestimated precipitation forcing in GLDAS. CLM produces relatively higher runoff among the GLDAS LSMs, which is closer to the *in-situ* discharge observations. Overall, the model biases show that the GLDAS generally underestimates runoff likely in part due to the underestimated precipitation forcing. It is also worth noting here that the discharge estimates obtained from the ECMWF LSM runoff are far superior to that of the GLDAS LSMs, for all stations considered, although the determination of the underlying reasons for the relative higher quality is beyond the stated scope of this study.

Figure 6 shows the evaluation metrics for the simulated river flow at multiple locations as a function of the meteorological forcing used. To be consistent, we focus on the Noah model with different meteorological forcing. The ECMWF-based simulations are also shown for reference, as done in the previous analysis. Clearly, the forcing also has an impact on the accuracy of the simulated flow, particularly in model bias. This impact on simulated flow is also to be expected because precipitation is a key driver of the terrestrial water cycle. **Figure 6** suggests that the updated precipitation forcing in GLDAS-2 helps reduce the aforementioned negative bias of the simulated river flow and improves model accuracy, which is evidenced by the high NSE values obtained. The performance of the ECMWF and GLDASv2.1 is similar in most cases. Yet, the ECMWF performs better for all *in-situ* stations except the Meghna, where the GLDASv2.1 metrics are slightly better. The accuracy of the simulated flow using GLDASv2.1 is greater than that of GLDASv2.0, and is followed by that of GLDAS.

The simulated flows are then evaluated for different spatio-temporal resolutions of the LSMs and for the spatial resolution of the river routing model (**Figure 7**). This analysis was carried out for 3-hourly and daily LSM data with 0.25 and 1° spatial resolution and using both fine and coarse resolutions for the river network used in the RAPID model. Surprisingly, despite these rather different spatio-temporal resolutions, **Figure 7** shows that the performance in all cases is very similar. The linearity of the Muskingum equations that drive the RAPID model may be the source of this similarity in discharge outputs. The analysis was repeated using a few smaller catchments (e.g., 2,000–5,500 km^2) within the Mekong basin (not shown here) and suggests some sensitivity of the simulated flow to varying spatial resolutions of LSM (i.e., 0.25 or 1°), which may be explained by the boundary of

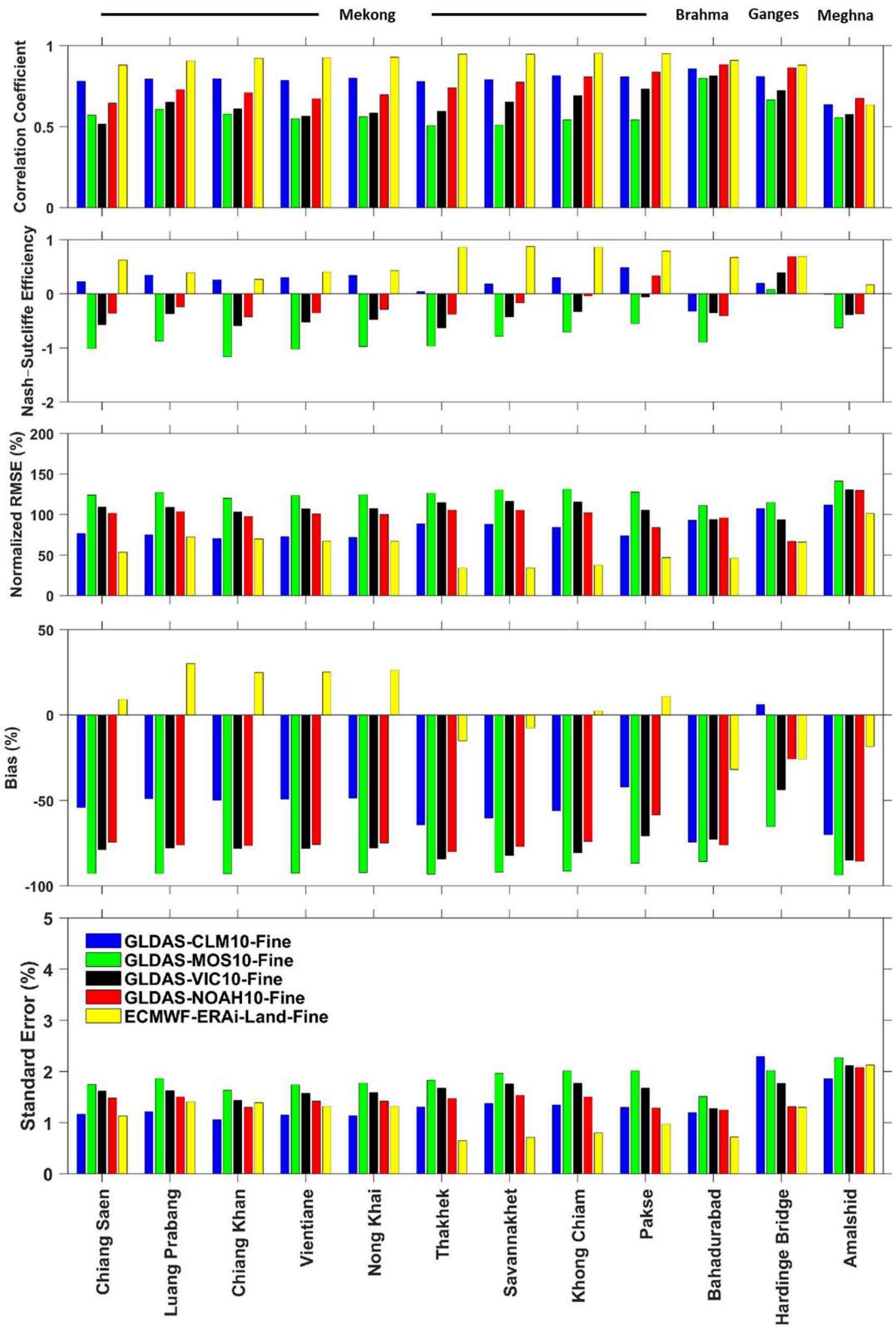


FIGURE 5 | Performance of simulated daily stream flow at different locations of the Mekong and GBM basin as a function of different LSMs. All the GLDAS LMSs in this experiment were forced with the same meteorological inputs. ERA-interim/Land from ECMWF (with different resolution and uses a different meteorological forcing) is shown here as a reference.

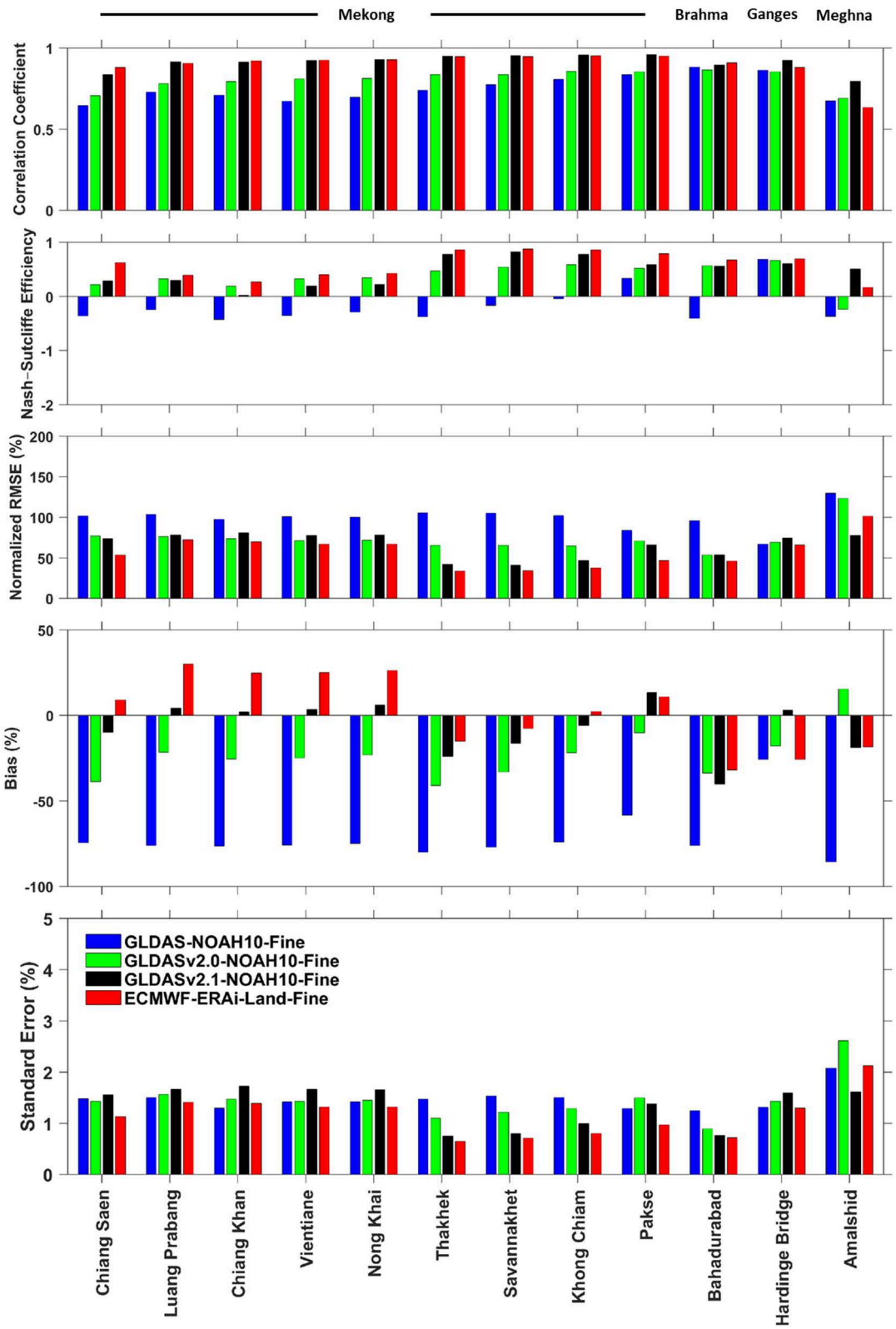


FIGURE 6 | Same as **Figure 5**, but as a function of different meteorological forcing in the same LSM (i.e., Noah model).

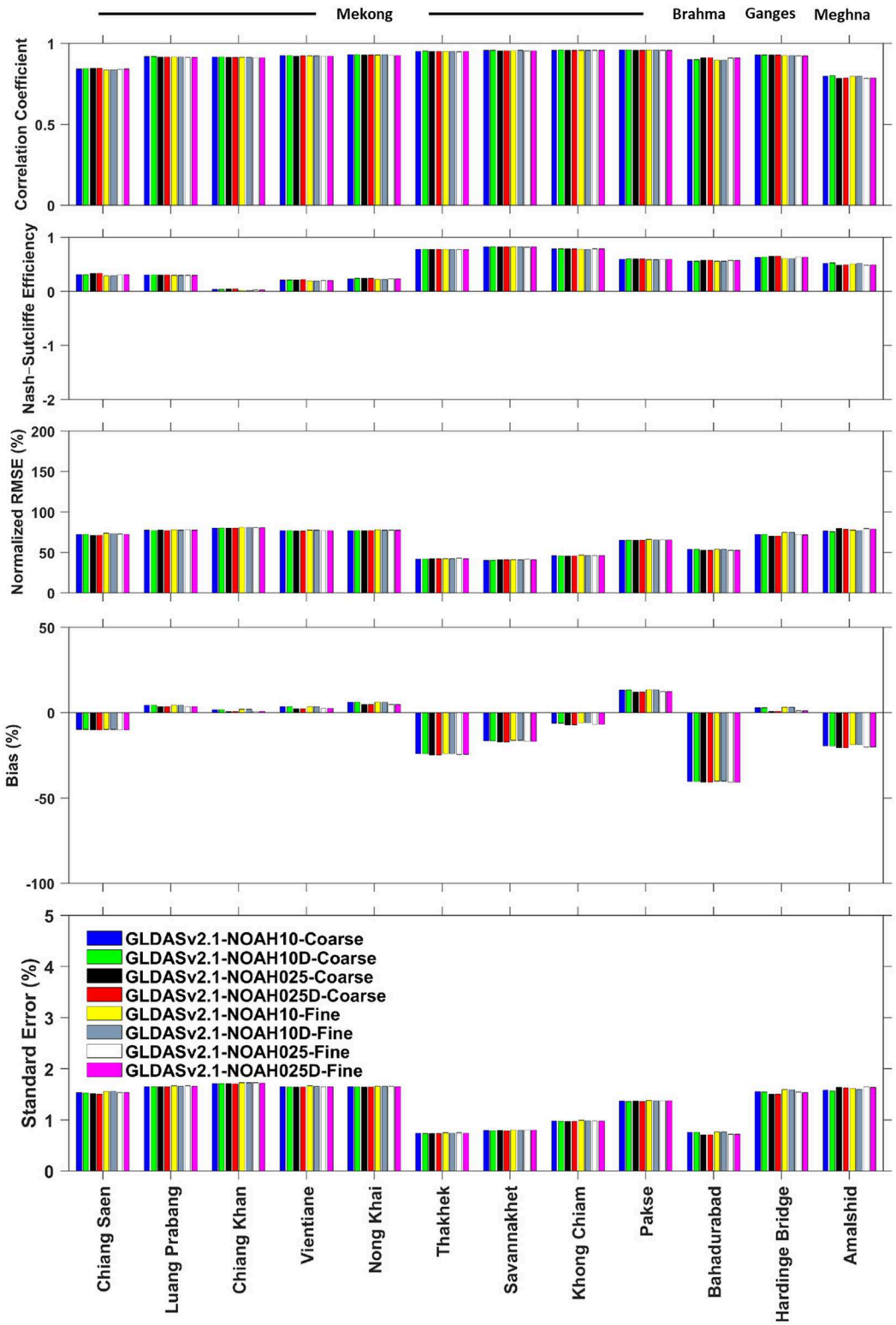


FIGURE 7 | Performance of simulated daily stream flow as a function of different spatio-temporal resolution of the same LSM using the same meteorological forcing (i.e., GLDASv2.1-NOAH). Here, NOAH10 and NOAH025 are for different spatial resolutions, 1 and 0.25°, respectively. The LSM outputs in daily scale, denoted here by "D," and the other test cases (i.e., without "D") are for 3-hourly LSM outputs. The last part of each test case name is representing the river network resolution (**Figure 1**) used for flow routing.

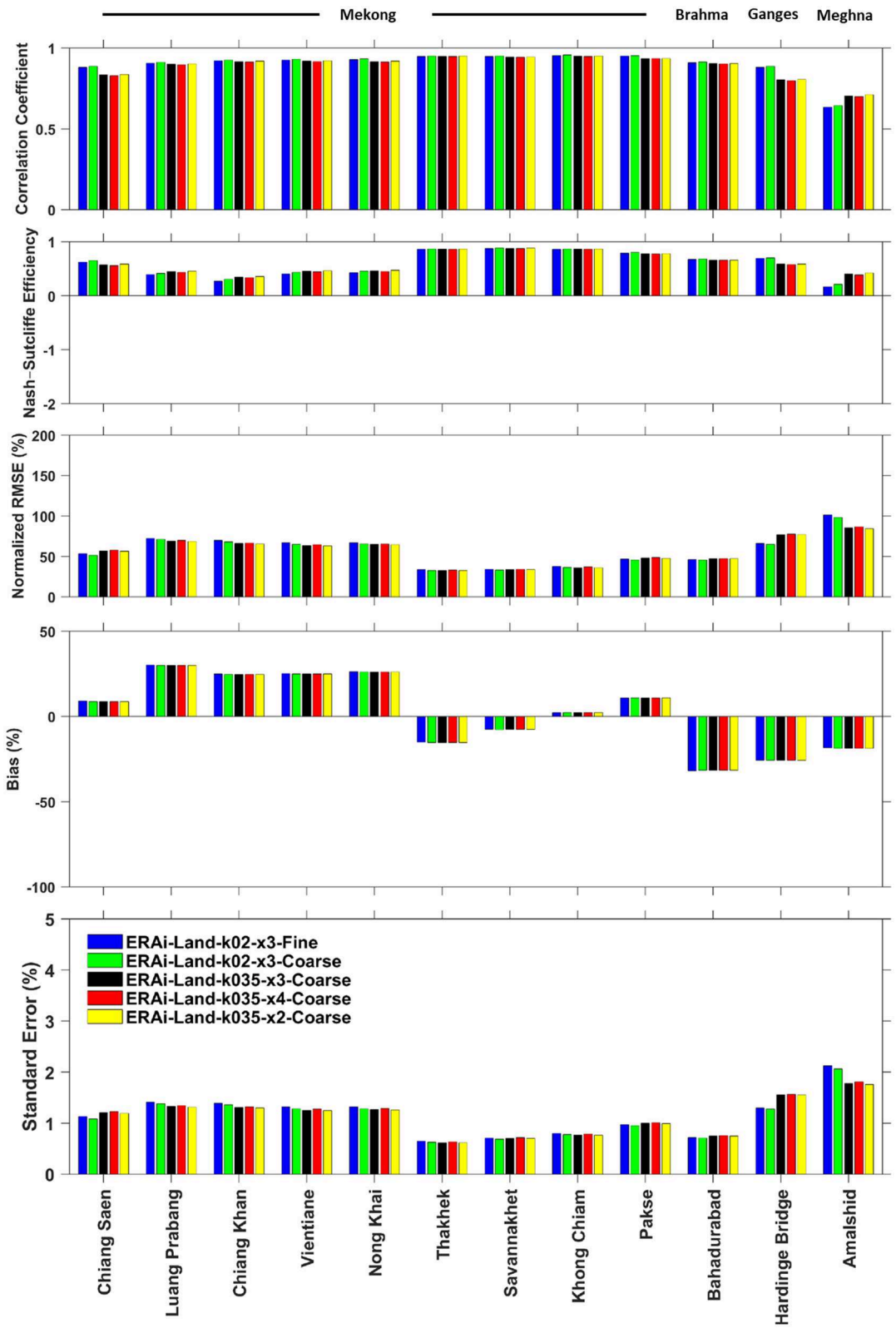


FIGURE 8 | Sensitivity of using different flow routing parameters at different locations of the Mekong and GBM basins. All the routing simulations were conducted with the same LSM outputs (i.e., ECMWF-ERA-interim/Land).

the smaller catchments being better defined by the fine resolution grid, also such effect remains limited.

The influence of the routing model parameters was also evaluated for the simulated river flow (**Figure 8**). The river flow was simulated using the ECMWF model runoff data with different weighting coefficient (x) and storage constant (k) scale factors. The analysis shows that the variation in simulated flow accuracy due to different routing model parameter is not as significant as the selection of the LSM or that of the meteorological forcing. Note here again that the number of test cases in this experiment (i.e., different routing model parameter) is different from the number of test cases considered for different LSMs and meteorological forcing. However, it is to be expected that the storage constant should have an influence on the timing of hydrograph, since this parameter is related to the time of concentration of the basin. This variation is not visible with the five metrics used in this analysis, possibly due to the large temporal window of the analysis (i.e., 6 year). Therefore, an additional metric was used here to determine the accuracy in the timing of the simulated hydrograph.

The lagged cross-correlation was hence calculated by applying a temporal offset to the simulated flow and determining the lag for which the cross-correlation with observed data was highest. The results of this analysis are shown in **Figure 9**, where the lag time of the simulated flows corresponding to the maximum correlation with the observed flows are reported. This analysis was carried out for the simulated flow with different routing model parameters (**Figure 9A**), as well as for different LSM with different meteorological forcing (**Figure 9B**). **Figure 9A** illustrates that the timing of hydrograph influenced by the storage constant (k), as expected. Therefore, while the routing model parameters are not significantly changing the accuracy of the simulated flow as measured by traditional metrics (**Figure 8**), the storage constant is important to fine-tune the hydrograph timing of the simulated flow (**Figure 9A**). This suggests that existing spatially-varying values of k (e.g., Allen et al., 2018) may lead also to improved peak timing. However, the diffusion parameter is relatively insignificant, as was already shown previously (Koussis, 1978). Furthermore, the same analysis for different LSMs with different meteorological forcing (**Figure 9B**) shows that the timing of hydrograph is also influenced by the selection of LSM, as it is by the meteorological forcing. Here again, the timing of hydrograph for the ECMWF experiments is relatively better than that of all versions of GLDAS LSMs

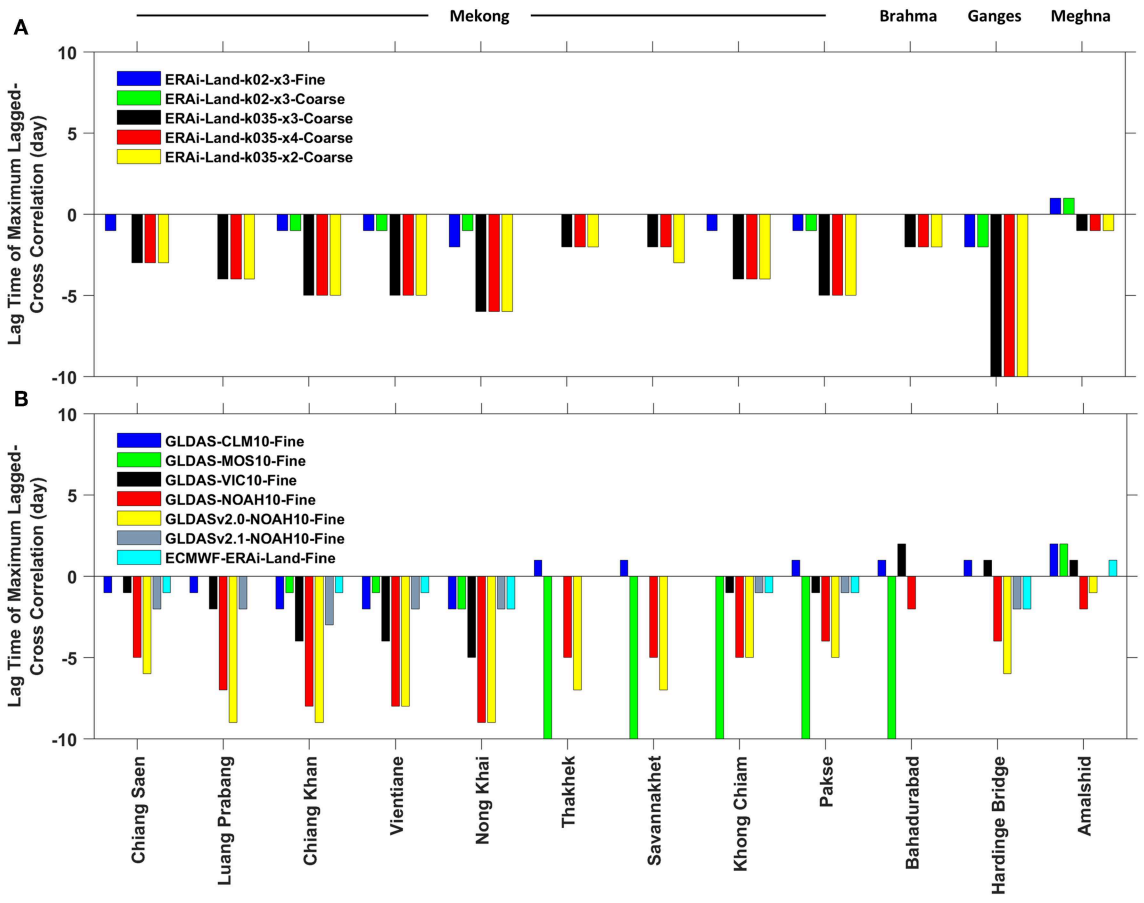


FIGURE 9 | Lag time corresponding to the maximum lagged cross correlation in between the observed and simulated flow at different locations of the Mekong and GBM basins, as a function of different **(A)** routing model parameters, **(B)** LSMs and input forcing.

in all discharge stations with the exception of the Meghna basin (**Figure 9B**).

The overall analysis of the available global LSM data therefore shows that GLDAS-2 and ECMWF performed relatively better than the earlier GLDAS. This relative superiority is reported here for mass balance residual errors (although with caveats for the absence of groundwater processes and the relatively-short study period), and for mean annual flow. The analysis of the RAPID simulated daily river flow using five traditional metrics (correlation, NSE, bias, RMSE, and standard error) with respect to different LSMs and different meteorological forcing together indicate that in most cases ECMWF outperforms the other datasets considered in this study. Note that the different ensemble sizes used for LSMs (five), precipitation (four), model spatiotemporal resolution (eight), and river routing parameters (four) is inconsistent among our experiments partly due to data availability. However, the demonstrated sensitivity obtained for available ensembles (**Figures 4–8**) remains remarkable. In addition, the analysis of lagged cross-correlations also suggests that the ECMWF-derived river flow more accurately maintains the timing of hydrograph. In general, the ECMWF's ERA-interim/Land runoff derived river flow shows the most optimal performance in our study of the Mekong and the GBM river basin, although the determination of the source for this superiority is beyond the scope of this paper. For further illustration, the simulated flow hydrographs from the ECMWF model along with the *in-situ* data at different locations of the aforementioned river basins are shown in **Figure 10**.

CONCLUSIONS

Global Land Surface Models (LSMs) have the potential to help fill the observational gap of ungauged and of transboundary river basins. These models can be particularly valuable to mitigate water resources challenges in the large and densely populated

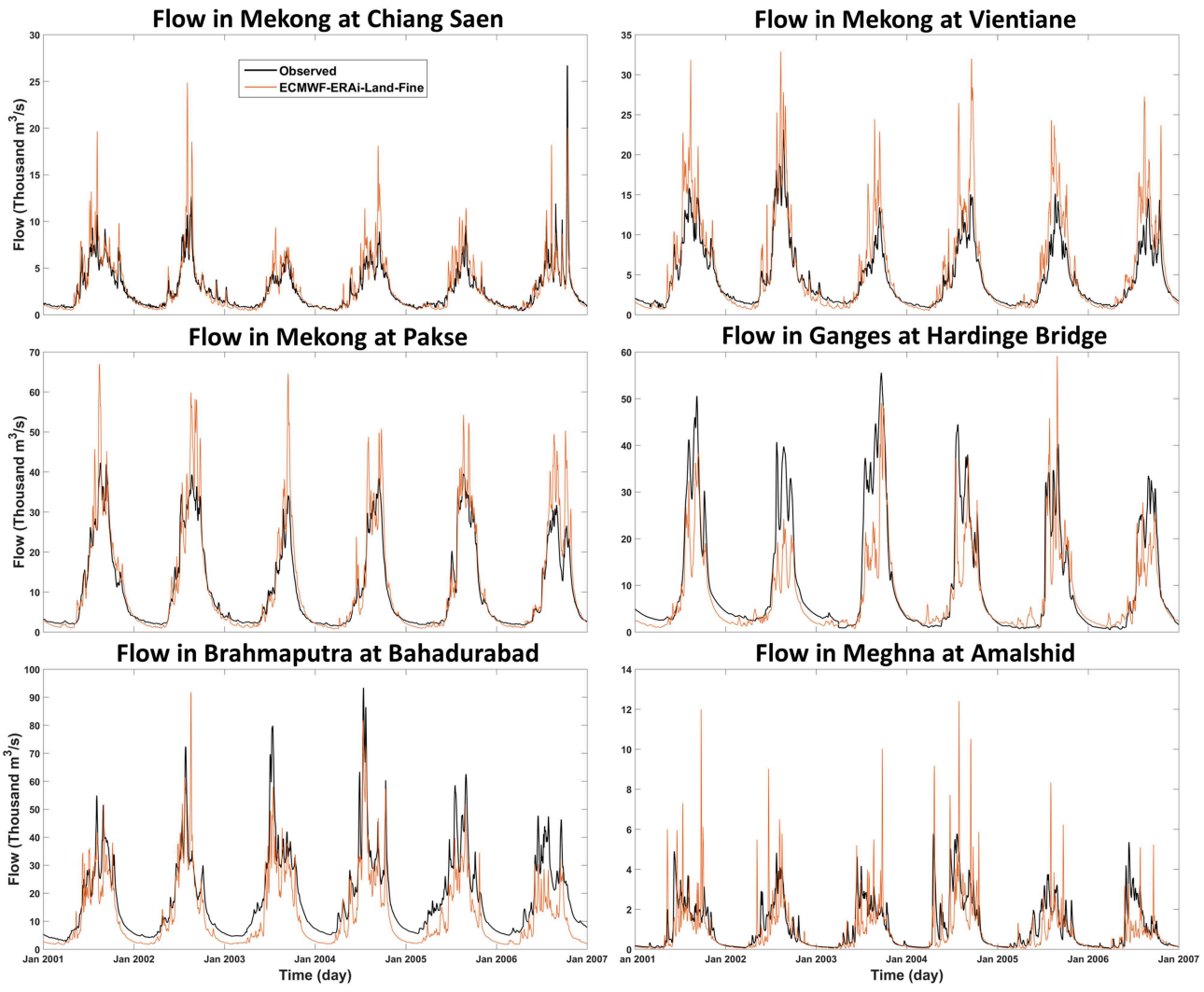


FIGURE 10 | Comparison of the simulated daily stream flow using the ECMWF-ERA-interim/Land outputs in the fine scale routing model (i.e., with fine river network) and *in-situ* discharge at different locations of the Mekong and GBM basins.

transboundary river basins of South and Southeast Asia because they provide valuable estimations of the hydrological fluxes and states. Thus, the investigation of the appropriate global LSM among the many available operational models is critical in order to make best use of these model outputs. The primary goal of this study is therefore to identify the most optimal global LSMs for the region as it pertains to river flow estimation for the region. Concurrently, it is of value to determine the factors influencing the variance in the quality of model simulated flow.

The first part of this study's analysis was carried out based on the mean annual water fluxes estimated by several global LSMs (i.e., GLDAS, GLDAS-2, ECMWF ERA-interim/Land) to generally illustrate model accuracy. The ratios of the model simulated mean annual evapotranspiration and runoff to precipitation was evaluated in this context. The investigation of available global LSMs was conducted with respect to *in-situ* discharge because these data are reliable and more frequently available than the distributed observed water fluxes. The analysis of the simulated flow was executed using four traditional metrics (correlation, Nash-Sutcliffe Efficiency, bias, standard error, and root-mean-square error) to evaluation the sensitivity of the simulated flow to the selection of the LSM, the meteorological forcing, the spatio-temporal resolutions of the models, and the routing model parameters. Finally, the lagged cross-correlation analysis was conducted to evaluate the accuracy in the timing of the simulated hydrograph.

The analyses based on the mean annual water fluxes indicate that GLDAS-2 and ECMWF show better consistency with observed data than the earlier version of GLDAS. The model estimated mean annual runoff from GLDAS-2 and ECMWF products are also show better agreement with the observed runoff. Among the four LSMs of the GLDAS, CLM may appear to perform relatively better in terms of the runoff, although this may be due to underestimated precipitation combined with underestimates of evapotranspiration. The four experimental cases for the sensitivity analysis reveal that simulated river flow is mostly influenced by the selection of the LSM and by its input meteorological forcing. In contrast, the impact of the spatio-temporal resolutions of the LSMs is much lower for large river basins. A similarly low sensitivity is shown in the case of different routing model parameters for the traditional discharge metrics used. One should note that the varying size of ensembles in our experiments was largely a result of data availability and may have an impact in the sensitivity analysis of this study although enforcing consistency in ensemble sizes is beyond our intended scope. However, the analysis of lagged cross-correlation suggests that the flow wave propagation time has notable impact on the timing of hydrographs. Overall, the discharge simulations using runoff from the ECMWF ERA-Interim/Land outperform those from all other tested LSMs in terms of the simulated river flow accuracy as well as the timing of hydrograph in our study of the GBM and the Mekong basins. Our results suggest that the accuracy of ECMWF-derived flows is then followed by that of GLDASv2.1 GLDASv2.0, and GLDAS. One notable limitation of this study is that it does not include the anthropogenic effects (e.g., water diversions, dams, or land use change) in the models or their analysis. However, previous studies argued that the Brahmaputra basin is relatively in pristine condition since there is no major human intervention in the river (Biancamaria et al., 2011). There are also several existing dams and barrages in the Mekong River basin, but the impact of water use (i.e., for irrigation) on the mean annual flow for these diversions is relatively insignificant, at least during the study period. Haddeland et al. (2006) reported that the total use of water in Mekong basin is only 2.3% of the mean annual flow. While this small water use portion is applicable to our study period, it is expected to grow for later dates as a few large dams are under construction on the mainstem of the Mekong river, which is expected to change the impact on mean annual flow from ∼2% in 2008 to ∼20% in 2025 (Hecht et al., 2019). Future similar studies could therefore consider the inclusion of these anthropogenic activities, particularly in the Mekong basin, but also in the GBM. It is important to note here that the determination of the underlying reasons for the relative superiority of ECMWF runoff in our study of South and Southeast Asia is beyond the scope of this paper, and that our results are likely to be geographically dependent. We do not therefore make any recommendation for other river basins.

DATA AVAILABILITY STATEMENT

The software used for this study can be found at https://github.com/c-h-david/rapid. GLDAS data are freely available with a NASA EarthData Login at https://disc.gsfc.nasa.gov. ERA-Interim/Land data are freely available with an ECMWF Login at https://apps.ecmwf.int/datasets/data/interim-land. Readers interested in obtaining daily discharge observations for our study domain and period may contact the Mekong River Commission and the Bangladesh Water Development Board.

AUTHOR CONTRIBUTIONS

CD and MM conceived this study. The analysis reported and the compilation of first full draft of the manuscript was performed by MS. Initial simulations and preliminary analysis was done by GA, who also wrote an early version of the introduction. XQ and EN provided the necessary data for the coarse resolution river routing model. All authors contributed to writing the manuscript.

FUNDING

MS, CD, and GA were supported by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA; including a grant from the NASA SERVIR Applied Sciences Team. The contribution from SERVIR-HKH hub was supported by USAID.

ACKNOWLEDGMENTS

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The authors

would like to thank the Bangladesh Water Development Board (BWDB) and the Mekong River Commission (MRC) for providing the observed data used in this study. Support from the SERVIR Coordination Office, particularly W. Lee Ellenburg, was critical to this study and is greatly appreciated. Valuable suggestions from the members of the Terrestrial Hydrology group of the Jet Propulsion Laboratory helped to enhance the study. The views and interpretations in this paper are those of the authors and are not necessarily attributable to BYU, ICIMOD, NASA, or USAID. Thank you to the editor and to the two reviewers for their suggestions which helped improve the initial version of this manuscript.

REFERENCES

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., et al. (2013). GloFAS–Global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci.* 17, 12293–12332. doi: 10.5194/hessd-9-12293-2012

Allen, G. H., David, C. H., Andreadis, K. M., Hossain, F., and Famiglietti, J. S. (2018). Global estimates of river flow wave travel times and implications for low-latency satellite data. *Geophys. Res. Lett*. 45, 7551–7560. doi: 10.1029/2018GL077914

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., et al. (2012). *ERA-Interim/Land: A Global Land-Surface Reanalysis Based on ERA-Interim Meteorological Forcing.* ERA Report Series. European Centre for Medium-Range Weather Forecasts (ECMWF). Available online at: https://www.ecmwf.int/node/7922 (accessed May 21, 2019)

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., et al. (2015). ERA-Interim/Land: a global land surface reanalysis data set. *Hydrol. Earth Syst. Sci*. 19, 389–407. doi: 10.5194/hess-19-389-2015

Berg, A. A., Famiglietti, J. S., Rodell, M., Reichle, R. H., Jambor, U., Holl, S. L., et al. (2005). Development of a hydrometeorological forcing data set for global soil moisture estimation. *Int. J. Climatol*. 25, 1697–1714. doi: 10.1002/joc.1203

Bi, H., Ma, J., Zheng, W., and Zeng, J. (2016). Comparison of soil moisture in GLDAS model simulations and *in situ* observations over the Tibetan Plateau. *J. Geophys. Res*. 121, 2658–2678. doi: 10.1002/2015JD024131

Biancamaria, S., Hossain, F., and Lettenmaier, D. P. (2011). Forecasting transboundary river water elevations from space. *Geophys. Res. Lett*. 38:L11401. doi: 10.1029/2011GL047290

Bonnema, M., and Hossain, F. (2017). Inferring reservoir operating patterns across the Mekong basin using only space observations. *Water Resour. Res*. 53, 3791–3810. doi: 10.1002/2016WR019978

Brouwer, R., Akter, S., Brander, L., and Haque, E. (2007). Socioeconomic vulnerability and adaptation to environmental risk: a case study of climate change and flooding in Bangladesh. *Risk Anal*. 27, 313–326. doi: 10.1111/j.1539-6924.2007.00884.x

CEGIS (2006). *Early Warning System*. Final Report to the Asian Development Bank. Centre for Environmental and Geographical Service (CEGIS).

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H. L., Koren, V., et al. (1996). Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res*. 101, 7251–7268. doi: 10.1029/95JD02165

Chen, Y., Yang, K., Qin, J., Zhao, L., Tang, W., and Han, M. (2013). Evaluation of AMSR-E retrievals and GLDAS simulations against observations of a soil moisture network on the Central Tibetan Plateau. *J. Geophys. Res*. 118, 4466–4475. doi: 10.1002/jgrd.50301

Chinnasamy, P., Maheshwari, B., and Prathapar, S. (2015). Understanding groundwater storage changes and recharge in Rajasthan, India through remote sensing. *Water* 7, 5547–5565. doi: 10.3390/w7105547

Dai, A., Qian, T., Trenberth, K. E., and Milliman, J. D. (2009). Changes in continental freshwater discharge from 1948 to 2004. *J. Clim.* 22, 2773–2792. doi: 10.1175/2008JCLI2592.1

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., et al. (2003). The common land model. *Bull. Am. Meteorol. Soc*. 84, 1013–1023. doi: 10.1175/BAMS-84-8-1013

David, C. H., Maidment, D. R., Niu, G. Y., Yang, Z. L., Habets, F., and Eijkhout, V. (2011). River network routing on the NHDPlus dataset. *J. Hydrometeorol.* 12, 913–934. doi: 10.1175/2011JHM1345.1

FAO (2016). *AQUASTAT Main Database.* Food and Agriculture Organization of the United Nations (FAO). Available online at: http://www.fao.org/nr/water/aquastat/data/query/index.html (accessed May 21, 2019).

Fread, D. L. (1993). "Flow routing," in *Handbook of Hydrology*, ed D. R. Maidment (New York, NY: McGraw-Hill), 10.7–10.8.

Haddeland, I., Lettenmaier, D. P., and Skaugen,. T. (2006). Effects of irrigation on the water and energy balances of the Colorado and Mekong River basins. *J. Hydrol.* 324, 210–223. doi: 10.1016/j.jhydrol.2005.09.028

Hanington, P., To, Q. T., Van, P. D. T., Doan, N. A. V., and Kiem, A. S. (2017). A hydrological model for interprovincial water resource planning and management: a case study in the Long Xuyen Quadrangle, Mekong Delta, Vietnam. *J. Hydrol*. 547, 1–9. doi: 10.1016/j.jhydrol.2017.01.030

Hecht, J. S., Lacombe, G., Arias, M. E., Dang, T. D., and Piman, T. (2019). Hydropower dams of the Mekong River basin: a review of their hydrological impacts. *J. Hydrol*. 568, 285–300. doi: 10.1016/j.jhydrol.2018.10.045

Hoang, L. P., Vliet, M. T. H., Kummu, M., Lauri, H., Koponen, J., Supit, I., et al. (2019). The Mekong's future flows under multiple drivers: how climate change, hydropower developments and irrigation expansions drive hydrological changes. *Sci. Tot. Environ.* 649, 601–609. doi: 10.1016/j.scitotenv.2018.08.160

Hossain, F., Siddique-E-Akbor, A. H., Mazumder, L. C., ShahNewaz, S. M., Biancamaria, S., Lee, H., et al. (2014). Proof of concept of an altimeter-based river forecasting system for transboundary flow inside Bangladesh. *IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens*. 7, 587–601. doi: 10.1109/JSTARS.2013.2283402

Hossain, F., Sikder, S., Biswas, N., Bonnema, M., Lee, H., Luong, N. D., et al. (2017). Predicting water availability of the regulated Mekong River basin using satellite observations and a physical model. *Asian J. Water Environ. Pollut*. 14, 39–48. doi: 10.3233/AJW-170024

Johnston, R., and Kummu, M. (2012). Water resource models in the Mekong basin: a review. *Water Resour. Manage*. 26, 429–455. doi: 10.1007/s11269-011-9925-8

Khandu, F. E., Schumacher, M., Awange, J. L., and Schmied, H. M. (2016). Exploring the influence of precipitation extremes and human water use on total water storage (TWS) changes in the Ganges-Brahmaputra-Meghna river basin. *Water Resour. Res*. 52, 2240–2258. doi: 10.1002/2015WR018113

Koren, V., Schaake, J., Mitchell, K., Duan, Q. Y., Chen, F., and Baker, J. M. (1999). A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *J. Geophys. Res*. 104, 19569–19585. doi: 10.1029/1999JD900232

Koster, R. D., and Suarez, M. J. (1996). *Energy and Water Balance Calculations in the Mosaic LSM*. NASA Tech. Memo. Vol. 9. Greenbelt, MD: NASA Goddard Space Flight Cent. Available online at: https://gmao.gsfc.nasa.gov/pubs/docs/Koster130.pdf (accessed May 21, 2019).

Koussis, A. D. (1978). Theoretical estimation of flood routing parameters. *J. Hydraul. Div*. 104, 109–115.

Kummu, M., Lu, X. X., Wang, J. J., and Varis, O. (2010). Basin-wide sediment trapping efficiency of emerging reservoirs along the Mekong. *Geomorphology* 119, 181–197. doi: 10.1016/j.geomorph.2010.03.018

Kummu, M., and Sarkkula, J. (2008). Impact of the Mekong river flow alteration on the tonle sap flood pulse. *AMBIO* 37, 185–192. doi: 10.1579/0044-7447(2008)37[185:IOTMRF]2.0.CO;2

Lakshmi, V., Fayne, J., and Bolten, J. (2018). A comparative study of available water in the major river basins of the world. *J. Hydrol.* 567, 510–532. doi: 10.1016/j.jhydrol.2018.10.038

Lehner, B., Verdin, K., and Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos Trans. Am. Geophys. Union*. 89, 93–104. doi: 10.1029/2008EO100001

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res*. 99, 14415–14428. doi: 10.1029/94JD00483

Maswood, M., and Hossain, F. (2015). Advancing river modelling in ungauged basins using satellite remote sensing: the case of the Ganges–Brahmaputra–Meghna basin. *Int. J. River Basin Manage.* 14, 103–117. doi: 10.1080/15715124.2015.1089250

McCarthy, G. T. (1938). "The unit hydrograph and flood routing," in *Conference of the North Atlantic Division* (New London, CT: U.S. Engineer Department), 1–19.

Mirza, M. M. Q. (2011). Climate change, flooding in South Asia and implications. *Region. Environ. Change* 11, 95–107. doi: 10.1007/s10113-010-0184-7

Mirza, M. M. Q., Warrick, R. A., and Ericksen, N. J. (2003). The implications of climate change on floods of the Ganges, Brahmaputra and Meghna rivers in Bangladesh. *Clim. Change* 57, 287–318. doi: 10.1023/A:1022825915791

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., et al. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res.* 109:D07S90. doi: 10.1029/2003JD003823

Murshed, S. B., and Kaluarachchi, J. J. (2018). Scarcity of fresh water resources in the Ganges Delta of Bangladesh. *Water Secur.* 4–5, 8–18. doi: 10.1016/j.wasec.2018.11.002

Nishat, B., and Rahman, S. M. M. (2009). Water resources modeling of the Ganges-Brahmaputra-Meghna river basins using satellite remote sensing data. *J. Am. Water Resour. Assoc.* 45, 1313–1327. doi: 10.1111/j.1752-1688.2009.00374.x

Orr, S., Pittock, J., Chapagain, A., and Dumaresq, D. (2012). Dams on the Mekong river: lost fish protein and the implications for land and water resources. *Glob. Environ. Change* 22, 925–932. doi: 10.1016/j.gloenvcha.2012.06.002

Pokhrel, Y., Burbano, M., Roush, J., Kang, H., Sridhar, V., and Hyndman, D. (2018). A review of the integrated effects of changing climate, land use, and dams on Mekong river hydrology. *Water* 10:266. doi: 10.3390/w10030266

Priya, S., Young, W., Hopson, T., and Avasthi, A. (2017). *Flood Risk Assessment and Forecasting for the Ganges-Brahmaputra-Meghna River Basins (English).* Washington, DC: World Bank Group. Available online at: http://documents.worldbank.org/curated/en/272611508255892547/Flood-risk-assessment-and-forecasting-for-the-Ganges-Brahmaputra-Meghna-River-basins (accessed May 21, 2019).

Purdy, A. J., David, C. H., Sikder, M. S., Reager, J. T., Chandanpurkar, H. A., Jones, N. L., et al. (2019). An open-source tool to facilitate the processing of GRACE Observations and GLDAS outputs: an evaluation in Bangladesh. *Front. Environ. Sci.* 7, 1–8. doi: 10.3389/fenvs.2019.00155

Räsänen, T. A., Koponen, J., Lauri, H., and Kummu, M. (2012). Downstream hydrological impacts of hydropower development in the upper Mekong basin. *Water Resour. Manage.* 26, 3495–3513. doi: 10.1007/s11269-012-0087-0

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., et al. (2004). The global land data assimilation system. *Bull. Am. Meteorol. Soc.* 85, 381–394. doi: 10.1175/BAMS-85-3-381

Rodell, M., McWilliams, E. B., Famiglietti, J. S., Beaudoing, H. K., and Nigro, J. (2011). Estimating evapotranspiration using an observation based terrestrial water budget. *Hydrol. Process.* 25, 4082–4092. doi: 10.1002/hyp.8369

Rodell, M., Velicogna, I., and Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature* 460, 999–1002. doi: 10.1038/nature08238

Rui, H., and Beaudoing, H. (2017). *README Document for GLDAS Version 1 Data Products.* NASA Goddard Earth Science Data Information and Services Center (GES DISC). Available online at: https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS_V1/README.GLDAS.pdf (accessed May 21, 2019).

Rui, H., and Beaudoing, H. (2019*). README Document for NASA GLDAS Version 2 Data Products.* NASA Goddard Earth Science Data Information and Services Center (GES DISC). Available online at: https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_3H.2.0/doc/README_GLDAS2.pdf (accessed May 21, 2019).

Sheffield, J., Goteti, G., and Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* 19, 3088–3111. doi: 10.1175/JCLI3790.1

Siddique-E-Akbor, A. H. M., Hossain, F., Sikder, S., Shum, C. K., Tseng, S., Yi, Y., et al. (2014). Satellite precipitation data–driven hydrological modeling for water resources management in the Ganges, Brahmaputra, and Meghna basins. *Earth Interact.* 18, 1–25. doi: 10.1175/EI-D-14-0017.1

Sikder, M. S., and Hossain, F. (2018). Improving operational flood forecasting in monsoon climates with bias-corrected quantitative forecasting of precipitation. *Int. J. River Basin Manage.* 1–11. doi: 10.1080/15715124.2018.1476368

Snow, A. D. (2015). *A new global forecasting model to produce high-resolution stream forecasts* [Master's thesis]. Brigham Young University, Provo, UT, United States.

Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., and Wilson, C. R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resour. Res.* 44:W02433. doi: 10.1029/2006WR005779

Tatsumi, K., and Yamashiki, Y. (2015). Effect of irrigation water withdrawals on water and energy balance in the Mekong river basin using an improved VIC land surface model with fewer calibration parameters. *Agric. Water Manage.* 159, 92–106. doi: 10.1016/j.agwat.2015.05.011

UNEP (2016). *Transboundary River Basins Status and Trends SUMMARY FOR POLICY MAKERS.* United Nations Environment Programme (UNEP). Available online at: http://www.geftwap.org/publications/river-basins-spm (accessed May 21, 2019).

Wang, F., Wang, L., Koike, T., Zhou, H., Yang, K., Wang, A., et al. (2011). Evaluation and application of a fine-resolution global data set in a semiarid mesoscale river basin with a distributed biosphere hydrological model. *J. Geophys. Res.* 116:D21108. doi: 10.1029/2011JD015990

Webster, P. J., Jian, J., Hopson, T. M., Hoyos, C. D., Agudelo, P. A., Chang, H. R., et al. (2010). Extended-range probabilistic forecasts of Ganges and Brahmaputra floods in Bangladesh. *Bull. Am. Meteorol. Soc.* 91, 1493–1514. doi: 10.1175/2010BAMS2911.1

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., et al. (2012b). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System Project Phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *J. Geophys. Res.* 117:D03110. doi: 10.1029/2011JD016051

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (2012a). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System Project Phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.* 117:D03109. doi: 10.1029/2011JD016048

Zaitchik, B. F., Rodell, M., and Olivera, F. (2010). Evaluation of the global land data assimilation system using global river discharge data and a source-to-sink routing scheme. *Water Resour. Res.* 46:W06507. doi: 10.1029/2009WR007811

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 05 November 2019
doi: 10.3389/fenvs.2019.00150

Land Cover Mapping in Data Scarce Environments: Challenges and Opportunities

*David Saah[1,2,3], Karis Tenneson[2], Mir Matin[4], Kabir Uddin[4], Peter Cutter[2,3], Ate Poortinga[2,3], Quyen H. Nguyen[3,5], Matthew Patterson[6], Gary Johnson[1], Kel Markert[7,8], Africa Flores[7,8], Eric Anderson[7,8], Amanda Weigel[7,8], Walter L. Ellenberg[7,8], Radhika Bhargava[9], Aekkapol Aekakkararungroj[2,5], Biplov Bhandari[2,5], Nishanta Khanal[3], Ian W. Housman[6], Peter Potapov[10], Alexandra Tyukavina[10], Paul Maus[6], David Ganz[11], Nicholas Clinton[4] and Farrukh Chishtie[2,3,5]**

[1] Geospatial Analysis Lab, University of San Francisco, San Francisco, CA, United States, [2] Spatial Informatics Group, LLC, Pleasanton, CA, United States, [3] SERVIR-Mekong, Bangkok, Thailand, [4] International Centre for Integrated Mountain Development, Kathmandu, Nepal, [5] Asian Disaster Preparedness Center, Bangkok, Thailand, [6] RedCastle Resources, Inc. Contractor to: USDA Forest Service Geospatial Technology and Applications Center (GTAC), Salt Lake City, UT, United States, [7] Earth System Science Center, The University of Alabama in Huntsville, Huntsville, AL, United States, [8] SERVIR Science Coordination Office, NASA Marshall Space Flight Center, Huntsville, AL, United States, [9] Google Inc., Mountain View, CA, United States, [10] Department of Geographical Sciences, University of Maryland, College Park, MD, United States, [11] RECOFTC—The Center for People and Forests, Bangkok, Thailand

OPEN ACCESS

Edited by:
Niall Patrick Hanan, New Mexico State University, United States

Reviewed by:
Francis Kwabena Dwomoh, University of Oklahoma, United States
Nathaniel K. Newlands, Agriculture and Agri-Food Canada (AAFC), Canada

***Correspondence:**
Farrukh Chishtie
farrukh.chishtie@adpc.net

Specialty section:
This article was submitted to Interdisciplinary Climate Studies, a section of the journal Frontiers in Environmental Science

Received: *02 June 2019*
Accepted: *17 September 2019*
Published: *05 November 2019*

Citation:
Saah D, Tenneson K, Matin M, Uddin K, Cutter P, Poortinga A, Nguyen QH, Patterson M, Johnson G, Markert K, Flores A, Anderson E, Weigel A, Ellenberg WL, Bhargava R, Aekakkararungroj A, Bhandari B, Khanal N, Housman IW, Potapov P, Tyukavina A, Maus P, Ganz D, Clinton N and Chishtie F (2019) Land Cover Mapping in Data Scarce Environments: Challenges and Opportunities. Front. Environ. Sci. 7:150. doi: 10.3389/fenvs.2019.00150

Land cover maps are a critical component to make informed policy, development, planning, and resource management decisions. However, technical, capacity, and institutional challenges inhibit the creation of consistent and relevant land cover maps for use in developing regions. Many developing regions lack coordinated capacity, infrastructure, and technologies to produce a robust land cover monitoring system that meets land management needs. Local capacity may be replaced by external consultants or methods which lack long-term sustainability. In this study, we characterize and respond to the key land cover mapping gaps and challenges encountered in the Lower Mekong (LMR) and Hindu Kush-Himalaya (HKH) region through a needs assessment exercise and a collaborative system design. Needs were assessed using multiple approaches, including focus groups, user engagement workshops, and online surveys. Efforts to understand existing limitations and stakeholder needs resulted in a co-developed and modular land cover monitoring system which utilizes state-of-the-art cloud computing and machine learning which leverages freely available Earth observations. This approach meets the needs of diverse actors and is a model for transnational cooperation.

Keywords: land cover/land use, GIS, remote sensing, Lower Mekong region, Hindu Kush region

1. INTRODUCTION

Accurate and timely land cover maps play a critical role in a variety of sectors in the developing world including food security, land use planning, hydrology modeling, and natural resource management planning. Countries like Cambodia and Vietnam suffer from substantial rice crop yield losses and understanding the spatial distribution of such variable yields are critical for agricultural planning for food security (Pandey et al., 2007; Saah et al., 2019). National

development plans use land cover as a basis for understanding changes in a country's natural capital, that in turn forms the basis for budget priorities and allocations (Tucker et al., 1985; Bounoua et al., 2002; Foley et al., 2005; Jung et al., 2006; Running, 2008). Land cover maps also underpin hydrology models that are used by governments to inform flood risk and preparedness in order to build resilience to climate change (Ge et al., 2007; Hibbard et al., 2010; Imaoka et al., 2010; Gong et al., 2013; Tolentino et al., 2016). Foresters use land cover maps to develop sustainable harvest management plans, integrate biodiversity conservation, and engage in climate finance initiatives such as Reducing Emissions from Deforestation and forest degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in Developing countries (REDD+) or Nationally Appropriate Mitigation Action (NAMA) (Buchanan et al., 2008; Hall et al., 2011; Potapov et al., 2019). However, many developing regions lack the coordinated capacity to produce timely, accurate, and temporally comparable geospatial data products sufficient to meet their management needs (Jha and Chowdary, 2007).

The Lower Mekong (LMR) and Hindu Kush-Himalaya (HKH) regions are both experiencing an acceleration in the rate of land cover change that is impacting the long-term sustainability of ecosystem services including food, water, and energy. Local decision makers are using infrequently updated national maps with no ability to monitor in a timely or integrated fashion. Furthermore, existing classification systems do not always meet the user group's needs, data products are often not widely shared between agencies and institutions, and accuracy assessment is often lacking. The users and developers of these maps are typically from different organizations, with different priorities and technical understandings. These differences pose a variety of challenges that often create roadblocks to the effective use of appropriate land cover data for policy formulation, planning, management, and other decision contexts. As a result, global land cover products are frequently used as the best available alternative when appropriate and timely maps are not available at the regional, national, or sub-national levels. The global products have their limitations in that they have been created using different sensors and different techniques and vary in spatial resolution and classification typology, and contain inconsistencies on global (Hansen and Reed, 2000; McCallum et al., 2006; Herold et al., 2008; Hansen et al., 2013) and regional (Fritz et al., 2010; Perera et al., 2010; Leinenkugel et al., 2013, 2014) scales. These inconsistencies hinder more widespread and effective use of land cover data to valuably contribute to policy formulation, planning, management, and other processes where more effective, transparent, and defensible decisions are known to lead to better real-world outcomes.

To understand and respond to the key gaps and challenges encountered in regional, national, and subnational land cover mapping efforts, we compiled existing information, undertook proactive and targeted needs assessment activities, and compiled other information relevant to this topic from a variety of ongoing activities and communications. This paper summarizes the details and findings of those efforts and proposes a framework for addressing the identified challenges. The purpose of this paper is to review the key challenges in land cover mapping and monitoring in the Lower Mekong and HKH region and to outline design principles for the co-development of a Regional Land Cover Monitoring System.

The process of cooperatively developing a consistent, flexible Regional Land Cover Monitoring System (RLCMS) across multiple nations in SE and South Asia will allow the creation of a unified pool of land cover data and land cover mapping/monitoring architecture that can be shared across agencies and countries. Such an RLCMS will also allow for consistent change analysis to be performed going into the future, and potentially over the entire historical satellite record. Such a system would represent a landmark improvement in the availability of remote sensing-based data products for use in planning and governance.

2. STUDY AREA AND HISTORICAL CONTEXT

SERVIR is a unique partnership between the U.S. Agency for International Development (USAID) and the U.S. National Aeronautics and Space Agency (NASA) focused on bringing space-based technologies to environmental decision makers in developing regions. SERVIR is represented in different parts of the world through regional hubs. The program is harnessing space-based remote sensing technology and open data to help address development challenges related to climate resilience. As a development program aimed at promoting use and development of regional geospatial data, SERVIR works in partnership with leading regional organizations to help countries in the hub regions use information provided by Earth observing satellites and geospatial technologies to manage climate risks and improve disaster preparedness. The program services are driven by a needs assessment of the partnering agencies. We selected the SERVIR hubs within the LMR and HKH regions to identify what challenges exist in creating and using land cover data products due to its multidimensional diversity that includes political differences, vegetation gradients, and a variety of needs for and uses of land cover (**Figure 1**). The Mekong hub was initiated in 2014, and the work during the first year was focused on conducting a needs assessment with organizations working in the partnering countries. The SERVIR HKH hub was established prior to the Mekong hub, and while their outreach methods varied slightly they also conduct needs assessments with their country stakeholders. Land cover/use mapping and monitoring challenges facing our stakeholders were compiled during the first year user needs assessment phase in the Mekong and the on-going needs assessment work in the HKH, and provides the basis for this manuscript.

2.1. The Lower Mekong Region (LMR)

The LMR covers a 1.9 million km^2 area, has a population of 240 million people, and includes five continental Southeast Asian countries: Cambodia, Laos, Myanmar, Thailand, and Vietnam (**Figure 1**). The LMR is largely tropical, with subtropical extensions in Laos, Myanmar, and Vietnam, and minor areas

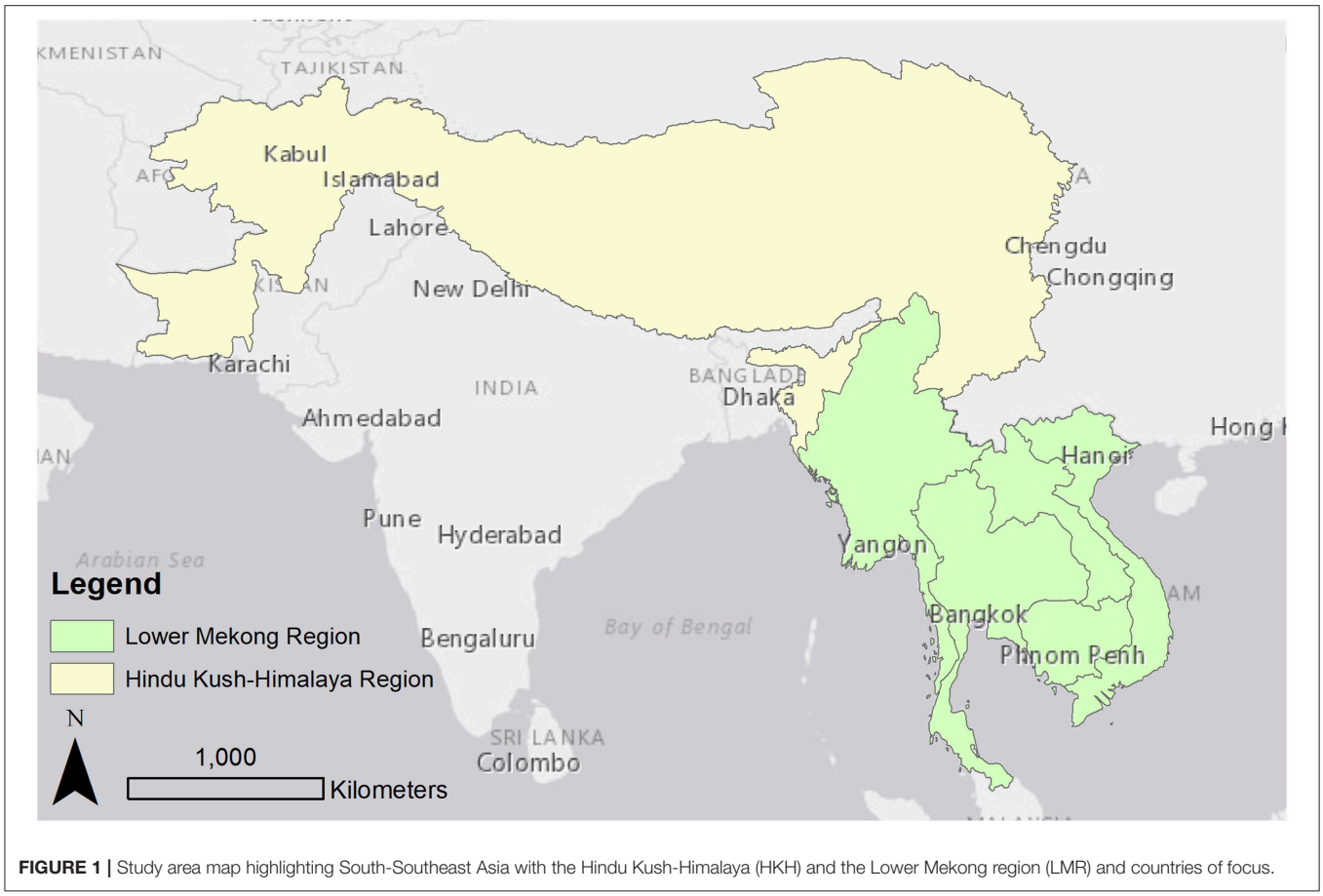


FIGURE 1 | Study area map highlighting South-Southeast Asia with the Hindu Kush-Himalaya (HKH) and the Lower Mekong region (LMR) and countries of focus.

of temperate vegetation in Northern Myanmar (Peel et al., 2007). The region falls within the Indo-Burma biodiversity hotspot and over three-quarters of the population depends directly or indirectly on agriculture (Kityuttachai et al., 2016) and other economic activities including tourism, forestry, fishing, manufacturing, and energy production (Costenbader et al., 2015). Between 1973 and 2009, LMR lost almost a third of its forest cover with national level reductions amounting to 43% in Vietnam and Thailand, 24% in Lao PDR and Myanmar, and 22% in Cambodia (Watch, 2015). The main proximate drivers of land cover change in the past decade included expansion of agriculture and plantation estates, development of infrastructure and settlements, extraction of minerals and gas, dam and water infrastructure development, logging, and forest fires (Curtis et al., 2018). Underlying drivers of land cover change included population and economic growth, often intensified by weak governance (Stibig et al., 2013; Costenbader et al., 2015; Poortinga et al., 2019).

2.2. The Hindu Kush Himalayan Region (HKH)

The HKH region extends 3,500 km over all or part of eight countries from Afghanistan in the west to Myanmar in the east (**Figure 1**). The eight countries include Afghanistan, Bangladesh, Bhutan, China, India, Myanmar, Nepal and Pakistan. The region has the highest number of mountain peaks in the world, including Mt. Everest. The predominant land cover of the region is grassland (54%), followed by agriculture (26%), and forest (14%) (Singh et al., 2011). It is the source of ten large Asian river systems and provides water, ecosystem services, and the basis for livelihoods to a population of around 210 million people. These river basins provide water to 1.3 billion people, a fifth of the world's population. It supports a rich variety of gene pools, species, ecosystems, and endemic species of global importance; making up four of the global biodiversity hotspots—Himalayan, Indo-Burma, Mountains of South-West China and Mountains of Central Asia (Myers et al., 2000).

Despite the emphasis on conservation, with 39% of land under protected areas (Chettri et al., 2008), the HKH region is facing significant deforestation (Ives and Messerli, 2003; Pandit et al., 2007). The dependency on fuel wood and timber from forests, population pressure for conversion of forest to agriculture and agriculture to built up area, shifting cultivation, and forest fire are among the many anthropogenic and natural causes of land cover change. The lack of historical land cover maps using a comparable and consistent methodology and classification scheme makes it difficult to analyze land cover changes within countries while inconsistency in typologies among different countries makes it difficult to develop regionally consistent land cover maps for regional analysis, studies and models.

3. METHODS

3.1. Stakeholder Engagement

There is a significant and growing body of evidence that meaningful participation of stakeholders in the identification of problems and the formulation of corresponding solutions has numerous benefits–especially in the context of complex, transdisciplinary challenges (Reed, 2008). To understand stakeholder needs in the HKH and LMR and focus their program strategies accordingly, each hub conducted both a general geospatial user needs assessment and a follow up workshop focused specifically on challenges regarding land cover mapping and monitoring.

3.1.1. Geospatial-Focused Needs Assessment

The Mekong hub completed the general needs assessment phase most recently, from late 2014 to mid-2015. Data was collected through (1) a literature review of relevant reports and articles, (2) in person consultations with regional stakeholders (in part during regional workshops and focus groups), and (3) an online questionnaire. The stakeholder outreach campaign completed by the HKH hub was similar. The specific aims of the assessment were to identify geospatial data and technology needs in the following thematic areas:

- Key areas where geospatial information is considered important for decision making such as land governance and management, water governance and management, climate change adaptation, disaster risk assessment etc.
- Capacity building requirements such as basic GIS skills, managing complex server structures etc.
- Availability of geospatial data such as Landsat remote sensing products, land cover maps, flood forecast maps, etc.
- Data sharing practices and protocols such as sharing of data from public agencies with the public, sharing of data between agencies, standards for metadata and data quality to facilitate sharing, etc.
- Access to geospatial tools and applications including decision support tools, online information portals, custom desktop applications, etc.

A snowball sample strategy was used, which entails building up the initial sample size via enrolments of relevant stakeholders made by participants during the course of the study (Biernacki and Waldorf, 1981). This resulted in a list of country and regional stakeholders produced from government agencies, academic and research institutions, non-governmental organizations (NGOs) and other civil society organizations, multi- and bilateral aid agencies, United Nations agencies and similar extra-national governance and support institutions, private sector entities, and individual citizens. The sample is not a probabilistic sample, however there was good representation from the targeted organizations by type (e.g., government, civil society) and by country. Further details are provided in **Appendix A**.

3.1.1.1. Focus groups and regional workshops

Stakeholder consultations were implemented in two formats and included the participation of 199 people from 128 organizations. Consultation trips to Cambodia, Vietnam, Myanmar, Thailand, and the Lao PDR were conducted between December 2014 and May 2015. These trips included roundtable discussions and direct meetings.

3.1.1.2. Online questionnaire

An invitation to complete an online questionnaire was circulated via e-mail to more than 300 potential respondents thought to have valuable perspectives on the above topics. The respondents were queried regarding their professional role, institutional affiliation, and GIS and remote sensing background; perspectives on geospatial data needs and gaps, data sharing challenges, technology issues, capacity needs and gaps, and decision support tools and applications. A special effort was made to reach stakeholders who could not attend the live stakeholder consultations by means of the questionnaire.

3.1.1.3. Literature review

Relevant published scientific reports, articles, and papers were used to complement the primary data collected through in-person consultations and the online questionnaire. The full list of documents reviewed is included in the **Appendix A**.

3.1.2. Land Cover Mapping Workshops

A series of four land cover mapping workshops were first conducted in the Mekong region from 2016 to 2017. The first workshop focused on user needs and objectives and outline a system design to address these challenges. The findings presented in this manuscript summarize the findings from this event. There were also three subsequent workshops: two production workshops and a product launch. Eighty stakeholders from the LMR contributed perspectives and ideas during these four regional land cover mapping workshops. Participants were involved in everything from conceptualization to classifications and algorithm development for emergent challenges from discussions.

The same approach of conducting at least four workshops began with regional partners in the HKH in 2018, with some slight adjustments. The initial needs assessment was a regional workshop that included participants from Afghanistan, Bangladesh, Nepal, and Myanmar. During the event the participants discussed and agreed on the conceptual framework and approach for typology and algorithm at regional scale. However, the two subsequent production workshops were organized as separate events with each mapping agency, as opposed to a series of integrated regional production workshops. These country level workshops are being conducted with larger groups of stakeholders to address the needs and approach for customizing the framework to need the country specific needs. This approach allowed for a greater number of participants from each agency, targeted the data generation focus in on one geospatial region, and helped garner agency ownership of the national product. At the culmination of the national workshops, the land cover products will be integrated into a regional data set.

Separate manuscripts are being prepared to describe the resulting land cover product and co-designed system architecture that resulted from these production workshops (Poortinga et al.,

2019; Potapov et al., 2019; Saah et al., 2019, in press). The current manuscript summarizes only the results regarding challenges and design principles from the geospatial needs assessment campaign and first land cover focused workshop.

4. RESULTS

The results are presented in two sections. First is a description of the major land cover mapping challenges facing government and civil society organizations in the LMR and HKH regions. These results summarize the findings from the geospatial needs assessment, workshop activities, and experiences with local land cover mapping efforts from multiple agencies, countries, and programs. Second, we present the co-designed solution that was developed during the land cover monitoring workshop aimed at designing solution(s) to address these challenges. The goal was to design a system which can be implemented in a long-term sustainable fashion.

4.1. Identification of the Challenges and Gaps in Land Cover Mapping and Monitoring Practices

Participants of the regional geospatial needs assessment identified two major types of challenges. The first was an insufficient commitment to land cover mapping efforts, and the second was the challenge accessing geospatial infrastructure. The first major challenge highlights the integration issues associated with people, finance, and priorities. These issues are common among many organizational structures but have specific implications when trying to develop land cover mapping products in the developing world.

4.1.1. Summary of Stakeholder Inputs

An online questionnaire was sent to 300 individuals within the LMR region with 55 responses, and 199 contributed to in-person meetings and discussion during the 2015 kick-off phase of the LMR SERVIR hub. Of the online survey respondents, 81% were familiar with GIS and internet mapping tools, while 60% reported familiarity with remote sensing. Similarly, 72% of respondents indicated that their organization acquired GIS data at least annually, while 54.5% indicated they acquired remotely sensed data at least annually. The majority (52 out of 53) of respondents indicated that the top priority application of geospatial data was for management purposes, including disaster reduction, early warning systems, and flood monitoring and management. Responses to gaps in data sharing and capacity building are depicted in **Figure 2** (see also Tables 1–5 in **Appendix B**). These results identify an understanding of and a need for geospatial data and training within the LMR region.

This online survey was not completed by individuals that attended in person consultation events, nor was it delivered in the HKH region. However, participants that attended the regional geospatial needs assessment workshops, focus groups, or the land cover monitoring workshop identified similar challenges.

Stakeholder involvement fostered participation which proved to be critical toward building a spirit of collaboration and trust. Based on the overall data gathered, the following salient challenges and gaps were identified which led to the co-development of a viable strategy to address key issues in land cover mapping.

4.1.2. Current Challenges and Gaps in Land Cover Mapping and Monitoring Practices

Key challenges found from the assessment are outlined as follows. In general, participants pointed out insufficient commitment to the provision of resources, building of capacity, and setting up of unrealistically tight timelines for work for geospatial applications. Within these identified gaps, lack of coordination was noted to lead to disconnected workflows which reinforced many of the other stated challenges. The interrelated shortages were found to be lack of data access, computing power, availability of technical capacity and proper procedures, as well as modeling and use of consistent land cover maps.

With respect to data challenges, lack of consistent (spatially and temporally), pre-processed satellite data was expressed as a common obstacle which oftentimes led to production of outdated maps. This also meant stakeholders were not able to meet user requirements or provided inconsistent products with respect to user needs. Another byproduct of these practices is the production of multiple maps for different agency needs which were not based on the same underlying data. Participants also cited political/bureaucratic issues which are often entangled with scientific procedures and decisions–often these limitations are undocumented or unacknowledged (e.g., reference data collection and compilation, mapping procedures, etc.). One example that was referred to frequently in the discussions was the resistance to using the updated Mekong River Commission (MRC) land cover maps, because they had not been produced under explicitly approved procedures. Further analysis of the stakeholder inputs also revealed a lack of proper understanding of error structures in the map making process and final map products. This finding can be attributed to a lack of proper communication of geospatial science and related technologies in the region.

Results from the needs assessment in HKH show that multiple agencies are involved in land cover mapping, and have inconsistencies related to the classification schema, underlying satellite datasets used, and interpretation methods. The inconsistencies across spatial and temporal dimensions create problems when comparing land cover products and change analyses. The topographic variability in the HKH region also poses a challenge for generating land cover interpretation and collection of field data for training and validation. Restrictions on data sharing vary between countries. While multiple agencies within the same nation are independently producing their own land cover maps, generating redundant data sets with conflicting statistics. The typologies used are also varied across different countries, which makes it difficult to harmonize the datasets across the region.

A crippling limitation was the provisioning of and access to resources. Resource limitations included financial as well as human resources. Well funded organizations often could not find the right personnel, others reported that adequately

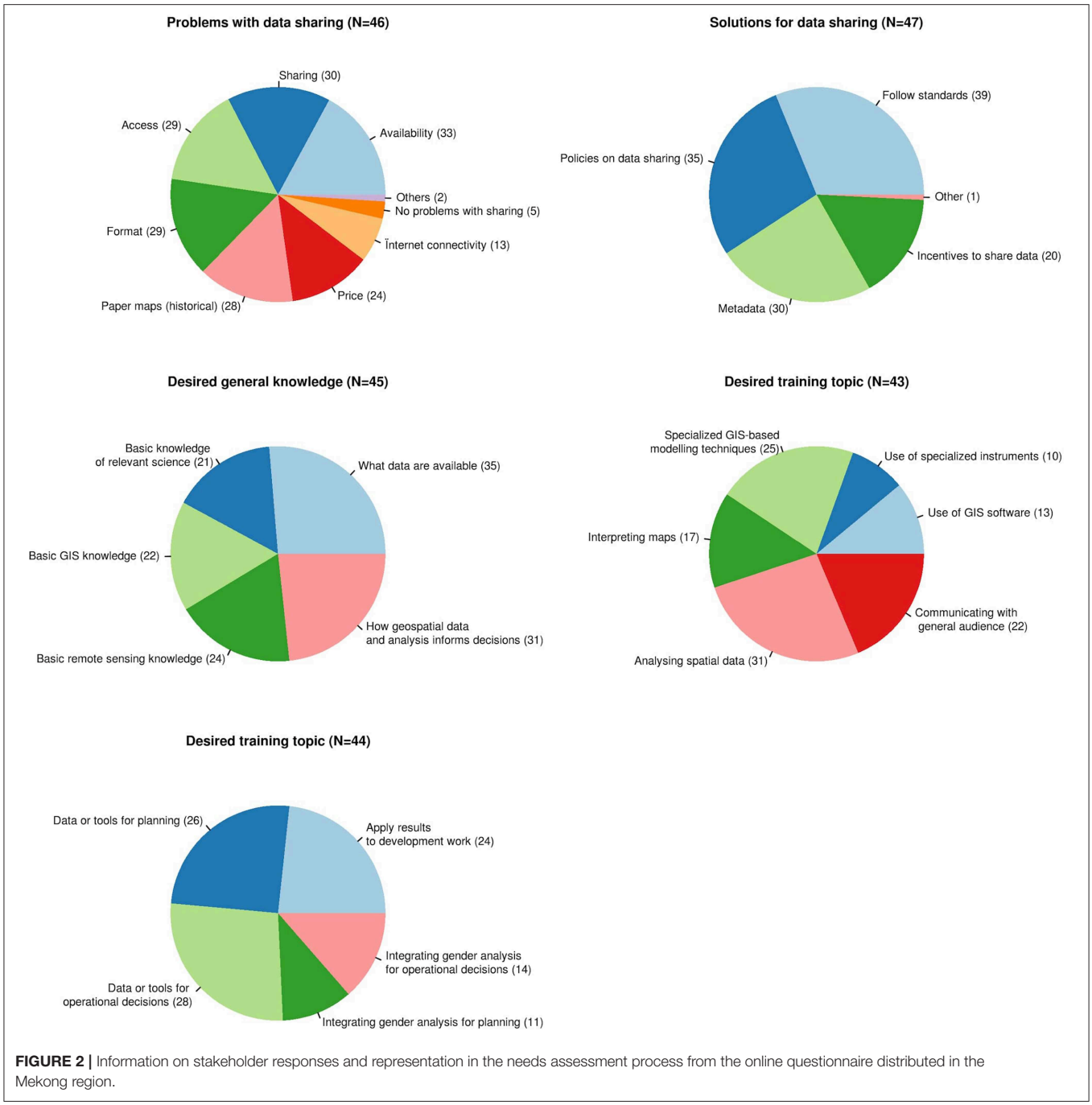


FIGURE 2 | Information on stakeholder responses and representation in the needs assessment process from the online questionnaire distributed in the Mekong region.

staffed organizations had a hard time retaining qualified team members. Land cover mapping often requires diverse teams that are technically skilled and understand departmental needs. We found that within the represented organizations, these individuals were often over-tasked and under-resourced. Also local land cover mapping programs are often partially resourced for a specific project or program related to a country's donor activities with no core funding for sustainable support of land cover monitoring programs.

Local land cover mapping professionals had a hard time remaining abreast of new data, emerging technologies, and the latest science. Many organizations had standard operating procedures developed by donor countries which relied on accessing commercial data, which have not been updated to take advantage of new data streams and information products. This has often left departments using dated data products that introduce time lags. Geospatial professionals often lacked access to emerging new technologies due to budgetary

constraints, licensing issues, or barriers in language. Professional geospatial growth was dependent on participation in donor-related activities focused on free and open source solutions. The same was also true for accessing data.

4.2. Enabling Connected Workflows: Co-development of a Land Cover Monitoring System

A key outcome of the stakeholder engagement process that aims to address the challenges stated above was the decision of stakeholder to collaborate on the co-development of a regional land cover monitoring system (RLCMS). This system can serve as a vehicle for a shared process of development and capacity building. Ultimately it serves as an example of a working land cover monitoring system based on best practices, and leverages state of the art technologies.

The first component of the system is the representation of a land cover scheme, or typology. The design of the typology is guided by the following list of 10 principles, which were established from the literature review (e.g., Running et al., 1994; Triepke et al., 2008; Lillesand et al., 2014) and group discussions:

1 Stakeholder engagement: Stakeholders identify their needs as a basis for the typology
2 Objective driven: A typology facilitates stakeholder objectives
3 Simple: A typology is no more complex than necessary to address stakeholder objectives
4 Exhaustive: Each location on the map is represented in the typology
5 Integrity among classes: Classes are mutually exclusive and have explicit class boundaries
6 Consistent: The typology is consistent from one area of the map to another, and from one generation of land cover mapping to another to support trends monitoring
7 Clear definitions: Map classes reflect measurable, diagnostic, biophysical features
8 Differentiates land use from land cover: The typology separates land use and land cover themes
9 Mappable: The typology is technologically and operationally feasible, for given budget and time constraints
10 Considers existing land cover schemes: Uses components of existing typologies whenever possible to maximize compatibility, shareability, and the use of available mapping technology, data, and applications.

Stakeholders further proposed that any land cover monitoring system appropriate for regional or national use meets (at minimum) the following design criteria:

1 Flexibility

- The system accommodates land cover typologies that vary by country by implementing a process to create and assemble biophysical land cover layers, referred to as primitives. These are continuous layers of biophysical attributes, such as fractional forest cover and forest height.
- Land cover primitives can be swapped for the most state-of-the-science product available at any time.

2 Consistency

- Every country uses the same primitives and the same assembly system, but can customize an assembly logic to create maps that accommodate regionally varying land cover semantics and objectives.

3 Based on remotely sensed data

- The system is data-driven with access to publicly available big Earth observing geodatasets, provided by cloud computing platforms.

4 Explicit quantification of uncertainty

- Monte-Carlo methods will be used in conjunction with independent validation data sets to incorporate uncertainty at the primitive level to provide pixel-based estimates of land cover uncertainty of the final map product.

5 Capacity building

- The collaborative nature of the system facilitates information and technology exchange.
- Free and broadly accessible, open tools are used wherever possible.

On the basis of these principles, the basic structure of a RLCMS typology and architecture was co-developed and a collaborative development process was completed through a series of four workshops and production segments from 2016 to 2017 in the Mekong. The HKH region began this process in 2018. During the first regional production workshop partners defined their end user objectives, a uniform land cover typology, data requirements, and laid out the collaborative design and production process (Steps 1 and 2 in **Figure 3**). The proposed land cover monitoring system production process includes six elements: typology development, land cover algorithms, reference data, accuracy assessment, assemblage of land cover algorithms, and production/metadata. At the conclusion of the first workshop, Mekong participants formed six co-development teams, each focusing on one of six key processes envisioned as central to the example system. However, the approach in the HKH was organized by as separate production events with each mapping agency.

In the Mekong, the development of each of the six land cover monitoring system components would be completed and refined during three follow up workshops. During the second regional workshop, prototype algorithms for generating land cover primitives were drafted along with documentation of the tools, procedures and reference data needed to successfully execute the algorithms (steps 3 and 4 in **Figure 3**). Participants worked on three objectives during the third workshop, including (1) the development of a flexible, customizable assembly protocol to combine thematic primitives maps into a final land cover classification with the users' specific target land cover classes, (2) specifying accuracy assessment procedures and (3) documenting end user needs to ensure the system responds (steps 5 and 6 in

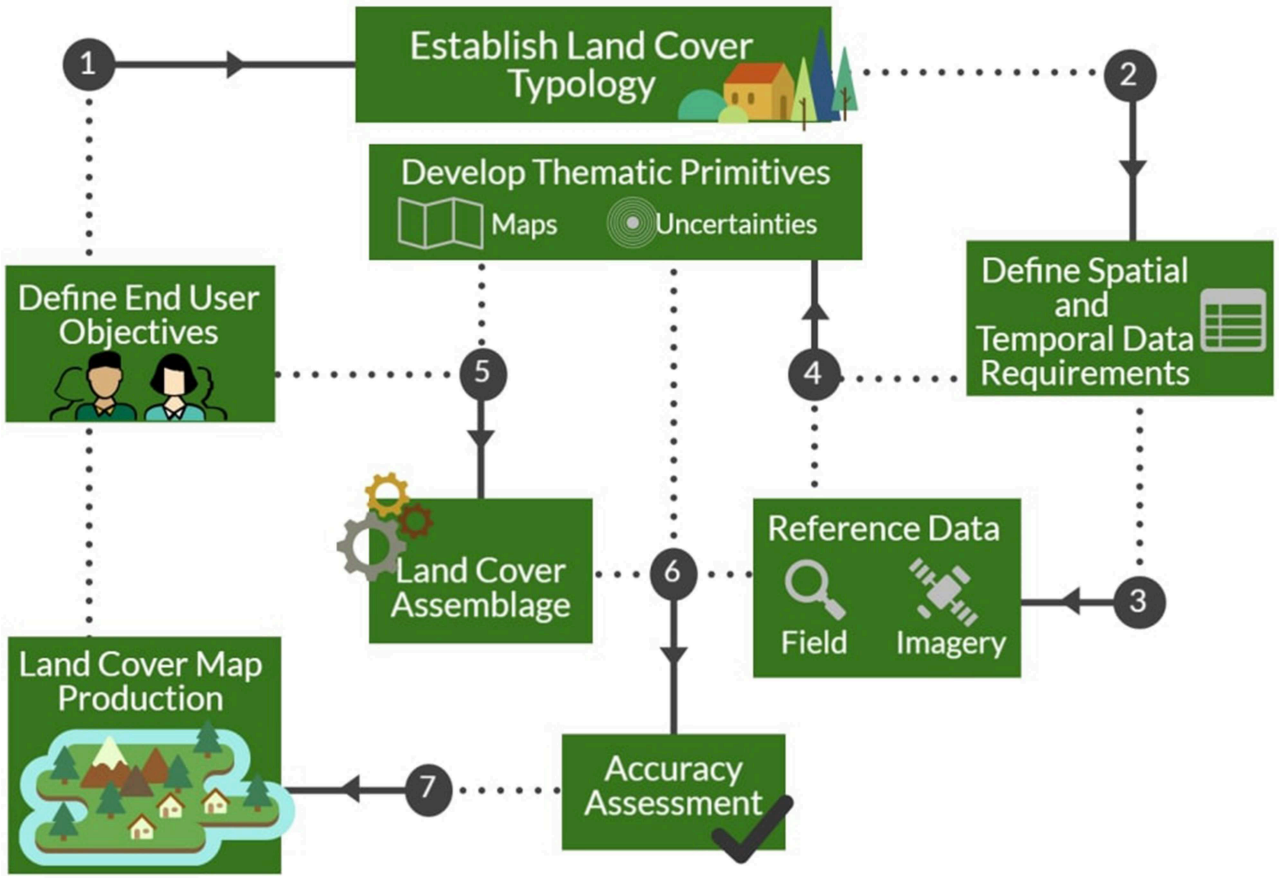


FIGURE 3 | Operational framework for a representative Regional Land Cover Monitoring System for the LMR. The arrows represent the direction of the process, while the dotted lines represent the iterative phases of the development of the system.

Figure 3). The final workshop culminated in the presentation of preliminary land cover maps and a road map to support end users customize the products to meet their specific needs. The HKH production phase also followed this three step workshop approach. However, their focus was on testing, refining, and, as necessary, customizing design principles and algorithms developed by the Mekong team to meet the needs of a new set of stakeholders. These production phases were co-implemented with the Mekong SERVIR hub to ensure that lessons learned were transferred between the two hubs to continue the joint improvement of the system.

4.3. Examples of the Operationalized System and Architecture

The system is implemented in Google Earth Engine, a freely available cloud computing infrastructure (Gorelick et al., 2017). For the first iteration, most primitives were built using a random forest classifier built primarily with data from the Landsat archive (Saah et al., in press). The Landsat program made the full archive of moderate resolution data with a long temporal range freely available (Wulder et al., 2012; Roy et al., 2014). However, because the system is modular other classification approaches are available and being tested, such as neural networks. Additionally, depending on the temporal requirements of the primitive layers, some are created from a data fusion between Landsat, MODIS (Justice et al., 1998), Sentinel-1 and Sentinel-2 (Berger et al., 2012), or climate hazards infrared precipitation (CHIRPS) (Funk et al., 2015) data sets, to name a few. For example, maps of rubber and palm oil plantations were developed using Sentinel-1 and Sentinel-2 data (Poortinga et al., 2019). The assembly logic includes a user-specified decision tree. The final land cover map product is created through a series of Monte Carlo iterations of the assembly logic and primitive error rates. Error rates are generated for both the primitive and land cover products using independent, probabilistic reference data sets. For the full details of the system architecture and co-designed products, please refer to Poortinga et al. (2019), Saah et al. (2019), Saah et al. (in press), and Potapov et al. (2019); other manuscripts are forthcoming. The Mekong data is available at https://rlcms-servir.adpc.net/en/. The launch date of the HKH data products is scheduled for 2020.

5. DISCUSSION

Land cover monitoring is a key activity which helps ensure health and sustainability of ecosystems. Earth observations and geospatial technology are playing an increasingly important role

in this regard (Poortinga et al., 2017; Simons et al., 2017). In this paper we have outlined a SERVIR-facilitated stakeholder engaged approach to identify needs and key challenges faced in the countries of the Lower Mekong and Hindu Kush Himalayan regions. There is a shared view among stakeholders in the region working on land cover mapping that there are significant gaps in the capacity, technical tools and systems, and data needed to undertake effective land cover mapping. This includes lack of commitment of resources and support by relevant authorities and disconnected workflows due to limited access to data, computing power, and availability of technical capacity. Similar challenges were echoed during the stakeholder workshops in the HKH region. Many projects have been conducted in HKH countries to develop land cover maps for specific target years. In most cases these maps were generated using different data sources and typologies that makes it difficult to understand the changes. Also the topographic and terrain heterogeneity in the HKH region put additional challenges for accurate land cover mapping and monitoring.

We found this approach useful as it fostered collaboration and trust, leading to the co-development of a RLCMS, which aims to address all of the issues raised, and provide greatly improved access to, and quality of land cover data products. However, data sharing challenges and perceptions about data uncertainty remain formidable obstacles to analysis and acceptance of spatial data products. The former is due to existing data sharing policies and the latter reflects the widely held view that error and/or uncertainty is a flaw rather than an inherent feature of both scientific measurement and modeling. Uncertainty tended to be viewed as a negative to be minimized by people on both sides of the decision making process, to the point that it may be under-represented or simply left out of how results are presented to policy makers. However, presenting uncertainty hand-in-hand with the costs of making poor decisions can be vital in using scientific data to inform decision making (Reckhow, 1994). Fischhoff and Davis (2014) presented a useful outline for how to productively think about and present uncertainty in a decision-making context: careful identification of the core question to be answered, assessment of the degree and nature of uncertainty around the methods used to answer the question, and conveying that uncertainty in a clear, precise way that addresses any gaps that might exist in the mental models of the world used by decision makers compared to scientists or analysts. Ultimately, this boils down to understanding the important role of uncertainty as a natural part of research and decision-making that depends on everything from the precise question asked, whether one method was chosen over another, the real-world conditions in which the work takes place, and the appropriateness of a given data product to answer the question being asked. This process can be facilitated through the development and application of spatial data standards, with a standardized representation and level of acceptable error that can be applied across many data products. This could be expressed in terms of standards for completeness of coverage, confidence intervals and root mean square error in position, or through other criteria.

Respondents in both the online questionnaire and in-person discussions mentioned problems with data accessibility and sharing. Data producers are concerned about data security and ownership leading data providers to be possessive or defensive of the use of their data products, keeping them within their departments and minimizing availability to other work groups or agencies. While surprisingly common at many levels of government around the world, this behavior is antithetical to effective resource management, and greatly impedes the ability of groups to work together effectively for large scale projects, such as MRV for compliance with international programs like REDD+. These views and behaviors toward data sharing ultimately lead to inconsistent data across regions inhibiting progress toward environmental monitoring and policy. Moving toward an open, non-proprietary model of cooperation and data sharing is essential in establishing resilient, high functioning institutional arrangements necessary for sustainable development and addressing climate change challenges (GFOI, 2016). Many of these gaps can be addressed through the application of well established prevailing best practices such as, (1) data standardization, (2) metadata collections, (3) common data formats, and (4) open and transparent methods (i.e., free and open source code). Promoting these best practices facilitates interoperability, data exchange among different institutions, and reduction of barriers to access data.

Fortunately, in many cases the cost and technical expertise needed to apply these technologies and practices to address gaps is generally decreasing. The availability of high-quality remotely sensed and GIS data has steadily increased with processing packages and large public data repositories such as that hosted by Google Earth Engine (GEE) now freely available to academic, NGO, and public sectors (Gorelick et al., 2017; Markert et al., 2018; Poortinga et al., 2018). Many new tools, sites, and organizations have emerged to increase the power and ease of web-mapping and web-based map and GIS services, while reducing the cost and other barriers to entry. Similarly, high quality technical training for use with tools like GEE, QGIS, Python, R, SEPAL, OpenFORIS, and other free or low cost spatial analytical tools is now increasingly available. In addition to these tools that can be used for self-directed learning and improvement, many organizations (including the authors of this paper) are working to produce workshops that develop new tools and help build institutional capacity.

6. CONCLUSION

In summary, our stakeholder based approach can potentially be applied in other contexts to foster collaborative work toward addressing similar issues and building partnerships and networks including institutions, agencies and experts across countries with shared goals. In our case, this development process was fostered by actively engaging regional end user groups in discussion, consultation and capacity building to ensure that the RLCMS was effectively embedded in policy formulation and implementation at all levels regionally, while also acting as a means of extending regional cooperation in land management.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We thank the participating agencies, civil society organizations, and university partners for sharing the key challenges they face in creating and using geospatial data to inform decision making and for scoping out design criteria to guide the collaborative development of a shared regional land cover monitoring system. Support for this work was provided through the joint US Agency for International Development (USAID) and National Aeronautics and Space Administration (NASA) initiative SERVIR and the USAID Regional Development Mission for Asia.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2019.00150/full#supplementary-material

REFERENCES

Berger, M., Moreno, J., Johannessen, J. A., Levelt, P. F., and Hanssen, R. F. (2012). Esa's sentinel missions in support of earth system science. *Remote Sens. Environ.* 120, 84–90. doi: 10.1016/j.rse.2011.07.023

Biernacki, P., and Waldorf, D. (1981). Snowball sampling: problems and techniques of chain referral sampling. *Sociol. Methods Res.* 10, 141–163. doi: 10.1177/004912418101000205

Bounoua, L., DeFries, R., Collatz, G. J., Sellers, P., and Khan, H. (2002). Effects of land cover conversion on surface climate. *Clim. Change* 52, 29–64. doi: 10.1023/A:1013051420309

Buchanan, G. M., Butchart, S. H., Dutson, G., Pilgrim, J. D., Steininger, M. K., Bishop, K. D., et al. (2008). Using remote sensing to inform conservation status assessment: estimates of recent deforestation rates on new britain and the impacts upon endemic birds. *Biol. Conservat.* 141, 56–66. doi: 10.1016/j.biocon.2007.08.023

Chettri, N., Shakya, B., Thapa, R., and Sharma, E. (2008). Status of a protected area system in the hindu kush-himalayas: an analysis of pa coverage. *Int. J. Biodivers. Sci. Manage.* 4, 164–178. doi: 10.3843/Biodiv.4.3:4

Costenbader, J., Broadhead, J., Yasmi, Y., and Durst, P. B. (2015). *Drivers Affecting Forest Change in the Greater Mekong Subregion (GMS): An Overview*. Rome: FAO.

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., and Hansen, M. C. (2018). Classifying drivers of global forest loss. *Science* 361, 1108–1111. doi: 10.1126/science.aau3445

Fischhoff, B., and Davis, A. L. (2014). Communicating scientific uncertainty. *Proc. Natl. Acad. Sci. U.S.A.* 111(Suppl. 4):13664–13671. doi: 10.1073/pnas.1317504111

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., et al. (2005). Global consequences of land use. *Science* 309, 570–574. doi: 10.1126/science.1111772

Fritz, S., See, L., and Rembold, F. (2010). Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. *Int. J. Remote Sens.* 31, 2237–2256. doi: 10.1080/01431160902946598

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* 2:150066. doi: 10.1038/sdata.2015.66

Ge, J., Qi, J., Lofgren, B., Moore, N., Torbick, N., and Olson, J. (2007). Impacts of land use/cover classification accuracy on regional climate simulations. *J. Geophys. Res.* 112, D05107, 1–12. doi: 10.1029/2006JD007404

GFOI (2016). *Integration of Remote-Sensing and Ground-Based Observations for Estimation of Emissions and Removals of Greenhouse Gases in Forests: Methods and Guidance from the Global Forest Observations Initiative, Edition* 2.0. Rome: Food and Agriculture Organization, 1–224.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., et al. (2013). Finer resolution observation and monitoring of global land cover: first mapping results with landsat tm and etm+ data. *Int. J. Remote Sens.* 34, 2607–2654. doi: 10.1080/01431161.2012.748992

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. doi: 10.1016/j.rse.2017.06.031

Hall, D., Hirsch, P., and Li, T. M. (2011). *Introduction to Powers of Exclusion: Land Dilemmas in Southeast Asia*. Singapore: National University of Singapore Press; Honolulu, HI: University of Hawaii Press.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science* 342, 850–853. doi: 10.1126/science.1244693

Hansen, M. C., and Reed, B. (2000). A comparison of the igbp discover and university of maryland 1 km global land cover products. *Int. J. Remote Sens.* 21, 1365–1373. doi: 10.1080/014311600210218

Herold, M., Mayaux, P., Woodcock, C., Baccini, A., and Schmullius, C. (2008). Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. *Remote Sens. Environ.* 112, 2538–2556. doi: 10.1016/j.rse.2007.11.013

Hibbard, K., Janetos, A., van Vuuren, D. P., Pongratz, J., Rose, S. K., Betts, R., et al. (2010). Research priorities in land use and land-cover change for the earth system and integrated assessment modelling. *Int. J. Climatol.* 30, 2118–2128. doi: 10.1002/joc.2150

Imaoka, K., Kachi, M., Fujii, H., Murakami, H., Hori, M., Ono, A., et al. (2010). Global change observation mission (gcom) for monitoring carbon, water cycles, and climate change. *Proc. IEEE* 98, 717–734. doi: 10.1109/JPROC.2009.2036869

Ives, J. D., and Messerli, B. (2003). *The Himalayan Dilemma: Reconciling Development and Conservation*. London; New York, NY: Routledge.

Jha, M. K., and Chowdary, V. (2007). Challenges of using remote sensing and gis in developing nations. *Hydrogeol. J.* 15, 197–200. doi: 10.1007/s10040-006-0117-1

Jung, M., Henkel, K., Herold, M., and Churkina, G. (2006). Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sens. Environ.* 101, 534–553. doi: 10.1016/j.rse.2006.01.020

Justice, C. O., Vermote, E., Townshend, J. R., Defries, R., Roy, D. P., Hall, D. K., et al. (1998). The moderate resolution imaging spectroradiometer (modis): land remote sensing for global change research. *IEEE Trans. Geosci. Remote Sens.* 36, 1228–1249. doi: 10.1109/36.701075

Kityuttachai, K., Heng, S., and Sou, V. (2016). "Land cover map of the lower mekong basin," in *Technical Paper* (Vientiane).

Leinenkugel, P., Kuenzer, C., Oppelt, N., and Dech, S. (2013). Characterisation of land surface phenology and land cover based on moderate resolution satellite data in cloud prone areas—a novel product for the mekong basin. *Remote Sens. Environ.* 136, 180–198. doi: 10.1016/j.rse.2013.05.004

Leinenkugel, P., Oppelt, N., and Kuenzer, C. (2014). A new land cover map for the mekong: Southeast Asia's largest transboundary river basin. *Pacif. Geograph.* 41, 10–14.

Lillesand, T., Kiefer, R. W., and Chipman, J. (2014). *Remote Sensing and Image Interpretation*. Hoboken, NJ: John Wiley & Sons.

Markert, K., Schmidt, C., Griffin, R., Flores, A., Poortinga, A., Saah, D., et al. (2018). Historical and operational monitoring of surface sediments in the lower mekong basin using landsat and google earth engine cloud computing. *Remote Sens.* 10:909. doi: 10.3390/rs10060909

McCallum, I., Obersteiner, M., Nilsson, S., and Shvidenko, A. (2006). A spatial comparison of four satellite derived 1 km global land cover datasets. *Int. J. Appl. Earth Observat. Geoinformat.* 8, 246–255. doi: 10.1016/j.jag.2005.12.002

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403:853. doi: 10.1038/35002501

Pandey, S., Bhandari, H. S., and Hardy, B. (2007). "Economic costs of drought and rice farmers' coping mechanisms: a cross-country comparative analysis," in *International Rice Research Institute* (Los Baños).

Pandit, M., Sodhi, N. S., Koh, L. P., Bhaskar, A., and Brook, B. W. (2007). Unreported yet massive deforestation driving loss of endemic biodiversity in Indian Himalaya. *Biodiver. Conservat.* 16, 153–163. doi: 10.1007/s10531-006-9038-5

Peel, M., Finlayson, B., and McMahon, T. (2007). Updated world map of the koppen-geiger climate classification. *Hydrol. Earth Syst. Sci.* 11, 1633–1644.

Perera, K., Herath, S., Apan, A., and Samarakoon, L. (2010). Mapping mekong land cover at 250m resolution without *in situ* observations. *Asian J. Geoinformat.* 10, 31–41.

Poortinga, A., Bastiaanssen, W., Simons, G., Saah, D., Senay, G., Fenn, M., et al. (2017). A self-calibrating runoff and streamflow remote sensing model for ungauged basins using open-access earth observation data. *Remote Sens.* 9:86. doi: 10.3390/rs9010086

Poortinga, A., Clinton, N., Saah, D., Cutter, P., Chishtie, F., Markert, K., et al. (2018). An operational before-after-control-impact (BACI) designed platform for vegetation monitoring at planetary scale. *Remote Sens.* 10:760. doi: 10.3390/rs10050760

Poortinga, A., Tenneson, K., Shapiro, A., Nquyen, Q., San Aung, K., Chishtie, F., et al. (2019). Mapping plantations in myanmar by fusing landsat-8, sentinel-2 and sentinel-1 data along with systematic error quantification. *Remote Sens.* 11:831. doi: 10.3390/rs11070831

Potapov, P., Tyukavina, A., Turubanova, S., Talero, Y., Hernandez-Serna, A., Hansen, M., et al. (2019). Annual continuous fields of woody vegetation structure in the lower mekong region from 2000-2017 landsat time-series. *Remote Sens. Environ.* 232:111278. doi: 10.1016/j.rse.2019.111278

Reckhow, K. H. (1994). Importance of scientific uncertainty in decision making. *Environ. Manage.* 18, 161–166. doi: 10.1007/BF02393758

Reed, M. S. (2008). Stakeholder participation for environmental management: a literature review. *Biol. Conservat.* 141, 2417–2431. doi: 10.1016/j.biocon.2008.07.014

Roy, D. P., Wulder, M., Loveland, T. R., Woodcock, C., Allen, R., Anderson, M., et al. (2014). Landsat-8: science and product vision for terrestrial global change research. *Remote Sens. Environ.* 145, 154–172. doi: 10.1016/j.rse.2014.02.001

Running, S. W. (2008). Ecosystem disturbance, carbon, and climate. *Science* 321, 652–653. doi: 10.1126/science.1159607

Running, S. W., Loveland, T. R., and Pierce, L. L. (1994). A vegetation classification logic-based on remote-sensing for use in global biogeochemical models. *Ambio* 23, 77–81.

Saah, D., Johnson, G., Ashmall, B., Tondapu, G., Tenneson, K., Patterson, M., et al. (2019). Collect earth: an online tool for systematic reference data collection in land cover and use applications. *Environ. Model. Softw* 118, 166–171. doi: 10.1016/j.envsoft.2019.05.004

Saah, D., Tenneson, K., Poortinga, A., Nquyen, Q., and Chishtie, F. (in press). Primitives as building blocks for constructing land cover maps. *Int. J. Appl. Earth Observat. Geoinformat.*

Simons, G., Poortinga, A., Bastiaanssen, W. G., Saah, D., Troy, D., Hunink, J., et al. (2017). *On Spatially Distributed Hydrological Ecosystem Services: Bridging the Quantitative Information Gap Using Remote Sensing and Hydrological Models*. FutureWater.

Singh, S. P., Bassignana-Khadka, I., Singh Karky, B., and Sharma, E. (2011). *Climate Change in the Hindu Kush-Himalayas: The State of Current Knowledge*. Technical report, International Centre for Integrated Mountain Development (ICIMOD).

Stibig, H., Achard, F., Carboni, S., Raši, R., and Miettinen, J. (2013). Change in tropical forest cover of Southeast Asia from 1990 to 2010. *Biogeosci. Discuss* 10, 12625–12653. doi: 10.5194/bgd-10-12625-2013

Tolentino, P. L., Poortinga, A., Kanamaru, H., Keesstra, S., Maroulis, J., David, C. P. C., et al. (2016). Projected impact of climate change on hydrological regimes in the philippines. *PLoS ONE* 11:e0163941. doi: 10.1371/journal.pone.0163941

Triepke, F. J., Brewer, C. K., Leavell, D. M., and Novak, S. J. (2008). Mapping forest alliances and associations using fuzzy systems and nearest neighbor classifiers. *Remote Sens. Environ.* 112, 1037–1050. doi: 10.1016/j.rse.2007.07.014

Tucker, C. J., Townshend, J. R., and Goff, T. E. (1985). African land-cover classification using satellite data. *Science* 227, 369–375. doi: 10.1126/science.227.4685.369

Watch, G. F. (2015). *Global Forest Watch*. Washington, DC: World Resources Institute. Available online at: http://www.globalforestwatch.org (accessed June 2015).

Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., and Woodcock, C. E. (2012). Opening the archive: how free data has enabled the science and monitoring promise of landsat. *Remote Sens. Environ.* 122, 2–10. doi: 10.1016/j.rse.2012.01.010

Conflict of Interest: KT, PC, AP, and RB are employed by spatial informatics group, LLC. NC is employed by Google. MP, IH, and PM are employed by RedCastle Resources.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 03 December 2019
doi: 10.3389/fenvs.2019.00186

Linking Earth Observations for Assessing the Food Security Situation in Vietnam: A Landscape Approach

Ate Poortinga[1], Quyen Nguyen[2], Karis Tenneson[1], Austin Troy[1,3], David Saah[1,4], Biplov Bhandari[2], Walter L. Ellenburg[5], Aekkapol Aekakkararungroj[2], Lan Ha[6], Hai Pham[7], Giang Nguyen[8,9] and Farrukh Chishtie[2]*

[1] Spatial Informatics Group, LLC, Pleasanton, CA, United States, [2] Asian Disaster Preparedness Center, Bangkok, Thailand, [3] Department of Urban and Regional Planning, University of Colorado Denver, Denver, CO, United States, [4] Geospatial Analysis Lab, University of San Francisco, San Francisco, CA, United States, [5] SERVIR Science Coordination Office, NASA Marshall Space Flight Center, Huntsville, AL, United States, [6] Institute of Water Resources Planning, Ministry of Agriculture and Rural Development, Hanoi, Vietnam, [7] Forest Inventory and Planning, Ministry of Agriculture and Rural Development, Hanoi, Vietnam, [8] Division of Nature, Forest and Landscape, KU Leuven, Leuven, Belgium, [9] Space Technology Institute, Vietnam Academy of Science and Technology, Hanoi, Vietnam

OPEN ACCESS

Edited by:
Niall Patrick Hanan, New Mexico State University, United States

Reviewed by:
Jefferson Metz Fox, East-West Center, United States
Stephen Leisz, Colorado State University, United States

***Correspondence:**
Ate Poortinga
apoortinga@sig-gis.com

Specialty section:
This article was submitted to Land Use Dynamics, a section of the journal Frontiers in Environmental Science

Received: *27 June 2019*
Accepted: *07 November 2019*
Published: *03 December 2019*

Citation:
Poortinga A, Nguyen Q, Tenneson K, Troy A, Saah D, Bhandari B, Ellenburg WL, Aekakkararungroj A, Ha L, Pham H, Nguyen G and Chishtie F (2019) Linking Earth Observations for Assessing the Food Security Situation in Vietnam: A Landscape Approach. Front. Environ. Sci. 7:186. doi: 10.3389/fenvs.2019.00186

Land cover change and its impact on food security is a topic that has major implications for development in population-dense Southeast Asia. The main drivers of forest loss include the expansion of agriculture and plantation estates, growth of urban centers, extraction of natural resources, and water infrastructure development. The design and implementation of appropriate land use policies requires accurate and timely information on land cover dynamics to account for potential political, economical, and agricultural consequences. Therefore, SERVIR-Mekong led the collaborative development of a Regional Land Cover Monitoring System (RLCMS) with key regional stakeholders across the greater Mekong region. Through this effort, a modular system was used to create yearly land cover maps for the period 1988–2017. In this study, we compared this 30-year land cover time-series with Vietnam national forest resources and agricultural productivity statistics. We used remote sensing-derived land cover products to quantify landscape changes and linked those with food availability, one of food security dimension, from a landscape approach perspective. We found that agricultural production has soared while the coverage of agricultural areas has remained relatively stable. Land cover change dynamics coincide with important legislation regarding environmental management and sustainable development strategies in Vietnam. Our findings indicate that Vietnam has made major steps toward improving its' food security. We demonstrate that RLCMS is a valuable tool for evaluating the relationship between policies and their impacts on food security, ecosystem services and natural capital.

Keywords: food security, land cover, earth observation, ecosystems, SERVIR, sustainability, land cover change, Mekong

1. INTRODUCTION

Land governance is at the center of development challenges in Southeast Asia: between 1973 and 2009 the lower Mekong sub-region lost almost a third of its forest cover (WWF, 2013). Such a drastic loss in an area of rapidly growing populations can result in dangerously reduced food security and ecosystem services. The main drivers of land cover change include expansion of agriculture and plantation estates, development of transportation, energy and water infrastructure, urbanization, extraction of natural resources, logging and forest fires (Curtis et al., 2018). Shifts in population dynamics coupled with economic growth are cited as underlying causes of land cover change, often intensified by weak governance (Bui et al., 2014; Stibig et al., 2014). Socioeconomic development, land cover and use and food security are all interconnected at different scales and have far-reaching impacts for not only local people, but the region at large. The development of appropriate land policies is therefore crucial for healthy economic and social development at the national and sub-national level.

Land cover maps are essential for assessing the five biophysical dimensions of food security (Bartholomé and Belward, 2005; Liu et al., 2008; See et al., 2015), which include estimations of cultivated area, crop type, cropping suitability, irrigation and water use, and crop yield. Traditional methods of acquiring information on crop yields and area were through in-person or field surveys using questionnaires, interviews, observations and sampling. These data were then combined and reported in statistical yearbooks. However, such methods vary depending upon skill level of surveyor, community boundaries, local traditions and definitions, and knowledge of those surveyed. Such an approach also requires significant time, personal equipment, and costs. To minimize the need for these resources, satellite data are often used to generate land cover maps (Friedl et al., 2010; Congalton et al., 2014). These maps, however, often lack ground-level validation, or the desired land cover typology is not detectable at the ideal spatial or temporal resolution. Moreover, the generation of satellite- based land cover maps requires technical expertise in the fields of remote sensing and information technology; such expertise is limited to a relatively small subset of people worldwide. New cloud-based geo-computational platforms enable a broader array of technical specialists to process large amounts of data without investing in expensive storage, extensive training and/or processing infrastructure (Gorelick et al., 2017; Markert et al., 2018; Poortinga et al., 2018).

To assess the food security within a country or smaller jurisdictional boundary, one must consider the landscapes and their impacts on food production. For example, forested highlands, which are often located upstream of agricultural areas, provide a wide variety of ecosystem services including raw materials, erosion control, water and nutrient regulation, and micro-climate regulation. These benefits are often conceptualized and quantified in an ecosystem service framework (Costanza et al., 1997; Daily et al., 1997; Stürck et al., 2014). Good environmental legislation must take these ecosystem services, and the dynamic trade-offs and synergies between them, into account. But to understand such dynamics, one must quantify the different dimensions of food security and their relations to the landscape. Current technologies provide capabilities for creating high resolution multi-temporal data series on land cover. These products have well known sources of error. For example, representing continuous land cover as discrete classes often results in misclassification at land type boundaries (Olofsson et al., 2014). Other errors in the data may include image aberrations from atmospheric conditions and flaws in the analysis methods (Foody, 2010). To accurately assess food security within a landscape, statistically robust, easily replicable and transparent approaches are required to translate pixel counts into unbiased estimates of change Olofsson et al. (2014).

This study used a landscape approach to characterize the trajectory and quantify change of two dimensions of food security in Vietnam: area under cultivation and crop types. The term "landscape approach" has been discussed since the 1980s, when the Global Landscape Forum defined it as "balancing competing land use and land cover demands in a way that is best for human well-being and the environment." This means that socio-economic development solutions must embed food and livelihoods, finance, rights, restoration and progress toward climate and development goals. Landscape approaches thus provide a framework for integrating policy and practice for multiple land uses within a given area, all while ensuring equitable and sustainable land use and improved measures to mitigate and adapt to environmental change (Reed et al., 2015).

We first assessed land cover change using Earth observation data from the Regional Land Cover Monitoring System (RLCMS). The RLCMS was developed by SERVIR-Mekong, a joint United States Agency for International Development (USAID) and National Aeronautics and Space Administration (NASA) collaborative project that aims to support development and sustainable landscape projects in the Mekong region. For this study, the RLCMS was used to produce annual land cover maps for the period from 1988 to 2017. The map products were used to characterize 30 years of changes in area under cultivation and crop type.

This Earth observation data was then coupled with crop productivity estimates from census data and estimates from statistical yearbooks. The crop type categories included in the assessment were aquaculture, rice, orchard and plantation forests, and other croplands. We applied best practices for area estimations to estimate coverage of crop commodities and other land cover classes over the same 30 years period. The assessment was applied at the national and, on a subset of the land, at the provincial level. Analyses at these two levels allowed us to demonstrate variation in observed trends across the country. We then qualitatively related these land cover and crop production trends to the adoption of key sustainable development policies in Vietnam.

We have been able to document what land covers are expanding, and what these expanding regions are replacing. This study demonstrated the capabilities of a cloud-based system like the RLCMS to accurately detects land cover dynamics, allowing developing counties like Vietnam to better address food security and nutrition needs of its' people.

2. MATERIALS AND METHODS

Vietnam's natural capital is provided by a total area of 33 million hectares of land and more than 3,000 km of coastline. The country is covered by a snaking and dense river and stream network, with an average total volume of surface water estimated at 830 billion cubic meters. It has rich forest resources, a wide range of minerals, fossil fuels, and a high level of biodiversity–with nearly 15 thousand species. The natural capital in Vietnam is of great value and plays an important role in the overall sustainable development strategy.

Vietnam is modernizing rapidly, with high rates of population growth and socioeconomic development since the promulgation of Doi Moi (Renovation) in 1986. As testament to this rapid growth, Vietnam's poverty rate fell from 58% in 1993 to 14.5% in 2008 (Begun, 2012). However, this swift development has negatively impacted the environment, leading to widespread land, forest and water overuse and degradation (Meyfroidt and Lambin, 2008). Deforestation, poor water quality, and decreased biodiversity are additional impacts.

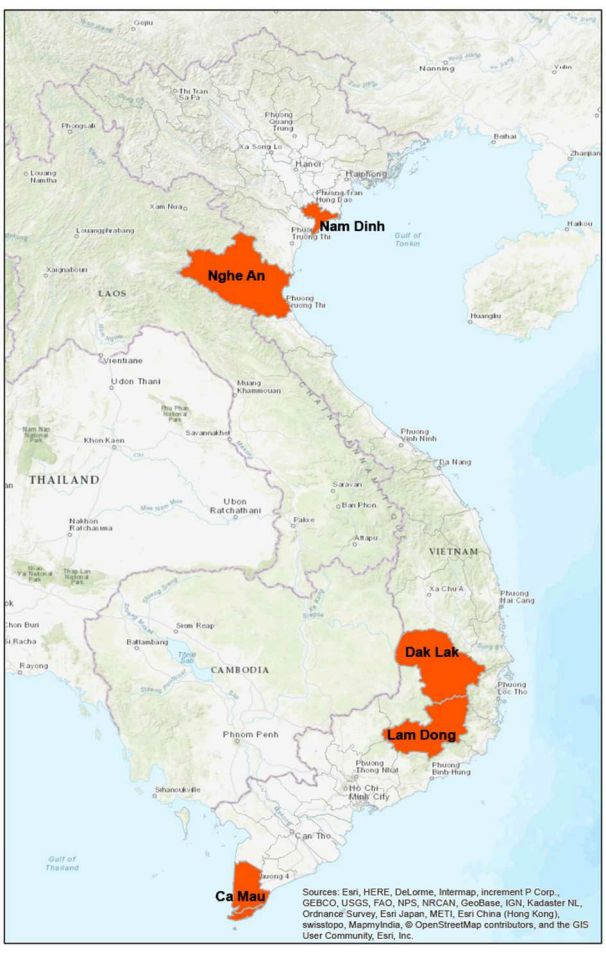


FIGURE 1 | Map of Vietnam with Nam Dinh, Nghe An, Dak Lak, Lam Dong, and Ca Mau indicated with orange colors.

Beginning in 1992 Vietnam started integrating sustainable development strategies into the country's legislative frameworks. In August 2004, Vietnam adopted the Strategic Orientation for Sustainable Development (Agenda 21), resulting in remarkable achievements in the economic, social, and environmental fields. Food security in Vietnam has improved significantly since 1986, with Vietnam becoming one of the biggest rice exporters in the world since 2010. The agricultural sector accounts for over 22% of the country's GDP, 30% of the country's exports, and 52% of the country's employment (IFAD, 2012). Agricultural growth has accelerated in recent decades through a combination of improved land use practices, better irrigation, technology adoption, and incentives, such as price liberalization and land titling, to encourage agricultural investment (Van Khuc et al., 2018). We study these dynamics in five provinces (**Figure 1**), which were selected to represent a range of agricultural production systems, biomes, and communities. A brief description of each province follows.

2.0.1. Nam Dinh

Nam Dinh province is part of the coastal area of northern Vietnam and located in the southern part of the Red River delta (Ha et al., 2018). The province has a total area of 1,676 km^2 and 72 km of coastline. Nam Dinh is surrounded by a dyke system that protects the land from inundation. Nam Dinh has alluvial, nutrient rich soil highly suitable for agriculture. The coastal province is well-known for its large Ramsar site wetland (Xuan Thuy National park), and its productive mangrove ecosystems. The province is frequently affected by natural disasters, including typhoons and floods. These affect the development of important economic sectors, including agriculture, aquaculture and fisheries. Economic development has also sparked debate about land reclamation into the sea and led to loss of mangrove forest area. Coastline erosion also has been impacting economic development in recent years.

2.0.2. Nghe An

Nghe An province is located in the north central coastal region of Vietnam. The total area of the province is 16,490 km^2. The provinces Eastern border is coastline with high mountain ranges in the west. The monsoon tropical climate supports a rich biodiversity. The total population of Nghe An was over 3.5 million in 2019. The Kinh people make up the majority group. The practice of shifting cultivation has led to an increase in fallow land over the past decades (Leisz, 2009). Urbanization and industrial zone development in the coastal zone area has increased rapidly. Recent numbers show that services and tourism contribute to over 40% of the GDP, whereas agriculture and forestry account for 20%.

2.0.3. Lam Dong

Lam Dong province is located in the southern part of Vietnam's central highland region and encompasses a total area of 9,764 km^2. There are 1.3 million people and 28 ethnic groups in this province. The majority of the population are Kinh people, at 97% of the population. The province is separated by the Di Linh and Lam Vien plateaus, which have altitudes ranging from 800 to

1,500 meters above sea level. The sloping terrain attracts many tourists. Lam Dong is the most biodiversity-rich province in the Central Highlands and south-central region. A large portion of the province is covered by natural or planted forests, that are home to a number of rare flora and fauna species. The province also contains a variety of soil types, which makes it suitable for industrial plant and crop cultivation as well as a source of mineral resources. However, the latter is a threat to the environment and biodiversity. After Vietnam released the Payment For Ecosystem Services (PFES) decree in 2009, Lam Dong was the first province that implemented the decree.

2.0.4. Dak Lak

Dak Lak province is located in the central highlands of Vietnam. The province has an average altitude of 600 m above sea level. The point with the highest elevation, at 2,442 m, is located in the Chu Yang Sin National Park. The province has a total area of 13,125 km^2 and a population of more than 2 million people from 44 ethnic groups. The province contains a dense river and stream network providing an abundance of surface water. Moreover, the province is known for rich forest resources. The soils in Dak Lak are of basalt origin, with high fertility which make them suitable for industrial crop plantations including coffee, pepper, and rubber.

2.0.5. Ca Mau

Ca Mau province is located in the Mekong Delta region, located in the southernmost part of Vietnam's inland territory. The province has a total area of 5,331 km^2 and supports a population of more than 1.2 million people. There are three main ethnic groups: the Kinh, Kh'mer, and Chinese. The province is situated in low, flat terrain which is frequently flooded. The average elevation is between 0.5 and 1.5 m above sea level. Ca Mau is a geologically young region, formed from silt and sediment accretion. The young, fertile lands support aquaculture, rice, mangroves, and brackish forests. In recent years a large proportion of rice paddies were converted into aquaculture uses.

2.1. Land Cover Maps

The land cover time series maps used in this study are from the Regional Land Cover Monitoring System (RLCMS) implemented across the Mekong region. They were created for the period from 1988 to 2017 using the Landsat legacy collection and machine learning methods in the Google Earth Engine open platform. The paper of Saah et al. (2020) provides specific technical details regarding the work flow and data processing of this system. We provide a short description below. Examples of the 1988 and 2017 maps are shown in **Figure 2**.

2.1.1. Annual Satellite Image Composites

The USGS Landsat 4, 5, 7, and 8 surface reflectance images were used to create land cover time series maps. These atmospherically corrected and orthorectified Landsat images are hosted in the Earth Engine data archive. Landsat 8 data were atmospherically corrected using the Landsat Surface Reflectance Code (LaSRC) (Holden and Woodcock, 2016; Roy et al., 2016a; Vermote et al., 2016), while the images from Landsat missions 4, 5, and 7 were corrected with LEDAPS (Vermote et al., 1997; Masek et al., 2006; Ju et al., 2012; Schmidt et al., 2013). All images come with a cloud, shadow, water and snow mask produced using CFMASK (Zhu and Woodcock, 2012), as well as a per-pixel saturation mask. Images from Landsat 7 ETM+ acquired after the 2003 Scan Line Corrector failure were not included in the analysis, as scan line effects were found to propagate through the data analysis into the final products. Additional image pre-processing steps were applied to account for sensor, solar, atmospheric, and topographic distortions (Young et al., 2017) and create a reliable and consistent time series. We applied shadow and cloud removal (Housman et al., 2018; Chastain et al., 2019), a bidirectional reflectance distribution function (BRDF) (Lucht et al., 2000; Roy et al., 2008, 2016b, 2017; Gao et al., 2014) and topographic correction (Smith et al., 1980; Justice et al., 1981; Teillet et al., 1982; Soenen et al., 2005).

2.1.2. Primitives

Primitives are the suite of biophysical and end member layers used to construct land cover products using a customizable assembly logic. Primitive examples include forest canopy height and fractional canopy cover (Potapov et al., 2019). The information represented by the suite of primitive layers is used to make decisions in a dichotomous key for land cover typing. To create the input primitives used in this study, we ran a supervised classification predicting the presence or absence of each land cover class. Reference data were merged with coincident values from the image stack to create a training data set used in a random forest classifier. Data for each class (presence observations) were combined with a random sample of all other classes (absence observations) to create a sample representing both classes for each primitive layer. The training data were used in a random forest model in R (Breiman, 2001; Liaw and Wiener, 2002; R Core Team, 2018) to determine the most important covariates. We used this information to select a subset of images and image derivatives, such as median normalized difference vegetation index (NDVI), from the full list of potential covariates. The random forest classifier was implemented on the subset of covariates in Google Earth Engine (GEE). It was trained with 100 decision trees and the output was saved using the probability mode option (instead of majority vote). The classifier was run on each annual image stack, creating a new time series stack of probability maps for each primitive class.

2.1.3. Assembly Logic and Monte-Carlo Simulations

The final land cover assemblage was created using a decision tree and Monte-Carlo simulation. The decision tree specifies the order and thresholds used to combine all primitives together into a final land cover map. This decision tree was run 100 times using a Monte-Carlo simulation process (Binder et al., 1993). In the Monte-Carlo process, random numbers were added to each of the primitives based on the error structure of the primitive layer. The final land cover map was the mode of the 100 simulations and a probability map which contained the count of the mode divided by the total number of model runs.

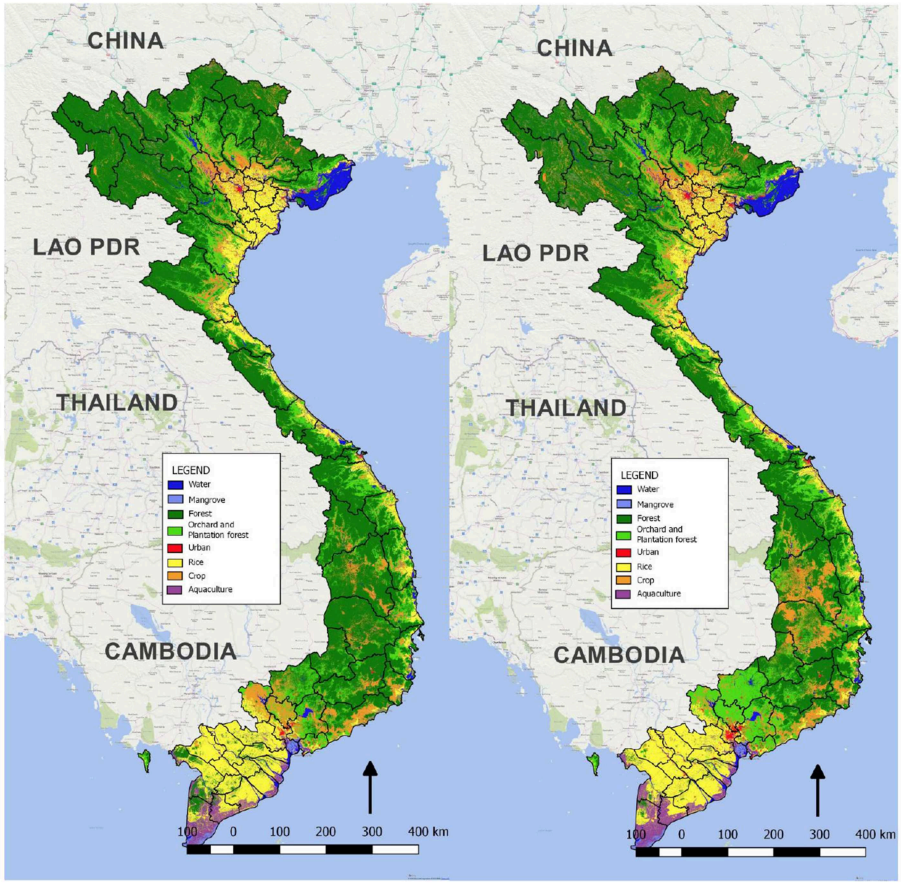


FIGURE 2 | Vietnam land cover map for the years 1988 **(Left)** and 2017 **(Right)**. Land cover maps of all years can be found online on the SERVIR-Mekong data portal, accessible at https://rlcms-servir.adpc.net/en/.

2.2. Typology

This study focuses on assessing the 8 main land cover types. Since definitions of land cover types vary between studies, we have included the precise definitions we used for each class below.

- Aquaculture is the farming of aquatic organisms, including fish, molluscs, crustaceans and aquatic plants. Areas defined as aquaculture consist of man-made pond systems on fresh or salt water surfaces, and can be permanently or temporarily flooded.
- Cropland is defined as land with temporary crops followed by harvest and a bare soil period Loveland and Belward (1997). These include single and multiple cropping systems. Example crops include cereals, oils seeds, vegetables, root crops and forages. Some areas include mixed rice and seasonal crop cultivars. Herbaceous and shrubby cultivated plants (e.g., tea and coffee) are included in this layer; but it excludes orchards, forest croplands, and forest plantations.
- Mangroves are defined as coastal sediment habitats with more than 10% woody vegetation canopy cover where the majority of cover is higher than 2 m (Loveland and Belward, 1997).
- Forest is land spanning more than 0.5 hectares with trees higher than 5 m and a canopy cover of more than 10%, or trees able to reach these thresholds *in situ*. It does not include land that is predominantly under agricultural or urban land use.
- Orchard or plantation forests are defined by FAO as land cultivated with perennial crops that occupy the land for long periods (Blanchez, 1997). Commercial tree crops in Vietnam include mainly rubber, cashew nut, and fruit trees. Forest plantations such as eucalyptus and acacias are included in this category.
- Rice paddies are defined as irrigated or flooded fields, or low land paddy fields where rice is intensively planted for more than 1 cycle per year (can be up to 2 or 3 cycles). Rice makes up the majority of vegetation cover.
- Surface water is defined as open water bodies larger than 30 m^2 which are open to the sky, including both fresh and saltwater (Pekel et al., 2016).
- Urban and built-up areas include cultural lands covered by buildings, roads, and other built structures.

2.3. Response Design for Area Estimation

Statistically robust and transparent methods to assess map accuracy and estimate an unbiased area (of change) are critical to ensure the integrity of land change statistics. Area estimations

from only counting pixels in land cover maps are often biased because of the inherent errors of the classification process (McRoberts, 2011). Errors are rarely equally distributed between land cover classes; instead there are specific land cover categories that are more difficult to separate than others. This is due to similarity and overlap in spectral and temporal signatures between these classes. Performance can also vary between ecoregions, provinces and even districts. These errors will propagate and create biased area estimations. Hence, based on these considerations, we apply the guidelines of Olofsson et al. (2013, 2014) to estimate map accuracy and provide unbiased area estimates. We apply a stratified random sample design to ensure that classes that cover a small portion of the total area are adequately represented in the sample, while still maintaining a manageable sample size. Equation (1) (Cochran, 1978) was used to estimate the total sample size.

$$n = \frac{(\sum W_i S_i)^2}{[S(\hat{o})]^2 + \frac{1}{N}\sum W_i S_i^2} \approx \left(\frac{\sum W_i S_i}{S(\hat{o})}\right)^2 \quad (1)$$

where:

n = Number of sample points
N = Number of units in the Region of Interest (ROI)
W_i = Mapped area of class i (proportion)
S_i = Standard deviation of stratum i (Equation 2)
$S(\hat{o})$ = Standard error of estimated overall accuracy.

The total sample size estimate is based on the assumption that N is large, which is valid for our present study. This approach was also suggested as a good practice in Olofsson et al. (2014) and Stehman (2009). The standard deviation of stratum i is,

$$S_i = \sqrt{U_i(1 - U_i)} \quad (2)$$

where:

U_i = User's accuracy (commission error).

Table 1 shows the number of samples used in this study, the proportional area, users' accuracy, and standard deviation for each land cover strata (Equation 2). The proportional area was estimated for the 2017 map and the users' accuracies were previously reported in Saah et al. (2020). Filling out Equation (1) using an S(o) of 0.01 leads to a total of 1,567 sample points. We applied a minimum threshold of 100 points for each land cover class, increasing the sample size of the surface water, mangrove, urban and built up and aquaculture strata. Without the minimum threshold we would have high uncertainty in the error estimates for these land cover strata due to their low occurrence in the landscape. The remaining points were proportionally allocated by area.

Area proportions were calculated using the confusion matrix from the validation data and Equation (3) (Stehman and Czaplewski, 1998; Olofsson et al., 2014).

$$\hat{p}_{ijy} = w_{iy}\frac{n_{ij}}{n_i} \quad (3)$$

TABLE 1 | Sampling strata for the different land cover classes.

	W_i	U_i	S_i	prop	*n*
Surface water	0.02	0.96	0.20	28	100
Mangrove	0.01	0.69	0.46	6	100
Forest	0.46	0.87	0.34	721	577
Orchard or Plantation Forest	0.19	0.45	0.50	303	242
Urban and Built up	0.02	0.96	0.20	29	100
Cropland	0.11	0.74	0.44	172	137
Rice	0.17	0.74	0.44	265	211
Aquaculture	0.03	0.75	0.43	44	100
Total	1.00			1,567	1,567

Prop indicates the area-proportional number of points, while n is the final number of points included in the sample per strata after taking into account the minimum size threshold of 100.

TABLE 2 | Number of validation points per province.

	Dak Lak	Ca Mau	Lam Dong	Nam Dinh	Nghe An
Aquaculture	–	335	–	100	100
Cropland	316	–	100	–	100
Forest	521	–	589	–	466
Mangrove	–	100	–	100	–
Plantations	122	100	166	-	136
Rice	–	102	–	242	100
Urban	100	100	100	100	100
Water	100	100	100	100	100
Total	1,159	837	1,038	642	1,102

Not all classes were included for every province as the areas were too small or not present.

where:

p_{ij} = Estimated area proportion for year y
W_i = Mapped area of class i (proportion) for year y
n_{ij} = Sample counts
n_i = Total sample counts of class i.

We created a separate response design for the five provinces as the area proportions vary between these regions. **Table 2** shows the number of points included for each class and province. It should be noted that not all classes are represented in each province or only cover a small portion of the total area. Hence, they were not included in the response design.

Reference data were collected in Collect Earth Online (CEO; Saah et al., 2019). CEO is a custom built, open-source, high resolution satellite image viewing and interpretation system, an on-line successor of Collect Earth (desktop) (Bey et al., 2016). No desktop installation is required and only an internet connection is needed. It includes a customizable widget interface where the project administrator can set up the display of image chips and time series plots using Google Earth Engine.

High resolution images are not available for each year of the analysis, therefore it was not possible to use this approach to validate the change areas for all years in our time series

TABLE 3 | Confusion matrix from the Vietnam country level map validation.

	Aqua	Crop	Forest	Mangr	Plant	Rice	Urban	Water	Users accuracy
Aquaculture	94	0	0	0	1	1	2	2	0.94
Cropland	0	107	0	0	11	1	2	0	0.88
Forest	0	18	497	0	31	0	2	2	0.90
Mangrove	4	0	8	73	3	0	1	3	0.79
Plantations	0	5	24	0	172	0	1	3	0.84
Rice	3	0	0	0	8	168	16	2	0.85
Urban	1	1	0	0	2	0	91	1	0.95
Water	0	0	0	0	0	1	0	96	0.99
Prod Accuracy	0.92	0.82	0.94	1.00	0.75	0.98	0.79	0.88	

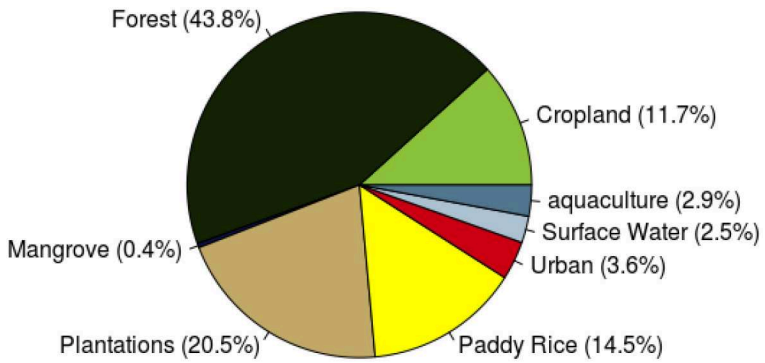


FIGURE 3 | Land cover area distribution of Vietnam in 2017.

assessment. Therefore, we integrated census data from the Vietnam national statistical yearbook to triangulate our findings with an independent data source.

3. RESULTS

The results of the national level map evaluation are displayed in **Table 3**. It is notable that the numbers do not match the number of samples defined in the response design; this is due to low-quality or a completely lack of high resolution satellite imagery. The overall accuracy of the national map was 89%, with users' accuracies of the land cover classes varying between 79% and 99% and producers' accuracies between 79% and 100%. Notable overlap was found between urban and rice, forest and plantations, and forest and cropland.

We used producers' accuracy to estimate the total area of each land cover class. The results for 2017 are shown in **Figure 3**. The largest land cover class across the country is forestland, followed by plantations and paddy rice. Cropland also accounts for a large portion, whereas aquaculture, surface water, urban, and mangrove forest constitute a small percentage of the total territory.

The 1988–2017 time series land cover maps were used for change analysis. **Figure 4** (**Supplemental Data**) shows the change in relation to the total area in 1988. The urban category gained the most area, followed by plantations and aquaculture. Rice and surface water expanded slightly but increases leveled off after 2010. Forest and mangrove areas declined, but the decline stabilized in 2000 (for forests) and in 2010 (for mangroves). The change trajectories for cropland are particularly interesting: they show a sharp increase beginning in 2000, a decline after 2003 and a stabilization after 2010. **Figure 4** shows that land cover change

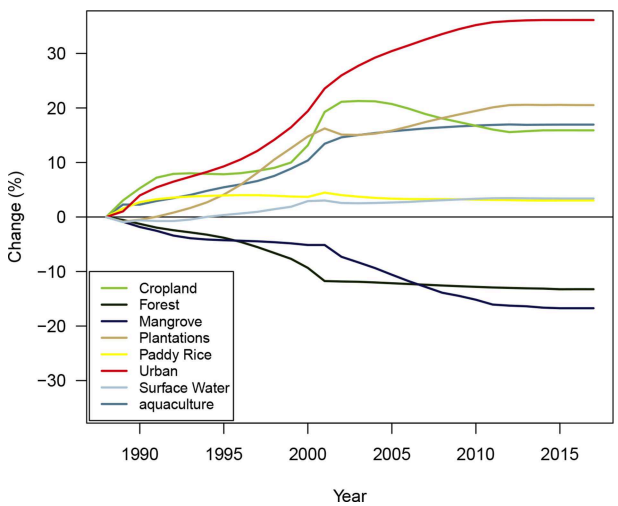


FIGURE 4 | The percent of change of coverage for each land cover class in Vietnam, using the 1988 area composition as the baseline.

dynamics in Vietnam were divided into three periods: (1) before 2000 when all land cover types changed rapidly; (2) a slow down in change between 2000 and 2005; and (3) stabilization of change from 2005 to 2010.

The change dynamics of **Figure 4** were compared with the numbers from the national statistical yearbook. **Figure 5** shows the reported area of cassava, sweet potatoes, maize, spring, winter and autumn paddy reported in the yearbook as well as the changes in detectable with the RLCMS. As the figure shows, the yearbook reports an increase in area of spring and autumn paddy, maize and cassava. The yearbook reports that the total cropland area increases, albeit at a greater magnitude than total cropland increased estimated from the RLCMS. This difference might indicate an intensification as only a small portion of the change is explained by an expansion of agricultural area.

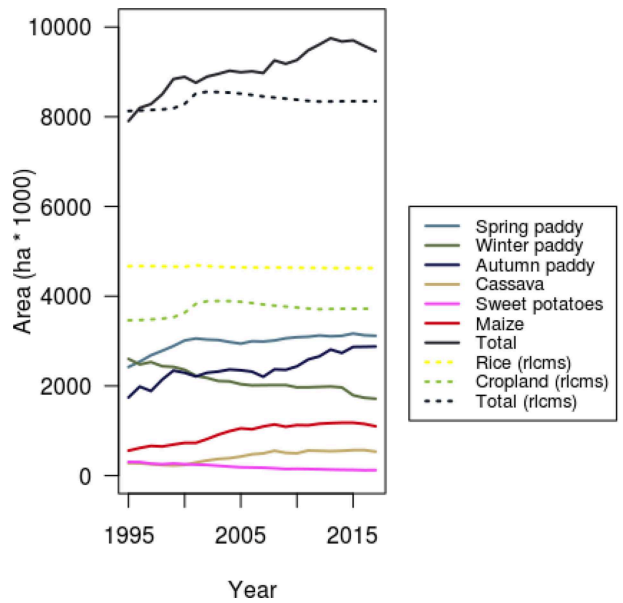


FIGURE 5 | Area from statistical yearbook and area calculated from the RLCMS landcover map.

The response design was then applied to each province. **Table 4** shows the province level users' and producers' accuracies. The highest overall accuracy was in Ca Mau, followed by Nghe An, Nam Dinh, Lam Dong and Dak Lak. Producers' accuracies were lower than the users' accuracies for plantations and crops. For the forest strata, producers' accuracies were higher than the users' accuracies.

Data from **Table 4** was used to map land cover change dynamics for each selected province in relation to the reported changes in cropland area and crop production from the statistical yearbook. **Figure 6** (**Supplemental Data**) plots the change dynamics between 1988 and 2017 as estimated by RLCMS and reported by the statistical yearbook for each province. **Figures 7–11** show the land cover changes at a 5 year interval between 1990 and 2017. Transitions from forest to cropland are most noticeable between 1990 and 2000. Charts show the estimated changes in area from the total area expressed as a percentage from the 1988 baseline. The statistics charts show the changes as reported in the statistical yearbook. The production numbers were also included (titled prod) and are expressed in thousand tons (kg).

The greatest increase in urban area was found to be in Nam Dinh, apparently at the expense of rice paddies. This is in agreement with the data from the statistical yearbook which also shows a decrease in rice area. Dak Lak province shows the highest increase in crop area, which is confirmed by the statistical data. It also shows a sharp decrease in forested area, as does Lam Dong province. In both provinces forest area was replaced by agricultural lands and tree plantations from 1988 to 2000, but this stabilized after 2000. Ca Mau province experienced rapid expansions in aquaculture ponds, mostly as a result of conversion from rice paddies and mangroves. This is well represented in both the RLCMS results and the statistical data. Land cover changes have been relatively less dynamic in Nghe An province, compared to other provinces; there were only slight increases in urban, rice, and forest areas.

The yearbook data on agricultural production levels tell a different story than the aerial coverage estimates. **Figure 6** shows that the production of almost all agricultural products

TABLE 4 | Overall user and producer accuracy for each province.

		Aqua	Crop	Forest	Mangrove	Plantations	Rice	Urban	Water	Overall
Dak Lak	*users*		0.87	0.75		0.74		0.95	1.00	0.82
	producers		0.74	0.92		0.64		0.83	0.93	
Ca Mau	*users*	0.96			0.85	0.93	0.74	0.94	0.80	0.91
	producers	0.91			1.00	0.90	0.93	0.94	0.80	
Lam Dong	*users*		0.90	0.81		0.90		0.93	0.94	0.86
	producers		0.72	0.98		0.61		0.91	0.96	
Nam Dinh	*users*	0.92			0.97		0.87	0.95	0.79	0.89
	producers	0.79			0.96		0.97	0.80	0.88	
Nghe An	*users*	0.89	0.88	0.94		0.81	0.85	0.85	0.97	0.90
	producers	0.96	0.87	0.95		0.76	0.85	0.88	0.90	

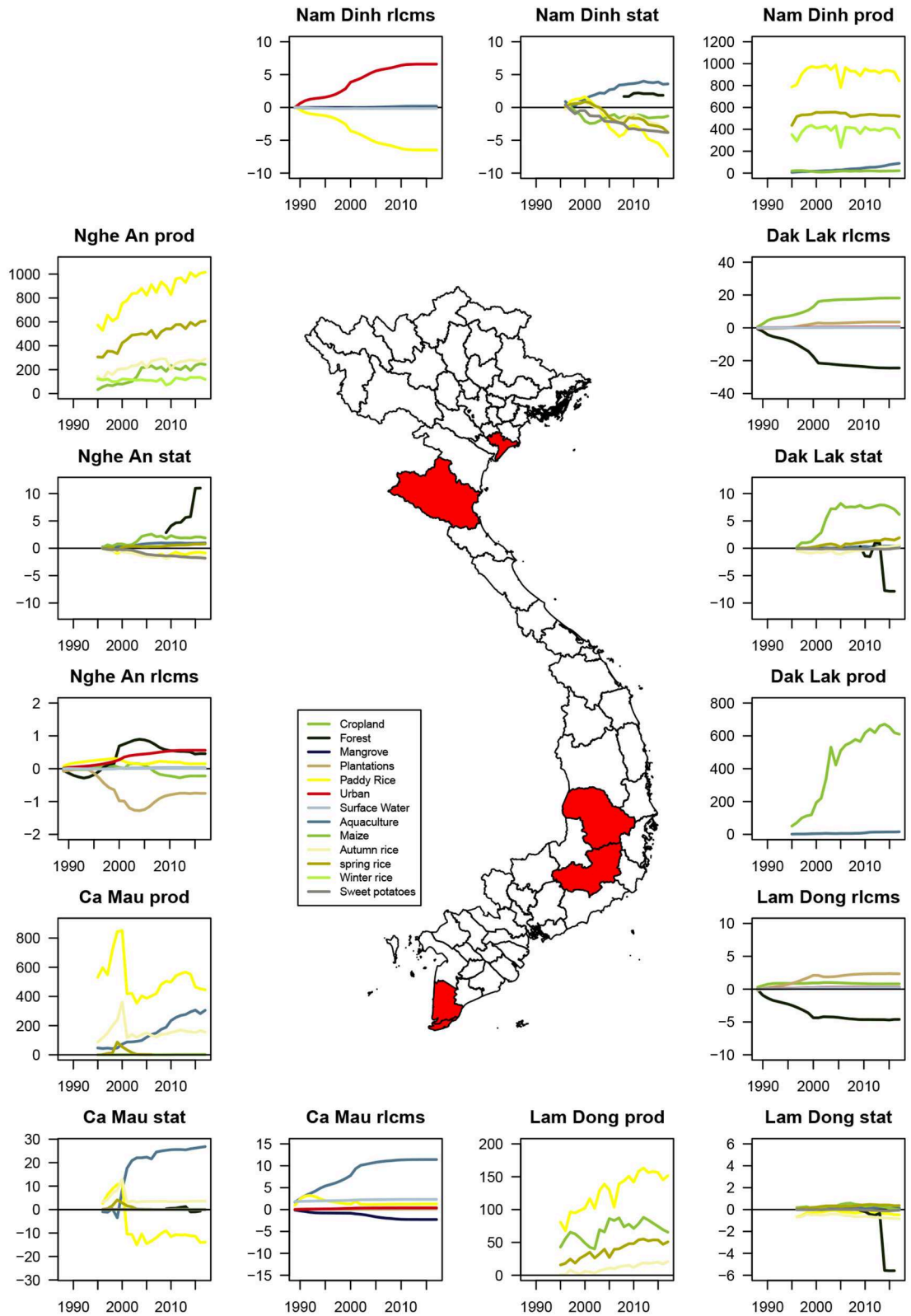


FIGURE 6 | The change in area from RLCMS, the statistical yearbook (stat) and the total production (prod) for different crop types. Change in area from RLCMS and the statistical yearbook are expressed as a percentage (%), the unit of production is in thousand tons.

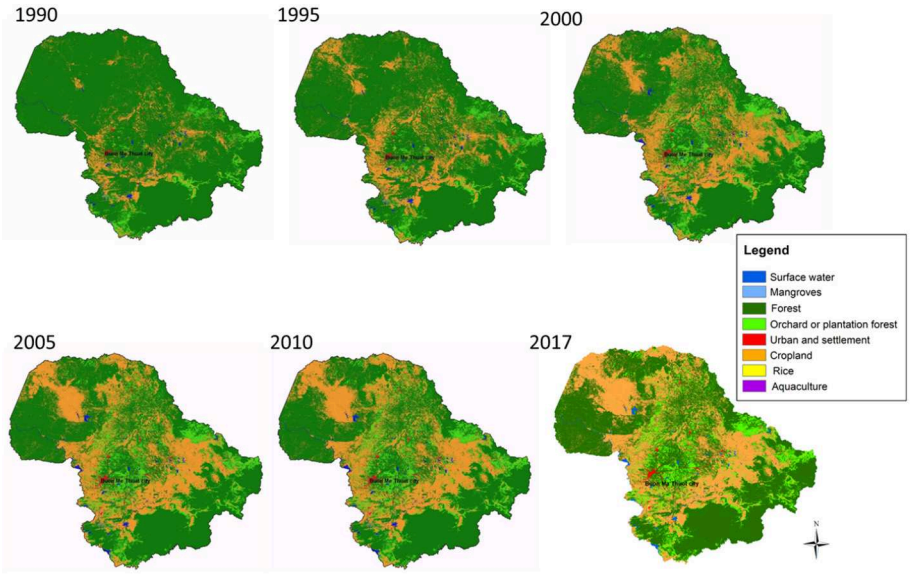


FIGURE 7 | Land cover change in Dak Lak with 5 year intervals from 1990 to 2017. The statistics can be found in **Supplementary Table 4**.

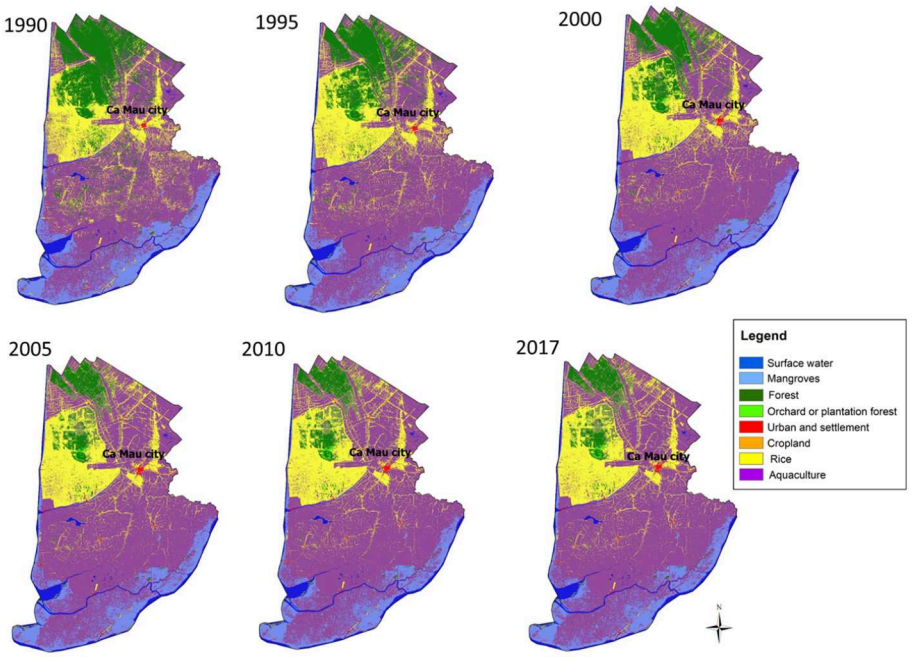


FIGURE 8 | Land cover change in Ca Mau with 5 year intervals from 1990 to 2017. The statistics can be found in **Supplementary Table 2**.

has increased across all provinces, without a proportional expansion in area under cultivation. Nam Dinh shows stable rice production, yet the area supporting rice cultivation decreased. Production levels in Nghe An province also increased while cultivated area decreased. For Dak Lak province we found an increase in maize production that coincides with areal expansion up to the year 2000, but after 2000 production continues to increased while area under cultivation does not. Aquaculture shows a linear increase in productivity over all years analyzed, while areal expansion stabilized after 2000.

4. DISCUSSION

In this study, we estimated national forest cover levels at 43.8%, which is close to the estimates of 42.7 and 44.6% reported by

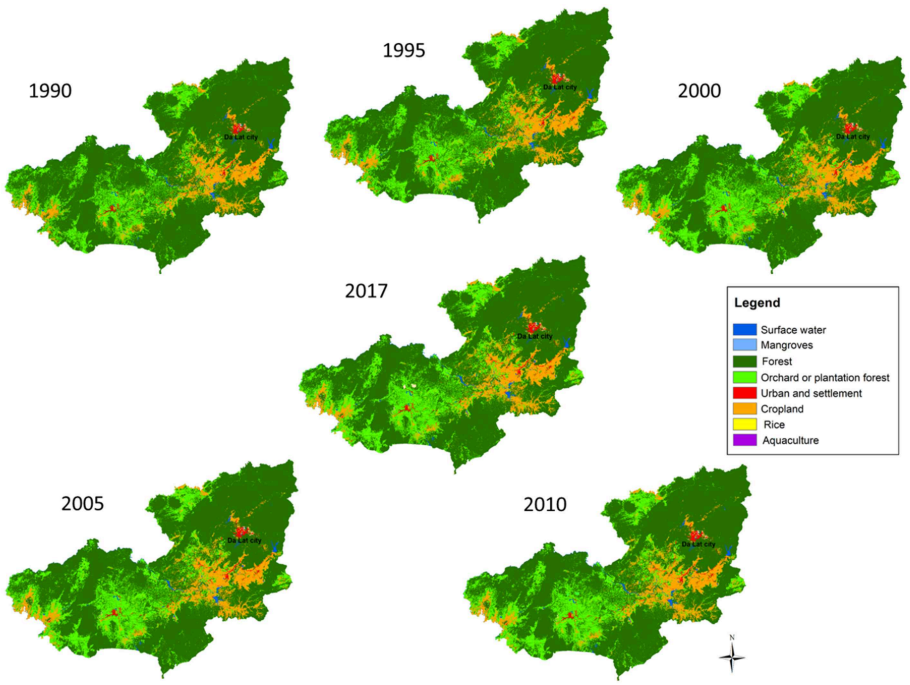


FIGURE 9 | Land cover change in Lam Dong with 5 year intervals from 1990 to 2017. The statistics can be found in **Supplementary Table 3**.

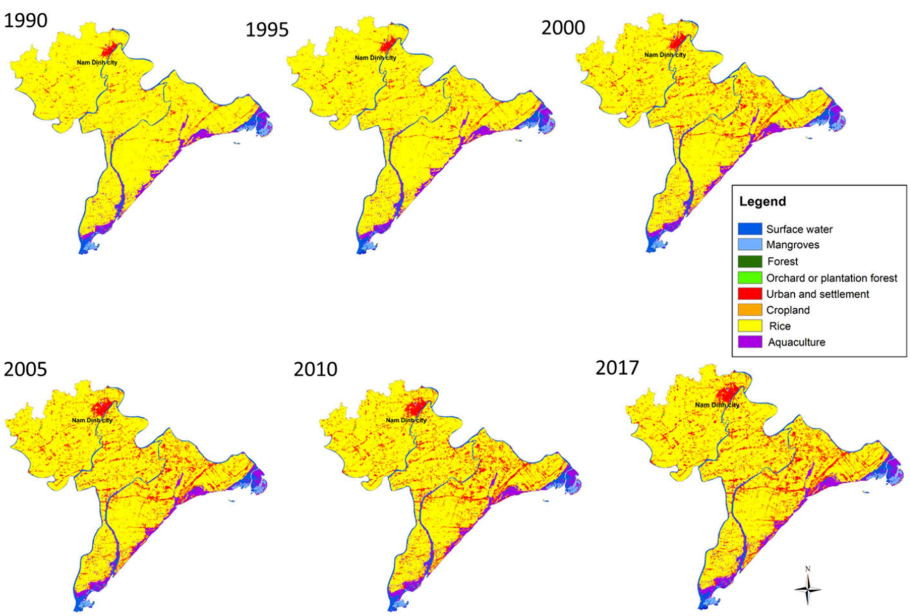


FIGURE 10 | Land cover change in Nam Dinh with 5 year intervals from 1990 to 2017. The statistics can be found in **Supplementary Table 6**.

the Food and Agricultural Organization of the United Nations (FAO) for the years 2010 and 2015 (FAO, 2015; Vogelmann et al., 2017). FAO also reports a sharp decline in primary forest between 1990 and 2000, and an increase in naturally regenerated forest from 2000. This is all in line with our analysis. Previous studies regarding the decrease in area of mangroves are also in line with our findings. Mangroves were estimated to cover around 1.2% of the country in 1943 (Maurand, 1943) but decreased to 0.42% in

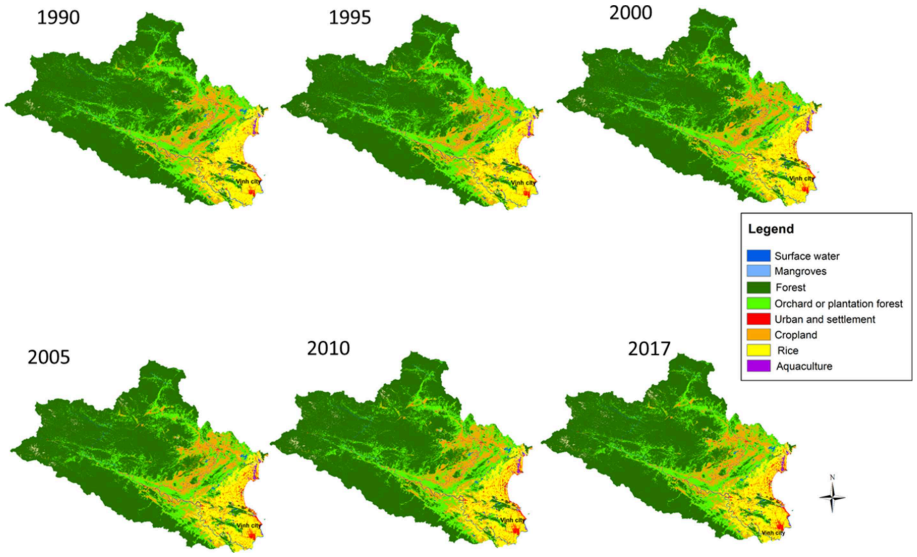


FIGURE 11 | Land cover change in Nghe An with 5 year intervals from 1990 to 2017. The statistics can be found in **Supplementary Table 5**.

2010 mostly because of the expansions of aquaculture and rice paddies (Richards and Friess, 2016; Grellier et al., 2017). These dynamics are clearly shown in **Figure 4**, but are particularly clear for Ca Mau province in **Figure 6**.

Applying best practices for data validation and area estimations using a stratified random sampling design are labor and resource demanding. This is because interpreting hundreds of data-points on high resolution satellite imagery is labor intensive, and acquiring high resolution satellite imagery is resource-demanding when large areas need to be covered. A limitation of this study is that validation was done using high resolution satellite imagery from only 2017, whereas the lower resolution time-series covers 30 years. Validation of the data over multiple years would further improve the quality of the area estimations. However, we attempted to triangulate our estimates by comparing the results to report from the national statistical yearbook, and we see good agreement between the two sources of crop coverage estimates.

Overall, our analysis shows a notable increase in production with little expansion of area under cultivation. These dynamics coincide with liberalized markets, improved standards, intensified production systems, more efficient fertilizer application, better land and water management practices, the new and more efficient plant varieties (OCDE, 2015; Nguyen et al., 2017; Xuan, 2018). These improvements happened because Vietnam adopted incentives in the agriculture sector such as the Vietnamese Good Agricultural Practices (VietGAP) and the Climate-Smart Agriculture (CSA) programs, which include a number of innovative technologies and practices that aim to increase agriculture production while preserving the environment.

Figure 12 provides a decadal timeline highlighting important legislation on environmental management in Vietnam. This timeline of improvement began when Vietnam participated in the 1992 Earth Summit in Brazil, where country committed to new sustainable development strategies. This helped the country to develop the foundation and legal framework for environmental and natural resource protection, the forest protection and development law, the environmental protection law, and the national plan on biodiversity and nature conservation, all of which were enacted between 1990 and 2000, and tied to economic development targets. As the results above show, these new policies led to a period of increased land cover transitions.

Between 2000 and 2010, the country began integrating the Strategic Orientation for Sustainable Development (Vietnam Agenda 21) into development policies, and also joined the World Trade Organization (WTO) and well as other international trade bodies. These memberships meant that Vietnam had to produce products that met requirements for world markets. International legislation was adapted to the Vietnamese context and integrated into national, ministerial, and local development and certification strategies. This ensured better linkages between economic, social, and environmental objectives, as well improved law enforcement, helping the country sustain healthier ecosystems.

Post-2010, Vietnam adopted the national Green Growth Development Strategy; this strategy defined important environmental policies, such as the national strategy on climate change mitigation and a nationwide policy on payments for forest environmental services (PFES). These new policies have undoubtedly promoted forest recovery, increased forest plantations and stabilized cropland expansion. These dynamics are also coupled with increasing food production, which is essential for national food security.

In particular, the PFES scheme seems to show promise for improved land use and socio-economic development. The implementation of the PFES program has mobilized

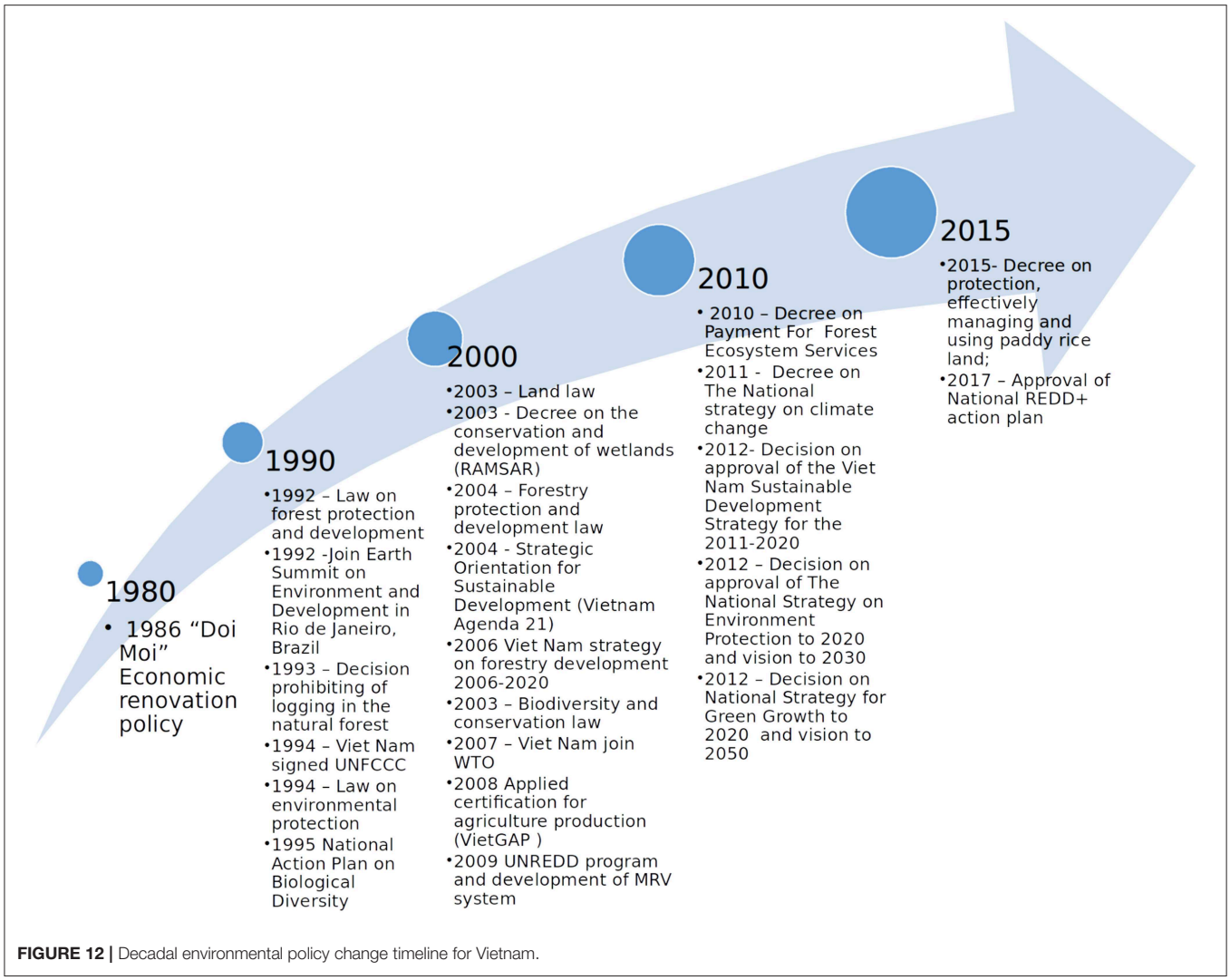


FIGURE 12 | Decadal environmental policy change timeline for Vietnam.

hundreds of thousands of households to protect and manage more than 5 million hectares of forestland in under 10 years. The program aims to promote sustainable livelihoods and incomes for local people who protect and manage forests. As a next step Vietnam is currently developing a monitoring framework to support effectiveness and transparency of PFES. Indicators in this framework include: forest area, forest quality, landscape, land use and land cover change dynamic, water quality, quantity and sedimentation. Monitoring change in these indicators will be essential for successful policy implementation.

Swidden agriculture was not included in the analysis as it is considered a land use type rather than land cover. Swidden cultivation has been a common land use systems in the tropics including Vietnam (Fox et al., 2000; Mertz et al., 2009; Van Vliet et al., 2012). Resolution 10 of 1988 and the Land Law of 1993 have played an important role in eliminate shifting cultivation by allocating fixed fields to shifting cultivators while encouraging them to cultivate these permanently in the northern and central mountainous region in Vietnam (Castella et al., 2006). Moreover, measures have been implemented on upland cultivation such as sloping agricultural land technology (SALT) to enhance the efficiency and crop productivity (Folving and Christensen, 2007). However, swidden agriculture is still practiced (Pham et al., 2018) and have evolved to accommodate the changing land laws and also to incorporate tree crops. Areas that were classified as forests in this study may be made up of trees that are part of managed regrowth in a fallow system. Similarly, croplands in the upland areas may be part of swidden systems. A land cover change analysis on using the RLCMS time series can provide more information on this.

In this paper we applied the landscape approach to analyze the land use and land cover change both nationally and in selected provinces in Vietnam. This approach allowed us to link physical land use change and changes in agricultural production to evolving policy and sustainable development strategies in Vietnam over that past 30 years. The assessment

shows how increasingly complex environmental, social and political challenges that transcend traditional management boundaries can be addressed through good sustainable development policies.

Ecosystems are directly affected by changes in land use and land cover. Such changes affect ecosystem services by increasing the availability of certain services while reducing others, thus reducing the ecosystems ability to support human needs. This in turn further accelerates ecological degradation and the decreases the resiliency of communities reliant on these services (Polasky et al., 2011). With this study we showed that land cover data and information provided by RLCMS can be used for relatively quick and efficient ecosystem service evaluation and assessments that can support better policy implementation and monitoring.

5. CONCLUSION

We cross-walked a remote sensing-derived 30 year land cover time-series with national statistical data of forest resources and agricultural productivity in Vietnam. We have shown that most land use and land cover changes occurred before the year 2000 and have stabilized since. This stabilization coincides with the implementation of important national environmental and sustainable development policies and strategies. We also show that agricultural production has soared while the coverage of agricultural areas has remained relatively steady. We demonstrated that RLCMS is an important tool to evaluate the effectiveness of sustainable development policies and the evaluation of ecosystem service and natural capital.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

AP, QN, BB, KT, DS, and FC contributed conception and design of the study. AP and QN organized the data collection. AP, QN, and FC performed the statistical analysis. AP and QN wrote the first draft of the manuscript. KT, DS, and FC wrote sections of the manuscript. All authors contributed to the manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

We thank SilvaCarbon, RDMA, USAID, NASA, Google, the SERVIR program, and Rachmat Mulia from ICRAF for supporting our efforts and providing valuable feedback. Support for this work was provided through the joint US Agency for International Development (USAID) and National Aeronautics and Space Administration (NASA) initiative SERVIR, Cooperative agreement number: AID-486-A-14-00002. We thank Jefferson Metz Fox, Stephen Leisz, and Niall Patrick Hanan for their constructive comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2019.00186/full#supplementary-material

REFERENCES

Bartholomé, E., and Belward, A. S. (2005). Glc2000: a new approach to global land cover mapping from earth observation data. *Int. J. Remote Sens.* 26, 1959–1977. doi: 10.1080/01431160412331291297

Begun, W. (2012). *Not Yet Done: Vietnam's Remarkable Progress on Poverty Reduction and the Emerging Challenges*. Hanoi: World Bank.

Bey, A., Sánchez-Paus Díaz, A., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., et al. (2016). Collect earth: land use and land cover assessment through augmented visual interpretation. *Remote Sens.* 8:807. doi: 10.3390/rs8100807

Binder, K., Heermann, D., Roelofs, L., Mallinckrodt, A. J., and McKay, S. (1993). Monte carlo simulation in statistical physics. *Comput. Phys.* 7, 156–157. doi: 10.1063/1.4823159

Blanchez, J, L. (1997). "Forest resources and roundwood supply in the Asia Pacific countries: situation and outlook to the Year 2010," in *Working Paper no: APFSOS/WP/17*. FAO; Forestry Policy and Planning Division.

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Bui, Y. T., Orange, D., Visser, S., Hoanh, C. T., Laissus, M., Poortinga, A., et al. (2014). Lumped surface and sub-surface runoff for erosion modeling within a small hilly watershed in northern vietnam. *Hydrol. Process.* 28, 2961–2974. doi: 10.1002/hyp.9860

Castella, J.-C., Boissau, S., Thanh, N. H., and Novosad, P. (2006). Impact of forestland allocation on land use in a mountainous province of vietnam. *Land Use Policy* 23, 147–160. doi: 10.1016/j.landusepol.2004.07.004

Chastain, R., Housman, I., Goldstein, J., and Finco, M. (2019). Empirical cross sensor comparison of sentinel-2A and 2B MSI, landsat-8 oli, and landsat-7 ETM+ top of atmosphere spectral characteristics over the conterminous United States. *Remote Sens. Environ.* 221, 274–285. doi: 10.1016/j.rse.2018.11.012

Cochran, W. G. (1978). "Laplace's ratio estimator," in *Contributions to Survey Sampling and Applied Statistics*, ed H. A. David (Academic Press), 3–10.

Congalton, R. G., Gu, J., Yadav, K., Thenkabail, P., and Ozdogan, M. (2014). Global land cover mapping: a review and uncertainty analysis. *Remote Sens.* 6, 12070–12093. doi: 10.3390/rs61212070

Costanza, R., d'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., et al. (1997). The value of the world's ecosystem services and natural capital. *Nature* 387:253. doi: 10.1038/387253a0

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., and Hansen, M. C. (2018). Classifying drivers of global forest loss. *Science* 361, 1108–1111. doi: 10.1126/science.aau3445

Daily, G. C. (1997). *Nature's Services: Societal Dependence on Natural Ecosystems*. Washington, DC: Island Press. 392.

FAO (2015). *Global Forest Resources Assessment 2015*. Technical report, Food and Agriculture Organization of the United Nations (FA).

Folving, R., and Christensen, H. (2007). Farming system changes in the Vietnamese uplands—using fallow length and farmers' adoption of sloping agricultural land technologies as indicators of environmental sustainability. *Geografisk Tidsskrift-Danish J. Geogr.* 107, 43–58. doi: 10.1080/00167223.2007.10801374

Foody, G. M. (2010). Assessing the accuracy of land cover change with imperfect ground reference data. *Remote Sens. Environ.* 114, 2271–2285. doi: 10.1016/j.rse.2010.05.003

Fox, J., Truong, D. M., Rambo, A. T., Tuyen, N. P., Cuc, L. T., and Leisz, S. (2000). Shifting cultivation: a new old paradigm for managing tropical forests.

BioScience 50, 521–528. doi: 10.1641/0006-3568(2000)050[0521:SCANOP]2.0.CO;2

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., et al. (2010). Modis collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote sens. Environ.* 114, 168–182. doi: 10.1016/j.rse.2009.08.016

Gao, F., He, T., Masek, J. G., Shuai, Y., Schaaf, C. B., and Wang, Z. (2014). Angular effects and correction for medium resolution sensors to support crop monitoring. *IEEE J. Select. Top. Appl. Earth Observat. Remote Sens.* 7, 4480–4489. doi: 10.1109/JSTARS.2014.2343592

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. doi: 10.1016/j.rse.2017.06.031

Grellier, S., Janeau, J.-L., Dang Hoai, N., Nguyen Thi Kim, C., Le Thi Phuong, Q., Pham Thi Thu, T., et al. (2017). Changes in soil characteristics and c dynamics after mangrove clearing (Vietnam). *Sci. Tot. Environ.* 593, 654–663. doi: 10.1016/j.scitotenv.2017.03.204

Ha, L., Bastiaanssen, W., Van Griensven, A., Van Dijk, A., and Senay, G. (2018). Calibration of spatially distributed hydrological processes and model parameters in swat using remote sensing data and an auto-calibration procedure: a case study in a Vietnamese River basin. *Water* 10:212. doi: 10.3390/w10020212

Holden, C. E., and Woodcock, C. E. (2016). An analysis of landsat 7 and landsat 8 underflight data and the implications for time series investigations. *Remote Sens. Environ.* 185, 16–36. doi: 10.1016/j.rse.2016.02.052

Housman, I., Chastain, R., and Finco, M. (2018). An evaluation of forest health insect and disease survey data and satellite-based remote sensing forest change detection methods: case studies in the United States. *Remote Sens.* 10:1184. doi: 10.3390/rs10081184

International Fund for Agricultural Development (IFAD) (2012). *Socialist Republic of Viet Nam Country Programme Evaluation*. Report No. 2606-VN.

Ju, J., Roy, D. P., Vermote, E., Masek, J., and Kovalskyy, V. (2012). Continental-scale validation of modis-based and ledaps landsat ETM+ atmospheric correction methods. *Remote Sens. Environ.* 122, 175–184. doi: 10.1016/j.rse.2011.12.025

Justice, C. O., Wharton, S. W., and Holben, B. (1981). Application of digital terrain data to quantify and reduce the topographic effect on landsat data. *Int. J. Remote Sens.* 2, 213–230. doi: 10.1080/01431168108948358

Leisz, S. J. (2009). Dynamics of land cover and land use changes in the upper CA river basin of Nghe An, Vietnam. *Japanese J. Southeast Asian Stud.* 47, 287–308. Available online at: https://www.jstage.jst.go.jp/article/tak/47/3/47_KJ00006032400/_article

Liaw, A., and Wiener, M. (2002). Classification and regression by randomforest. *R News* 2, 18–22. Available online at: https://cran.r-project.org/web/packages/randomForest/citation.html

Liu, J., Fritz, S., Van Wesenbeeck, C., Fuchs, M., You, L., Obersteiner, M., et al. (2008). A spatially explicit assessment of current and future hotspots of hunger in Sub-Saharan Africa in the context of global change. *Glob. Planet. Change* 64, 222–235. doi: 10.1016/j.gloplacha.2008.09.007

Loveland, T. R., and Belward, A. (1997). The IGBP-DIS global 1km land cover data set, DISCover: first results. *Int. J. Remote Sens.* 18, 3289–3295. doi: 10.1080/014311697217099

Lucht, W., Schaaf, C. B., and Strahler, A. H. (2000). An algorithm for the retrieval of Albedo from space using semiempirical BRDF models. *IEEE Trans. Geosci. Remote Sens.* 38, 977–998. doi: 10.1109/36.841980

Markert, K., Schmidt, C., Griffin, R., Flores, A., Poortinga, A., Saah, D., et al. (2018). Historical and operational monitoring of surface sediments in the lower mekong basin using landsat and google earth engine cloud computing. *Remote Sens.* 10:909. doi: 10.3390/rs10060909

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., et al. (2006). A landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geosci. Remote Sens. Lett.* 3, 68–72. doi: 10.1109/LGRS.2005.857030

Maurand, P. (1943). *L'Indochine Forestiére : les Forêts d'Indochine, Exploitation, Défrichement, Maménagement, Reconstitution des Forêts*. Hanoï: Imprimerie d'Extrême Orient, 22–26.

McRoberts, R. E. (2011). Satellite image-based maps: scientific inference or pretty pictures? *Remote Sens. Environ.* 115, 715–724. doi: 10.1016/j.rse.2010.10.013

Mertz, O., Padoch, C., Fox, J., Cramb, R. A., Leisz, S. J., Lam, N. T., et al. (2009). Swidden change in Southeast Asia: understanding causes and consequences. *Hum. Ecol.* 37, 259–264. doi: 10.1007/s10745-009-9245-2

Meyfroidt, P., and Lambin, E. F. (2008). Forest transition in vietnam and its environmental impacts. *Glob. Change Biol.* 14, 1319–1336. doi: 10.1111/j.1365-2486.2008.01575.x

Nguyen, T., Roehrig, F., Grosjean, G., Tran, D., and Vu, T. (2017). *Climate Smart Agriculture in Vietnam*. Hanoi: International Center for Tropical Agriculture (CIAT).

OCDE, O. (2015). *Agricultural Policies in Viet Nam 2015*. Organisation for Economic Co-operation and Development. doi: 10.1787/9789264235151-en

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* 148, 42–57. doi: 10.1016/j.rse.2014.02.015

Olofsson, P., Foody, G. M., Stehman, S. V., and Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* 129, 122–131. doi: 10.1016/j.rse.2012.10.031

Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature* 540:418. doi: 10.1038/nature20584

Pham, T., Moeliono, M., Wong, G. Y., Brockhaus, M., and Dung, L. N. (2018). The politics of swidden: a case study from Nghe An and Son La in Vietnam. *Land Use Policy*. doi: 10.1016/j.landusepol.2017.10.057

Polasky, S., Carpenter, S. R., Folke, C., and Keeler, B. (2011). Decision-making under great uncertainty: environmental management in an era of global change. *Trends Ecol. Evol.* 26, 398–404. doi: 10.1016/j.tree.2011.04.007

Poortinga, A., Clinton, N., Saah, D., Cutter, P., Chishtie, F., Markert, K. N., et al. (2018). An operational before-after-control-impact (BACI) designed platform for vegetation monitoring at planetary scale. *Remote Sens.* 10:760. doi: 10.3390/rs10050760

Potapov, P., Tyukavina, A., Turubanova, S., Talero, Y., Hernandez-Serna, A., Hansen, M., et al. (2019). Annual continuous fields of woody vegetation structure in the lower mekong region from 2000-2017 landsat time-series. *Remote Sens. Environ.* 232:111278. doi: 10.1016/j.rse.2019.111278

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reed, J., Deakin, L., and Sunderland, T. (2015). What are 'integrated landscape approaches' and how effectively have they been implemented in the tropics: a systematic map protocol. *Environ. Evid.* 4:2. doi: 10.1186/2047-2382-4-2

Richards, D. R., and Friess, D. A. (2016). Rates and drivers of mangrove deforestation in southeast asia, 2000–2012. *Proc. Natl. Acad. Sci. U.S.A.* 113, 344–349. doi: 10.1073/pnas.1510272113

Roy, D. P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., et al. (2008). Multi-temporal modis–landsat data fusion for relative radiometric normalization, gap filling, and prediction of landsat data. *Remote Sens. Environ.* 112, 3112–3130. doi: 10.1016/j.rse.2008.03.009

Roy, D. P., Kovalskyy, V., Zhang, H., Vermote, E. F., Yan, L., Kumar, S., et al. (2016a). Characterization of landsat-7 to landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sens. Environ.* 185, 57–70. doi: 10.1016/j.rse.2015.12.024

Roy, D. P., Li, J., Zhang, H. K., Yan, L., Huang, H., and Li, Z. (2017). Examination of sentinel-2a multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize msi reflectance to nadir brdf adjusted reflectance. *Remote Sens. Environ.* 199, 25–38. doi: 10.1016/j.rse.2017.06.019

Roy, D. P., Zhang, H., Ju, J., Gomez-Dans, J. L., Lewis, P. E., Schaaf, C., et al. (2016b). A general method to normalize landsat reflectance data to NADIR BRDF adjusted reflectance. *Remote Sens. Environ.* 176, 255–271. doi: 10.1016/j.rse.2016.01.023

Saah, D., Johnson, G., Ashmall, B., Tondapu, G., Tenneson, K., Patterson, M., et al. (2019). Collect Earth: an online tool for systematic reference data collection in land cover and use applications. *Environ. Model. Softw.* 118, 166–171. doi: 10.1016/j.envsoft.2019.05.004

Saah, D., Tenneson, K., Poortinga, A., Nguyen, Q., Chishtie, F., San Aung, K., et al. (2020). Primitives as building blocks for constructing land cover maps. *Int. J. Appl. Earth Observ. Geoinform.* 85:101979. doi: 10.1016/j.jag.2019.101979

Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., and Gao, F. (2013). *Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description*. Technical report, US Geological Survey.

See, L., Fritz, S., You, L., Ramankutty, N., Herrero, M., Justice, C., et al. (2015). Improved global cropland data as an essential ingredient for food security. *Glob. Food Secur.* 4, 37–45. doi: 10.1016/j.gfs.2014.10.004

Smith, J., Lin, T. L., and Ranson, K. (1980). The lambertian assumption and landsat data. *Photogrammetr. Eng. Remote Sens.* 46, 1183–1189.

Soenen, S. A., Peddle, D. R., and Coburn, C. A. (2005). SCS+ C: a modified sun-canopy-sensor topographic correction in forested terrain. *IEEE Trans. Geosci. Remote Sens.* 43, 2148–2159. doi: 10.1109/TGRS.2005.852480

Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. *Int. J. Remote Sens.* 30, 5243–5272. doi: 10.1080/01431160903131000

Stehman, S. V., and Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sens. Environ.* 64, 331–344. doi: 10.1016/S0034-4257(98)00010-8

Stibig, H.-J., Achard, F., Carboni, S., Rasi, R., and Miettinen, J. (2014). Change in tropical forest cover of Southeast Asia from 1990 to 2010. *Biogeosciences* 11:247. doi: 10.5194/bg-11-247-2014

Stürck, J., Poortinga, A., and Verburg, P. H. (2014). Mapping ecosystem services: the supply and demand of flood regulation services in europe. *Ecol. Indicat.* 38, 198–211. doi: 10.1016/j.ecolind.2013.11.010

Teillet, P., Guindon, B., and Goodenough, D. (1982). On the slope-aspect correction of multispectral scanner data. *Can. J. Remote Sens.* 8, 84–106. doi: 10.1080/07038992.1982.10855028

Van Khuc, Q., Tran, B. Q., Meyfroidt, P., and Paschke, M. W. (2018). Drivers of deforestation and forest degradation in Vietnam: an exploratory analysis at the national level. *For. Policy Econ.* 90, 128–141. doi: 10.1016/j.forpol.2018.02.004

Van Vliet, N., Mertz, O., Heinimann, A., Langanke, T., Pascual, U., Schmook, B., et al. (2012). Trends, drivers and impacts of changes in swidden cultivation in tropical forest-agriculture frontiers: a global assessment. *Glob. Environ. Change* 22, 418–429. doi: 10.1016/j.gloenvcha.2011.10.009

Vermote, E., Justice, C., Claverie, M., and Franch, B. (2016). Preliminary analysis of the performance of the landsat 8/oli land surface reflectance product. *Remote Sens. Environ.* 185, 46–56. doi: 10.1016/j.rse.2016.04.008

Vermote, E. F., Tanré, D., Deuze, J. L., Herman, M., and Morcette, J.-J. (1997). Second simulation of the satellite signal in the solar spectrum, 6s: an overview. *IEEE Trans. Geosci. Remote Sens.* 35, 675–686. doi: 10.1109/36.581987

Vogelmann, J., Khoa, P., Lan, D., Shermeyer, J., Shi, H., Wimberly, M., et al. (2017). Assessment of forest degradation in Vietnam using landsat time series data. *Forests* 8:238. doi: 10.3390/f8070238

WWF (2013). *Ecosystems in the Greater Mekong: Past Trends, Current Status, Possible Futures*. Greater Mekong: WWF.

Xuan, V. T. (2018). "Rice production, agricultural research, and the environment," in Vietnam's Rural Transformation, eds B. J. T. Kerkvliet and D. G. Porter (Abingdon: Routledge), 185–200.

Young, N. E., Anderson, R. S., Chignell, S. M., Vorster, A. G., Lawrence, R., and Evangelista, P. H. (2017). A survival guide to landsat preprocessing. *Ecology* 98, 920–932. doi: 10.1002/ecy.1730

Zhu, Z., and Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in landsat imagery. *Remote Sens. Environ.* 118, 83–94. doi: 10.1016/j.rse.2011.10.028

Conflict of Interest: AP, KT, AT, DS, and FC are employed by spatial informatics group, LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 11 December 2019
doi: 10.3389/fenvs.2019.00191

Operational Flood Risk Index Mapping for Disaster Risk Reduction Using Earth Observations and Cloud Computing Technologies: A Case Study on Myanmar

Kittiphong Phongsapan[1], Farrukh Chishtie[1], Ate Poortinga[2], Biplov Bhandari[1], Chinaporn Meechaiya[1], Thannarot Kunlamai[1], Khun San Aung[1], David Saah[2,3], Eric Anderson[4,5], Kel Markert[4,5], Amanda Markert[4,5] and Peeranan Towashiraporn[1]*

[1] Asian Disaster Preparedness Center, Bangkok, Thailand, [2] Spatial Informatics Group, LLC, Pleasanton, CA, United States, [3] Geospatial Analysis Lab, University of San Francisco, San Francisco, CA, United States, [4] Earth System Science Center, The University of Alabama in Huntsville, Huntsville, AL, United States, [5] SERVIR Science Coordination Office, NASA Marshall Space Flight Center, Huntsville, AL, United States

OPEN ACCESS

Edited by:
Niall Patrick Hanan,
New Mexico State University,
United States

Reviewed by:
Nidhi Nagabhatla,
United Nations University Institute for Water Environment and Health,
Canada
Sharad Kumar Jain,
National Institute of Hydrology, India

***Correspondence:**
Farrukh Chishtie
farrukh.chishtie@adpc.net

Specialty section:
This article was submitted to Freshwater Science, a section of the journal Frontiers in Environmental Science

Received: 19 July 2019
Accepted: 20 November 2019
Published: 11 December 2019

Citation:
Phongsapan K, Chishtie F, Poortinga A, Bhandari B, Meechaiya C, Kunlamai T, Aung KS, Saah D, Anderson E, Markert K, Markert A and Towashiraporn P (2019) Operational Flood Risk Index Mapping for Disaster Risk Reduction Using Earth Observations and Cloud Computing Technologies: A Case Study on Myanmar. Front. Environ. Sci. 7:191. doi: 10.3389/fenvs.2019.00191

People, livelihoods, and infrastructure in Myanmar suffer from devastating monsoonal flooding on a frequent basis. Quick and effective management of flood risk relies on planning and preparedness to ensure the availability of supplies, shelters and emergency response personnel. The mandated government agency Department of Disaster Management (DDM) as well as local and international organizations play roles in producing, disseminating, and using accurate and timely information on flood risk. Currently, systematic flood risk maps are lacking, which leaves DDM to rely on inconsistent historic reports and local knowledge to inform their emergency planning. Although these types of knowledge are critical, they can be complemented to reduce bias and human error to planning processes and decisions. As such, the present situation has led to ineffective distribution of emergency response resources prior to flooding, leaving vulnerable populations less-than-prepared for inevitable flood events. Given these issues, we have developed a flood risk decision-support tool in collaboration with DDM. The tool uses surface water maps developed by the Joint Research Center (JRC), which were derived from more than 30 years of Landsat imagery. We have also incorporated population data, land cover data, and other information on flood exposure and vulnerability to create the first scalable and replicable Flood Risk Index (FRI) for flood risk reduction in Myanmar.

Keywords: flood frequency, remote sensing, water management, Google Earth Engine, disaster preparedness, Myanmar, disaster management, earth observations

1. INTRODUCTION

Floods are considered to be one of the most recurrent natural hazards which can rapidly become significant disasters. Impacts of floods are amplified in the wake of increased vulnerability due to many factors such as rapid land cover changes (Markert et al., 2018b), urbanization, and changing climate (Tolentino et al., 2016). According to a recent United Nations Office for Disaster Risk

Reduction report, floods were reported to be the most frequent disasters, where 3,148 occurrences accounted for 43.4% of total types of disasters from 1998 to 2017. The damages from these events were also found to be the largest at an estimated 2 billion lives affected, which accounted for 45% of the total impacts from all disasters (Wannous and Velasquez, 2017). Developing countries are particularly vulnerable to floods due to the lack of resources to prevent, mitigate, and adequately respond to floods (Adger et al., 2003; Douglas et al., 2008; Poortinga et al., 2017).

Given the occurrence and widespread damages due to floods, it becomes imperative to address flood risks from a disaster risk reduction (DRR) approach. In this regard, a variety of scientific approaches are being used for mapping flood hazard and informing risk assessments. Risk assessment is a key component in risk management and reduction in the broader sense. Disaster management aims to avoid or reduce potential risks from floods and assure immediate and appropriate response to flood events. Furthermore, effective disaster management enables rapid and effective recovery after a flood event. Four phases of disaster management have been used by governments, including mitigation, preparation, response, and recovery (Thieken et al., 2007; Carter, 2008). Implementing these four components in a disaster risk reduction approach is expected to increase resilience and reduce economic and human losses.

Flood risk is a function of spatio-temporal hazard of floods, exposure to floods, and vulnerability to floods (UNISDR, 2011). Flood hazard is defined by the spatial extent and temporal frequency of flood events themselves (Winsemius et al., 2013). Exposure is considered by the intersection of the hazard with the people and assets who may experience the hazard. Vulnerability refers to the susceptibility of those people and objects to potential loss and is defined by their intrinsic characteristics (Alexander, 2002; Plate, 2002). Flood risk is a combination of the magnitude and frequency of the hazard, along with the vulnerability of people and assets exposed to floods (Alexander, 2002). Hence, understanding of the geographic location and extents is an essential input into any flood risk assessment.

Winsemius et al. (2013) identified the probability density of flood hazard, socio-economic indicators, resilience, and adaptive capacity as main components of their flood risk framework. Reducing the probability of a flood hazard affecting populations is therefore a straightforward way to reduce flood risk. Probability of flood can be derived from historic records and flood forecasting systems (Carsell et al., 2004; Verkade and Werner, 2011). However, forecasting systems adequate for local and national disaster management often rely on complex models that require extensive inputs and computational power. Developing countries often lack the capacity, infrastructure and data to run such sophisticated models. Despite the complex nature of flood events, knowledge of the location and extent of floods is often concentrated in specific flood prone areas. Thorough analysis of historical data is therefore crucial in complementing existing knowledge to better identify flood prone areas.

It has been recognized that socio-economic and vulnerability data are crucial components in disaster risk reduction (Gornitz, 1991). Studies of e.g., Abuodha and Woodroffe (2006) and Boruff et al. (2005) include examples of vulnerability analyses that include statistical data on education, family structure, and social dependence in a robust and consistent manner. Such vulnerability indices are useful in distinguishing the relative vulnerabilities of different areas to disasters (Balica et al., 2012). Recognition of the spatio-temporal dimensions in the local context are important in determining the degree of flood exposure and vulnerability. Geographic Information Systems (GIS) are useful in managing and analyzing data from different sources to map and understand the spatio-temporal dynamics of flood risk.

Recent advances in the field of EO have resulted in technologies and products that make data more easily accessible for non-experts. Examples are cloud-based platform for planetary-scale environmental data analysis (Gorelick et al., 2017) that allow for the development of real-time applications to monitor environmental conditions (Simons et al., 2017; Markert et al., 2018b; Poortinga et al., 2018, 2019). Moreover, there are a large variety of readily available EO derived global products on surface water extent (Pekel et al., 2016), rainfall (Funk et al., 2015), surface elevation (Farr et al., 2007; Tadono et al., 2016), and others. These data can be used directly without experience in processing raw satellite data. These products are a great resource in a wide variety of disciplines including hazard and risk mapping.

In this study we present an innovative approach to flood risk mapping in a disaster risk reduction framework, leveraging open data and state-of-the-art cloud computing technologies. We present a framework to map spatially explicit flood hazard, exposure and vulnerability, and to merge those data into a single flood risk index (FRI). The study is presented in the context of Myanmar, a developing nation that faces many challenges in the field of disaster risk reduction. The work conducted was under the auspices of the SERVIR-Mekong project, which is a collaborative venture between the US National Aeronautics and Space Administration (NASA) and the US Agency for International Development. Given the mandate of serving the Lower Mekong countries, SERVIR-Mekong presently addresses the needs of Cambodia, Laos, Myanmar, Thailand and Vietnam. Driven by user-based needs, this program responds to local issues via provision of cutting edge EO, science and associated technologies as solutions to development challenges. Key among these needs are requirements for addressing various aspects of floods, which includes hazard and risk mapping, monitoring, and forecasting, via provision of publicly available EO within user-defined web based applications. Toward addressing flood hazard mapping, SERVIR-Mekong has developed a Historical Flood Analysis (HFA) product which was modified to meet the flood hazard mapping requirement by the Department of Disaster Management (DDM), Myanmar. This product is based on open source approach and is available here: https://hfa.adpc.net/en/.

2. MATERIALS AND METHODS

2.1. Study Area

Myanmar is located in Southeast Asia and shares its borders with China, Lao People's Democratic Republic,Thailand, Bangladesh

and India (**Figure 1**). Myanmar's climate is classified as a tropical monsoon climate and experiences heavy rainfall events. It has high humidity and an annual average temperature ranging from 22 to 27°C. The monsoon gives variations in the water levels of the Ayeyarwaddy, Chindwin, Thanlwin, and Sittoung, which are the four main rivers in the country (Taft and Evers, 2016). These rivers are a vital source for drinking water, transportation, irrigation and generation of electricity. However, due to regular flooding events in the monsoon season, effective disaster management is a major concern which requires a systematic approach to address the impacts of these events. Toward meeting these aims, in the following sections, we provide details of an operational index developed for flood risk assessment for Myanmar.

2.2. Flood Risk Assessment

Flood risk is calculated from various sources of information including satellite-derived data products and area-aggregated statistical data. Most data layers are normalized so that data from different sources can be compared qualitatively. The final risk map is calculated using Equation (1), multiplying the flood hazard, flood exposure, and flood vulnerability indices. **Figure 2** shows that different data layers are used to calculate the hazard, exposure, and vulnerability and how those layers are combined into the final flood risk index. The flood risk index (FRI) is calculated from the risk map, in which risk values are aggregated over a township. We describe how the different indices are calculated in the following sections.

$$FRI = FHI \times FEI \times FVI \tag{1}$$

where:

FHI = Flood frequency (-)
FEI = Exposure (-)
FVI = Vulnerability (-).

2.3. Flood Hazard Index

Historical trends in flooding are important for understanding the current risk and what might happen in the future (Klis et al., 2005). As field data are scattered and difficult to obtain in Myanmar, we used remote sensing derived products to estimate flood hazard from historical data. The JRC global water dataset was used to generate flood frequency maps across Myanmar. The JRC team developed a method to calculate water pixels from Landsat satellite imagery. The imagery is going through a sequence of steps where they detect water while accounting for false positives including shadow effects. The JRC Monthly Water History (V1) was used in this study. The dataset contains monthly layers of the location and temporal distribution of surface water from 1984 to 2015. The data contains information on (0) no data, (1) not water and (2) water. The flood frequency for any given period is calculated by dividing the number of water observation by total number of observation where no data is not taken into account. We used all available Landsat data in the JRC tool as historical occurrence contain valuable information on the probability of occurrence. As the data-series contains monthly layers, different time-slices such as months or seasons could also be investigated.

Permanent water was removed from the data in order to only include flood events. The United Nations Institute for Training and Research (UNITAR) provides data on natural disasters through the Operational Satellite Applications Programme (UNOSAT) including flood maps. The UNOSAT flood map is a well recognized data source and was used to distinguish permanent from temporary water. We used the 2015 data for the comparison. **Figure 3** shows both our water occurrence map and the UNOSAT permanent water data. It can be seen that pixels with high water occurrence values are marked as permanent water in the UNOSAT data. By cross-walking the data, we found 82% a suitable threshold to distinguish permanent from temporary water in our study area. Permanent waters were masked out to distinguish them from flooded areas.

2.4. Flood Exposure Index

Exposure to floods is defined by the assets and values located in flood-prone areas (Jongman et al., 2014). We separated between assets that are under direct threat from flood (i.e., land use) and the distance from assets that offer potential relief to people exposed to floods. In the category of assets we used a landcover map containing information on urban, cropland and rice. We consider these classes important for supporting livelihood which could be negatively impacted by floods. In the other category we calculated distance from hospitals, schools and roads as they provide shelter to the people in case of flood emergencies. The population density was taken into account as the most important factor. The calculation for exposure is shown in Equation (2) and the data sources are shown in **Table 1**. The different data-layers are described in the next sections.

$$FEI = W_p \frac{C + 0.5R + U + S_d + H_d + R_d}{5.5} \tag{2}$$

where:

FEI = Flood Exposure Index (-)
W_p = Population data (-)
C = Cropland (-)
R = Rice (-)
U = Urban (-)
S_d = School distance (-)
H_d = Hospital distance (-)
R_d = Road distance (-).

2.4.1. Land Cover

Buildings and agricultural lands are directly affected by floods due to loss of property and means of production. Whereas floods have played an important role in traditional agricultural systems (Van Liere, 1980), they have also caused severe damage to the major crops and threatening the food security of large regions (Del Ninno et al., 2003). We used the SERVIR-Mekong cropland, rice and urban probability layers from the regional land cover monitoring system (https://rlcms-servir.adpc.net/en/). These yearly maps were created from the Landsat legacy

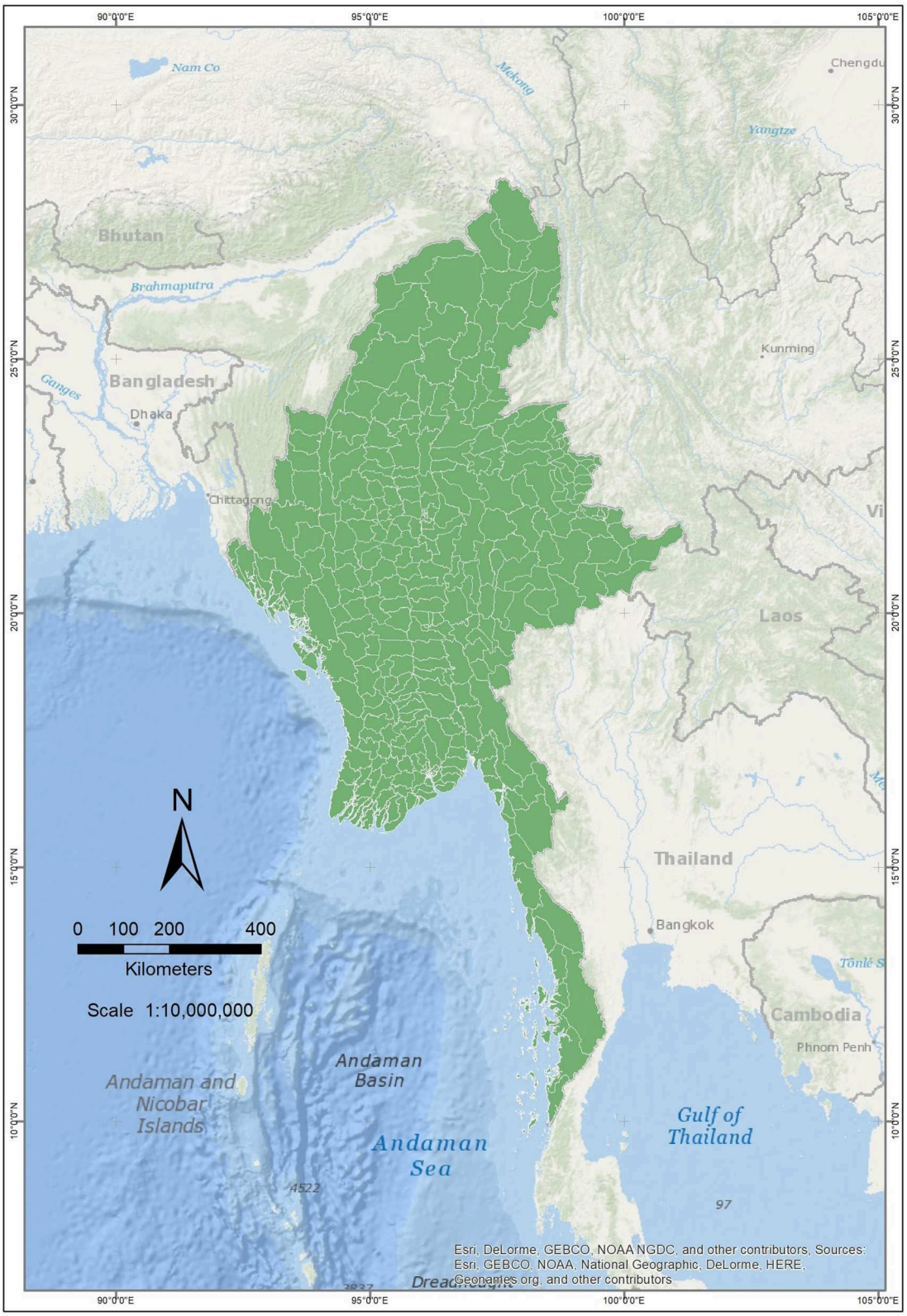


FIGURE 1 | Political map of Myanmar/Southeast Asia.

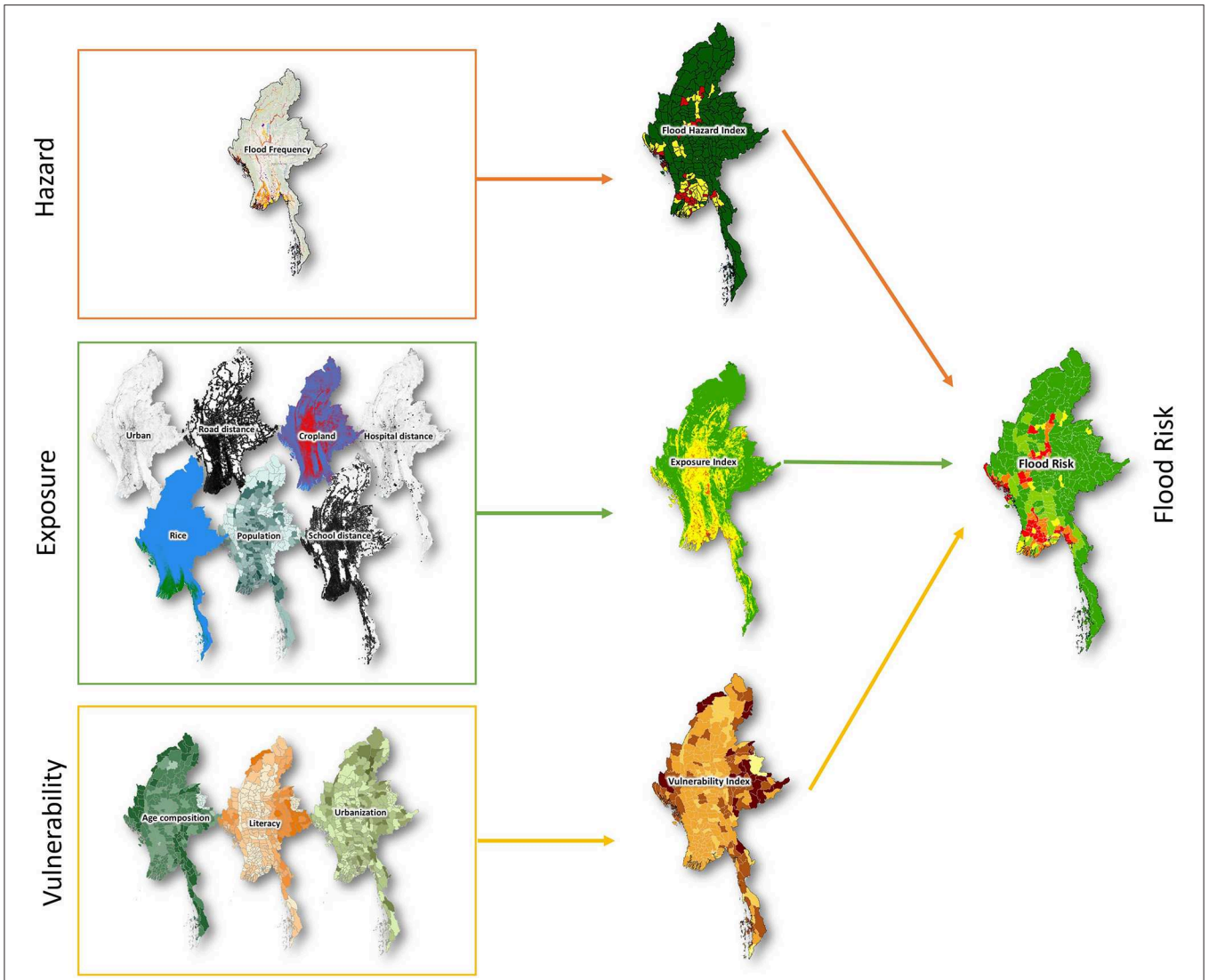


FIGURE 2 | Schematic overview of how the indices on Hazard, Exposure, and Vulnerability are combined into the flood risk index.

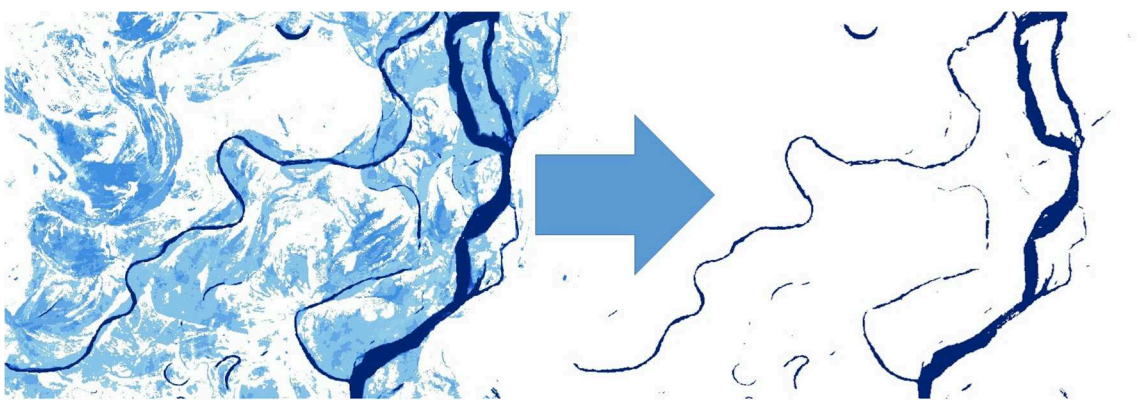


FIGURE 3 | The water occurrence map calculated from the JRC surface water data (left) and UNOSAT permanent water map data on the right.

TABLE 1 | Data sources for the Flood Exposure Index (FEI; Equation 2).

Layer	Data sources
Population	World population data (Stevens et al., 2015)
Cropland	SERVIR-Mekong Land cover data (Saah et al., 2020)
Rice	SERVIR-Mekong Land cover data (Saah et al., 2020)
Urban	SERVIR-Mekong Land cover data (Saah et al., 2020)
School	Myanmar Information Management Unit (MIMU, 2019)
Hospitals	Open Street Map (OpenStreetMap Contributors, 2017)
Roads	Open Street Map (OpenStreetMap Contributors, 2017)

archive using reference data from the field and collected through high resolutions satellite imagery (Saah et al., 2019) in a machine learning algorithm. The probability layers were scaled between 0 and 1. The rice map was multiplied by a 0.5 fraction as the negative consequences of floods on rice are generally less severe in comparison to other land cover types.

2.4.2. Infrastructure

Infrastructure is a key component in disaster risk reduction as failure or capacity degradation directly affects the community (Luathep et al., 2013). We selected hospitals, schools and roads as the key components for the risk assessment. Roads are the lifeline for communities of goods and services and enable people to evacuate from the area in case of emergency. Hospitals also provide direct relief to crisis situations for affected people in the neighborhood. Schools play an important role in raising awareness among students, teachers, and parents (Strike, 2000). The school can also serve as a shelter when the integrity of the building compromised during the flood. We used the Open Street Map data vector data and calculated the shortest distance to a feature on a 30 by 30 meter grid. The maximum distance was set to 10 km. These data were normalized to values between 0 and 1.

2.4.3. Population

Inundation of densely populated areas poses many threats including the loss of human lives and property as well as the spread of infectious diseases (Levy et al., 2016). Population exposure was calculated with the Worldpop dataset. We used the 2015 national totals adjusted to match UN population division estimates (Stevens et al., 2015). The product was created using a random forest regression tree-based mapping approach integrating census and a wide range of remotely-sensed and geospatial datasets.

2.5. Flood Vulnerability Index

Vulnerability reduction and increasing resilience are key components in disaster risk reduction. A first step is to analyze the current vulnerability of a community, township or region to floods. However, there are a wide variety of definitions of vulnerability including a variety of different indicators (e.g., Cannon, 1994; Pelling et al., 2004; Borden et al., 2007). We include three main socio-economic indicators of literacy, age composition and urbanization, as data on other indicators were found to be scarce. These data were collected from (http://www.dop.gov.mm/en) and contain data on the township level. All data were scaled between 0 and 1 using the maximum value. This approach was suited to be appropriate as it reflects the relative vulnerability on a country level. Higher literacy was considered to decrease vulnerability. In the age composition young and old people were considered to be more vulnerable. Rural people were also considered to be more vulnerable than people in the city. The composition of these vulnerabilities are formulated in an equation shown below (Equation 3).

$$FVI = \frac{A + U + L}{3} \tag{3}$$

where:

FVI = Vulnerability Index (-)
A = Age composition (-)
U = Urbanization (-)
L = Literacy (-).

2.6. Flood Risk Index

The flood risk data were aggregated based on administrative boundaries, which was used for risk index calculations. Data aggregation on the administrative level was done to align it with the scales of social-economic including information on social vulnerability, environmental vulnerability and capacity. Another advantage of data aggregation is the reduction of uncertainty related to individual pixels.

We aggregated the pixel based maps on flood hazard and exposure into administrative boundaries (with township boundary representing a particular flood hazard index) by summing up all the pixel values and then divided this by the township area. The FRI Equation (1) and figure below describe the calculation of the flood hazard index for individual Township. Flood risk per pixel is represented by F_i, N is the number of pixels per township and A the area of Township (km^2). A graphical presentation is shown in **Figure 4**, where two townships are compared. It can be seen that township A has a higher FHI than township B.

$$FHI = \frac{\sum_{i=1}^{n} F_i N_i}{A} \tag{4}$$

2.7. Computational Framework

The workflow was built using Google Earth Engine (GEE). This is an online platform that applies cloud computing and storage frameworks to allow for parallel calculations of large geospatial datasets (Gorelick et al., 2017). The archive contains a large amount of Earth observations data such as the JRC global surface water product. Detailed information on the GEE can be found in the website (https://earthengine.google.com/). The GEE has a Python Application Programming Interface (API) which can be used to develop web applications. The web application provides an interface to the data while calculations and visualizations are done in real time. All data layers were ingested into the GEE which enables users to investigate the separate risk components and apply different weighing factors in the flood risk analysis.

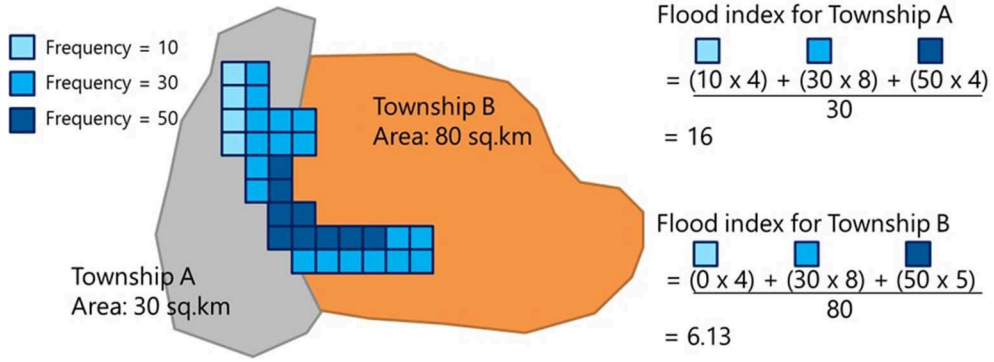


FIGURE 4 | Example of flood hazard index calculation for individual Townships.

3. RESULTS

The flood hazard map containing the flood frequencies for the period 1984–2015 is shown in **Figure 5**. Areas with high flood occurrence can be found near the Irrawaddy river and in the delta and near the coast. The central part of the country also shows areas with historical floods. It can be noted that recently constructed dams are visible in the map.

The exposure map is shown in **Figure 6**. It can be seen that the vegetated highlands have a low exposure while the higher exposure can be found in the agricultural and population centers. The coastal areas of the Rakhine State, located at the western coast also show high exposure rates. Naypyitaw is the modern capital of and Yangon, the former capital also show high exposure risk. We also found higher exposure risk along the Irrawaddy river. This river is one of the least regulated rivers in Asia (Hedley et al., 2010; Taft and Evers, 2016) and known for its dynamic morphology.

The flood vulnerability map composed from age composition, literacy, and urbanization, is shown in **Figure 7**. The map shows the risk in different shades of brown and districts with no information are shown in white. In contrast to the flood frequency and exposure map, the vulnerability map is calculated on a district level. Shan State in the eastern part of the country bordering China and Thailand shows a particular high vulnerability. But also districts in western Myanmar show high vulnerability. The lowlands show a generally low to medium level of vulnerability.

The flood risk map for the country on a pixel level is shown in **Figure 8**. It can be seen that the Irrawaddy delta and coastal regions in western Myanmar have the highest risk. But also riverine areas throughout the country show higher risk. The high risk in central Myanmar is particularly notable. It can also be seen that the eastern and western provinces, which have vulnerability show a low risks because of the low hazard and exposure rates.

We calculated the FHI for all townships in Myanmar and normalized the values between 0 an 100 using the minimum and maximum values. We then created three categories of low, moderate and high to classify township in an easy to understand manner. Classification was done by means of data exploration and expert knowledge. The threshold classification are <5 percent, 5–10 percent and more than 10 percent. Those percentage represents low, moderate and high classification respectively. The resulting map is shown in **Figure 9**. It can be seen that districts with a high risk are concentrated in the deltas and central parts of the country. Population centers such as Yangon and Mandalay were found to have a high risk whereas Naypyitaw has a low risk.

4. DISCUSSION

There is very limited detailed knowledge on river basins and water dynamics in Myanmar (Taft and Evers, 2016). Although at a relatively coarse resolution, this study and the tool provides valuable information on water dynamics and potential human impacts on a district level. Applications of the tool were identified from various organizations and departments in Myanmar. They included the Department of Urban and Housing Development (DUHD) can use the flood frequency results as one of their consideration on selecting construction site locations to avoid frequently flooded areas. The Directorate of Water Resources and Improvement of River Systems (DWIR) may potentially use the flood hazard index to identify townships along rivers where the level of hazard is high to consider for constructing of the mitigation infrastructure. The Department of Meteorology and Hydrology (DMH) are interested to use flood frequency maps for selecting locations of new meteorological stations. The Department of Disaster Management (DDM) wants to use the flood hazard index to adjust the amount of the relief items for the stockpiling. The threshold value of flood hazard index is applicable only for Myanmar. Different countries may need to re-adjust the threshold with the expert consultation.

A limitation of this study it that we have not performed any validation of the data to quantify the uncertainty and errors. This could not be done as there very limited field data

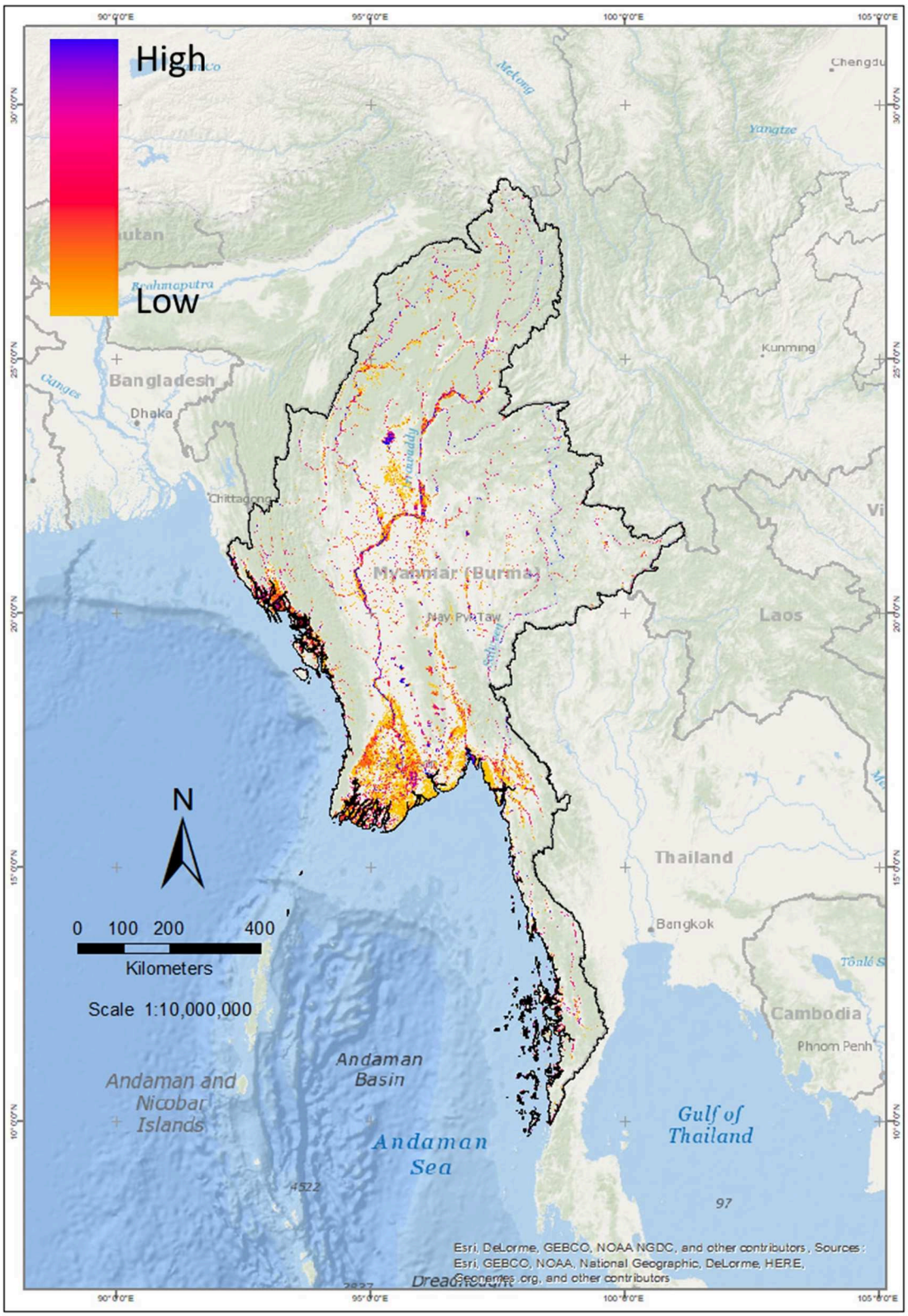


FIGURE 5 | Flood frequency map of Myanmar calculated from the JRC surface water data.

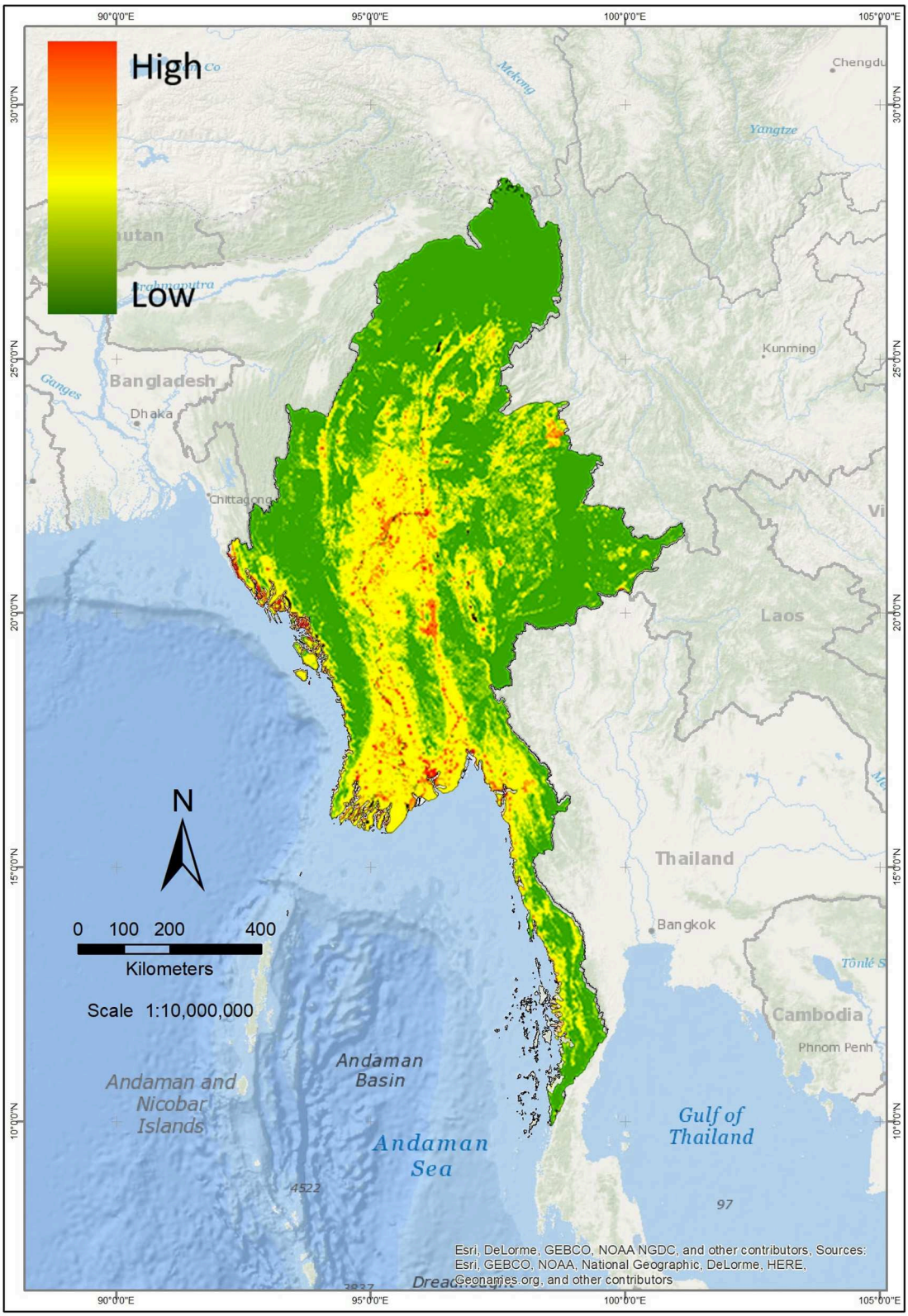


FIGURE 6 | Flood exposure map of Myanmar. Flood exposure includes land cover, infrastructural, and population data.

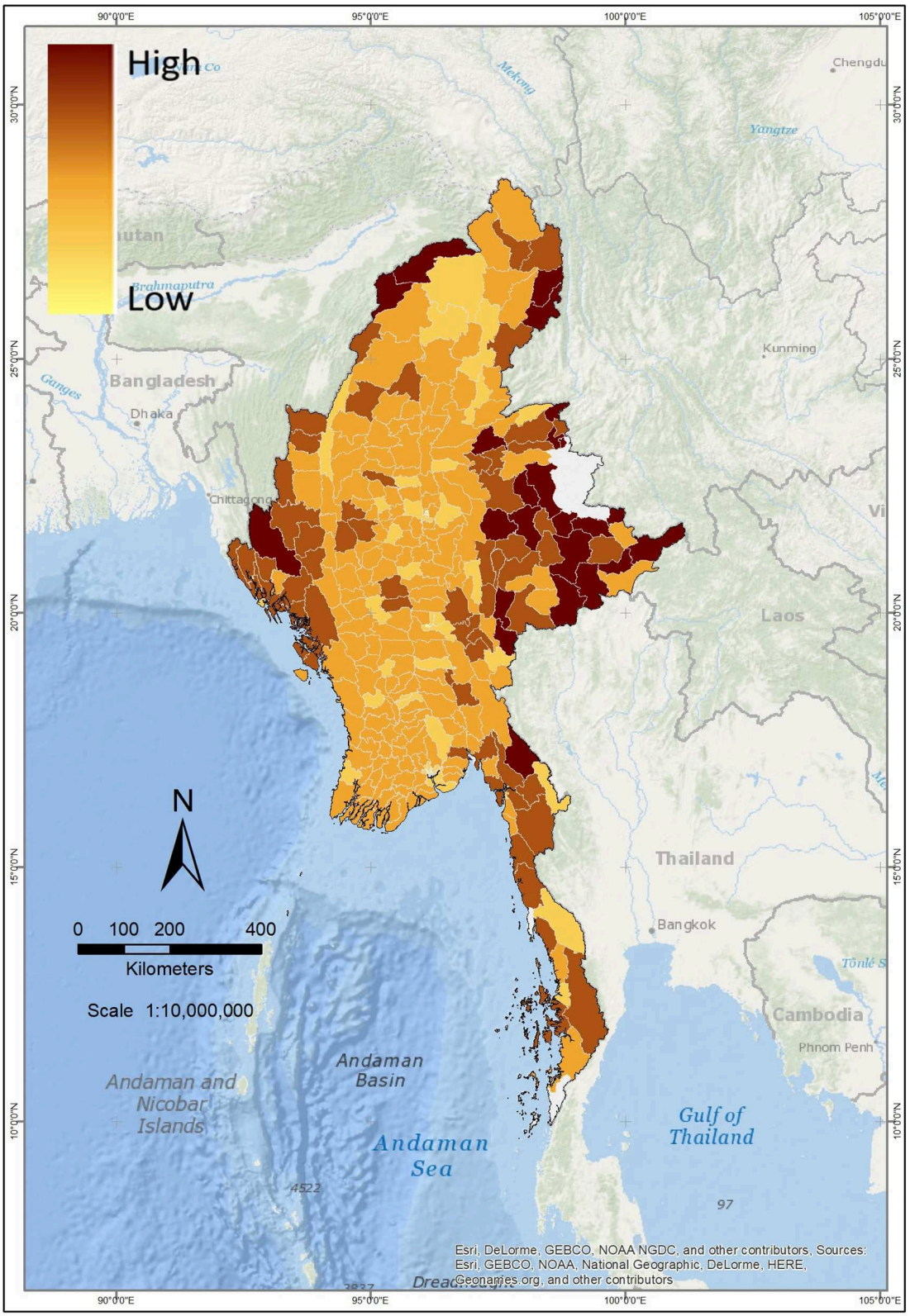


FIGURE 7 | Flood vulnerability on a township level for Myanmar. Vulnerability was calculated from literacy, population composition, and urbanization.

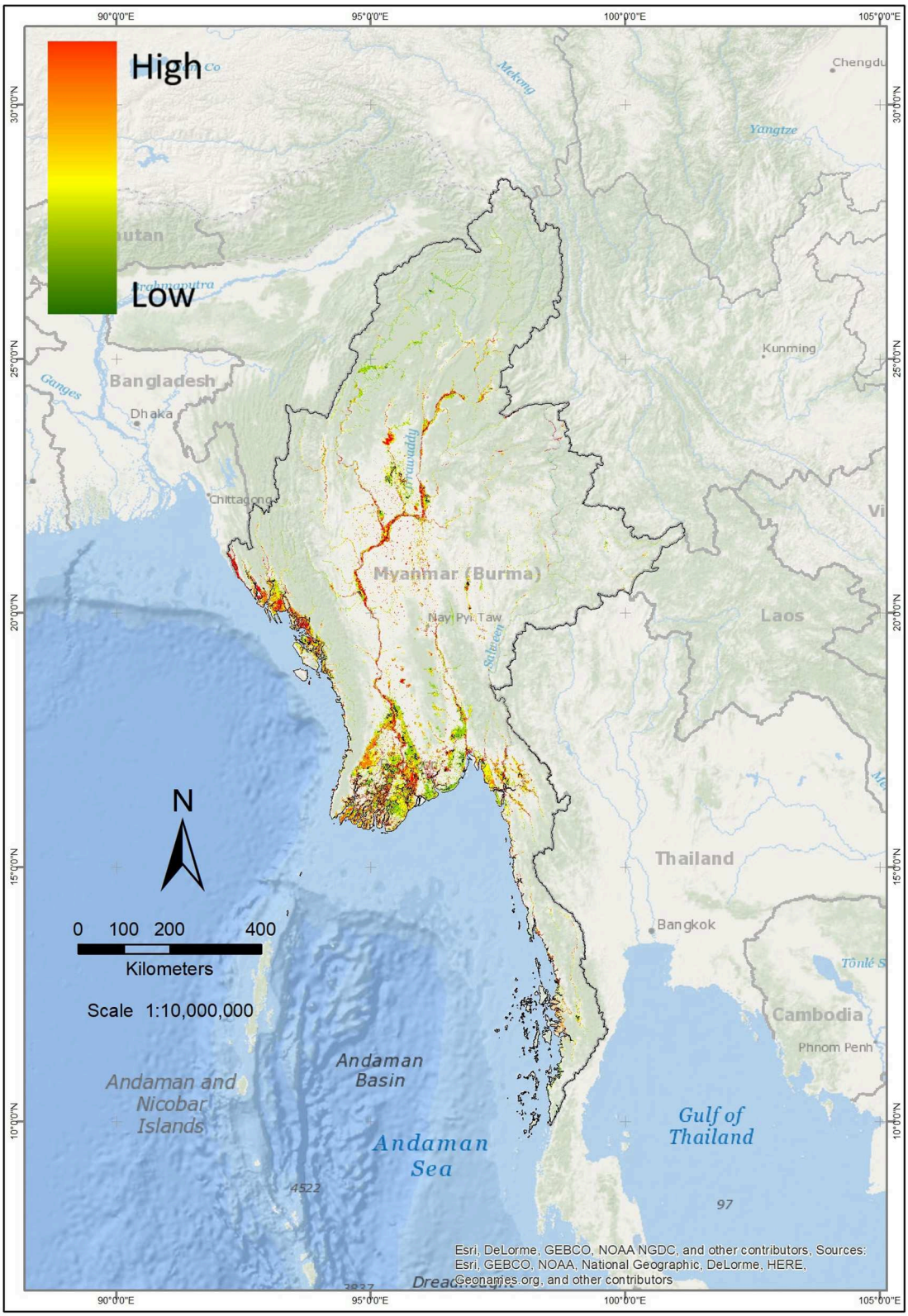


FIGURE 8 | Flood risk at the country-wide scale for Myanmar.

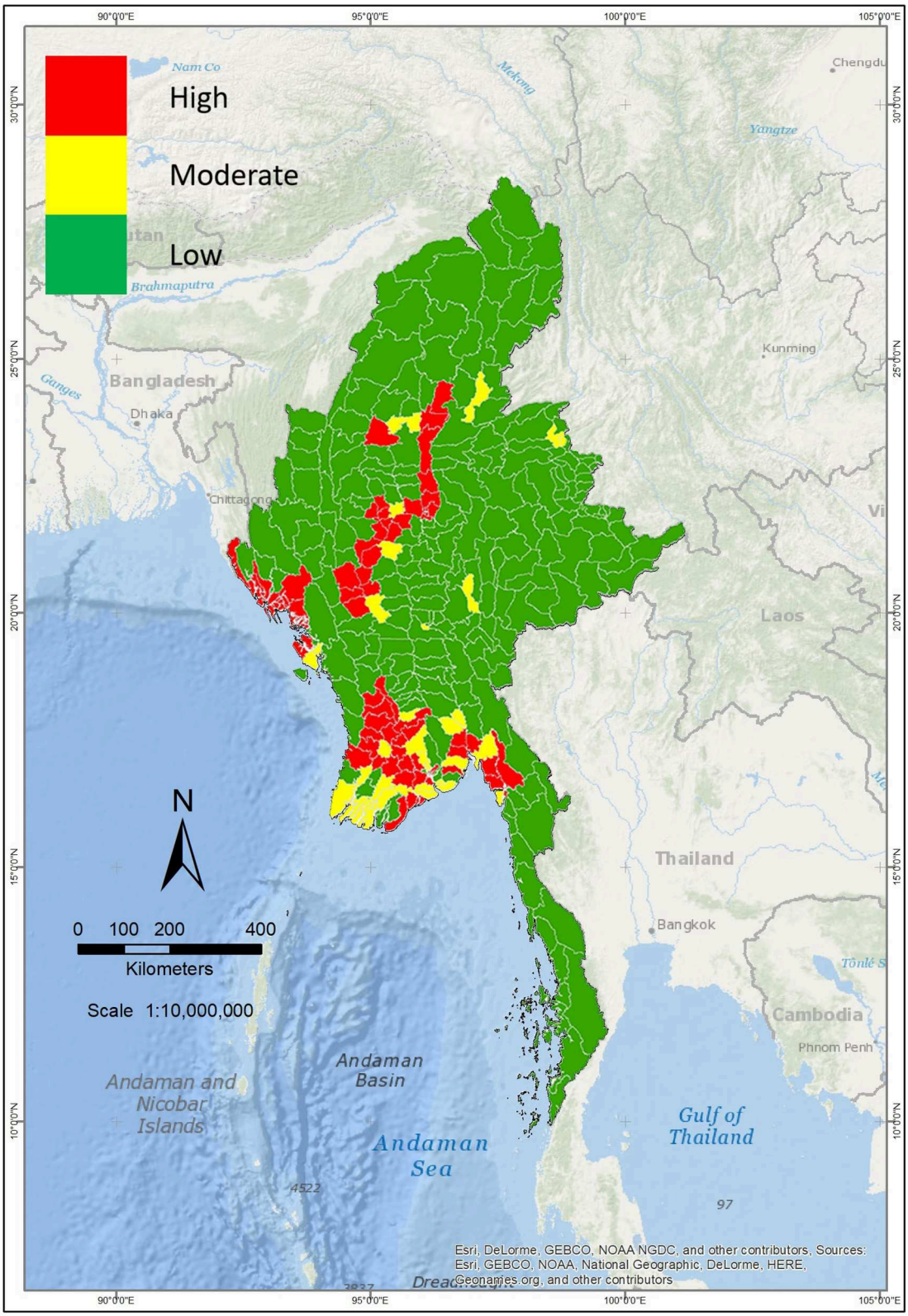


FIGURE 9 | Flood risk index at a township level in Myanmar.

available from Myanmar. This was also identified by previous studies of for example Salmivaara et al. (2013), Varis et al. (2012), and Taft and Evers (2016). Remote sensing can fill an important data gaps, but has limitations. For example, Landsat has a revisit schedule of 16 days, and does not capture surface data in cloudy conditions. Hence, some historical flood events might not be captured, or the maximum extents of historical floods may be underrepresented. The MODIS archive could add valuable information to the system as it contains daily imagery for the period 2000–2018. With the recent availability of the European Space Agency (ESA) Copernicus Sentinel-1a/b satellites that acquire synthetic aperture radar (SAR) data, observations of flood extent during cloudy conditions will improve. The forthcoming NASA-ISRO SAR (NISAR) mission will acquire additional SAR data. Radarsat-1 archived data and ALOS PALSAR-1 data have also recently become available and could add to the flood record. Furthermore, including data fusion methods that incorporate multiple sensors (both optical and SAR data) can potentially increase temporal resolution across cloudy regions (Markert et al., 2018a), thus improving surface water observations which in turn can improve the flood frequency analysis.

In this work, we have developed a novel approach to distinguish surface water from floods. However, the JRC tool Pekel et al. (2016) does not separate water in paddy from other floods; therefore, a process to distinguish standing water in paddy field from floods that impact populations, infrastructure, and other land use types, would enhance the FHI. Additionally, the current FHI data derived from the JRC surface water data is only available from 1984 to 2015. Further improvements to this approach are planned including the update of the JRC tool to the present date for the Lower Mekong Region and beyond. Inputs to each of the indices were available at different spatial resolutions. The FHI relied on Landsat pixels at a 30m horizontal resolution. The FEI used 30 m land cover data and 100 m gridded population data. Inputs to the FVI were available at the township level, and the final FRI was produced at the township level.

The method to calculate FHI could consider other parameters that influence floods (e.g., rainfall intensity, elevation, land use, distance from drainage network). Other approaches could capture more sudden, fleeting flash floods that are unlikely to appear in the Landsat-based observations. Therefore, the final FRI produced in this study likely represents riverine and coastal floods, and not flash or finer scale urban floods.Future studies could integrate potential damage assessments as inferred from flood depth and duration as well (Oddo et al., 2018). While GIS allows for overlaying and combination of disparate data layers, it does not solve the decision-making processes involved in assigning weights or significance to quantitative and qualitative data. Roy (2018) discusses multi-criteria analysis (MCA) approach in applying geospatial indices toward reducing flood risk in India, the lessons of which could be applied to risk mapping and related decisions in Myanmar.

New cloud based geo-computational technologies make it easy to develop online tools that perform real time calculation. This enables end users with limited knowledge in the field of remote sensing and GIS to perform analysis which were beyond their reach before. This is particularly useful in the context of developing countries and disaster risk reduction. The end-user will be able to select and prioritize variables of the model and assign different weight to them. Results can be calculated and displayed in real time in a spatial explicit manner.

5. CONCLUSIONS

For the purposes of incorporating EO into planning and risk reduction, it is important to distinguish between flood and surface water, and our approach achieves this objective for the DDM in Myanmar. We have applied an index-based approach to reclassify historic flood frequency into a flood hazard index. We have also created geospatial exposure and vulnerability indices using socioeconomic and land use data. Combined, this resulted in a nationwide flood risk map, summarized at the township scale for Myanmar as depicted in **Figure 9**. Such knowledge aids in a more objective and complete understanding of historic flood patterns, which may inform annual budget decisions on the pre-allocation of flood relief supplies before each monsoon season. Further, this new characterization of overall flood risk by township pinpoints critical areas for additional disaster risk reduction investments.

Further refinement of the present work are planned in order to address the limitations outlined in section 3.1 above. This includes an update of the JRC dataset including data from additional sensors such as Sentinel-2, inclusion of other layers of locally available exposure and vulnerability data (including different types of social and socio-economic data) as are made available in the collaboration with DDM and other participating departments in Myanmar.

The risk map developed in this work and in collaboration with DDM is now being considered by other relevant departments including the Road Transportation Administration Department (RTAD) in Myanmar. With the feedback from end-user's, we expect to refine and build upon methods for improved decision support and planning for flood hazards.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

KP, AP, BB, DS, and FC contributed to the conception and design of the study. KP organized the data collection. KP, AP, BB, and FC

performed the statistical analysis. KP, AP, and FC wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank the support and feedback from Myanmar's Department of Disaster Management (DDM) on the tool and analysis. A special thanks is given to the Google Earth Engine team for their support and use of the platform. Support for this work was provided through the joint US Agency for International Development (USAID) and National Aeronautics and Space Administration (NASA) initiative SERVIR-Mekong, Cooperative Agreement Number: AID-486-A-14-00002. Individuals affiliated with the Earth System Science Center, University of Alabama in Huntsville (UAH) are funded through the NASA Applied Sciences Capacity Building Program, NASA Cooperative Agreement: NNM11AA01A.

REFERENCES

Abuodha, P. A. O., and Woodroffe, C. D. (2006). *Assessing Vulnerability of Coast to Climate Change: A Review of Approaches and Their Application to the Australian Coast*. University of Wollongong, Faculty of Science Papers.

Adger, W. N., Huq, S., Brown, K., Conway, D., and Hulme, M. (2003). Adaptation to climate change in the developing world. *Prog. Dev. Stud.* 3, 179–195. doi: 10.1191/1464993403ps060oa

Alexander, D. (2002). *Principles of Emergency Planning and Management*. London: Oxford University Press.

Balica, S. F., Wright, N. G., and van der Meulen, F. (2012). A flood vulnerability index for coastal cities and its use in assessing climate change impacts. *Nat. Hazards* 64, 73–105. doi: 10.1007/s11069-012-0234-1

Borden, K. A., Schmidtlein, M. C., Emrich, C. T., Piegorsch, W. W., and Cutter, S. L. (2007). Vulnerability of US cities to environmental hazards. *J. Homel. Secur. Emerg. Manage.* 4, 1–21. doi: 10.2202/1547-7355.1279

Boruff, B. J., Emrich, C., and Cutter, S. L. (2005). Erosion hazard vulnerability of us coastal counties. *J. Coastal Res.* 21, 932–942. doi: 10.2112/04-0172.1

Cannon, T. (1994). Vulnerability analysis and the explanation of 'natural' disasters. *Disast. Dev. Environ.* 1, 13–30.

Carsell, K. M., Pingel, N. D., and Ford, D. T. (2004). Quantifying the benefit of a flood warning system. *Nat. Hazards Rev.* 5, 131–140. doi: 10.1061/(ASCE)1527-6988(2004)5:3(131)

Carter, W. N. (2008). *Disaster Management: A Disaster Manager's Handbook*. Manila: Asian Development Bank. Available online at: http://hdl.handle.net/11540/5035

Del Ninno, C., Dorosh, P. A., and Smith, L. C. (2003). Public policy, markets and household coping strategies in bangladesh: Avoiding a food security crisis following the 1998 floods. *World Dev.* 31, 1221–1238. doi: 10.1016/S0305-750X(03)00071-8

Douglas, I., Alam, K., Maghenda, M., Mcdonnell, Y., McLean, L., and Campbell, J. (2008). Unjust waters: climate change, flooding and the urban poor in africa. *Environ. Urban.* 20, 187–205. doi: 10.1177/0956247808089156

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007). The shuttle radar topography mission. *Rev. Geophys.* 45, 1–13. doi: 10.1029/2005RG000183

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., et al. (2015). The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes. *Sci. Data* 2:150066. doi: 10.1038/sdata.2015.66

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. doi: 10.1016/j.rse.2017.06.031

Gornitz, V. (1991). Global coastal hazards from future sea level rise. *Global Planet. Change* 3, 379–398. doi: 10.1016/0921-8181(91)90118-G

Hedley, P. J., Bird, M. I., and Robinson, R. A. (2010). Evolution of the irrawaddy delta region since 1850. *Geograph. J.* 176, 138–149. doi: 10.1111/j.1475-4959.2009.00346.x

Jongman, B., Koks, E. E., Husby, T. G., and Ward, P. J. (2014). Increasing flood exposure in the Netherlands: implications for risk financing. *Nat. Hazards Earth Syst. Sci.* 14, 1245–1255. doi: 10.5194/nhess-14-1245-2014

Klis, H., Baan, P. J. A., and Asselman, N. E. M. (2005). *Historische analyse van de gevolgen van overstromingen in Nederland: een globale schatting van de situatie rond 1950, 1975 en 2005*. Q4005.

Levy, K., Woster, A. P., Goldstein, R. S., and Carlton, E. J. (2016). Untangling the impacts of climate change on waterborne diseases: a systematic review of relationships between diarrheal diseases and temperature, rainfall, flooding, and drought. *Environ. Sci. Technol.* 50, 4905–4922. doi: 10.1021/acs.est.5b06186

Luathep, P., Suwanno, P., and Taneerananon, P. (2013). "Identification of critical locations in road networks due to disasters," in *Proceedings of the Eastern Asia Society for Transportation Studies*, Vol. 9 (Hat Yai), 206.

Markert, K., Chishtie, F., Anderson, E. R., Saah, D., and Griffin, R. E. (2018a). On the merging of optical and sar satellite imagery for surface water mapping applications. *Results Phys.* 9, 275–277. doi: 10.1016/j.rinp.2018.02.054

Markert, K., Schmidt, C., Griffin, R., Flores, A., Poortinga, A., Saah, D., et al. (2018b). Historical and operational monitoring of surface sediments in the lower mekong basin using landsat and google earth engine cloud computing. *Remote Sens.* 10:909. doi: 10.3390/rs10060909

Myanmar Information Management Unit (MIMU). *Formal Sector School Location Upper Myanmar* (2019). Available online at: http://geonode.themimu.info/ (accessed May 20, 2019).

Oddo, P. C., Ahamed, A., and Bolten, J. D. (2018). Socioeconomic impact evaluation for near real-time flood detection in the lower mekong river basin. *Hydrology* 5:23. doi: 10.3390/hydrology5020023

OpenStreetMap Contributors (2017). *Planet Dump*. Retrieved from: https://planet.osm.org

Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature* 540:418. doi: 10.1038/nature20584

Pelling, M., Maskrey, A., Ruiz, P., Hall, P., Peduzzi, P., Dao, Q. H., et al. (2004). *Reducing Disaster Risk: A Challenge for Development*. New York, NY: United Nations Development Programme, Bureau for Crisis Prevention and Recovery.

Plate, E. J. (2002). Flood risk and flood management. *J. Hydrol.* 267, 2–11. doi: 10.1016/S0022-1694(02)00135-X

Poortinga, A., Bastiaanssen, W., Simons, G., Saah, D., Senay, G., Fenn, M., et al. (2017). A self-calibrating runoff and streamflow remote sensing model for ungauged basins using open-access earth observation data. *Remote Sens.* 9:86. doi: 10.3390/rs9010086

Poortinga, A., Clinton, N., Saah, D., Cutter, P., Chishtie, F., Markert, K. N., et al. (2018). An operational before-after-control-impact (baci) designed platform for vegetation monitoring at planetary scale. *Remote Sens.* 10:760. doi: 10.3390/rs10050760

Poortinga, A., Tenneson, K., Shapiro, A., Nquyen, Q., San Aung, K., Chishtie, F., et al. (2019). Mapping plantations in Myanmar by fusing landsat-8, sentinel-2 and sentinel-1 data along with systematic error quantification. *Remote Sens.* 11:831. doi: 10.3390/rs11070831

Roy, P. S. (2018). Flood risk assessment using multi-criteria analysis: a case study from kopili river basin, assam, india. *Geomat. Nat. Hazards Risk* 9, 79–93. doi: 10.1080/19475705.2017.1408705

Saah, D., Johnson, G., Ashmall, B., Tondapu, G., Tenneson, K., Patterson, M., et al. (2019). Collect earth: an online tool for systematic reference data collection in land cover and use applications. *Environ. Modell. Softw.* 118, 166–171. doi: 10.1016/j.envsoft.2019.05.004

Saah, D., Tenneson, K., Poortinga, A., Nguyen, Q., Chishtie, F., Aung, K. S., et al. (2020). Primitives as building blocks for constructing land cover maps. *Int. J. Appl. Earth Observ. Geoinform.* 85:101979. doi: 10.1016/j.jag.2019.101979

Salmivaara, A., Kummu, M., Keskinen, M., and Varis, O. (2013). Using global datasets to create environmental profiles for data-poor regions: a case from the Irrawaddy and Salween River basins. *Environ. Manage.* 51, 897–911. doi: 10.1007/s00267-013-0016-x

Simons, G., Poortinga, A., Bastiaanssen, W. G., Saah, D., Troy, D., Hunink, J., et al. (2017). *On Spatially Distributed Hydrological Ecosystem Services: Bridging the Quantitative Information Gap Using Remote Sensing and Hydrological Models*. Wageningen: FutureWater.

Stevens, F. R., Gaughan, A. E., Linard, C., and Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS ONE* 10:e0107042. doi: 10.1371/journal.pone.0107042

Strike, K. A. (2000). Schools as communities: four metaphors, three models, and a dilemma or two. *J. Philos. Educ.* 34, 617–642. doi: 10.1111/1467-9752.00198

Tadono, T., Nagai, H., Ishida, H., Oda, F., Naito, S., Minakawa, K., et al. (2016). "Generation of the 30 m-mesh global digital surface model by alos prism," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Prague), 41. doi: 10.5194/isprs-archives-XLI-B4-157-2016

Taft, L., and Evers, M. (2016). A review of current and possible future human–water dynamics in Myanmar's river basins. *Hydrol. Earth Syst. Sci.* 20, 4913–4928. doi: 10.5194/hess-20-4913-2016

Thieken, A. H., Kreibich, H., Müller, M., and Merz, B. (2007). Coping with floods: preparedness, response and recovery of flood-affected residents in germany in 2002. *Hydrol. Sci. J.* 52, 1016–1037. doi: 10.1623/hysj.52.5.1016

Tolentino, P. L. M., Poortinga, A., Kanamaru, H., Keesstra, S., Maroulis, J., David, C. P. C., et al. (2016). Projected impact of climate change on hydrological regimes in the philippines. *PLoS ONE* 11:e0163941. doi: 10.1371/journal.pone.0163941

UNISDR (2011). *Global Assessment Report on Disaster Risk Reduction: Revealing Risk, Redefining Development.*

Van Liere, W. J. (1980). Traditional water management in the lower mekong basin. *World Archaeol.* 11, 265–280. doi: 10.1080/00438243.1980.9979766

Varis, O., Kummu, M., and Salmivaara, A. (2012). Ten major rivers in monsoon Asia-Pacific: an assessment of vulnerability. *Appl. Geogr.* 32, 441–454. doi: 10.1016/j.apgeog.2011.05.003

Verkade, J., and Werner, M. (2011). Estimating the benefits of single value and probability forecasting for flood warning. *Hydrol. Earth Syst. Sci.* 15, 3751–3765. doi: 10.5194/hess-15-3751-2011

Wannous, C., and Velasquez, G. (2017). "United Nations Office for Disaster Risk Reduction (UNISDR)-UNISDR's Contribution to Science and Technology for Disaster Risk Reduction and the Role of the International Consortium on Landslides (ICL)," in *WLF 2017:Advancing Culture of Living with Landslides*, eds K. Sassa, M. Mikoš, and Y. Yin (Cham: Springer), 99–115. doi: 10.1007/978-3-319-59469-9_6

Winsemius, H., Van Beek, L., Jongman, B., Ward, P., Bouwman, A. (2013). A framework for global river flood risk assessments. *Hydrol. Earth Syst. Sci.* 17, 1871–1892. doi: 10.5194/hess-17-1871-2013

Conflict of Interest: FC, AP, and DS are employed by Spatial Informatics Group, LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

TECHNOLOGY AND CODE
published: 11 December 2019
doi: 10.3389/fenvs.2019.00187

Application of MODIS NDVI for Monitoring Kenyan Rangelands Through a Web Based Decision Support Tool

Lilian Ndungu[1]*, *Maungu Oware*[1], *Steve Omondi*[1], *Anastasia Wahome*[1], *Robinson Mugo*[1] *and Emily Adams*[2]

[1] SERVIR-Eastern and Southern Africa, Regional Centre for Mapping Resource for Development, Nairobi, Kenya, [2] SERVIR Science Coordination Office, Huntsville, AL, United States

OPEN ACCESS

Edited by:
Niall Patrick Hanan,
New Mexico State University,
United States

Reviewed by:
Randall Boone,
Colorado State University,
United States
Julius Anchang,
New Mexico State University,
United States

***Correspondence:**
Lilian Ndungu
lndungu@rcmrd.org

Specialty section:
This article was submitted to Land Use Dynamics, a section of the journal Frontiers in Environmental Science

Received: 31 May 2019
Accepted: 11 November 2019
Published: 11 December 2019

Citation:
Ndungu L, Oware M, Omondi S, Wahome A, Mugo R and Adams E (2019) Application of MODIS NDVI for Monitoring Kenyan Rangelands Through a Web Based Decision Support Tool. Front. Environ. Sci. 7:187. doi: 10.3389/fenvs.2019.00187

The Kenyan rangelands contribute significantly to the country's GDP through livestock production and tourism. With dependence on rain-fed pastures, climate variability coupled with human induced factors such as overgrazing have adversely affected the rangeland ecosystems. And while indigenous communities and conservation experts already use their knowledge of the landscape to make decisions, this information is usually localized. Earth observation imagery provides a bigger picture that can complement indigenous knowledge and improve decision making. This research leverages on data from the MODIS receiver located at the Regional Centre for Mapping of Resources for Development (RCMRD) to develop the indices for the web-based Rangelands Decision Support Tool (RDST). The tool (RDST) automates data processing and provides an easy to use interface for accessing indices for rangeland monitoring. MODIS Normalized Difference Vegetation Index (NDVI), anomalies and deviation indices are provided on the tool at decadal, monthly, and seasonal time steps. Users begin their assessments by selecting their monitoring units and an NDVI index that responds to their specific questions. These questions respond to assessing current conditions, monitoring trends and changes in vegetation, and evaluating proxies for drought conditions. The information can then be overlaid with other ancillary datasets (roads, water sources, invasive species, protected areas, place names, conflict areas, migration routes), for context. At the click of a button, the information can be downloaded as a map for further analysis or application in sub regional decision making. Information and maps generated by this tool are being used decision making tool by rangeland managers in the counties and in other management units (conservancies and ranches). Specifically, inform adjustments to existing grazing plans, managing movement of livestock from designated grazing areas in wet and dry season, monitoring the success of rehabilitation efforts and resilience of the rangeland ecosystems, monitoring drought, managing scarce water resources, and monitoring the spread of invasive species. Successful implementation and application for decision making has relied heavily on local indigenous knowledge

and capacity building on use of the earth observation indices. The SERVIR project service planning engagement approach was used in engagements with stakeholders. This improved their participation in co-development of the tool and indices; and in adoption of the tools for decision making.

Keywords: ASALS, rangelands, NDVI, MODIS, vegetation indices, Kenya

INTRODUCTION

Rangelands make up 80% of Kenya's land mass which is classified as arid and semi-arid. The rangelands generate 90% of the tourism revenue, host 70% of the country's livestock, and are home to ∼10 million people. Approximately 75 to 85% of the population's livelihoods depends on rainfed pasture production (Government of Kenya, 2012; REGLAP, 2012).

The rangelands are under continuous threat due to climate variability, coupled with rapidly growing livestock and human populations. Extreme events are also increasing in intensity and frequency, resulting in notable decline in productivity in the rangelands due to shorter recovery time (Landmann and Dubovyk, 2014). Drought has been considered to be a major factor triggering land degradation processes especially in the rangelands. Studies have stressed the importance of drought episodes in explaining the occurrence of major degradation in some regions (Indeje et al., 2005; Vågen et al., 2014; Vicente-Serrano et al., 2015; Kyuma et al., 2016).

Degradation when not managed, leads to desertification and colonization of the degraded lands by invasive species. Between 1997 and 2000, severe degradation was estimated to have grown from 23 to 30% in Kenya (Mulinge et al., 2015). The increase in degradation in Kenya can be linked to a combination of factors such as unsustainable land use/land cover changes. Some of those, are caused by human factors such as felling of trees for charcoal and timber, overgrazing, exceeding the carrying capacity of the land and encroachment into protected areas (Mulinge et al., 2015).

Degraded lands in Kenya and especially in the North, when continually put into production, without restoration or other conservation measures, can become irreversibly unproductive, jeopardizing the livelihoods of millions of people who depend on these systems (Herrick et al., 2013; Vågen et al., 2014; Winowiecki et al., 2018). Current adaptation strategies that include shifts from purely pastoral to agro-pastoral systems have resulted in increased conflict as livestock and wildlife migration routes become closed due to changes in land ownership.

Despite these challenges, the role of the rangelands in contributing to food security in Kenya is set to grow with the changing dietary wants especially due to urbanization (Herrero et al., 2010; Silvestri et al., 2012). In developing countries, as economies grow, the dietary wants of the population are changing from cereal based to meat based, giving rise to the projected need for sustainable livestock systems to meet the needs of the population in what was termed as a "livestock revolution" (Delgado et al., 2001). For example, the total consumption of meat and milk is projected to grow by 119 and 74%, respectively by 2020 (Alexandratos and Bruinsma, 2015). The Kenyan rangelands have a diverse potential that is not fully utilized in livestock production and tourism, but also in renewable energy, natural resources, and resilient communities (Government of Kenya, 2012).

Recognizing the critical role that the rangeland plays in Kenya's economy, organized groups have been established to develop mechanisms for improving the health of the rangelands. This has led to growing investments in the rangeland's different actors. Private and donor partners have implemented resilience programs to promote rangeland conservation in the counties. The United States Agency for International Development (USAID), has been working through the Partnerships for Resilience and economic growth (PREG) and through collaborations such as RCMRD and its SERVIR East & Southern Africa project, to promote development of technologies, products and tools for decision making.

Locally, organized groups have been established to develop mechanisms for improving the health of the rangelands. These include conservancies such as the Northern Rangelands Trust (NRT) which brings together 35 community led conservancies to collectively implement initiatives for sustainable management of the rangeland ecosystems. This is achieved through efforts such as reseeding and practicing planned rotational grazing. The conservancies use emerging technologies and the indigenous knowledge of the rangelands landscape, to implement measures geared toward improving the rangelands ecosystem.

The government of Kenya also set up a Ministry for the Development of Northern Kenya and other Arid Lands (MDNKOAL) in recognition of the prolonged under-investment and inequalities in the rangelands. This was in cognizant of the fact that the failure to specifically fast track and prioritize developments in the rangelands was affecting achievement of development agendas such as Kenya's Vision 2030 among others (Government of Kenya, 2012).

Successful management of the rangelands requires reliable and relevant information on vegetation changes to quantify the

Abbreviations: ASAL, Arid and Semi-Arid Lands; FEWSNET, Famine Early Warning Network; GDP, Gross Domestic Product; GHA, Greater Horn of Africa; LWF, Laikipia Wildlife Forum; MAM, March April May; MDNKOAL, Ministry for the Development of Northern Kenya and other Arid Lands; MODIS, Moderate Resolution Imaging Spectro-radiometer; MVC, Maximum Value Composite; NASA, National Aeronautical Space Agency; NDMA, National Drought Management Authority; NDVI, Normalized Difference Vegetation Index; NRT, Northern Rangelands Trust; OND, October, November and December; PREG, Partnerships for Resilience and economic growth; RCMRD, Regional Centre for Mapping of Resources for Development; RDST, Rangelands Decision Support Tool; REGLAP, Regional Learning and Advocacy Programme; SDG, Sustainable Development Goals; USAID, United States Agency for International Development; VCI, Vegetation Condition Index; VI, Vegetation indices; WMS, Web Map Services.

effects of these changes on the rangeland ecosystem. Intensive and long-term ground-based assessments can produce fine scale, accurate and localized information. However, they are expensive, time and labor intensive; and do not provide full coverage and visualization of the spatial dynamism in rangeland ecosystem. Increasing availability in multi-temporal and multi-spatial satellite imagery provides an avenue for their application to provide an in-depth understanding of the vegetation dynamics at different scales and time steps (Trodd and Dougill, 1998). NDVI has been widely used to monitor vegetation changes in the arid and semi-arid lands. These changes include vegetation response and improvement in vigor following rainfall events. Further, the changes in vegetation abundance and changes in total plant cover and greenness due to utilization as livestock graze and move through the landscape or monitoring units can be observed using earth observation data (Trodd and Dougill, 1998).

Rationale

Different online data catalogs and repositories provide satellite imagery, while others host static maps on vegetation condition and indicators of rangeland productivity. These repositories do not provide users with the freedom to create their own outputs. Sometimes, the information required to make a decision is in different locations. The developed Rangelands Decision Support Tool (RDST) seeks to facilitate near real time assessment and monitoring of rangeland resources by developing a web-based tool that aggregates key indicators of rangeland productivity with ancillary data and allow for integration of user selected indicators to produce maps at different administrative and conservancy boundaries.

The greatest challenge in uptake of technologies and tools meant for decision making is sustainability of initiatives and especially ensuring data recency, coupled with insufficient capacity in understanding and interpreting information. The RDST has been simplified using easy to understand color schemes, simplified interface and detailed information snippets that allow users from different backgrounds such as grazing coordinators and rangeland managers to utilize the tool, generate the maps and make informed assessments. All data processing is automated.

Continuous engagement with stakeholders was done to build their capacity in application of the indices integrated in the RDST for decision making. During training, practical application-based scenarios were developed to demonstrate application of the tool to inform decision making. Specifically, in monitoring of vegetation condition changes in dry and wet seasons, monitoring the effectiveness of grazing plans, assessment of pasture availability by delineating usable pastures from invasive species, monitoring trends and assessing the success of interventions for rehabilitation or improvement of the rangelands.

STUDY AREA

While data was generated for the entire country, the application of the tool for decision making focuses on 24 counties (**Figure 1**) which comprise the arid and semi-arid areas (ASALS) where aridity levels range between 30 and 100%. Rainfall in the arid areas ranges between 150 and 550 mm per year, and in semi-arid between 550 and 850mm per year. These areas are characterized by high temperatures throughout the year with high evapotranspiration rates. The ASALs experience high variability in the frequency and intensity of rainfall events.

Population in the rangelands varies by region. For example population density ranges from 1 or 2 people per km^2 in parts of Turkana and Marsabit to 358 people per km^2 in parts of Kilifi. Areas where conservancies, community and private monitoring units have been established have been highlighted in **Figure 1**. These include the NRT group of conservancies and Laikipia Wildlife Forum (LWF) monitoring units.

MATERIALS

Data and Data Sources

Satellite Imagery

Vegetation indices (VI) are transformations of multi spectral data that enhance information about the vegetation properties and reduce effect of external factors such as atmospheric effects. Identification of the different vegetation types due to their properties is based on spectral signatures. The VI quantifies the amount and vigor of vegetation (Xu and Guo, 2015). They normally change over time with the seasonal behavior of vegetation and decrease in cases of vegetation stress. Dead or light vegetation reflect more red light while green vegetation absorbs red light. Many factors affect VI values like plant photosynthetic activity, total plant cover, biomass, plant and soil moisture, and plant stress.

This study utilized the Normalized Difference Vegetation Index (NDVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS) receiver, located at RCMRD. Modis NDVI was selected due to correlation with many ecosystem attributes, availability of the data and coverage over the study area at 250 m resolution. The dataset, which is freely available, can be acquired daily, however a Maximum Value Composite (MVC) every 10 days is preferred to allow for error reduction and corrections due to cloud cover and other distortions. Because of its ease of use and relationship to many ecosystem parameters (Landmann and Dubovyk, 2014), NDVI has seen widespread use in rangeland ecosystems such as monitoring vegetation dynamics or plant phenological changes over time, estimating biomass production g/m^2, monitoring grazing impacts or attributes related to grazing management (e.g., stocking rates), assessing changes in rangeland condition, vegetation or land cover classification and carbon sequestration (Indeje et al., 2005; Baumann, 2009; Landmann and Dubovyk, 2014).

Ancillary Data

Relevant ancillary data to assist in interpretation and contextualization of the NDVI indices was prepared. In consultation with the stakeholders, administrative, and monitoring boundaries (counties, wards, conservancies were

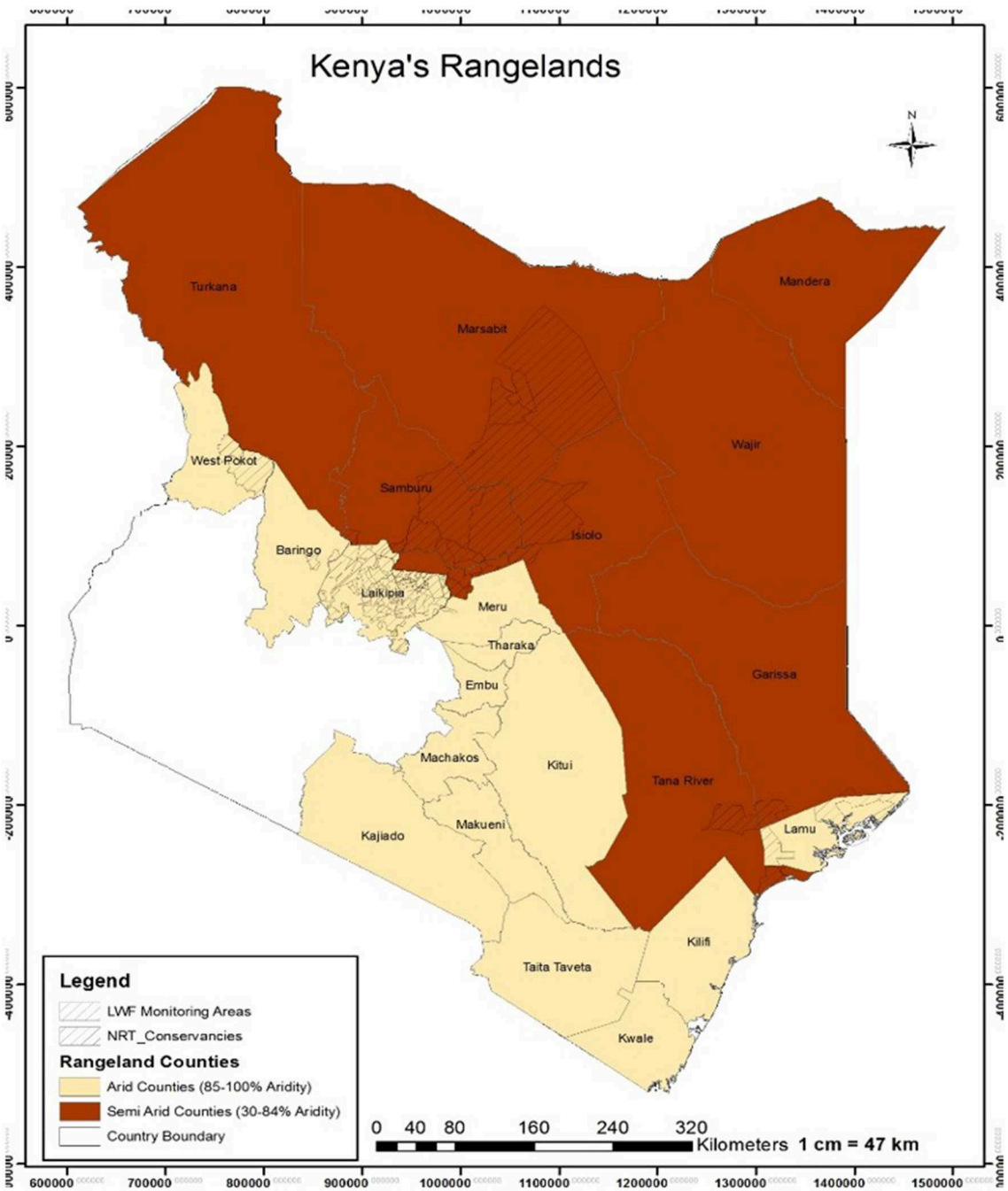


FIGURE 1 | Study area.

added to the web-based tool. Stakeholders provided shapefiles of their monitoring units such as grazing blocks that were divided into wet and dry grazing blocks and buffer zones. Other datasets included rivers, roads, surface water points (dams and water pans), conflict zones, migration routes and protected areas. Invasive species maps showing location of *Opuntia* and *Prosopis Juliflora* with a time stamp based on date of data collection are included in the tool (see **Supplementary Table 1**).

METHODS

The methodology workflow is illustrated in **Figure 2**.

Data Acquisition and Processing

MODIS Terra NDVI is acquired between 8 a.m. and 9:30 p.m. since the passes within this time frame produced clear images with less cloud cover. The raw files are automatically projected

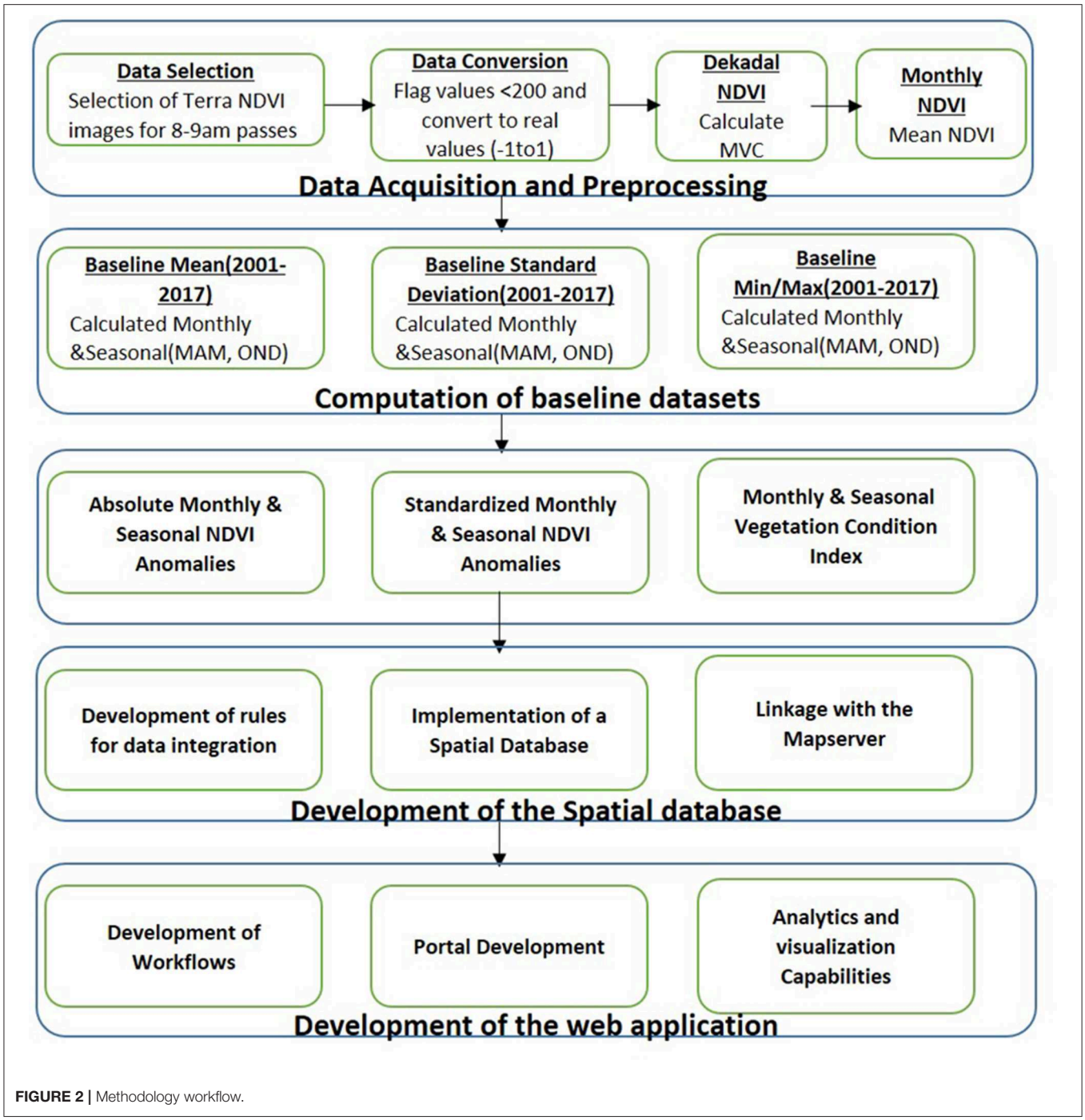


FIGURE 2 | Methodology workflow.

to WGS84 coordinate system and then clipped using the Kenya national boundary shapefile. The physical NDVI values were then computed from the digital numbers. Some digital values were used to flag specific pixels: 255 for invalid pixels, 254 for water/sea, 253 for snow/ice, 252 for cloud/shadow, and 251 for missing data. Ten-day Maximum Value Composites (MVC) were computed from the daily files. NDVI values range between −1 and 1. The values lower than 0.1 typically correspond to areas with little or no vegetation (rocks, ice, and desert). Moderate values (∼0.2 and 0.3) correspond to shrub and grasslands and high values (0.5 and above) typically correspond to dense vegetation. Monthly and seasonal NDVI was also generated. The seasons were loosely defined for the long rains (March, April and May) and short rains (October, November, and December).

At a monthly and seasonal interval, NDVI absolute anomalies, standard anomalies and the Vegetation Condition Index (VCI) were also generated. Anomalies are quantitative measurements applied to assess how the greenness of the vegetation varies over space and time. The standard anomalies represent the difference between the current values and the long-term mean.

They highlight deviations from normal vegetation development portraying both positive and negative changes that give a qualitative indication of how good or bad the current season or month is compared with other similar periods. Below-normal conditions relate to less than normal conditions such as drought or stress while above normal conditions relate to increased/improved productivity. Anomalies help highlight areas of potential concern and take into account the variations over time (dekad, month, and season) in productivity of the vegetation hence giving the bigger picture. Continuous persistence of negative anomalies is a red flag especially due to dependence on rain-fed pastures for livelihoods in the rangelands. VCI was also calculated due to its ability to detect and spatially delineate anomalies in vegetation condition and growth, both in extension and intensity which is useful for monitoring vegetation changes within a season for early warning.

1. Monthly NDVI ($NDVI_{i,m,y}$) for each Monitoring Unit (MU) i in month m and year y is obtained by averaging the 3 dekadal values in each month:

$$NDVI_{i,m,y} = \sum_{d \in m} NDVI_{i,d,m,y}$$

2. NDVI Absolute Anomalies ($AANDVI_{i,m,y}$) for each MU i in month m and year y is obtained by subtracting monthly NDVI from a baseline monthly mean (2001 to 2017)

$$AANDVI_{i,m,y} = NDVI_{i,m,y} - E(NDVI_{i,m})$$

3. Standard Anomalies (Z-score) of the NDVI for each MU i in each month m and each year y ($NDVI_{i,m,y}$) were calculated with a baseline monthly mean (2001 to 2017) mean (E) and standard deviation (SD) of the particular MU and month:

$$SANDVI_{i,m,y} = \frac{NDVI_{i,m,y} - E(NDVI_{i,m})}{SD(NDVI_{i,m})}$$

4. VCI for each MU i in each month m and each year y ($NDVI_{i,m,y}$) were calculated with a baseline monthly minimum(min) and maximum(max) from 2001 to 2017

$$VCI_{i,m,y} = \frac{NDVI_{i,m,y} - min(NDVI_{i,m})}{\max\left(NDVI_{i,m}\right) - max(NDVI_{i,m})}$$

Similar computations were applied for computation of the seasonal variables with long term mean and standard deviation calculated for the months within the season.

Development of the Spatial Database

A relational database was designed and implemented with rules for integration of spatial and non-spatial data effected. The database was implemented in PostgreSQL/PostGIS. PostGIS is an Extender for PostgreSQL that allows its ability to manage geospatial data within a relational database structure. The database was then populated with both raster and vector datasets.

Development of the Web Application

The web application was built using JavaScript libraries—ExtJS, OpenLayers, and jQuery. Extjs renders the viewport containing the panels, forms, toolbars and menus. Web maps were rendered using Openlayers which connects to the Web Map Services (WMS) published via Geoserver and ArcGIS Server. The vector and raster datasets found on the application reside in a Geoserver where they are updated frequently. The ArcGIS server provided the web processing service which enables on-the-fly generation of downloadable maps. Both the front-end and backend applications were deployed on the web by the Apache2 web server running in Ubuntu Linux OS. Communication between the front-end and backend was enabled via Ajax requests and responses with a REST API. JSON was used as the data format during information exchange.

Automation

Python scripts were developed to automate the data processing chain and automatic publishing of final indices. This ensures that data is automatically available for users at preset intervals at the beginning of a new dekad, month, or season. A parallel dataset is downloaded as backup to ensure that local servers' failure does not affect timely availability of the indices.

Navigation

The Rangelands decision support tool interface is illustrated in **Figure 3**. It allows users to select indices, the necessary ancillary data, and produce downloadable maps. The system is able to link user selection in the production of final maps providing the user with the freedom of working with specific inputs from the wide array of indices and ancillary data provided at a user defined boundary.

The SERVIR Service Planning Approach

Development of the tool adopted a consultative needs assessment method, where rangeland actors were involved to identify gaps and challenges in their decision-making processes. Initial consultations engaged three groups from NRT, LWF and FEWSNET (Famine Early Warning Network). The need for a spatially and temporally flexible dynamic tool was identified and the stakeholders were involved in co-development of the tool. An understanding of information flow in decision making and identification of key actors in the process informed the selection of indices, temporal and spatial extents and monitoring units. By developing stakeholder maps, critical rangeland actors were identified and engaged and this led to a growth in the number users for the tool. With the expected growth, processing of the indices was done for the entire ASAL region to allow for new users to plug in their monitoring units and use the tool.

Through the PREG partnership, Kenya RAPID, SERVIR E&SA, and Mercy Corps Livestock Market Systems program, supported five counties (Turkana, Isiolo, Marsabit, Garisa, and Wajir) to digitize their grazing plans and rehabilitation areas. This will enable them use to the tool to monitor the rangelands and assess the success of interventions. With an understanding of the multiple stakeholders with different

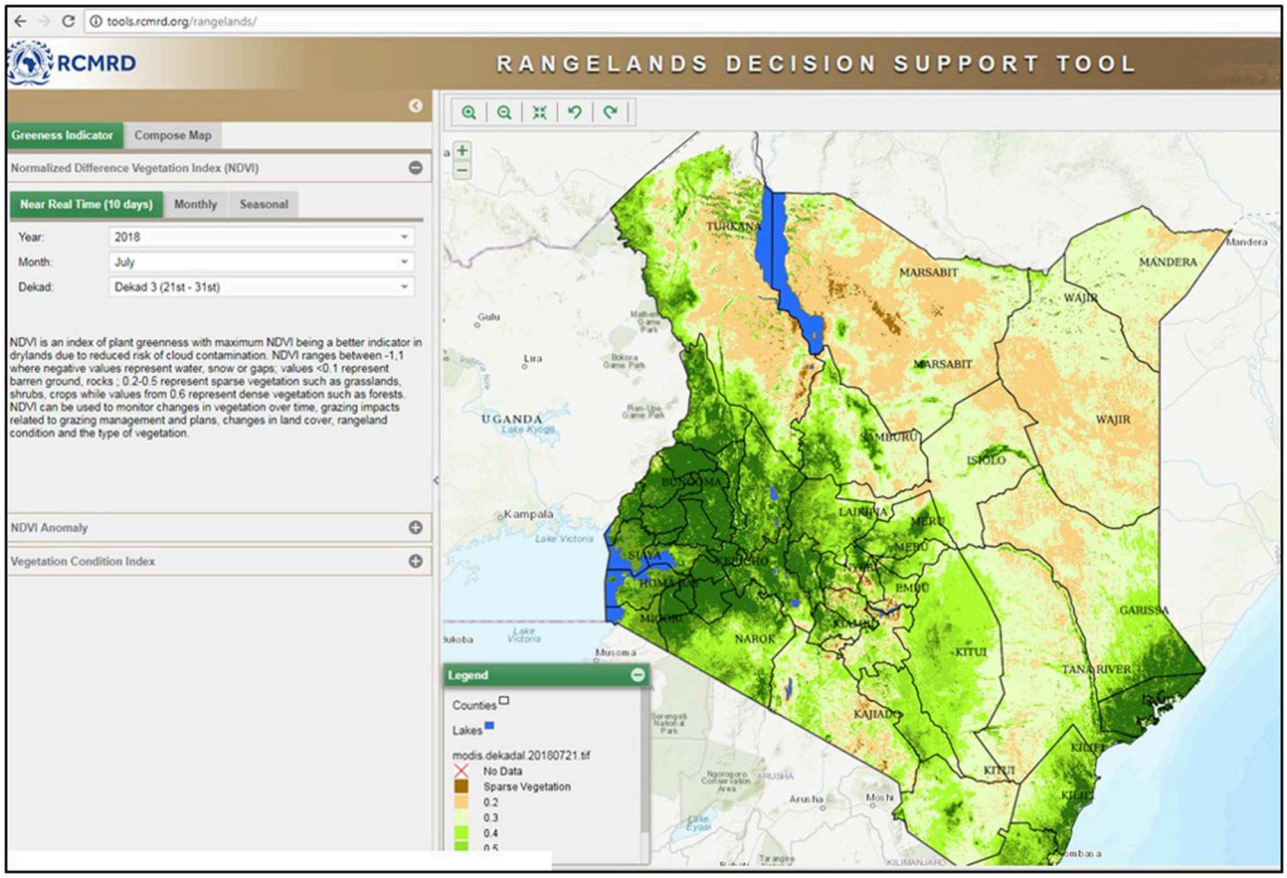


FIGURE 3 | The rangeland decision support tool interface.

backgrounds, the tool's main goal was to provide an easy to use interface, processed indices for different monitoring needs, easily interpretable and relatable legends where green colors represent improving vegetation conditions and browning represent declining conditions; and the flexibility to allow users select ancillary data from an array of relevant indicators to produce maps at the click of a button. To support understanding of the indices and correct interpretation, awareness and capacity building trainings were conducted.

The trainings included all relevant rangeland actors such as researchers, data producers, rangeland managers, rangeland coordinators, and county officers. The approach on capacity building required users to apply the tool in solving challenges or informing relevant decisions. The participants were required to identify management problems or concerns they were experiencing and apply the tool to assess/monitor the situation and come up with actionable information that would be presented management for reporting, or to influence/emphasize on a suggested plan of action, or even to an understanding of conditions being experienced. The use cases focused on short term and long-term trends assessments, monitoring vegetation recovery or decline, and monitoring success of interventions.

RESULTS

Context

Application of the tool for decision making requires interpretation within the right context. Some of the considerations include an understanding of the response of the Kenyan ASALS to the challenges being experienced. Recurrent droughts that leave no time for recovery and diminishing resources for grazing due to privatization of land and agro pastoralism has contributed to increasing the strain on available resources, resulting in degradation and colonization of invasive species in degraded lands (**Figure 4**).

Further there is need to interpret negative changes within the right context. Negative values or browning can be caused by normal seasonal changes such as the loss of annuals which leave the land bare. Identification of degradation should be done by use of multiple indices to observe trends of vegetation changes over time. Greening in the rangelands can also be attributed to colonization by invasive species.

Careful interpretation of the VCI is called for since it can show very unpredictable and sharply fluctuating results for regions with rather constant vegetation index profiles, such as warm and cold deserts and equatorial forests and therefore users should apply it in assessments in areas with clear phenological profiles.

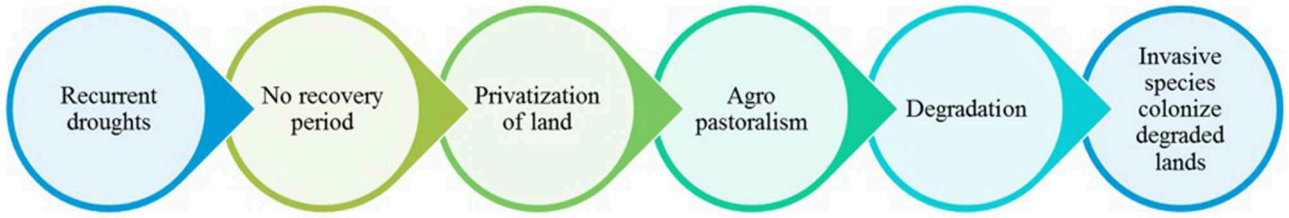


FIGURE 4 | Contextualizing interpretation of rangeland conditions.

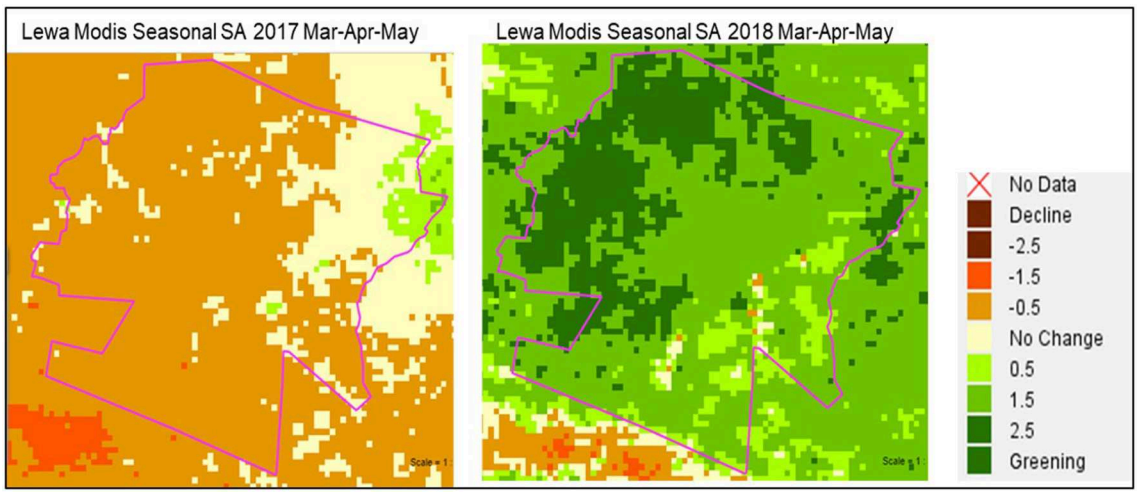


FIGURE 5 | Standardized anomalies for lewa conservancy during the 2017 and 2018 long rains season.

Application of the Tool for Addressing Specific Management and Monitoring Challenges

The tool has been successfully applied for monitoring by NRT and LWF groups of conservancies. Specifically, monitoring utilization of grazing resources over time in the wet and dry grazing areas by using the tool to inform movement of livestock from one paddock to the next. This resulted in time saved and efficiency by eliminating the need to frequently send scouts to monitor vegetation changes in the grazing blocks. The tool has also been used to monitor improvements or decline in vegetation productivity, an assessment of degradation or resilience of the landscape and the success of investments in conservation and rehabilitation of the rangeland's ecosystem. NRT and LWF provided their grazing blocks and rehabilitation areas shapefiles as an input to the tool. With capacity building, the stakeholders realized the value of spatially viewing the health of the ecosystem which allows them to identify hotspots where interventions should be prioritized. By using the tool, the stakeholders were able to assess the success of their grazing plans and identify areas where they need to be reviewed. Specific application examples are provided below.

Assessment of the Resilience of Lewa Conservancy

Concerns that the conservancy was experiencing degradation raised during a drought year (2017) led to the assessment to ascertain if the loss in vegetation productivity was linked to drought or degradation. An assessment of the vegetation conditions from 2016 to 2018 using standardized anomalies for the long rains season (MAM) (**Figure 5**) found that the decline in vegetation productivity was linked to the drought and the conservancy demonstrated full recovery when the drought ended and the long rains commenced. The information was submitted to management to array concerns on the efficacy of management strategies implemented as the ecosystem was resilient and recovered quickly after the major drought.

Assessment of the Drought Events in Ol-Pejeta Conservancy

Assessment of drought impact on vegetation conditions from onset to recovery over several years was assessed in Ol Pejeta Conservancy using absolute anomalies. Feedback from the conservancy confirmed that there were four prolonged dry spells between 2006 and 2017 which affected the communities. From the maps it was clear to see that drought events are becoming

more frequent with the sequence alternating between a good (wet) year and a dry (drought) year leaving little time for total recovery. The maps were developed and used for reporting to management. **Figure 6** shows the drought progression between 2009 and 2011, and 2017 to 2017 Long Rains Season in Ol-Pejeta Conservancy using Absolute Anomalies.

Assessment Human Induced Degradation on Mathews Range Forest Reserve

Mathews mountain range in Namunyak Conservancy in Samburu County, is protected forest reserve made of an expanse of lush indigenous forest which hosts abundant populations of wildlife and rare plant species (Numanyak Wildlife Conservancy, 2010).

Due to increased frequency of drought events with little time for recovery, strain on the protected resource is clearly visible in **Figure 7**, even in good years such as 2011 and 2012 where degradation (dark brown) can be observed. Discussions with conservancy managers confirmed that the situation was as a result of an influx of livestock from the North of Kenya, resulting in exceeded carrying capacity and encroachment by livestock on the protected resource. Further due to the impact of the droughts on livelihoods of the communities, cutting of trees for charcoal burning after the drought in 2010 contributed to the observed trends. The assessment further correlated to evaluation of the forest, which established that parts of the forest were experience slight to severe degradation (Numanyak Wildlife Conservancy, 2010) due to human activities such as cutting of trees and overgrazing.

Prompting a Paradigm Shift From "Business as Usual" to Rehabilitate Degraded Areas in Il Ngwesi Conservancy

The NRT's Il Ngwesi conservancy has established rotational grazing with clear delineation of wet and dry grazing areas. However, with frequent droughts, the capacity of the ecosystem is strained and recovery, even when the drought ends is slow. Further encroachment by invasive species such as the Prosopis Juliflora, further reduces available pastures. The NRT team used the maps to verify both the loss in vegetation productivity, coupled with their knowledge on the ground, to delineate extent of invasive species. From the maps developed, the proposed a review of the current grazing plans which outlined the need to prioritize and set aside part of the conservancy for rehabilitation through reseeding and clearing of the invasive species to promote recovery. VCI maps were developed for the short rains season to assess the situation (**Figure 8**).

From their grazing map in **Figure 9**, Region A represents Mukogondo forest which is a dry season grazing resource. Region C represents the area where degradation notable by continued

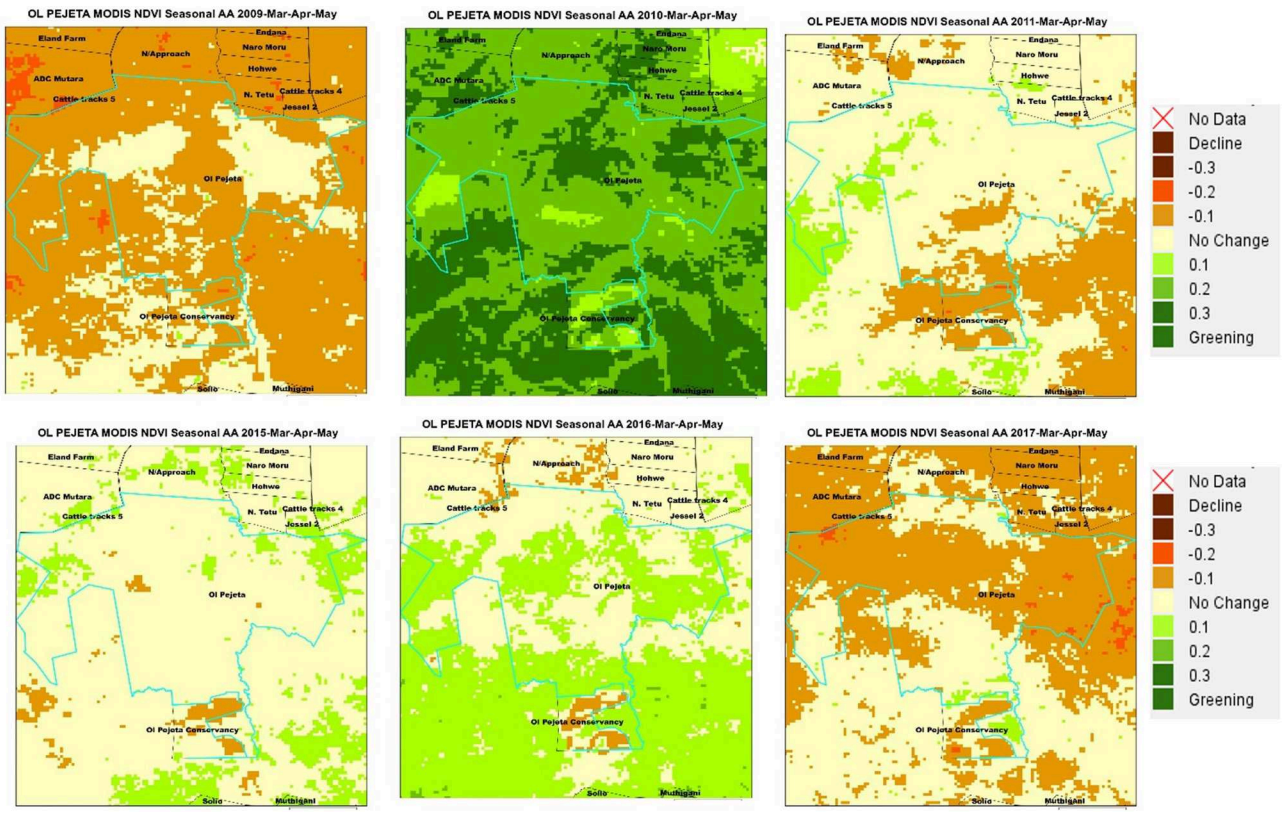


FIGURE 6 | Drought progression between 2009 and 2011, and 2017 to 2017 long rains season in ol-pejeta conservancy using absolute anomalies.

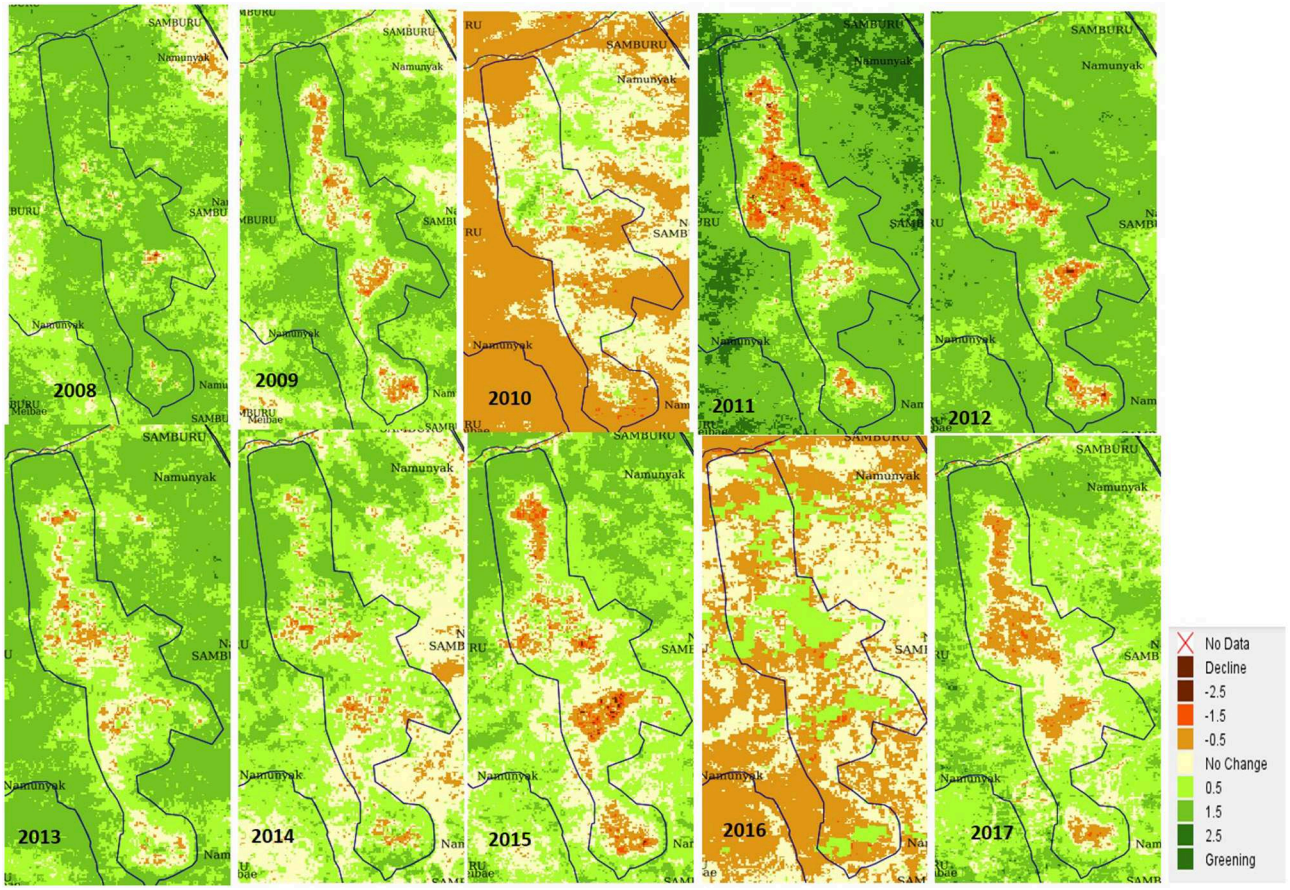


FIGURE 7 | Implication of prolonged drought events on Mathews Range Forest Reserve.

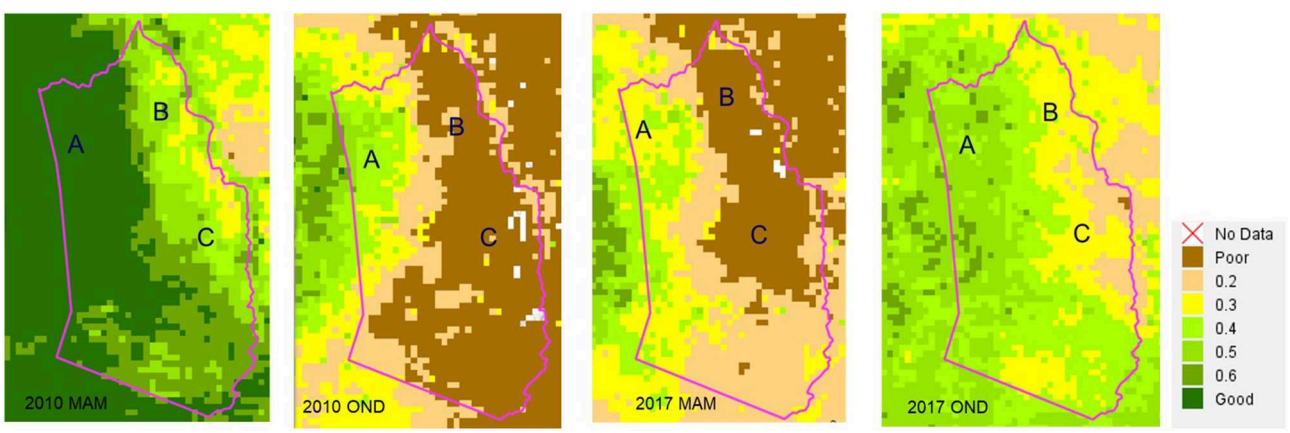


FIGURE 8 | Seasonal VCI for IlNgwesi on two drought years during the long rains (MAM) and Short rains (OND) seasons.

vegetation productivity decline over time has been recorded. Toward the north in Region B is the Ol-Donyiro area where the invasive species (prosopis Juliflora) has dominated. Region B&C which are generally drier represent the wet season grazing area. To promote recovery in B&C the grazing coordinator, proposed that those areas should be set aside for rehabilitation, with clearing of invasive species done in Area B and reseeding in B & C.

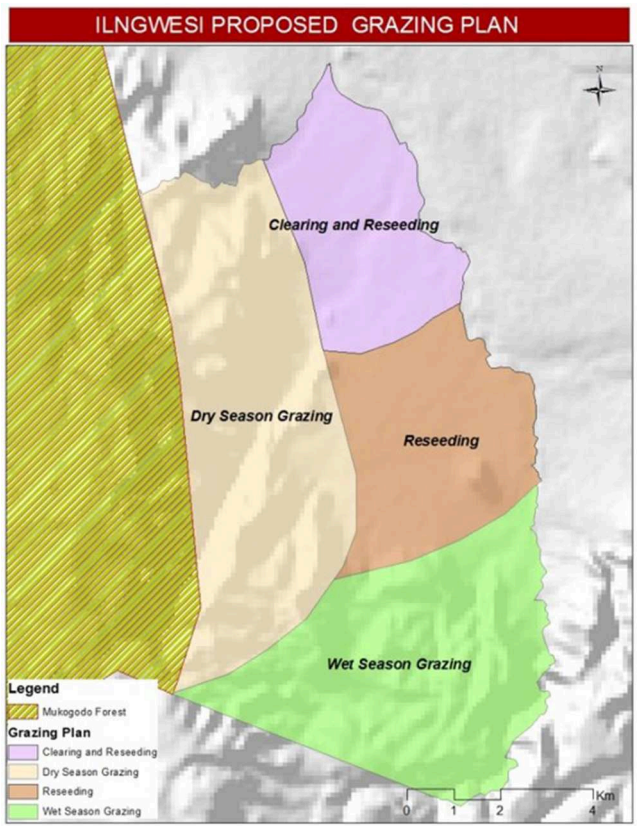


FIGURE 9 | Proposed revised grazing plan for Il Ngwesi Conservancy.

CONCLUSIONS

Sample case studies demonstrate the application of the tool for decision making in the rangelands. Specifically, the application of the tool for monitoring changes in vegetation condition due to events such as drought, with local knowledge identification of human induced and climate induced changes in vegetation, and the assessment of the ability of a rangeland ecosystem to recover after drought events.

Clearly, supplementary local and indigenous knowledge was critical in the assessments. Emphasis is therefore given on the combination of local indigenous knowledge of the rangeland landscape coupled with the web-based tool indices for improved decision making. During the discussions with stakeholders, it was clear that the tool can support assessment of the success of current management plans and identify areas where interventions were not successful for review. For example, due to recurrent dry spells, some areas in Namunyak conservancy were becoming degraded and the grazing coordinators developed maps that demonstrated these negative changes. They planned to present the maps to management to suggest a review of the management plans in terms of livestock movements in wet and dry season to ensure recovery. The tool in such case was used to prompt action by reviewing the management plans. The tool was also applied to allay alarm and panic by the management on alleged degradation in Lewa since the maps provided proof that ecosystem recovered well and quickly after drought.

RECOMMENDATIONS

While the application of the tool for decision making is clear, further improvements are required to improve its capabilities. Based on feedback from stakeholders, addition of a monitoring component on surface water is under development to complement the vegetation indices. Addition of a smoothing algorithm is also required to improve the indices especially VCI. Development of an NDVI utilization component to support improved assessments per land cover class within a monitoring unit is required. This would allow quantification of the changes in different classes of interest for different monitoring needs. Data collection should be done for validation of the vegetation indices.

DATA AVAILABILITY STATEMENT

The datasets generated and processed for this research can be accessed through the web based decision support tool at http://tools.rcmrd.org/rangelands or the mini viewer at http://apps.rcmrd.org/rangelands/.

AUTHOR CONTRIBUTIONS

LN led the development of the web-based decision support tool and writing this paper, with review from RM, AW, and EA. MO was the lead web developer, with SO leading development of scripts for automation.

FUNDING

The tool has been developed with funding from USAID under the SERVIR Eastern and Southern Africa (ESA) project at the Regional Centre for Mapping of Resources for Development.

ACKNOWLEDGMENTS

Participation and contribution from rangeland stakeholders have provided invaluable input in the development and application of the tool for decision making. Kenya RAPID provided support in engaging stakeholders in their implementing counties and promoting use of the tool for decision making. Special thanks to NRT, LWF, and specifically Lewa (Dr. David Kimiti) and Ol-Pejeta (Benard Chiira) for contributing information on application of the tool through the case studies. Ngugi Kimani who supported the integration of user input in map composing function on the RDST.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2019.00187/full#supplementary-material

REFERENCES

Alexandratos, N., and Bruinsma, J. (2015). *World Agriculture Towards 2030/2050 the 2012 Revision*. ESA Working Paper No 12–03.

Baumann, G. (2009). *How to assess rangeland condition in semiarid ecosystems? The indicative value of vegetation in the High Atlas Mountains, Morocco* (Dissertation). Mathematisch-Naturwissen-Schaftlichen Fakultät, Universität Köln, Cologne, Germany.

Delgado, C. L., Rosegrant, M. W., and Meijer, S. (2001). "Livestock to 2020: the revolution Continues," *Paper Presented at the Annual Meetings of the International Agricultural Trade Research Consortium (IATRC)* (Auckland). Available online at: http://ageconsearch.umn.edu/bitstream/14560/1/cp01de01.pdf

Government of Kenya (2012). *Vision 2030 Development Strategy for Northern Kenya and Other Arid Lands*. Final Report. Government of Kenya.

Herrero, M., Ringler, C., van de Steeg, J., Thornton, P., Zhu, T., Omolo, E., et al. (2010). *Climate Variability and Climate Change and Their Impacts on Kenya's Agricultural Sector*. Nairobi: ILRI.

Herrick, J. E., Urama, K. C., Karl, J. W., Boos, J., Johnson, M.-V. V., Shepherd, K. D., et al. (2013). The global land-potential knowledge system (LandPKS): supporting evidence-based, site-specific land use and management through cloud computing, mobile applications, and crowdsourcing. *J. Soil Water Conserv.* 68, 5A−12A. doi: 10.2489/jswc.68.1.5A

Indeje, M., Ogallo, L. J., Davies, G., Dilley, M., and Anyamba, A. (2005). Predictability of the normalized difference vegetation index in Kenya and potential applications as an indicator of Rift Valley fever outbreaks in the greater. *J. Clim.* 19, 1673–1687. doi: 10.1175/JCLI3708.1

Kyuma, R. K., Wahome, R. G., Kinama, J. M., and Wasonga, V. O. (2016). Temporal relationship between climate variability, prosopis juliflora invasion and livestock numbers in the drylands of magadi, Kenya. *Afr. J. Environ. Sci. Technol.* 10, 129–140. doi: 10.5897/AJEST2015.2034

Landmann, T., and Dubovyk, O. (2014). Spatial analysis of human-induced vegetation productivity decline over eastern Africa using a decade (2001-2011) of medium resolution MODIS time-series data. *Int. J. Appl. Earth Observ. Geoinformation* 33, 76–82. doi: 10.1016/j.jag.2014.04.020

Mulinge, E., Gicheru, P., Muriithi, F., and Kihiu, E. (2015). "Economics of land degradation and improvement in Kenya," in *Economics of Land Degradation and Improvement - A Global Assessment for Sustainable Development,* 1–686. doi: 10.1007/978-3-319-19168-3

Numanyak Wildlife Conservancy (2010). *Mathews Forest Range Ecological Assessment Summary Report.* Nairobi.

REGLAP (2012). *Key Statistics on the Drylands of Kenya, Uganda and Ethiopia, REGLAP Secretariat, October, 2012.* Nairobi: REGLAP Secretariat.

Silvestri, S., Bryan, E., Ringler, C., Herrero, M., and Okoba, B. O. (2012). Climate change perception and adaptation of agro-pastoral communities in Kenya. *Reg. Environ. Change* 12, 791–802. doi: 10.1007/s10113-012-0293-6

Trodd, N. M., and Dougill, A. J. (1998). Monitoring vegetation dynamics in semi-arid African Rangelands: use and limitations of earth observation data to characterize vegetation structure. *Appl. Geogr.* 18, 315–330. doi: 10.1016/S0143-6228(98)00024-1

Vågen, T. G., Winowiecki, L. A., Tondoh, J. E., Desta, L. T., and Gumbricht, T. (2014). Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma* 263, 216–225. doi: 10.1016/j.geoderma.2015.06.023

Vicente-Serrano, S. M., Cabello, D., Tomás-Burguera, M., Martín-Hernández, N., Beguería, S., Azorin-Molina, C., et al. (2015). Drought variability and land degradation in semiarid regions: assessment using remote sensing data and drought indices (1982-2011). *Remote Sens.* 7, 4391–4423. doi: 10.3390/rs70404391

Winowiecki, L., Vågen, T.-G., Kinnaird, M. F., and O'Brien, T. G. (2018). Application of systematic monitoring and mapping techniques: assessing land restoration potential in semi-arid lands of Kenya. *Geoderma* 327, 107–118. doi: 10.1016/j.geoderma.2018.04.017

Xu, D., and Guo, X. (2015). Some insights on Grassland health assessment based on remote sensing. *Sensors* 15, 3070–3089. doi: 10.3390/s150203070

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 30 January 2020
doi: 10.3389/fenvs.2020.00004

Toward Operational Mapping of Woody Canopy Cover in Tropical Savannas Using Google Earth Engine

Julius Y. Anchang[1], Lara Prihodko[2], Wenjie Ji[1], Sanath S. Kumar[1], C. Wade Ross[1], Qiuyan Yu[1], Brianna Lind[1], Mamadou A. Sarr[3], Abdoul A. Diouf[3] and Niall P. Hanan[1]*

[1] Plant and Environmental Sciences, New Mexico State University, Las Cruces, NM, United States, [2] Animal and Range Sciences, New Mexico State University, Las Cruces, NM, United States, [3] Centre de Suivi Ecologique, Dakar, Senegal

Savanna woody plants can store significant amounts of carbon while also providing numerous other ecological and socio-economic benefits. However, they are significantly under-represented in widely used tree cover datasets, due to mapping challenges presented by their complex landscapes, and the underestimation of woody plants by methods that exclude short stature trees and shrubs. In this study, we describe a Google Earth Engine (GEE) application and present test case results for mapping percent woody canopy cover (%WCC) over a large savanna area. Relevant predictors of %WCC include information derived from radar backscatter (Sentinel-1) and optical reflectance (Sentinel-2), which are used in conjunction with plot level %WCC measurements to train and evaluate random forest models. We can predict %WCC at 40 m pixel resolution for the full extent of Senegal with a root mean square error of ∼8% (based on independent sample evaluation). Further examination of model results provides insights into method stability and potential generalizability. Annual median radar backscatter intensity is determined to be the most important satellite-based predictor of %WCC in savannas, likely due to its relatively strong response to non-leaf structural components of small woody plants which remain mostly constant across the wet and dry season. However, the best performing model combines radar backscatter metrics with optical reflectance indices that serve as proxies for greenness, dry biomass, burn incidence, plant water content, chlorophyll content, and seasonality. The primary use of GEE in the methodology makes it scalable and replicable by end-users with limited infrastructure for processing large remote sensing data.

Keywords: earth observation, vegetation structure, Copernicus Sentinel data, cloud-computing, machine learning

OPEN ACCESS

Edited by:
Mark A. Drummond, United States Geological Survey (USGS), United States

Reviewed by:
Perrine Hamel, Stanford University, United States
Sónia Maria Carvalho Ribeiro, Federal University of Minas Gerais, Brazil

****Correspondence:***
Julius Y. Anchang
anchang@nmsu.edu

Specialty section:
This article was submitted to Land Use Dynamics, a section of the journal Frontiers in Environmental Science

Received: *01 June 2019*
Accepted: *09 January 2020*
Published: *30 January 2020*

Citation:
Anchang JY, Prihodko L, Ji W, Kumar SS, Ross CW, Yu Q, Lind B, Sarr MA, Diouf AA and Hanan NP (2020) Toward Operational Mapping of Woody Canopy Cover in Tropical Savannas Using Google Earth Engine. Front. Environ. Sci. 8:4. doi: 10.3389/fenvs.2020.00004

INTRODUCTION

The Ecological and Socio-Economic Importance of Woody Vegetation in Savannas

Woody plants are an important component of terrestrial ecosystems; they play a major role in carbon, nutrient and hydrological cycles (Vitousek, 1982; Jackson et al., 2002; Huxman et al., 2005) and provide habitat for other species (Ratter et al., 1997). Savannas occur across tropical, sub-tropical and temperate latitudes, and feature the co-dominance of woody and herbaceous plant forms (Werner, 2009). While they may exhibit

on average lower woody densities relative to closed canopy forests, their vast spatial extent (up to 40% of the Earth's terrestrial surface, depending on definition; see Scholes and Archer, 1997; Bond and Midgley, 2000; Sankaran et al., 2005; Ratnam et al., 2011, and references therein) indicates an important role in global and regional carbon storage. As such, a lack of detailed information on woody vegetation cover in savannas contributes uncertainties in current carbon stock estimates and constrains scientific understanding of the role they may play in long-term climate change. In socio-economic terms, trees and shrubs are key to the basic livelihoods of millions living in regions such as the West African Sahel. The predominantly agro-pastoralist societies rely on woody biomass for energy (fuelwood and charcoal) and food (including livestock browse) (Wessels et al., 2013; Hanan, 2018). An accurate quantitative and spatially explicit assessment of woody vegetation cover in such regions is thus crucial to local, regional and global efforts aimed at understanding and combating the effects of climate change through carbon sequestration, reducing food insecurity through extensive livestock systems, and promoting sustainable land use practices.

Challenges to Large Area Mapping of Woody Vegetation in Savannas

Savannas have complex landscapes; a picture of individual shrubs and trees or discontinuous tree canopies against a background of grassland and/or cultivated surfaces comes to mind. This picture is often the result of millennia of climate and human induced changes that continuously alter landscapes on short- and long-term basis (Behling and Hooghiemstra, 1999; Werner, 2009). Medium and coarse resolution land cover maps that categorize individual pixels into unique land cover/land use types cannot adequately separate different vegetation types in such areas. At the same time, using commercial very high resolution (VHR) imagery (e.g., <1 m) for detailed mapping remains impractical beyond the local scale, due to the insufficient coverage of currently archived data, high cost of new large area acquisitions, and steep computation and storage requirements for processing.

In lieu of categorical land cover maps, Vegetation Continuous Fields (VCF) land cover products have been developed to fill the need for more detailed vegetation mapping over large areas; by providing sub-pixel/fractional estimates of specific canopy properties (e.g., percent tree cover) (Defries et al., 2000b; Hansen et al., 2002). VCF methodology generally involves using satellite-derived metrics as discriminants in an empirical model that is calibrated with continuous-scale measurements obtained from the field (or from higher resolution imagery) (Foody and Cox, 1994; Hansen et al., 1996; Defries et al., 2000a; Gessner et al., 2013; Baumann et al., 2018). Existing global VCF products are however not ideal for estimating canopy properties in drylands and savannas. For example, the annual MODerate Resolution Imaging Spectrometer (MODIS) VCF product (MOD44B) (Hansen et al., 2002), and the Landsat equivalent (Sexton et al., 2013), are designed to only represent tree canopies with certain characteristics based on the definition of forests provided by the Food and Agricultural Organization (e.g., canopy height > 5 m) (FAO, 2000). Consequently, these datasets are known to significantly underestimate woody plant cover in dry savannas, where trees and shrubs tend to be of considerably lower stature (**Figure 1**, as well as Gessner et al., 2013; Brandt et al., 2016a).

The Use of Satellite Remote Sensing in Mapping Woody Cover in Savannas

Satellite-obtained optical reflectance data have been used for decades to map and monitor vegetated surfaces over large extents (Matthews, 1983; DeFries et al., 1995; Hansen et al., 2002). This is largely due to their ability to provide scalable spectral information relevant to plant canopies such as greenness, leaf area index and phenology (Zhang et al., 2003; Xie et al., 2008). The extensive spatiotemporal coverage and free availability of data from sensors such as MODIS and Landsat, and more recently Copernicus Sentinel instruments, has also contributed to making remote sensing data a far less costly and laborious option to derive global and regional vegetation maps. However, particularly in tropical and subtropical regions, these benefits are counteracted by problems associated with cloud coverage, atmospheric scattering, background reflectance, and saturation of optical indices in more densely vegetated landscapes (Huete et al., 1997).

In the context of open savannas and woodlands, phenology and seasonality metrics derived from optical reflectance data have been used to discriminate woody and non-woody plant forms. Brandt et al. (2016a) demonstrated this using dry-season integrated Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), derived from MODIS and SPOT-VEGETATION (SPOT-VGT), to estimate woody plant cover at 1km resolution scale for the West African Sahel. In this region, annual vegetation productivity is strongly controlled by precipitation which typically falls within a short window (3–5 months) (Nicholson and Webster, 2007). Herbaceous vegetation would typically green-up and senesce during this brief period, while trees and shrubs can flush leaves before the rains and often retain leaves for some months into the dry season (Hiernaux et al., 1994). In theory, this means woody vegetation may be distinguished from herbaceous vegetation using time series optical reflectance data to detect differences in phenophases between woody and non-woody plants. However, such precise phenology information is harder to obtain using remote sensing data at finer (sub-100 m) scales, where operational instruments have lower revisit times; and also, where the differences in phenophases among woody species (see Brandt et al., 2016a,b) becomes more visible, making it harder to collectively separate them from herbaceous vegetation.

Unlike optical reflectance, microwave (radar) backscatter is largely insensitive to atmospheric/cloud conditions and, depending on wavelength, can be sensitive to the seasonally invariant structural components of trees and shrubs. For example, the leaves of small trees and shrubs would be largely transparent to Sentinel-1 C-band radar backscatter (∼5 cm wavelength) but these wavelengths are sensitive to the stems and branches (Flores-Anderson et al., 2019). Thus, the addition

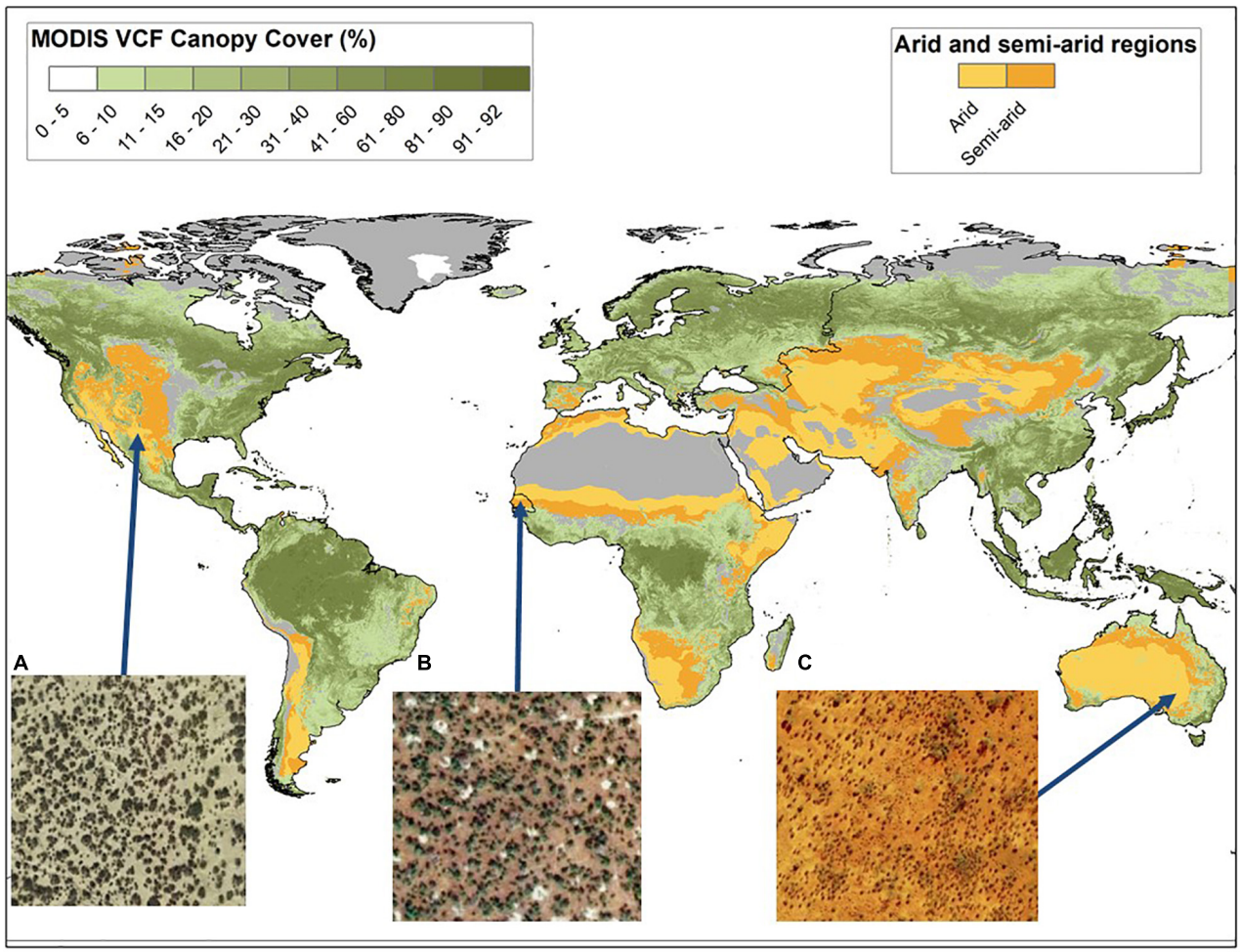


FIGURE 1 | An illustration of how the MODIS Vegetation Continuous Fields Tree Cover data (MOD44B) effectively excludes global dryland regions (yellow and orange) where VCF generally indicates canopy cover is 0–5%. Closer examination (inset images **A**, **B**, and **C**) suggests these areas can contain significant amounts of woody cover in the form of (smaller) trees and shrubs. Imagery Source: ESRI Basemaps, DigitalGlobe.

of radar backscatter should improve our ability to discriminate woody vegetation in savannas with or without the benefit of knowledge on tree-grass phenological differences. This was demonstrated in a recent study that used ALOS PALSAR L-band backscatter to map woody vegetation in a Southern African savanna (Urbazaev et al., 2015), and another study that showed that fusion of Sentinel-1 (C-band) radar backscatter with Landsat 8 optical reflectance data significantly improved the accuracy of mapping tree and shrub cover in South America (Baumann et al., 2018).

Objectives

Against this background, our main objective is to describe the methodology and present prototype results for mapping %WCC over a large, predominantly savanna region in West Africa. Our approach combines relevant metrics, derived from both radar backscatter and optical reflectance data, as empirical correlates of %WCC at medium resolution (40 m) in an ensemble decision tree (random forest) model. Furthermore, we interpret our test case results to answer the following questions: (1) Which earth observation metrics contribute most to accurate predictions of %WCC in tropical savannas? (2) Given the 'black box' nature of machine learning models used for prediction, can we discern meaningful (statistical or mechanistic) relationships between important predictors and %WCC? (3) How does our derived %WCC compare with other similarly published datasets for the region? Answers to these questions should indicate how our approach contributes a new and useful tool for reliably mapping savanna tree and shrub cover at relatively fine scales using remote sensing data, and its transferability to other regions.

Local, national and regional institutions in developing regions such as West Africa face tremendous difficulties in the operational use of remote sensing data for environmental and natural resource monitoring. These difficulties mostly arise from the technical requirements (e.g., internet bandwidth, computer processing power, and storage capacity) required for handling large volumes of satellite data. In anticipation of these difficulties, and to make our approach accessible for implementation in

other regions, our methodology relies on the use of geospatial cloud computing resources provided by Google Earth Engine (GEE; Gorelick et al., 2017), with automated and documented workflows that facilitate local adaptation by relevant stakeholder communities. This framework is scalable and repeatable and aims to support mapping in arid and semiarid regions around the world with woody plant canopy/height varying from open/short (<5 m) to closed/tall (>5 m).

MATERIALS AND METHODS

Study Area

Our study area is the full geographic extent of Senegal in West Africa (**Figure 2**). Senegal has a bio-climatic gradient ranging from the arid Sahel in the north (mean annual precipitation or MAP < 300 mm, with low %WCC and shrubs and trees generally 1–2 m tall), through the semi-arid Sudano-Sahelian zone (300 mm < MAP < 900 m, a blend of woodland and grassland savannas, with shrubs and trees 2–6 m tall), to the humid Guinean savanna/forest mosaic south of The Gambia (MAP > 900 mm, trees up to ∼10 m tall) (Kumar et al., 2019). Geomorphologically, Senegal is a mostly flat country with a maximum elevation less than 200 m above sea-level (Diouf et al., 2015; Anchang et al., 2019). The northern, central and eastern parts of Senegal are the main zones of pastoral activity and have been divided into 4 ecological zones reflecting soil and land use characteristics: Sandy Ferlo (north), Ferruginous Ferlo (north east), mixed agro-pastoral zone (central) and savanna-woodland transition zone (East) (see **Figure 2**, as well as Diouf et al., 2015). Senegal has an estimated population of 16 million people, with a rural populace that is principally engaged in rain-fed agriculture and pastoral (livestock grazing) activities, both of which directly impact woody resource availability and sustainability. Urban populations also exert notable influence on woody resources through a high household demand for fuelwood and charcoal (2002 Census data from Minnesota Population Center, 2018).

Software Tools

The methodology described in this paper was designed and implemented mostly using Google Earth Engine (GEE; Gorelick et al., 2017), a cloud-based computing platform that allows for planetary scale geospatial data retrieval, processing and analyses. It can be accessed programmatically using a Java code editor browser interface or a Python application programing interface. GEE significantly lowers the technical and infrastructural requirements for geospatial analysis of large areas, as the 'heavy lifting' is carried out by server-side functions. It currently boasts an impressive and constantly improving library of free earth observation/geospatial datasets and open source analytic tools. We used GEE for all satellite data retrieval and preprocessing, model training and validation, and deriving final %WCC maps (see **Supplementary Material** for details on how to access project code materials and an online demo).

Collect Earth Online (CEO)[1] is another free online tool featured in our methods. It is a browser-based adaptation of the Collect Earth desktop tool (Bey et al., 2016). CEO allows for augmented visual analysis of VHR imagery to derive land cover data at the field/plot scale. We used CEO to acquire supplementary measurements for model training and validation.

The use of offline (i.e., desktop) tools was intentionally limited. ArcGIS Pro (ESRI, 2017) was used to prepare GIS point shapefiles of field/CEO data for upload to GEE environment (any open source GIS software, e.g., QGIS, could also be used for this). Python machine learning libraries were also used to reproduce model results on the desktop environment and to access advanced model utilities not currently available in GEE.

Methodology

Workflow Description

Figure 3 illustrates the steps (preprocessing, compositing, and modeling) employed in mapping %WCC from combined optical-radar remote sensing data.

Data Preparation

Plot level woody canopy cover

We combined plot level measurements obtained from the field and from VHR imagery to train and validate empirical models for predicting %WCC. Field data were collected in 2015 from 24 field sites located in relatively homogenous landscapes in the Northern, Central and Eastern regions of Senegal (**Figure 2**; green dots). These sites are part of a long-term *in situ* biomass monitoring effort by the Ecological Monitoring Center (*Centre de Suivi Ecologique* or CSE) located in Dakar, Senegal. Each site is a 1km transect along which four (4) circular plots with radius varying from ∼19 m (totaling 0.5 hectare per site) to ∼28 m (totaling 1 hectare per site) are placed at regular intervals (200, 400, 600, and 800 m), giving a total of 96 plots with data available for our analysis. %WCC is assessed within each plot at the end of the growing season (i.e., peak canopy greenness) every 2 years, through an exhaustive inventory process that includes, amongst other things, measuring the diameter along two axes of the visible crown surface area of every woody plant (Diouf and Lambin, 2001). Normally, the plot data are aggregated for each site to provide estimates of %WCC at the hectare (ha) scale. For our purposes, however, the circular plots are at a scale well-matched to that of medium resolution (<100 m) remote sensing data. Thus, for this study we utilized the non-aggregated (i.e., plot level) measurements obtained during the 2015 field campaign.

Given that the field sites are largely restricted to the drier parts of Senegal (the Ferlo regions, **Figure 2**), field measured %WCC was mostly in the 0 – 60% range. We thus additionally sampled the southern more humid/forested portion of the country (i.e., southeast Senegal and the Casamance region located south of The Gambia). This was done to acquire additional %WCC measurements at the field plot scale, using VHR image data, for reliable predictions in the 60 – 100% range (**Figure 2**; blue

[1] https://collect.earth/

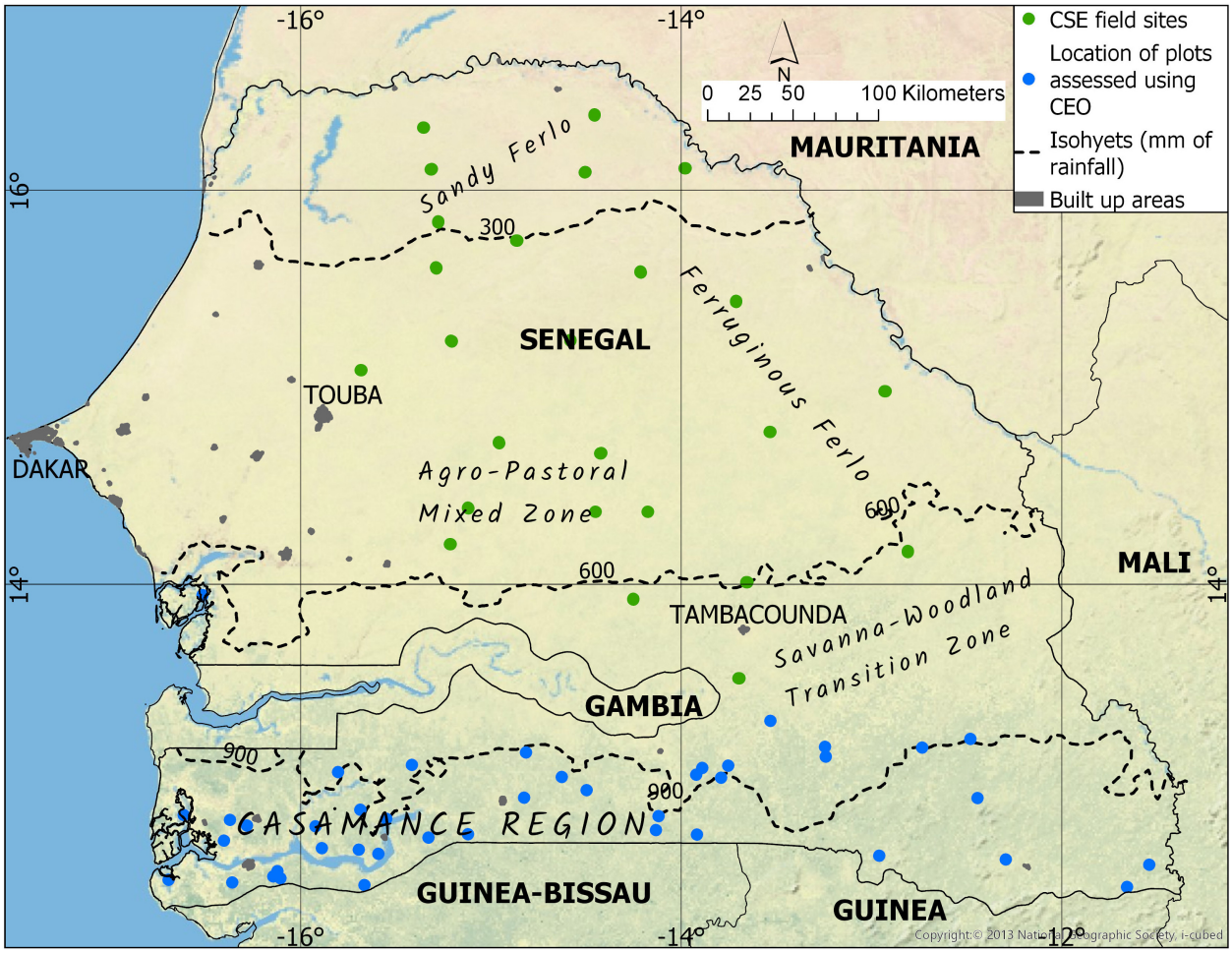


FIGURE 2 | Map of Study Area (Senegal). *Centre de Suivi Ecologique* (CSE) field sites indicated in green. Locations of plots analyzed using very high resolution (VHR) imagery within Collect Earth Online (CEO) indicated in blue. Isohyets (mean annual precipitation in mm for the years 1981–2018) were derived using Climate Hazards Infra-Red Precipitation with Stations (CHIRPS) rainfall data (Funk et al., 2015).

dots). In total 200 random locations were chosen within this region to be assessed using the CEO tool. To be consistent as possible with both the CSE field data and satellite data (described hereafter), we set up a 40 by 40-m rectangular plot at each random point location and filtered the DigitalGlobe imagery to the years 2015–2017. Each plot was populated with gridded sample points spaced 5 m apart inclusive of plot edges (i.e., 9 × 9 = 81 sample points per plot) (**Figure 4**). The %WCC fraction within a plot was determined by labeling each sample point as either covered by a tree/shrub or not and obtaining a tally for the entire plot (1 labeled point = 1/81 or ∼1.23% of cover). To ensure accurate results in the CEO analysis, we only assessed plots with the highest visual quality in the DigitalGlobe imagery available for the given years, leading to a very low retention rate (47/200 or ∼25% of plot data retained). Added to the 96 CSE field plots, this gave us a total of 143 plot level measurements to be used in calibrating our %WCC prediction models.

Satellite data

Using the code editor (Java) interface of GEE, we retrieved all Sentinel-1 (C-band synthetic aperture radar) and Sentinel-2 (optical reflectance) data, acquired at 12- and 5-day intervals respectively within a 3-year period (01 January 2015 – 31 December 2017), and covering the entirety of Senegal (Copernicus Sentinel Data, 2015). We used these data to create gap-free annual composite metrics, aggregated to 40 m spatial scale, that would serve as empirical correlates of per-pixel %WCC. Tree cover mapping studies have shown that using multi-year imagery leads to more accurate and stable predictions than using single date imagery (Karlson et al., 2015; Urbazaev et al., 2015; Brandt et al., 2016a). Longer-term composites are less susceptible to variations in image acquisition conditions that would otherwise influence backscatter/reflectance from single-date imagery or very short-term composites.

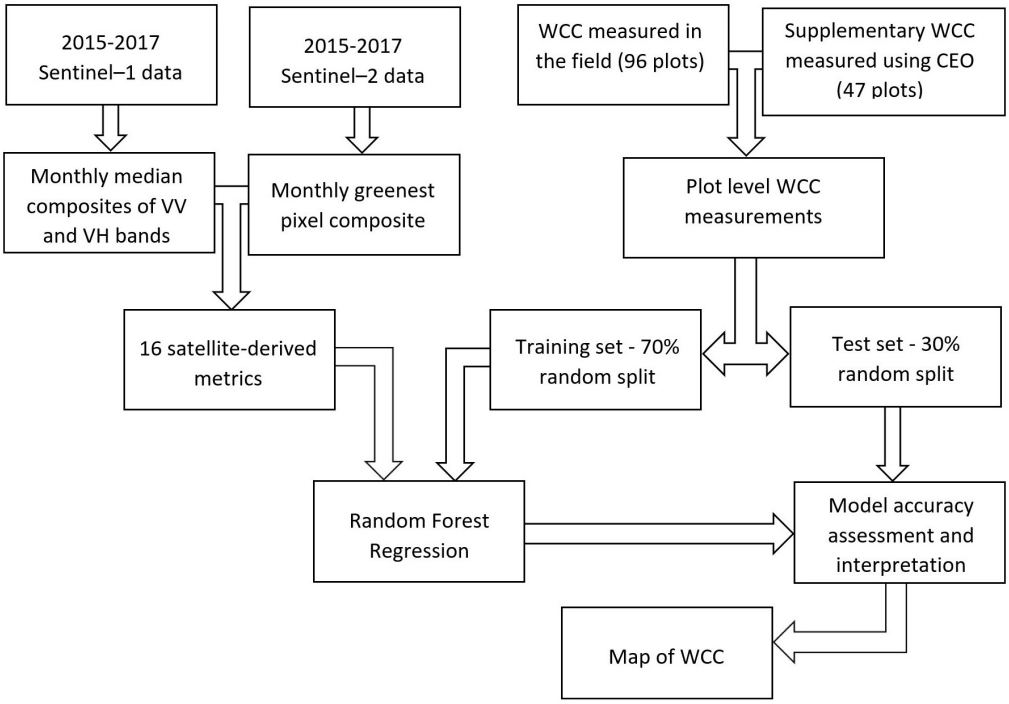


FIGURE 3 | Methodology used to model woody canopy cover (%WCC) from remote sensing data. CEO, Collect Earth Online; VV, vertical-vertical polarization; VH, vertical-horizontal polarization.

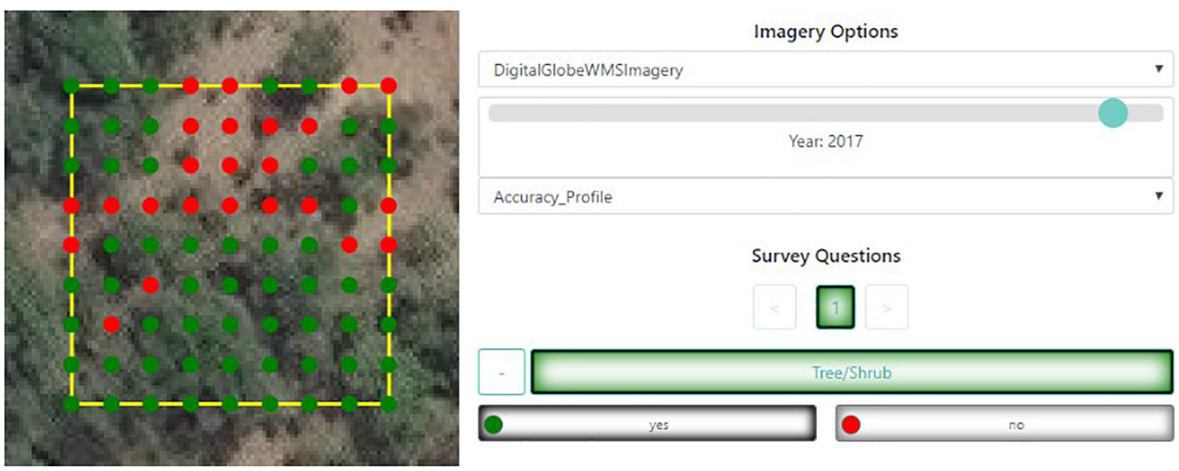


FIGURE 4 | Example of plot level assessment of %WCC using Collect Earth Online (https://collect.earth/). Plot dimensions are 40 m by 40 m with 9 × 9 (81) sample points spaced 5 m apart. In this illustration, the proportion of 'green' dots provides an estimate of percent %WCC within the plot. Base VHR image shown is a 2017 acquisition (Courtesy of DigitalGlobe).

For Sentinel-1 (S1), we specifically used the Interferometric Wide (IW) Ground Ranged Detected (GRD) high resolution (10m) product, with both vertical-vertical (VV) and vertical-horizontal (VH) polarization, acquired in the ascending orbit. S1-GRD data in GEE is already pre-processed as follows: orbital file application, thermal noise removal, radiometric calibration, and terrain correction. Our workflows additionally applied a spectral noise filtering function using the Enhanced Lee Speckle Filter (Lopes et al., 1990) and an incidence angle correction function to minimize inter-scene variation.

For Sentinel-2 (S2), we used the Level 1C (non-atmospherically corrected) product, retrieving spectral bands in the visible (10 m), near infra-red (NIR) (10 m), red-edge (RE) (20 m), and short-wave infrared (SWIR) (20 m) electromagnetic region. We note that GEE has already ingesting atmospherically corrected (Level 2A) S2 products as of early

2019. However, these do not yet offer enough temporal coverage for our current analysis and are reserved for future iterations. For the present analysis, however, we performed the following steps to minimize atmospheric effects and improve the overall quality of composite metrics: (i) cloud and cloud shadow masking, (ii) bi-directional reflectance distribution function (BRDF) correction (Roy et al., 2017a,b), (iii) geometric and topographic correction, (iv) monthly 'greenest' (i.e., maximum NDVI) compositing, (v) use of 'median' instead of 'mean' when averaging to minimize effects of temporal outliers.

Satellite-Derived Metrics as Empirical Determinants of %WCC

A total of 16 remote sensing metrics were generated as a single multiband composite image, at a 40 m resolution scale, and used to provide independent variables for predicting %WCC. These metrics were chosen to be relevant in sensing diverse vegetation properties, including those that are useful in discriminating woody from non-woody plants (**Table 1**). Most importantly, they could be derived directly from S1 and S2 data. The pixel values of each band were then extracted for the 143 plot locations to produce a single data table for modeling.

TABLE 1 | Satellite-derived metrics evaluated for predicting %WCC.

Variable	Description	Units	Relevance	References
med_vv	Median radar backscatter (VV)	Decibels (dB)	Detection of canopy structural and water content properties.	Urbazaev et al., 2015; Baumann et al., 2018; Flores-Anderson et al., 2019
med_vh	Median radar backscatter (VH)			
std_vv	Standard deviation of radar backscatter (VV)		Standard deviation values could capture seasonality to help discriminate woody-herbaceous signals.	
std_vh	Standard deviation of radar backscatter (VH)			
med_nd	Median of normalized difference vegetation index (NDVI) (NDVI = ρNIR-ρRED/ρNIR + ρRED)	Unitless	Detection of green vegetation material	Tucker, 1979; Karlson et al., 2015
med_ndw	Median of normalized difference water index (NDWI) (NDWI = ρNIR-ρSWIR1/ρNIR + ρSWIR1)	Unitless	Detection of plant water content and stress	Gao, 1996
med_ndre	Median of Red-Edge normalized difference Index (RE-NDVI = ρRE2-ρRE1/ρRE2 + ρRE1)	Unitless	Detection of leaf chlorophyll content	Sims and Gamon, 2002
med_swir21	Median of ratio of shortwave infra-red bands (SWIR21 = ρSWIR2/ρSWIR1)	Unitless	Empirical surrogate for cellulose absorption index used to detect senescent biomass high in cellulose content Also potentially correlates to burn activity	Key and Benson, 2005; Guerschman et al., 2009; Hill et al., 2016
max_nd	Maximum of NDVI	Unitless	Intra annual/seasonal leaf phenology to help discriminate woody-herbaceous signal	DeFries et al., 1995; Gessner et al., 2013; Karlson et al., 2015; Brandt et al., 2016a
Min_nd	Minimum of NDVI			
rge_nd	Range of NDVI			
dry_nd	Dry season NDVI (Median NDVI of driest 3 months)			
wet_nd	Wet season NDVI (Median NDVI of greenest 3 months)			
wet_minus_dry_nd	Difference in wet and dry NDVI			
med_red	Median of red band	reflectance	Brightness indices	(Baumann et al., 2018)
med_nir	Median of near infra band		related to albedo change with vegetation cover	

VV, vertical-vertical polarization; VH, vertical-horizontal polarization; NIR, near infra-red; RE1, Sentinel-2 red edge band 1 (~704 nm); RE2, Sentinel-2 red edge band 2 (~740 nm); SWIR1, Sentinel-2 short wave infra-red 1 (~1610 nm); SWIR2, Sentinel-2 short wave infra-red 2 (~2200 nm); ρ, reflectance. Although potentially relevant to tree canopy albedo, metrics derived from blue and green S-2 bands were excluded to minimize effect of atmospheric noise in model input reflectance data.

Random Forest Model Training and Validation

We used random forest regression (Breiman, 2001) within GEE to predict percent %WCC (response variable) from our selected satellite-derived metrics (independent variables, **Table 1**). The random forest technique belongs to the family of ensemble decision tree models where final predictions are obtained by averaging the predictions of multiple individual regression trees. Generalization error in random forests is minimized by increasing diversity among the tree population through the random subsampling of observations and variables (features). Current implementation of the machine learning algorithms in GEE is relatively 'barebones,' with missing utilities such as the ability to visualize variable importance and partial dependence on the fly. As such we also replicated the modeling exercise in a desktop environment using the random forest regressor function provided in the Python machine learning library (Scikit-learn) (Pedregosa et al., 2011). To allow use of python tools to evaluate and fine-tune GEE based models, we established reproducibility of results between GEE and Python for the same data and basic model parameters.

A 70%/30% random split was used to create independent training and test samples, respectively, for the random forest model. Hyper-parameter tuning of the model was done by trial and error: sequentially altering individual parameters (e.g., number of trees) and observing the effect on training root mean square error. We eventually settled on the following model settings for our case of Senegal: 150 trees, minimum leaf prediction size of 4, 1/3 of variables randomly selected per tree, and a bag fraction of 0.9 (i.e., fraction of training sample randomly chosen with replacement for each tree model). The choice of a high bag fraction was necessary due to the relatively small size of our training sample (i.e., 70% of only 143 plot level measurements) ensuring enough data for model learning while allowing for some cross validation within the model. Eventually, independent model validation was done by calculating the root mean square error (RMSE) of predictions on the test sample (plot observations not exposed to model fitting and tuning).

Model Interpretation

In addition to our main objective of accurate %WCC mapping, we also sought to identify the most important satellite-derived determinants of savanna %WCC and the possible causal (statistical or mechanistic) relationships driving the model. We used two model-agnostic interpretation tools (i.e., tools that are

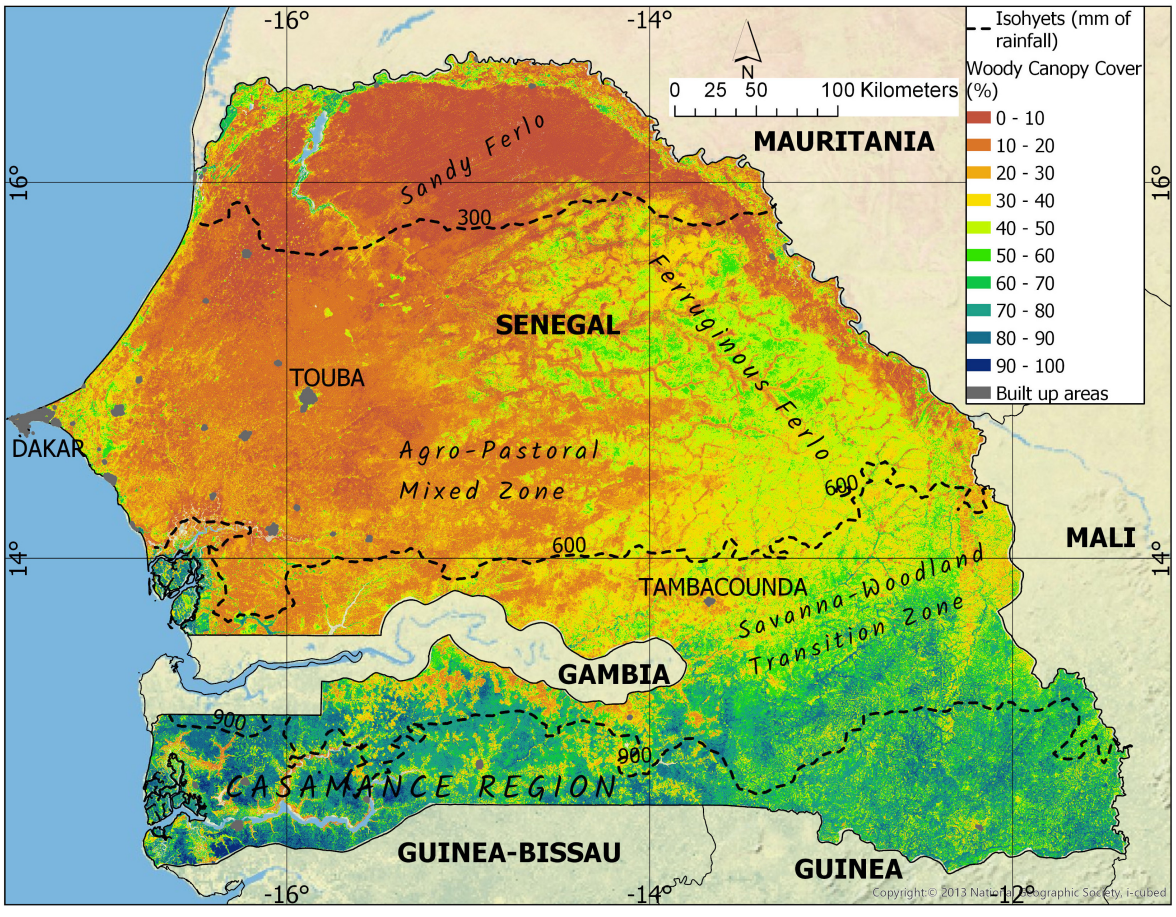


FIGURE 5 | Map of percent %WCC in Senegal, predicted from combined radar and optical remote sensing metrics using a random forest model trained and evaluated using field and VHR imagery data (**Figure 1**).

not specific to any given model), namely variable permutation importance (Breiman, 2001) and the Accumulated Local Effects (ALE) (Apley, 2016; Molnar, 2019). Permutation importance measures the mean decrease in model accuracy (mean increase in prediction error) when a specific variable is excluded, by shuffling its original values to create noise drawn from the same distribution (Breiman, 2001). We calculated permutation importance using only the test sample data, in order to place it in the context of model generalization (Molnar, 2019). By contrast, the ALE plot shows the locally averaged marginal effect of a specific independent variable on predictions of the dependent variable and is an improvement over the more commonly used partial dependence plot (Friedman, 2001). For each observation of a specific predictor (i.e., *x*-value), the ALE plot calculates the average change in the target prediction within a local multidimensional window around *x*-value. This alleviates the requirement for model covariates to be uncorrelated and provides an unbiased visualization of the shape (e.g., linear, monotonic) of the individual predictor-response space. We used ALE plots to infer underlying statistical and/or biophysical drivers of the model, using our knowledge of savannas and radar/optical remote sensing principles.

Finally, we used several sub-models to examine the value of the optical-radar data fusion approach. A separate model was developed for each of the following sets of independent variables: (1) only radar-based metrics, (2) only optical-based metrics, (3) only radar-based median metrics, (4) only optical-based median metrics, (5) only radar-based inter-annual variability metrics, (6) only NDVI seasonality metrics. We compared the performance (RMSE) between sub-models and with the full model to determine which combinations of metrics were most optimal for accurately predicting %WCC.

RESULTS AND DISCUSSION

%WCC in Senegal

Using all available predictor variables, our random forest model was able to predict %WCC for Senegal at 40 m spatial resolution (**Figure 5**) with a high degree of accuracy (training sample RMSE of ∼5%, independent test sample RMSE of ∼8%, **Figure 6**). A visual examination of the final map showed the distribution of predicted %WCC in Senegal to be consistent with what we expect of woody resource distributions across climatic, biogeographic and anthropogenic gradients in Senegal. With the exception of riparian vegetation and irrigated agriculture along the Senegal River in the northern border with Mauritania, predicted %WCC in Senegal generally followed a latitudinal (southward increasing) gradient, supporting the ecological postulation that maximum %WCC in African savannas is constrained by precipitation levels (Sankaran et al., 2005).

In northern Senegal (the region known as the Sandy Ferlo), low %WCC cover was predicted by our model (mostly < 10%, **Figure 5**), consistent with the expectation that low rainfall (MAP ∼300 mm or less) limits the establishment and maintenance of woodland systems. However, despite the low cover, other time-series studies have inferred long-term gains in woody vegetation

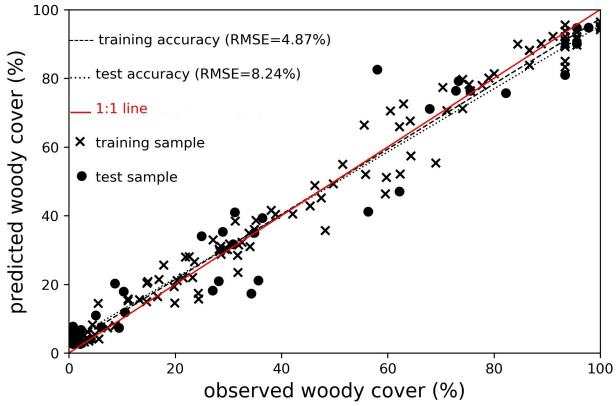


FIGURE 6 | Random forest model training (*N* = 104) and validation (*N* = 39) accuracy for %WCC estimates in Senegal.

in this area, mostly as recovery from 1970s/80s drought events (Kaptué et al., 2015; Anchang et al., 2019), but also due to relatively low human influence in areas that are largely not suitable for agriculture (Brandt et al., 2017). Our ability to detect low cover of trees and shrubs in this region can play a key role in supporting such conclusions in future studies.

In the Sudano-Sahel savanna ecoregion, predicted %WCC was noticeably lower on the western side (mostly < 30%, **Figure 5**), a likely result of the greater population density and prevalence of agricultural activities in this area. Western Senegal is home to large urban population centers like Dakar and Touba, as well numerous other built-up settlements (**Figure 4**, gray colored areas) which exert more pressure on local woody resources due to the greater demand for wood products such as fuelwood and charcoal. By contrast, in the eastern Ferruginous Ferlo and savanna-woodland transition zones (**Figure 5**), predicted %WCC increased to intermediate levels (30 – 60%). In this area, where the urban footprint is considerably less, %WCC is more strongly influenced by herbivory (grazing) and fire activity (Kahiu and Hanan, 2018b).

In the regions of Senegal south of The Gambia (i.e., Casamance region), and furthest to the southeast of Senegal, where MAP > 900 mm, mapping results showed highest levels of %WCC, with pockets of >80% cover predicted in some areas. These high levels of %WCC are evident in the southwestern corner, particularly in the Saloum delta area and along the Casamance River, where riparian/mangrove systems abound. Results also showed, however, instances of fragmentation (breaks in high %WCC) in the southern forested landscape, notably in the southcentral zone (**Figure 5**), a likely outcome of forest clearance for cultivation.

Interpreting Model Predictions of %WCC From Satellite-Derived Metrics

Variable Importance

We used variable importance (permutation importance) scores obtained from the fitted random forest model to examine the importance of individual metrics in accurately predicting %WCC

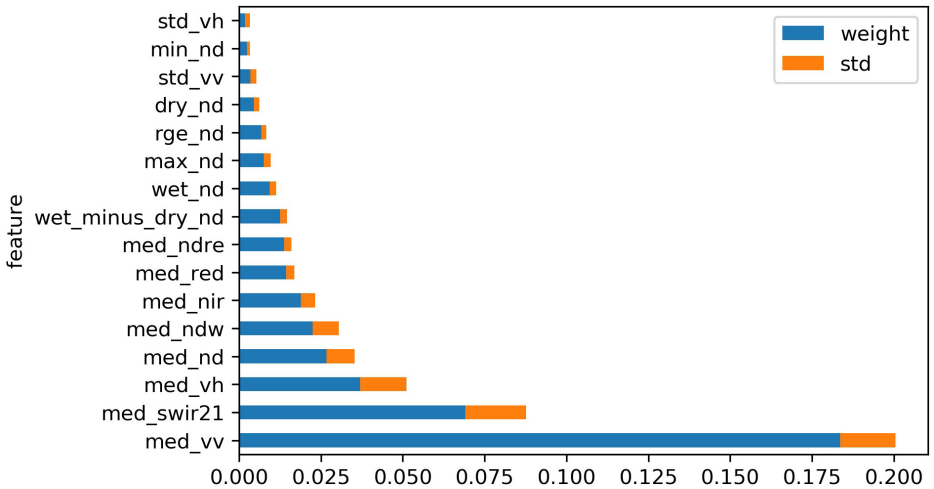


FIGURE 7 | Variable importance in predicting %WCC (refer to **Table 1** for full description of abbreviated variable names). Importance weights (*x*-axis; blue color) reflect the mean increase in test sample prediction error (with associated standard deviation indicated in orange) arising from iterative permutations of each predictor variable (*y*-axis).

in the validation sample (**Figure 7**). Results indicate that the median of VV and VH backscatter (med_vv and med_vh), the median of SWIR21 (med_swir21), and the median of NDVI (med_nd), were the most important predictors. Meanwhile, the standard deviation of backscatter (std_vv and std_vh), the minimum of NDVI (min_nd), and the median NDVI of the driest quarter of the year (dry_nd) were the least important.

The importance of VV backscatter in predicting woody cover in our model (more than twice as important as the runner-up, med_swir21, **Figure 7**) is supported by previous savanna mapping studies. Co-polarized radar (i.e., HH and VV polarization) is known to be sensitive to non-leaf canopy components (such as branches and stems) compared to cross-polarized radar (HV and VH polarization), and hence should be more effective in sensing woody vegetation in the dry regions like the Sahel where most trees are absent leaves at some point during the prolonged dry season (Urbazaev et al., 2015). The second most influential variable was the median of SWIR21, the ratio SWIR band 2/SWIR band 1 (see **Table 1**). Sentinel 2-derived SWIR21 is analogous to the MODIS-derived SWIR32 (ratio of MODIS SWIR bands 3 and 2), which has been found to be correlated with cellulose absorption index (CAI), derived from hyperspectral data and used primarily in remote sensing of dry/senescent biomass (Guerschman et al., 2009; Hill et al., 2016, 2017). We therefore postulate that SWIR21 in our model correlates to the abundance and persistence of (dry) herbaceous biomass, and that its relatively high importance for predicting %WCC arises indirectly from the competitive interactions between trees and grasses in mesic savannas (Scholes and Archer, 1997; Dohn et al., 2013; Kahiu and Hanan, 2018a).

An interesting perspective gleaned from the variable importance plot is that metrics measuring annual central tendency (i.e., 'median') in remote sensing data appeared to be more useful than metrics measuring intra-annual variability (seasonality) (e.g., standard deviation of radar backscatter and the maximum, minimum, range of NDVI). In an attempt to explain this, we consider that the most seasonally variant component of savanna trees and shrubs are the leaves, which are mostly transparent to C-band radar (l∼5 cm) (Flores-Anderson et al., 2019) and depending on phenology (i.e., level of deciduousness), would mostly senesce and fall off during the long dry season. The intra-annual variability of radar backscatter thus offers little in the way of discriminating woody canopies due to the (mostly) unchanging nature of non-leaf components. At the same time, the prevalence of deciduous woody species with varying phenologies (Brandt et al., 2016a,b) combined with long dry periods weakens the model's ability to discriminate woody plants collectively from herbaceous vegetation based on inter-seasonal variations in 'greenness' (NDVI). We anticipate however that precise phenology metrics that capture the onset/length of the greening and leafing (e.g., the MODIS Land Surface Phenology product or MCD12Q2.006; Ganguly et al., 2010; Friedl et al., 2019) may be more useful for mapping %WCC in savannas. However, for our application, these metrics would need to be derived at a spatial scale similar to Sentinel data.

Relationship Between Predicted %WCC and Important Satellite-Derived Variables

We used ALE plots to examine the relationship between individual independent metrics and %WCC predictions (**Figure 8**). ALE plots measure the sensitivity of the dependent variable to a specific predictor, by averaging the change in prediction derived using all values of other variables found within a local window. The ALE plot for Median VV backscatter showed a strong monotonic association with %WCC prediction. The average change in predicted %WCC generally increased with the value of median VV backscatter, most strongly between the values of ∼−16 dB (below which we expect sample plots observations to be mostly void of trees and shrubs) and ∼−10 dB (above which we expect woody canopy saturation) (**Figure 8A**).

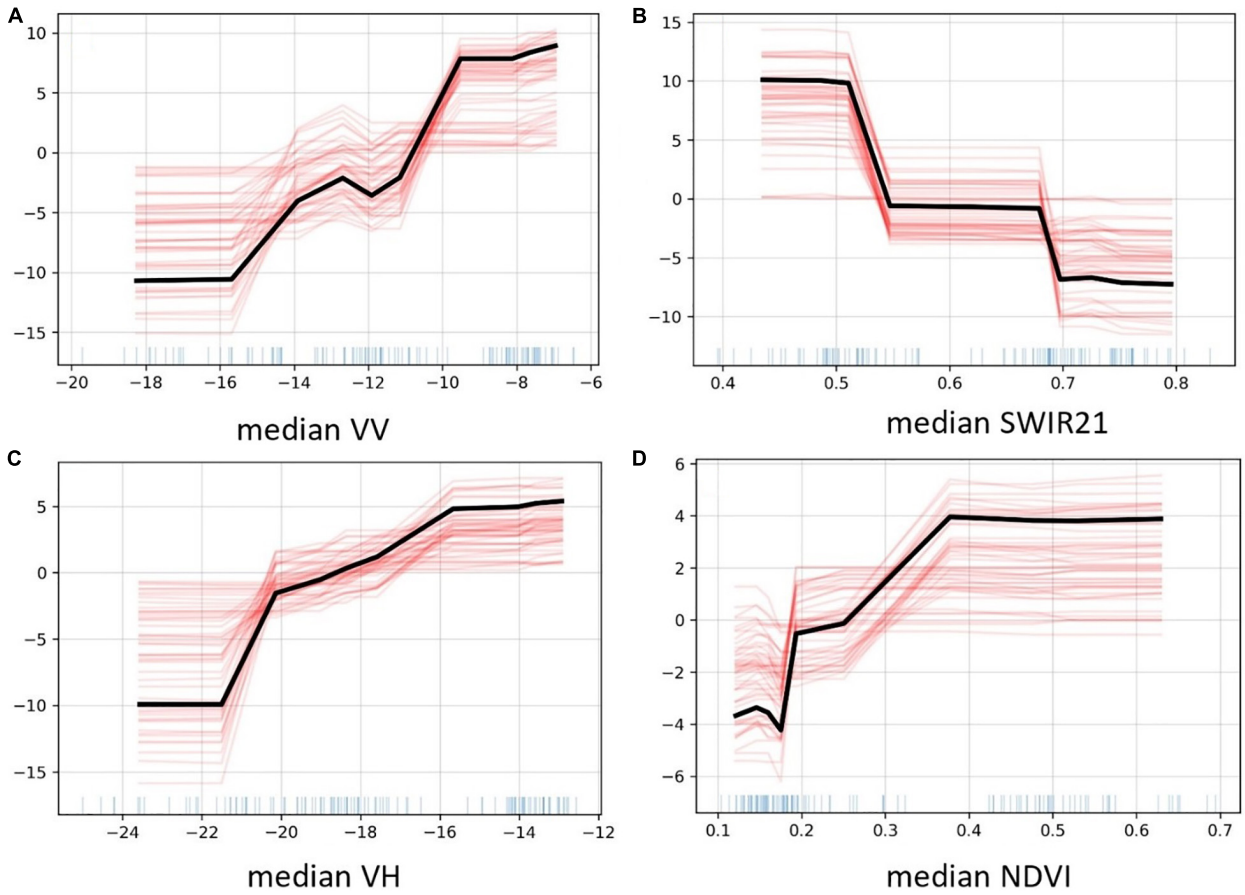


FIGURE 8 | Plots showing Accumulated Local Effects (ALE) plots of percent woody canopy cover (%WCC) prediction sensitivity to the 4 most important independent variables: **(A)** median of VV backscatter in dB, **(B)** median of SWIR2/SWIR1 band ratio, **(C)** median of VH backscatter in dB, **(D)** median of NDVI. *Y*-axis indicates average change in %WCC prediction. Red lines are plots for 50 different Monte Carlo samples drawn from the model training data, with black lines showing response without Monte Carlo sampling. Blue markings on *x*-axis show the distribution of each independent variable.

A similar relationship was observed between change in predicted %WCC and the median of VH backscatter (**Figure 8C**), although the latter's effect on increasing %WCC predictions begins to saturate at much lower values. This supports our findings (from variable importance analysis) of the greater effectiveness of VV polarized radar to sensing woody canopies.

The ALE plot of med_swir21 (median of SWIR2/SWIR1 band ratio) revealed a strong negative effect on %WCC predictions (**Figure 8B**), also supporting its relatively high position (second) in variable importance ranking. As we have previously explained, SWIR21 is used in this case a proxy for detecting dry herbaceous biomass, and so its annual median value should positively correlate with grass cover and production in a given location, and hence may negatively correlate with %WCC due to the dynamics of tree-grass competition (Dohn et al., 2013). Incidentally, the numerator of the SWIR21 ratio (i.e., SWIR band 2, ∼2200 nm wavelength) is also used to calculate normalized burn ratio (NIR-SWIR/NIR + SWIR) in which low values are used detect large area burn scars (Key and Benson, 2005). By extension, this means the model could be picking up the likely correlation between SWIR21 and persistent burning activity, which also negatively impacts woody cover (Scholes and Archer, 1997). The ability of tree-based models to incorporate such latent information and handle interactions is what makes them powerful (though not always transparent) tools for predictions.

As would be expected, median NDVI was positively correlated with %WCC prediction, though with a noticeable saturation at NDVI > ∼0.35 (**Figure 8D**). This is a reminder of the important, but limited, role of spectral indices that respond to only green material in vegetation; they become less effective if used without other data sources in the prediction of the abundance of woody components of the landscape.

Differences in the Performance of Sub-Categories of Satellite-Derived Metrics

The best overall model in terms of accuracy was the full model with all variables present (lowest test sample RMSE of 8.2%, **Figure 9A**), supporting the assertion that combining radar and optical data sources allows for a more accurate mapping of trees and shrubs (Baumann et al., 2018). It also means less important variables were still useful in minimizing the model generalization error and need not be excluded. Although the

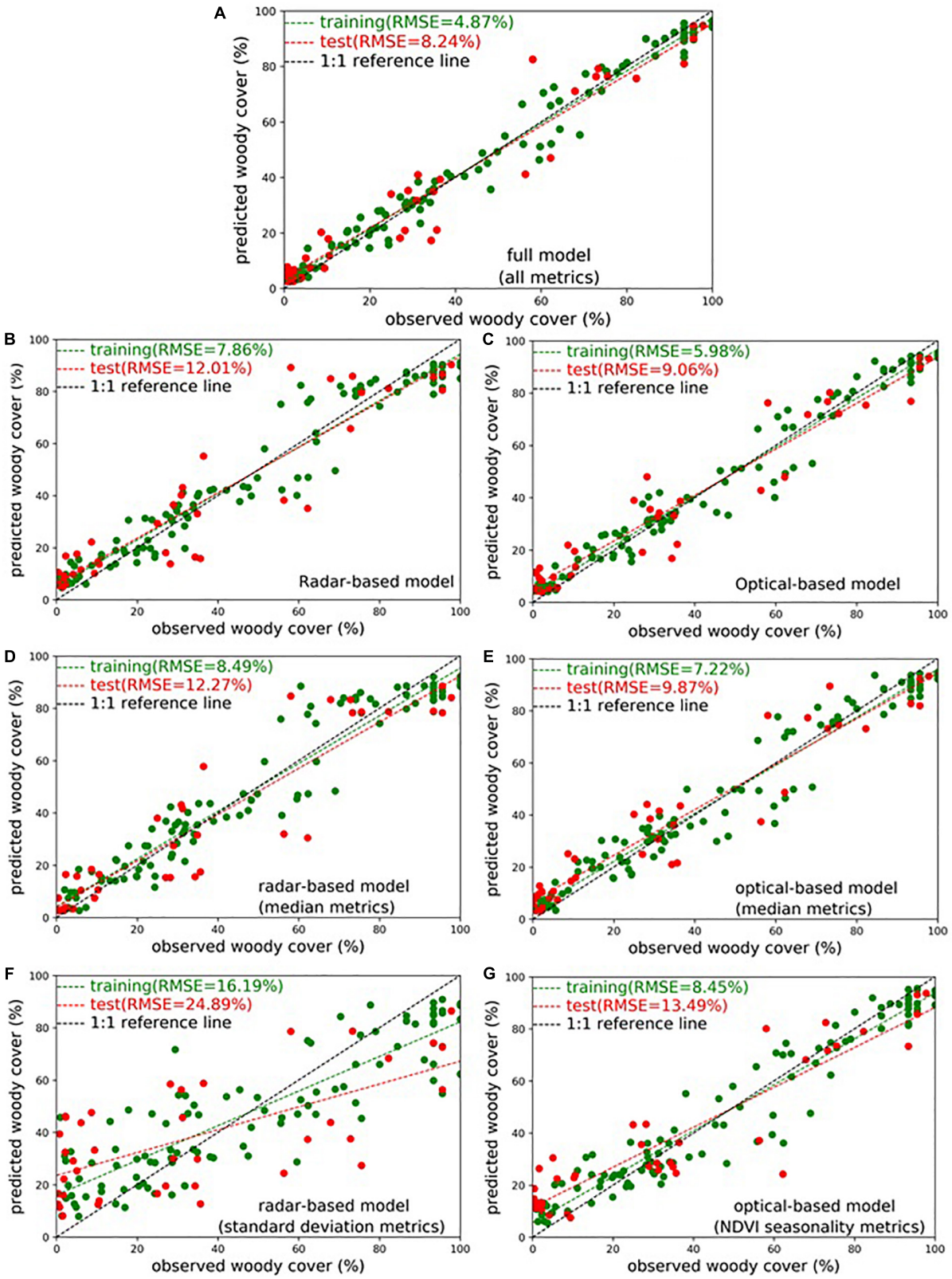


FIGURE 9 | Woody canopy cover (%WCC) prediction accuracy (training = green, validation = red) for models fitted using different groupings of satellite-derived variables: **(A)** Full model with all metrics, **(B)** all radar-based metrics, **(C)** all optical-based metrics, **(D)** radar-based median metrics, **(E)** optical-based median metrics, **(F)** radar-based standard deviation metrics, **(G)** optical-based (NDVI) seasonality metrics.

median backscatter metrics individually ranked high in predictive importance, models using only radar-based metrics did not achieve the highest model accuracy (test sample RMSE of ∼12%, **Figures 9B,D**). In fact, models using optical reflectance metrics, with or without NDVI-based seasonality metrics, collectively outperformed the radar-only models (test sample RMSE of ∼9% and ∼10%, **Figures 9C,E**, respectively). A possible explanation is that the optical reflectance metrics collectively provide more diverse information that is always useful in an ensemble tree-based model. The overall weakness of metrics only capturing intra-annual variability or seasonality in both backscatter and greenness is once again evident (**Figures 9F,G**). As we have explained previously, variability in radar backscatter (particularly the VV band) will not produce a strong signal for discriminating woody canopies as the most seasonally variable component (i.e., small leaves than senesce and fall off during the dry season) are mostly transparent to C-band radar. However, the usefulness of variables that capture annual variability of NDVI for woody-grass differentiation depends on the relative abundance of woody species with long versus short leaf production periods.

Comparisons With Other Existing Woody Canopy Cover Datasets

Our derived %WCC map for Senegal reveals minor similarities and very strong differences when compared to other currently available datasets. Just like its MODIS counterpart, Landsat VCF tree cover data is based on methodology that only considers woody plants greater than 5 m tall, and as such underestimates canopy cover in the savannas by significant margins (in this case indicating < 10% for most of Senegal, and <20% for even the southern densely forested mangrove region in the Saloum delta, **Figure 10C**).

Much closer to the estimates in this study are those by Brandt et al. (2016a), who used FAPAR phenology metrics and regression models to predict mean 2009-2013%WCC at 1km

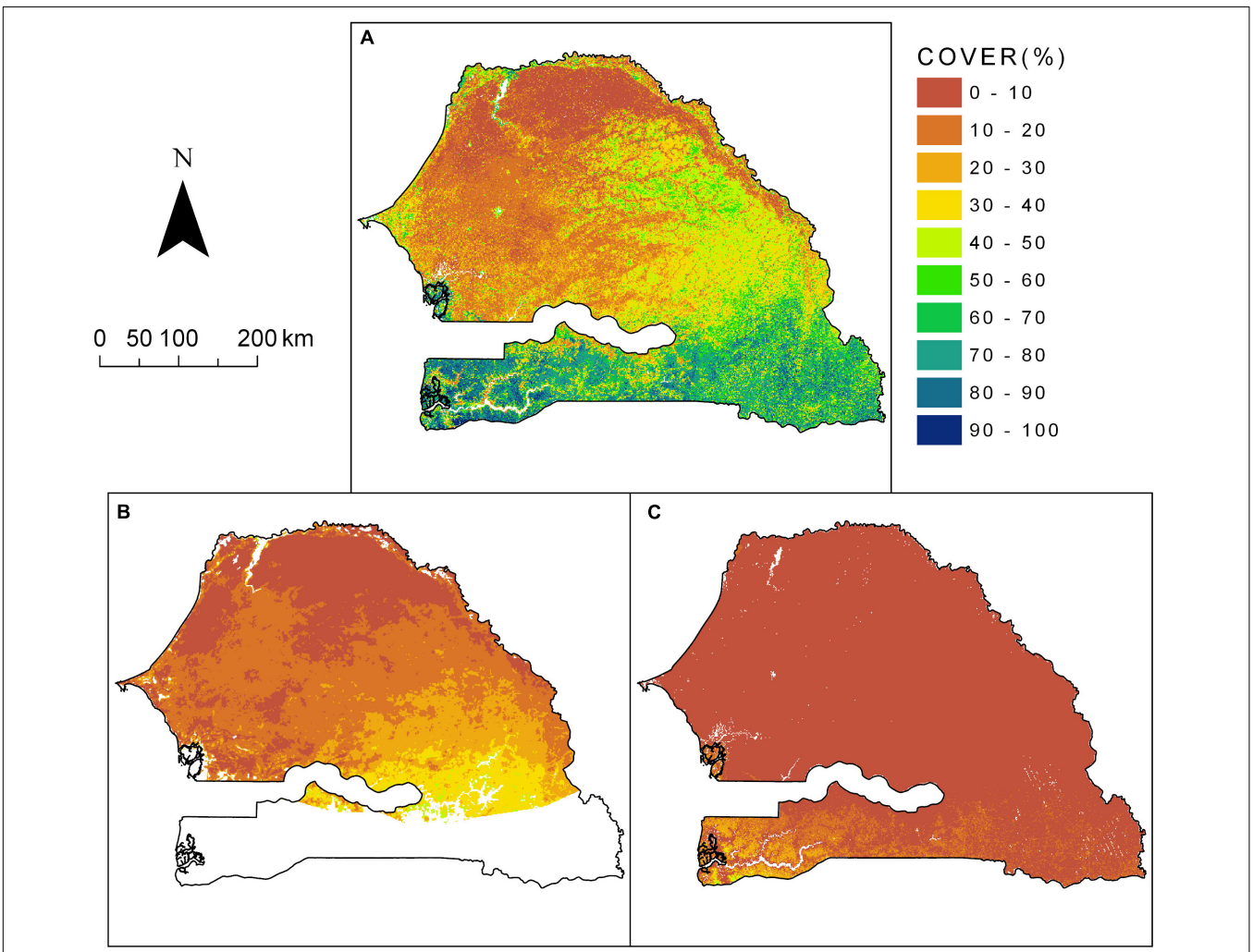


FIGURE 10 | Comparison of different woody cover datasets currently available for Senegal: **(A)** ∼2016 %WCC derived in this study, **(B)** 2009 – 2013 mean %WCC estimated by Brandt et al. (2016a) for the West African Sahel using FAPAR phenology metrics, **(C)** 2015 Landsat VCF tree cover by Sexton et al. (2013). All raster data shown above are resampled to 1 km cell size for direct comparisons and classified using the same color legend.

resolution for the entire West African Sahel (Senegal subset shown in **Figure 10B**). It is worth noting that both studies share commonality in the field datasets used for modeling, with differences in acquisition years and spatial aggregation scale (individual plot versus site scale, see description of field data in section "Data preparation"). Agreements can be seen to a certain extent in the northern and western parts of Senegal, where both products report mostly < 20% of %WCC (compare **Figures 10A,B**). However, although both maps show a steady southward increase in woody cover, a divergence in the values is noticeable in the Ferruginous Ferlo to the east, and the woodlands to the southeast, with differences of up to 20% or more for the same locations. This likely stems from the significant difference in the spatial scale of the models used to derive both maps (40 m vs. 1 km). Our 40 m scale allows us to detect fine but clearly visible patterns of woody occurrence in otherwise open landscapes. For example, for a small area in the eastern Ferlo (**Figure 11**), we correctly detected high canopy cover (50–70%) for a subset of 40 m pixels, leading to a higher (and potentially more accurate) average %WCC estimate when scaled to a larger (e.g., 1 km) area (**Figure 11B**).

Mapping at a coarse spatial resolution may also weaken the ability of phenological metrics, as used by Brandt et al. (2016a), to discriminate woody canopies in principally deciduous landscapes. At 1 km scale, it would be challenging to capture variability in leaf production patterns among woody species. The relative dominance of a specific phenological type (e.g., an *Acacia* sp. with short duration in annual leaf production) would influence model predictions of canopy cover for the area. Meanwhile, the lack of precise phenology metrics in our Sentinel-based model is offset by the higher spatial resolution and the inclusion of radar backscatter metrics which is less sensitive to seasonal leaf dynamics.

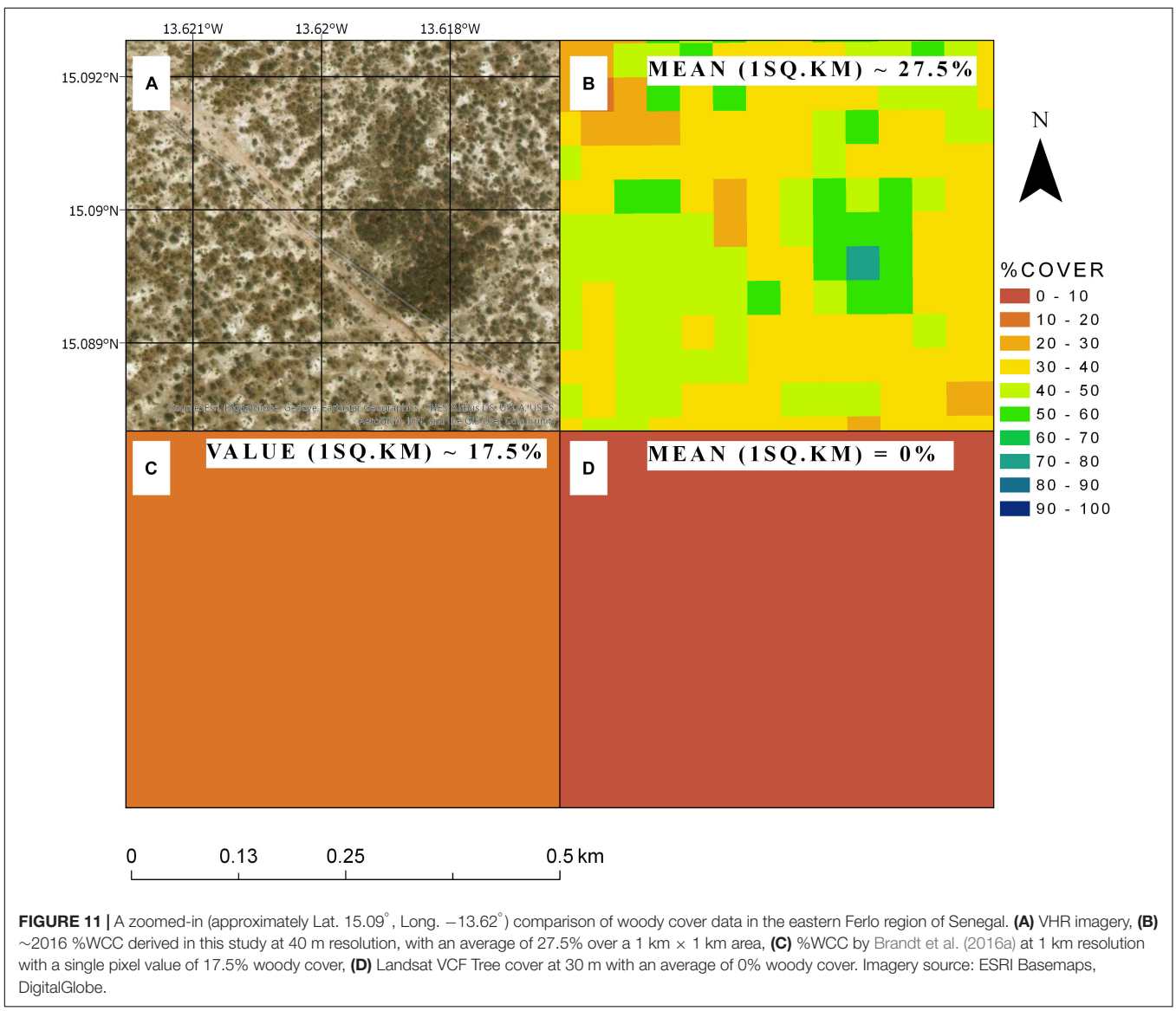


FIGURE 11 | A zoomed-in (approximately Lat. 15.09°, Long. −13.62°) comparison of woody cover data in the eastern Ferlo region of Senegal. **(A)** VHR imagery, **(B)** ~2016 %WCC derived in this study at 40 m resolution, with an average of 27.5% over a 1 km × 1 km area, **(C)** %WCC by Brandt et al. (2016a) at 1 km resolution with a single pixel value of 17.5% woody cover, **(D)** Landsat VCF Tree cover at 30 m with an average of 0% woody cover. Imagery source: ESRI Basemaps, DigitalGlobe.

CONCLUSION

In this paper we described an approach to efficiently and accurately map percent %WCC over large savanna areas. We relied on the use of radar backscatter and optical reflectance metrics as empirical predictors of %WCC in a random forest model. The workflow was implemented almost entirely in Google Earth Engine to leverage the computational power and ease of data-access made possible in geospatial cloud computing applications. The intent being to facilitate adoption by potential users in other countries and regions where low internet bandwidth and limited access to compute facilities might otherwise prevent similar analysis.

Using the full extent of Senegal as a test area, we were able to predict percent %WCC at 40 m resolution with a high degree of accuracy (RMSE of 8.2% with test plot predictions). Exploring the inner workings of the model revealed that median radar backscatter with 'vertical-vertical' (VV) polarization was the most important metric in predicting %WCC, while metrics measuring intra-annual or inter-seasonal variation in satellite-derived information were the weakest; a likely result of the phenological variability within local savanna woody species in West Africa, which complicates their collective separation from herbaceous cover at fine scales solely based on seasonal leaf dynamics. However, our results confirm the findings of other studies that fusion of backscatter and optical reflectance data allows for the most effective mapping of tree and shrub canopies in savannas.

Earth observation data is increasingly important to developing countries seeking more cost-effective tools for monitoring and evaluation in the context of reducing emissions from deforestation and degradation (REDD+), monitoring resource use, and assessing progress toward sustainable development goals and targets. However, countries in regions such as West Africa continue to face challenges in large-scale operational use remote sensing data. In the case of vegetation mapping, the advent of improved and freely available satellite imagery such as the European Space Agency's Copernicus Sentinel data, and cloud computing technologies such as Google Earth Engine, can significantly impact the accessibility of larger datasets and more complex analysis approaches. This will increase capacity of local and regional stakeholders for environmental and natural resource management.

The approach described in this study is designed to fill the need for woody resource mapping tools tailored for tropical savanna regions, where current datasets tend to underperform. Our methodology can be applied across multiple geographic scales, from local to national and regional levels, and is transferable to other areas, ideally using local data and expertise to calibrate and validate predictive models. The remote sensing and cloud-computing approach advocated here makes for a highly automated, scalable and repeatable tool that can allow management agencies and scientists to implement activities for carbon and woody resource monitoring, adapted to regional conditions and local stakeholder needs.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed in this study are available in the following shared Google Earth Engine asset folder: https://code.earthengine.google.com/?asset=users/savannalabnmsu/treecover_data, and also upon request to the corresponding author (anchang@nmsu.edu).

AUTHOR CONTRIBUTIONS

JA wrote the manuscript, conducted data processing and analysis, and developed all scripts and assets. LP and NH acquired funding, conceived the study, and contributed to the manuscript editing. MS and AD helped prepare field data used for modeling and contributed to the manuscript editing. WJ, SK, CR, QY, and BL contributed to the manuscript editing.

FUNDING

This research was funded by the National Aeronautics and Space Administration SERVIR Applied Science Team grant number NNX16AN30G, and the National Aeronautics and Space Administration NASA Carbon Cycle Science program grant number NNX17AI49G.

ACKNOWLEDGMENTS

We acknowledge the field/technical staff of the *Centre de Suivi Ecologique* (CSE) in Dakar, Senegal, for the acquisition and provision of *in situ* woody canopy cover data used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2020.00004/full#supplementary-material

REFERENCES

Anchang, J. Y., Prihodko, L., Kaptué, A. T., Ross, C. W., Ji, W., Kumar, S. S., et al. (2019). Trends in woody and herbaceous vegetation in the Savannas of West Africa. *Remote Sens.* 11:576. doi: 10.3390/rs11050576

Apley, D. W. (2016). Visualizing the effects of predictor variables in black box supervised learning models. *arXiv:1612.08468 [Preprint]* Available at: http://arxiv.org/abs/1612.08468 (accessed May 29, 2019).

Baumann, M., Levers, C., Macchi, L., Bluhm, H., Waske, B., Gasparri, N. I., et al. (2018). Mapping continuous fields of tree and shrub cover across the Gran Chaco using Landsat 8 and Sentinel-1 data. *Remote Sens. Environ.* 216, 201–211. doi: 10.1016/j.rse.2018.06.044

Behling, H., and Hooghiemstra, H. (1999). Environmental history of the Colombian savannas of the Llanos Orientales since the last glacial maximum from lake records El Pinal and Carimagua. *J. Paleolimnol.* 21, 461–476.

Bey, A., Sánchez-Paus Díaz, A., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., et al. (2016). Collect earth: land use and land cover assessment through augmented visual interpretation. *Remote Sens.* 8:807. doi: 10.3390/rs8100807

Bond, W. J., and Midgley, G. F. (2000). A proposed CO2-controlled mechanism of woody plant invasion in grasslands and savannas. *Glob. Change Biol.* 6, 865–869. doi: 10.1046/j.1365-2486.2000.00365.x

Brandt, M., Hiernaux, P., Rasmussen, K., Mbow, C., Kergoat, L., Tagesson, T., et al. (2016a). Assessing woody vegetation trends in Sahelian drylands using MODIS based seasonal metrics. *Remote Sens. Environ.* 183, 215–225. doi: 10.1016/j.rse.2016.05.027

Brandt, M., Hiernaux, P., Tagesson, T., Verger, A., Rasmussen, K., Diouf, A. A., et al. (2016b). Woody plant cover estimation in drylands from Earth Observation based seasonal metrics. *Remote Sens. Environ.* 172, 28–38. doi: 10.1016/j.rse.2015.10.036

Brandt, M., Rasmussen, K., Peñuelas, J., Tian, F., Schurgers, G., Verger, A., et al. (2017). Human population growth offsets climate-driven increase in woody vegetation in sub-Saharan Africa. *Nat. Ecol. Evol.* 1:0081. doi: 10.1038/s41559-017-0081

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Copernicus Sentinel Data (2015). *Google Earth Engine*. Paris: European Space Agency.

DeFries, R., Hansen, M., and Townshend, J. (1995). Global discrimination of land cover types from metrics derived from AVHRR pathfinder data. *Remote Sens. Environ.* 54, 209–222. doi: 10.1016/0034-4257(95)00142-5

Defries, R. S., Hansen, M. C., and Townshend, J. R. G. (2000a). Global continuous fields of vegetation characteristics: a linear mixture model applied to multi-year 8 km AVHRR data. *Int. J. Remote Sens.* 21, 1389–1414. doi: 10.1080/014311600210236

Defries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., and Loveland, T. R. (2000b). A new global 1-km dataset of percentage tree cover derived from remote sensing. *Glob. Change Biol.* 6, 247–254. doi: 10.1046/j.1365-2486.2000.00296.x

Diouf, A., and Lambin, E. F. (2001). Monitoring land-cover changes in semi-arid regions: remote sensing data and field observations in the Ferlo, Senegal. *J. Arid Environ.* 48, 129–148. doi: 10.1006/jare.2000.0744

Diouf, A. A., Brandt, M., Verger, A., Jarroudi, M. E., Djaby, B., Fensholt, R., et al. (2015). Fodder biomass monitoring in Sahelian Rangelands using phenological metrics from FAPAR time series. *Remote Sens.* 7, 9122–9148. doi: 10.3390/rs70709122

Dohn, J., Dembélé, F., Karembé, M., Moustakas, A., Amévor, K. A., and Hanan, N. P. (2013). Tree effects on grass growth in savannas: competition, facilitation and the stress-gradient hypothesis. *J. Ecol.* 101, 202–209. doi: 10.1111/1365-2745.12010

ESRI (2017). *ArcGIS Pro*. Redlands, CA: Environmental Systems Research Institute.

FAO (2000). *FRA 2000: On Definitions of Forest and Forest Change. Forest Resource Assessment Programe* Working Paper No. # 33. Rome: Food and Agriculture Organization.

Flores-Anderson, A. I., Herndon, K. E., Thapa, R. B., and Cherrington, E. (2019). *The SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*, 1st Edn. Huntsville, AL: NASA Marshall Space Flight Center.

Foody, G. M., and Cox, D. P. (1994). Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions. *Int. J. Remote Sens.* 15, 619–631. doi: 10.1080/01431169408954100

Friedl, M., Gray, J., and Sulla-Menashe, D. (2019). *MCD12Q2 MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V006* [Data set]. NASA EOSDIS Land Processes DAAC. Sioux Falls, SD: USGS Earth Resources Observation and Science (EROS) Center. doi: 10.5067/modis/mcd12q2.006

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Statistics* 29, 1189–1232.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* 2:150066. doi: 10.1038/sdata.2015.66

Ganguly, S., Friedl, M. A., Tan, B., Zhang, X., and Verma, M. (2010). Land surface phenology from MODIS: characterization of the collection 5 global land cover dynamics product. *Remote Sens. Environ.* 114, 1805–1816. doi: 10.1016/j.rse.2010.04.005

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266. doi: 10.1016/S0034-4257(96)00067-3

Gessner, U., Machwitz, M., Conrad, C., and Dech, S. (2013). Estimating the fractional cover of growth forms and bare surface in savannas. a multi-resolution approach based on regression tree ensembles. *Remote Sens. Environ.* 129, 90–102. doi: 10.1016/j.rse.2012.10.026

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. doi: 10.1016/j.rse.2017.06.031

Guerschman, J. P., Hill, M. J., Renzullo, L. J., Barrett, D. J., Marks, A. S., and Botha, E. J. (2009). Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sens. Environ.* 113, 928–945. doi: 10.1016/j.rse.2009.01.006

Hanan, N. P. (2018). Agroforestry in the Sahel. *Nat. Geosci.* 11, 296–297. doi: 10.1038/s41561-018-0112-x

Hansen, M., Dubayah, R., and DeFries, R. (1996). Classification trees: an alternative to traditional land cover classifiers. *Int. J. Remote Sens.* 17, 1075–1081. doi: 10.1080/01431169608949069

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., Dimiceli, C., and Carroll, M. (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sens. Environ.* 83, 303–319. doi: 10.1016/S0034-4257(02)00079-2

Hiernaux, P. H., Cisse, M. I., Diarra, L., and De Leeuw, P. N. (1994). Seasonal fluctuations in the leafing of Sahelian trees and shrubs. consequences for the quantification of forage resources. *J. Livest. Vet. Med. Trop.* 47, 117–125.

Hill, M. J., Zhou, Q., Sun, Q., Schaaf, C. B., and Palace, M. (2017). Relationships between vegetation indices, fractional cover retrievals and the structure and composition of Brazilian Cerrado natural vegetation. *Int. J. Remote Sens.* 38, 874–905. doi: 10.1080/01431161.2016.1271959

Hill, M. J., Zhou, Q., Sun, Q., Schaaf, C. B., Southworth, J., Mishra, N. B., et al. (2016). Dynamics of the relationship between NDVI and SWIR32 vegetation indices in southern Africa: implications for retrieval of fractional cover from MODIS data. *Int. J. Remote Sens.* 37, 1476–1503. doi: 10.1080/01431161.2016.1154225

Huete, A. R., Liu, H., and van Leeuwen, W. J. D. (1997). "The use of vegetation indices in forested regions: issues of linearity and saturation," in *Proceedings of the IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing – A Scientific Vision for Sustainable Development*, Vol. 4 (Singapore: Institute of Electrical and Electronics Engineers), 1966–1968. doi: 10.1109/IGARSS.1997.609169

Huxman, T. E., Wilcox, B. P., Breshears, D. D., Scott, R. L., Snyder, K. A., Small, E. E., et al. (2005). Ecohydrological implications of woody plant encroachment. *Ecology* 86, 308–319. doi: 10.1890/03-0583

Jackson, R. B., Banner, J. L., Jobbágy, E. G., Pockman, W. T., and Wall, D. H. (2002). Ecosystem carbon loss with woody plant invasion of grasslands. *Nature* 418, 623–626. doi: 10.1038/nature00910

Kahiu, M. N., and Hanan, N. P. (2018a). Estimation of woody and herbaceous leaf area index in sub−Saharan Africa using MODIS data. *J. Geophys. Res. Biogeosci.* 123, 3–17. doi: 10.1002/2017jg004105

Kahiu, M. N., and Hanan, N. P. (2018b). Fire in sub−Saharan Africa: the fuel, cure and connectivity hypothesis. *Glob. Ecol. Biogeogr.* 27, 946–957. doi: 10.1111/geb.12753

Kaptué, A. T., Prihodko, L., and Hanan, N. P. (2015). On regreening and degradation in Sahelian watersheds. *Proc. Natl. Acad. Sci. U.S.A.* 112, 12133–12138. doi: 10.1073/pnas.1509645112

Karlson, M., Ostwald, M., Reese, H., Sanou, J., Tankoano, B., and Mattsson, E. (2015). Mapping tree canopy cover and aboveground biomass in Sudano-Sahelian woodlands using Landsat 8 and random forest. *Remote Sens.* 7, 10017–10041. doi: 10.3390/rs70810017

Key, C. H., and Benson, N. C. (2005). *Landscape Assessment: Ground Measure of Severity, the COMPOSITE BURN Index; and Remote Sensing of Severity, the Normalized Burn Ratio. FIREMON: Fire Effects Monitoring and Inventory*

System, 2004. Ogden, UT: USDA Forest Service, Rocky Mountain Research Station.

Kumar, S. S., Hanan, N. P., Prihodko, L., Anchang, J., Ross, C. W., Ji, W., et al. (2019). Alternative vegetation states in tropical forests and savannas: the search for consistent signals in diverse remote sensing data. *Remote Sens.* 11:815. doi: 10.3390/rs11070815

Lopes, A., Touzi, R., and Nezry, E. (1990). Adaptive speckle filters and scene heterogeneity. *IEEE Trans. Geosci. Remote Sens.* 28, 992–1000. doi: 10.1109/36.62623

Matthews, E. (1983). Global vegetation and land use: new high-resolution data bases for climate studies. *J. Clim. Appl. Meteorol.* 22, 474–487. doi: 10.1175/1520-0450(1983)022<0474:gvalun>2.0.co;2

Minnesota Population Center (2018). *Integrated Public Use Microdata Series, International: Version 7.1*. Minneapolis, MN: *IPUMS*. doi: 10.18128/d020.v7.1

Molnar, C. (2019). *Interpretable Machine Learning*. Available at: https://christophm.github.io/interpretable-ml-book/ (accessed May 28, 2019).

Nicholson, S. E., and Webster, P. J. (2007). A physical basis for the interannual variability of rainfall in the Sahel. *Q. J. R. Meteorol. Soc.* 133, 2065–2084. doi: 10.1002/qj.104

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Ratnam, J., Bond, W. J., Fensham, R. J., Hoffmann, W. A., Archibald, S., Lehmann, C. E. R., et al. (2011). When is a 'forest' a savanna, and why does it matter? *Glob. Ecol. Biogeogr.* 20, 653–660. doi: 10.1111/j.1466-8238.2010.00634.x

Ratter, J. A., Ribeiro, J. F., and Bridgewater, S. (1997). The Brazilian cerrado vegetation and threats to its biodiversity. *Ann. Bot.* 80, 223–230. doi: 10.1006/anbo.1997.0469

Roy, D. P., Li, J., Zhang, H. K., Yan, L., Huang, H., and Li, Z. (2017a). Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sens. Environ.* 199, 25–38. doi: 10.1016/j.rse.2017.06.019

Roy, D. P., Li, Z., and Zhang, H. K. (2017b). Adjustment of Sentinel-2 multi-spectral instrument (MSI) red-edge band reflectance to nadir BRDF adjusted reflectance (NBAR) and quantification of red-edge band BRDF effects. *Remote Sens.* 9:1325. doi: 10.3390/rs9121325

Sankaran, M., Hanan, N. P., Scholes, R. J., Ratnam, J., Augustine, D. J., Cade, B. S., et al. (2005). Determinants of woody cover in African savannas. *Nature* 438, 846–849. doi: 10.1038/nature04070

Scholes, R. J., and Archer, S. R. (1997). Tree-grass interactions in savannas. *Annu. Rev. Ecol. Syst.* 28, 517–544. doi: 10.1146/annurev.ecolsys.28.1.517

Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., et al. (2013). Global, 30-m resolution continuous fields of tree cover: landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *Int. J. Digit. Earth* 6, 427–448. doi: 10.1080/17538947.2013.786146

Sims, D. A., and Gamon, J. A. (2002). Relationships between leaf pigment content and optical reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* 81, 337–354. doi: 10.1016/S0034-4257(02)00010-X

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150. doi: 10.1016/0034-4257(79)90013-0

Urbazaev, M., Thiel, C., Mathieu, R., Naidoo, L., Levick, S. R., Smit, I. P. J., et al. (2015). Assessment of the mapping of fractional woody cover in southern African savannas using multi-temporal and polarimetric ALOS PALSAR L-band images. *Remote Sens. Environ.* 166, 138–153. doi: 10.1016/j.rse.2015.06.013

Vitousek, P. (1982). Nutrient cycling and nutrient use efficiency. *Am. Nat.* 119, 553–572. doi: 10.1086/283931

Werner, P. A. ed. (2009). *Savanna Ecology and Management: Australian Perspectives and Intercontinental Comparisons*. Hoboken, NJ: John Wiley & Sons.

Wessels, K. J., Colgan, M. S., Erasmus, B. F. N., Asner, G. P., Twine, W. C., Mathieu, R., et al. (2013). Unsustainable fuelwood extraction from South African savannas. *Environ. Res. Lett.* 8:014007. doi: 10.1088/1748-9326/8/1/014007

Xie, Y., Sha, Z., and Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *J. Plant Ecol.* 1, 9–23.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., Gao, F., et al. (2003). Monitoring vegetation phenology using MODIS. *Remote Sens. Environ.* 84, 471–475.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 05 February 2020
doi: 10.3389/fenvs.2020.00007

Hyperspectral Satellite Remote Sensing of Water Quality in Lake Atitlán, Guatemala

Africa I. Flores-Anderson[1,2], Robert Griffin[1,2], Margaret Dix[3], Claudia S. Romero-Oliva[3], Gerson Ochaeta[4], Juan Skinner-Alvarado[5], Maria Violeta Ramirez Moran[5], Betzy Hernandez[1,2], Emil Cherrington[1,2], Benjamin Page[6] and Flor Barreno[5]*

[1] Earth System Science Center, University of Alabama in Huntsville, Huntsville, AL, United States, [2] NASA-SERVIR Science Coordination Office, Huntsville, AL, United States, [3] Centro de Estudios de Atitlán, Universidad del Valle de Guatemala, Solola, Guatemala, [4] Autoridad para el Manejo del Lago de Peten Itza, Flores, Guatemala, [5] Authority for the Sustainable Management of Lake Atitlán Basin and its Environment (AMSCLAE), Solola, Guatemala, [6] Water Resources Center, University of Minnesota, Minneapolis, MN, United States

In this study we evaluated the applicability of a space-borne hyperspectral sensor, Hyperion, to resolve for chlorophyll a (Chl *a*) concentration in Lake Atitlan, a tropical mountain lake in Guatemala. *In situ* water quality samples of Chl *a* concentration were collected and correlated with water surface reflectance derived from Hyperion images, to develop a semi-empirical algorithm. Existing operational algorithms were tested and the continuous bands of Hyperion were evaluated in an iterative manner. A third order polynomial regression provided a good fit to model Chl *a*. The final algorithm uses a blue (467 nm) to green (559 nm) band ratio to successfully model Chl *a* concentrations in Lake Atitlán during the dry season, with a relative error of 33%. This analysis confirmed the suitability of hyperspetral-imagers like Hyperion, to model Chl *a* concentrations in Lake Atitlán. This study also highlights the need to test and update this algorithm with operational multispectral sensors such as Landsat and Sentinel-2.

Keywords: hyperspectral remote sensing, water quality, chlorophyll a concentration, Lake Atitlán, Guatemala

OPEN ACCESS

Edited by:
Niall Patrick Hanan, New Mexico State University, United States

Reviewed by:
Marc Peipoch, Stroud Water Research Center, United States
Qiuyan Yu, New Mexico State University, United States

****Correspondence:***
Africa I. Flores-Anderson
africa.flores@nasa.gov

Specialty section:
This article was submitted to Freshwater Science, a section of the journal Frontiers in Environmental Science

Received: *15 June 2019*
Accepted: *19 January 2020*
Published: *05 February 2020*

Citation:
Flores-Anderson AI, Griffin R, Dix M, Romero-Oliva CS, Ochaeta G, Skinner-Alvarado J, Ramirez Moran MV, Hernandez B, Cherrington E, Page B and Barreno F (2020) Hyperspectral Satellite Remote Sensing of Water Quality in Lake Atitlán, Guatemala. Front. Environ. Sci. 8:7. doi: 10.3389/fenvs.2020.00007

INTRODUCTION

Fresh water bodies provide multiple services ranging from recreation to ecological and economical. In Guatemala, the combination of poor development planning, lack of sewage treatment infrastructure, and overuse of land for agriculture with absent soil protective practices, has led to the degradation of inland water bodies (Perez Gudiel, 2007; Pérez et al., 2011; Romero-Oliva et al., 2014). Lake Atitlán, located in the highlands of Guatemala (14.68 N, 91.16 W) exemplifies a fresh water body subjected to pressures that have increased over the years. Lake Atitlán is the second most visited tourist attraction in the country, as such represents the livelihood of communities located around it. The 15 municipalities in Lake Atitlán's watershed hold about 368,000 inhabitants (INE, n.d.; AMSCLAE, 2017), from which about 170,000 inhabitants are surrounding the lake (INE, n.d.; SIGSA, 2019). In October 2009 the lake experienced a never before seen algal bloom that lasted about 2 months. At its highest point this algal bloom covered about 40% of the 132 square kilometers lake's surface. The bloom was caused by cyanobacteria, first tentatively identified as *Lyngbya robusta* (Rejmánková et al., 2011) and later as *Limnoraphis robusta* (Komárek et al., 2013). This bloom affected the local economy as tourism came to a halt. In addition, the lake is a direct source of drinking water, use with no purification, for two of the municipalities around the lake (Romero Santizo, 2009; Dix et al., 2012b).

The data available for the lake show its degradation overtime (Chandra et al., 2013; AMSCLAE, 2017). Lake Atitlán is considered a unique example of an oligotrophic lake, given its unusual transparencies (Secchi disk transparency) that in the 1960s were as high as 20 m (Weiss, 1971) and in 2010 as high as 15 m (Dix et al., 2012a). The scientific data available for the lake, even though existent, is confined to a few monitoring points measured over a few months per year. Due to limited resources the data is not published on the regular basis. The two most complete studies published for Lake Atitlán's water quality are from Weiss (1971) and Dix et al. (2012a) and both are based on *in situ* standard water quality monitoring methods. Both studies report and provide evidence of the oligotrophic condition of the lake during the dry season. However, there exists a significant time gap between both studies. Weiss and Dix's year long studies showed that the optical properties of Lake Atitlán such as transparency (measured by Secchi disk), vary by season. The rainy season had lower transparencies than the dry season. Weiss found maximum transparencies of 22 m in February 1969, in a station equidistant from San Pablo, San Pedro, and San Marcos towns and Dix et al. (2012a) maximum of 15.5 m in March 2010, at the center of the lake in the station Center Weiss G.

The limited published information available undermines the understanding of the factors promoting the degradation of the lake and the intrinsic behavior and dynamics of Lake Atitlán. Despite recent efforts to invest more in science and technology in Guatemala (UNESCO, 2015), funding for water quality monitoring is minimal. In addition, standard methods to monitor water quality are generally expensive, time consuming and require special equipment and trained personnel, which consequently provides data with limited spatial coverage and temporal frequency (Palmer et al., 2015). Earth observing satellites, on the other hand, can provide a cost-effective solution to Guatemalan authorities and academia to complement water quality measurements on the ground. Several studies exemplified how satellite data can be used for water quality monitoring in inland water bodies, particularly algal bloom monitoring (Vincent et al., 2004; Ogashawara and Moreno-Madriñán, 2014; Watanabe et al., 2015; Page et al., 2018). During Lake Atitlán's algal bloom in 2009 images from NASA Earth observing satellites, such as Landsat, Advance Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and Earth Observing 1 (EO-1) were used to estimate the extent and progression of the algal bloom and provided authorities and general public a complete picture of the event (SERVIR, 2009). In 2011, satellite images were also used to study water surface extent and other minor algal blooms events for the same lake (SERVIR, 2011). Thus far, the common factor among the different analyses using satellite imagery for this region has been their qualitative nature. Still, satellite remote sensing can be more beneficial and actually represent a reliable, quantitative source of water quality parameters that can effectively complement the *in situ* data collected. In addition, the use of satellite imagery to estimate water quality parameters can increase the timeliness of information and can also provide estimates for the entire water body and not only from single points of measurement. To evaluate if the combination of *in situ* measurements and satellite imagery can enhance our understanding of lake's Atitlán dynamics, the present study is aimed to evaluate satellite remote sensing to estimate water quality parameters, specifically chlorophyll-a (Chl *a*) concentration. Chlorophyll concentration is an indirect measurement of phytoplankton biomass (Schalles, 2006). Chl *a* is a dominant light harvesting pigment and is universally present in eukaryotic algae and cyanobacteria (Rowan, 1989). Therefore, all algae, whether toxic or non-toxic, have Chl *a*. In this study we evaluate the applicability of a space-borne hyperspectral sensor, Hyperion, to resolve for Chl *a* concentration. Hyperion was a hyperspectral imager on board of the EO-1 satellite. It is foreseen that this study will contribute to the transition of remote sensing applications from a qualitative to a quantitative nature for algal bloom monitoring and assessment in Guatemala. In addition, this study provides valuable information on the capabilities of hyperspectral satellite data for Chl *a* concentration retrieval, relevant to ongoing and future hyperspectral satellite missions.

Satellite Remote Sensing and Chlorophyll Algorithms

The Chl *a* algorithms in ocean waters are based on a simple interaction of phytoplankton density with water, in which usually blue to green band ratios have a robust and sensitive relation to Chl *a* during low concentrations 1–30 mg/m^3 situations. This relationship becomes less sensitive at higher Chl *a* concentrations (above 30 $mg/m^3 Chla$) and is highly compromised by the effects of colored dissolved organic matter (CDOM) in turbid and optically complex waters (Schalles, 2006). According to Schalles (2006) and Mobley (1995) (1) extremely low Chl *a* concentrations $< 2\ mg/m^3 Chla$ show higher reflectances in the blue part of the spectrum (400–500 nm) and reflectance decreases as wavelength increases, with extremely low reflectance values, near to 0, in the NIR (700–800 nm); (2) Chl *a* concentrations between 2 and 30 mg/m^3 show higher reflectances in the green (500–600 nm) and red bands (600–700 nm), with peak reflectance in the green part of the spectrum; and (3) higher Chl *a* concentrations, $>$ 300 mg/m^3, show peak reflectances in the NIR and minimum high in the green part of the spectrum, the blue and red bands show low reflectances. These principles are used to select bands and develop algorithms to retrieve Chl *a* from satellite images, since it is evident that spectral signature changes depending on the content of Chl *a* in water. Usually local-based algorithms are needed for inland water bodies, and they vary significantly from one site to another since their development is based on the specific optical constituents of a water body. The measurement of Chl *a* in water is commonly used (a) as an indicator to monitor water quality programs in coastal and inland waters, (b) in surveillance programs of harmful algal blooms, (c) and in ecological studies of phytoplankton biomass and productivity (Jordan et al., 1991; Morrow et al., 2000). Moreover, Chl *a* has also been used as an indicator of cyanobacteria (Ogashawara and Moreno-Madriñán, 2014). Satellite remote sensing has been used for decades to estimate Chl *a* concentrations, most notably with operational applications in the oceans (Mobley, 1995; O'Reilly

et al., 1998; Schalles, 2006; Hu et al., 2012). However, significant progress has been made in applying satellite remote sensing in inland water bodies with positive outcomes as described in Palmer et al. (2015) and Bukata (2013). The main challenge to use remote sensing is to isolate the Chl *a* signal from other cell components and other optically active compounds and the effects of the vertical distribution variation of chlorophyll in the water column.

The first satellite sensor developed to evaluate water properties, particularly Chl *a* concentration was the Coastal Zone Color Scanner (CZCS) on board Nimbus 7 and launched in late 1978. A two-band ratio of 443–550 nm was calibrated and routinely used for Chl *a* estimation (O'Reilly et al., 1998). Later, another two operational sensors were also designed to monitor Chl *a* estimations using bands in the blue and the green regions (Sea-viewing Field of view sensor—SeaWIFS- and Moderate resolution Imaging Spectroradiometer — MODIS-) (O'Reilly et al., 1998; Schalles, 2006). The current operational algorithm used for Chl *a* estimation has been updated for the sixth time (version 6) and is generated by the NASA Ocean Biology Processing Group (OBPG). These algorithms are based on a multi-band optimization procedure called OC4 (for Ocean Color 4) and their approach is termed Maximum Band Ratio (MBR).

These operational algorithms are based on comparing blue to green ratios. The largest value of the ratios is used in a fourth order polynomial regression equation as the exponential term in a power function equation. These exponential equations best represent the sigmoidal relationship between Chl *a* and band ratio calculations (O'Reilly et al., 1998).

The operational algorithms for Chl *a* concentration estimations are based on blue to green band ratios and have been generated for oceanic waters which color is dominated by phytoplankton. The good performance of blue and green ratios in oceanic waters is due to the general tendency that as the phytoplankton concentration increases, reflectance decreases in the blue (400–515 nm) and increases in the green (515–600 nm) (Kirk, 1994).

Evaluating all the algorithms utilized to estimate Chl *a* concentration, we can deduct that the majority of passive remote sensing chlorophyll algorithms use either (a) blue to green band ratio, or (b) a NIR/red band ratio, or (c) spectral curvature or slope at different regions of the spectrum to estimate Chl *a*, this last one uses three bands.

In summary, Chl *a* algorithms develop for: (a) pythoplankton dominated-waters are based on blue to green band ratios, and (b) for optically complex waters are based on either a two-band ratio using NIR and red bands or the spectral curvature approach. Even though these approaches are mostly for oceanic waters, they represent the basis to inland fresh water bodies. Moreover, these approaches can be extrapolated onto inland fresh water bodies, per appropriate calibration and validation with *in-situ* data. Per previous studies in the lake (Weiss, 1971; Dix et al., 2012a; Chandra et al., 2013), we can deduct that the color of Lake Atitlán's waters is dominated mostly by pythoplankton during the dry season and during the rainy season, given all the runoff and sediment deposited in the lake, the waters become more optically complex, with a mix of constituents affecting the color of the water.

Area of Study

Lake Atitlán is a tropical mountain lake located in the Department of Sololá, Guatemala at 14.70°N, 91.19°W. Its origin is volcanic and is situated within a caldera that was formed 84,000 years ago (Chesner and Halsor, 2010). Located at 1,565 masl, Lake Atitlàn has a volume of 24 km^3, a maximum depth greater than 300 m with an average of 188 m, and a surface area of about 132 Km^2. Lake Atitlàn is surrounded by three volcanoes and forms part of an endorheic basin, where the point of discharge is the lake and there is not an obvious outflow. However, Weiss (1971) proposed the possibility that the lake discharges through subterranean passages into River Madre Vieja, on the Pacific slope drainage, since chemical water characteristics were similar between the lake and river waters. river waters. **Figure 1** shows the unique topography of the basin and its location.

METHODS

Remote-sensing spectra was collected from Hyperion satellite images and correlated with *in situ* measurements of Chl *a* concentration. Below we explain how *in situ* measurements were acquired, how satellite data was processed and how the algorithm was developed and tested.

Field Measurements

The Centro de Estudios Atitlán from the Universidad del Valle de Guatemala (UVG) collected 40 *in situ* measurements of Chl *a* concentration in synchronization with Hyperion overpasses between January and April 2013. We use the term "*in situ*" in this paper for data that was collected in the field. These months represent the dry season in Lake Atitlán and consequently the months in which the lake's water is the clearest (Weiss, 1971; Dix et al., 2012b). Samples were collected in the field, placed in a cooler with ice and transported to the laboratory where they were filtered the same day. A standard volume of water (180 ml) was filtered through Whatman GFF filters with 7 micrometer mesh and 25 mm diameter. The filters were individually packaged in aluminum foil and frozen for 24 h. Chlorophyll measurements were carried out using 20 ml methanol to extract the pigment during 12–24 h refrigeration in the dark. Readings were carried out using a Turner fluorometer following Standard Method (Eaton, 2005) and based on Ritchie (2006).

The Chl *a* measurements used in this study were collected at the same time than the Secchi disk transparencies. In addition, AMSCLAE supported the collection of *in situ* data, measuring Secchi disk transparencies. A total of five field campaigns were carried out between UVG and AMSCLAE to collect *in situ* measurements of Chl *a* and Secchi disk transparency. **Figure 2** shows the locations of these *in situ* measurements.

The extremely low values of chlorophyll concentration measured during the dry season in Lake Atitlán at the subsurface level (below limit of detection) provided the initial foundation to determine the depth for the *in situ* Chl *a* measurements,

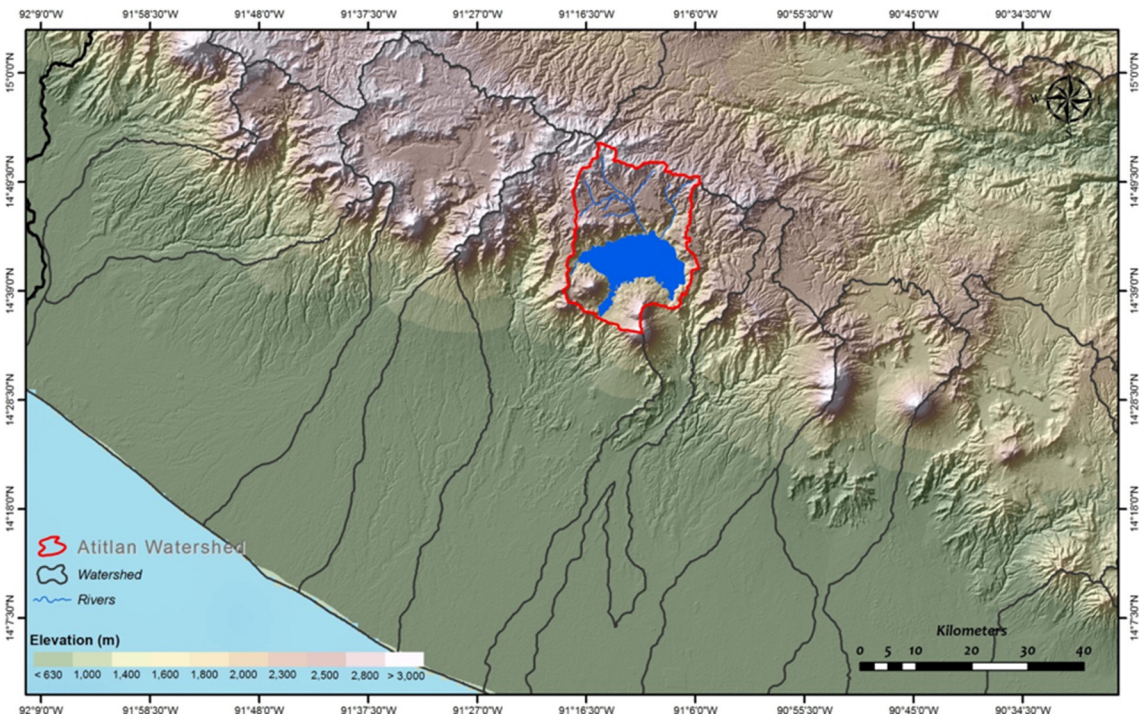


FIGURE 1 | Map of Lake Atitlán Watershed. Topography of the Lake Atitlán watershed located on the highlands of Guatemala. Data source: ASTER GDEM, MAGA. ASTER GDEM is a product of METI and NASA (NASA et al., 2009).

which was based on less than the minimum transparency depth, in this study *in situ* samples of Chl *a* were measured at 3 m depth. Since there are no profile measurements of chlorophyll concentration or any other optically active components available that corresponded to coinciding overpass Hyperion imagery, it is not possible to determine the influence of the vertical structure of chlorophyll concentration on the surface reflectance measured by the satellite. This is a limitation of this study. However, calculation of Kd490 based on Mueller (2000) derived an average value of 0.3/m, and a median of 0.2/m. These Kd490 values indicate that the intensity of the visible light in the blue to green region of the spectrum will be reduced about 0.3 units per meter. Therefore, at 3m depth there is still available light in the blue and green region of the spectrum in the water column. As described by Stramska and Stramski (2005), even the operational empirical algorithms for chlorophyll retrieval (Ocean color algorithms) are affected to an unknown degree by the nonuniformity of Chl *a* profiles.

Satellite Data

Hyperion was a hyperspectral imager on board of the satellite Earth Observing-1 (EO-1), launched in 2000 as part of a 1-year technology validation/demonstration mission (U.S. Geological Survey, 2018). The EO-1 mission was undertaken originally to meet the needs of Landsat continuity program. After the baseline mission of EO-1 was accomplished, NASA approved the Extended Mission operations phase, with the objective of maximizing the use of EO-1 data, in December 2001, and this ran through early 2017, when the satellite was decommissioned. Hyperion has continuous spectral bands of about 10 nm width that cover from 0.4 to 2.5 μm of the electromagnetic spectrum, containing 220 spectral bands. The images have 30 m spatial resolution and cover a swath of 7.6 km. The EO-1 satellite flew in formation with the Landsat-7 satellite in a sunsynchronous, 705 km orbit with an equatorial crossing time 1 min later than that of Landsat-7 (Liao et al., 2000). Level 1 Gst data of Hyperion was used for this study. Level 1 Gst is radiometrically corrected and resampled for geometric correction and registration to a geographic map projection (USGS, 2006). There were two main reasons to use Hyperion in this study, first for research purposes, to assess suitability of a hyperspectral sensor to retrieve Chl *a* concentration and second, due to the ability to task image acquisitions. Hyperion satellites images were tasked at time of *in situ* data collection during cloud free conditions. A strong coordination effort to acquire *in situ* data and satellite data at the same time allowed us to generate a sound scientific dataset suitable for algorithm development.

Atmospheric Correction

The digital numbers obtained from the Hyperion satellite imagery were first converted to top of atmosphere (TOA) radiances and then to reflectance. Reflectance is a dimensionless value obtained from the ratio between the upwelling radiance emittance and the incoming radiant flux (irradiance).

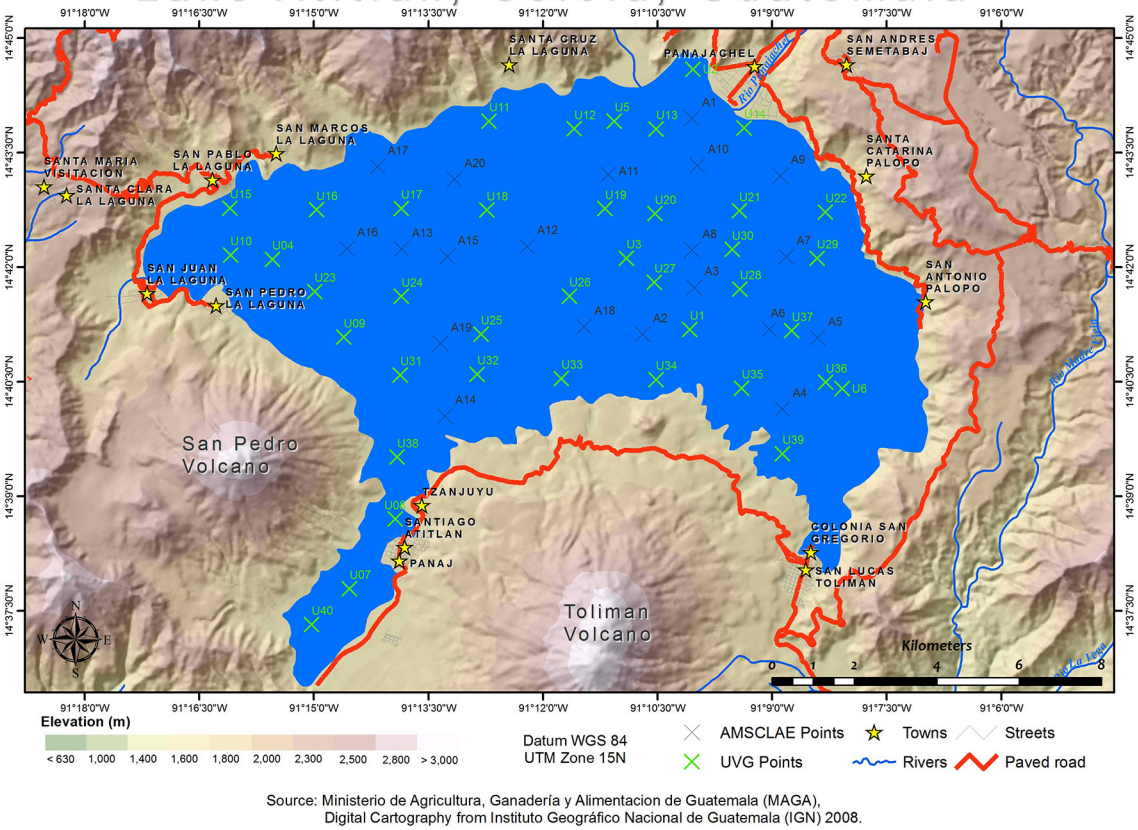


FIGURE 2 | Location of *in situ* samples. Samples were collected by UVG and AMSCLAE. Topography of the Lake Atitlán watershed. Data source: ASTER GDEM, MAGA. ASTER GDEM is a product of METI and NASA (NASA et al., 2009).

The spectral radiance is obtained by radiometrically calibrating the digital number collected in Hyperion Level 1 data. Hyperion data are scaled to limit the amount of saturation and storage space (Beck, 2003). The digital values of the Level 1 product are 16-bit radiances and are stored as a 16-bit signed integer (Beck, 2003). The visible and NIR infrared bands are divided by 40 and the SWIR bands by 80 to de-scale the data and obtain the radiance in $W/m^2 sr\mu m$. Then, Hyperion top of the atmosphere (TOA) reflectance was calculated using the following equation:

$$\rho TOA = (\pi * L_\lambda * d^2)/(ESUN_\lambda * cos\phi_s) \quad (1)$$

where: ρTOA= measured spectral radiance in $W/m^2 sr\mu m$ d = Earth-Sun distance in astronomical units. These values were obtained from the Earth-Sun distance table provided by U.S. Geological Survey (2018) $ESUN_\lambda$ = Mean solar exoatmospheric irradiance phi_s = Solar zenith angle in degrees, obtained from the Hyperion imagery metadata.

The ρTOA calculated using 1 was transformed to surface reflectance to be compared with *in situ* values, for this a radiative model was used. The second simulation of satellite signal in the solar spectrum-vector (6SV) was the radiative transfer model used to account for the atmospheric effects on the signal recorded by Hyperion satellite sensor. According to Vermote et al. (2006) the 6SV radiative transfer code "is the most widely used, rigorously validated, and heavily documented radiative transfer code known in the scientific remote-sensing community." In addition, 6S has been successfully used for remote sensing of water quality (Potes et al., 2012; Shang and Shen, 2016; Markert et al., 2018), which makes it a suitable method for this analysis.

Python code was used to transform ρTOA to surface remote sensing reflectance using 6SV. P6S was run for satellite data accounting for each *in situ* sample.

The following criteria were used to compare satellite and *in situ* data following (Le et al., 2013): (1) both measurements (satellite and *in situ* sample) were acquired within a narrow window of ±3 h (Bailey and Werdell, 2006); and (2) a mean, maximum and median value from a 3 × 3 pixel box centered at the *in situ* sample site was used to filter sensor and algorithm noise (Hu et al., 2001). All the *in situ* samples were acquired in locations where the shallow bottom (depths < 2 m) was not a problem.

Testing Chlorophyll a Algorithms

Existing operational algorithms were tested, including the default blue to green ratio OCx algorithms (O'Reilly et al., 1998; Werdell

TABLE 1 | List of algorithms tested for Chlorophyll *a* estimations.

Reference	Model (bands based on references)	Equation
Blue to green ratio		
O'Reilly et al. (1998), Werdell and Bailey (2005)	$X = 443/550$	$Chla = a + bX$ (linear)
	$X = 490/550$	$Chla = a + bX$
	$X = 510/550$	$Chla = a + bX$
Red to NIR ratio		
Gitelson et al. (2009)	$X = 675/710$	$Chla = a + bX$
Fluorescence line height (FLH)		
Gower et al. (1999)	$X = SS(\lambda) = R_2 - R_1 - (R_3 - R_1)(\lambda_2 - \lambda_1)/(\lambda_3 - \lambda_1)$ $R_2 = 685\,nm$	$Chla = a + bX$
Maximum Chlorophyll Index (MCI)		
Gower et al. (2005)	$X = SS(\lambda) = R_2 - R_1 - (R_3 - R_1)(\lambda_2 - \lambda_1)/(\lambda_3 - \lambda_1)$ $R_2 = 710\,nm$	$Chla = a + bX$
OC_x		
O'Reilly et al. (1998), Werdell and Bailey (2005)	$X = R = log_{10}(R_{550}^{443} > R_{550}^{490})$	$Chla = 10(a - bX + cX^2 + dX^3 - eX^4)$

Band information provided in nm and as Surface Reflectance (SR), OC_x = Ocean Color algorithm.

and Bailey, 2005), the red to NIR band ratios and the three-band approach for spectral shape, specifically the Fluorescence Line Height (FLH) and Maximum Chlorophyll Index (MCI), mathematical formulations are presented in **Table 1**. FLH as described by Gower et al. (1999, 2005) and Wynne et al. (2008) uses a central band at 685nm and it measures the fluorescence of Chl *a*, which produces a narrow peak at this part of the spectrum for Chl *a* concentrations up to 30 mg/m^3. Above this concentration the absorption by water and Chl *a* pigments combine to shift the peak to longer wavelengths (706 nm at 300 mg/m^3) (Gower et al., 2005), which is the central band used in MCI. These last two algorithms use the reflectance height relative to a baseline formed linearly between two neighboring bands which are distributed evenly, hence the priority on the central band. The continuous bands of Hyperion were evaluated in an iterative manner to determine optimal position that would be used in the band ratio approach. Linear, power and polynomial equations were tested to find the best statistical correlations between *in situ* Chl *a* and satellite reflectance. For the algorithm development 75% of the *in situ* measurements (30 samples) were used and for the algorithm evaluation the remaining 25% (10 samples) were used.

Regression Analysis

From the total (40) *in situ* Chl *a* samples one was diskarded since it did not match the correspondent Hyperion image. From the resulting 39 samples, 30 were selected randomly for the algorithm development. The least square method was used for the regression analysis which was tested using linear and polynomial fits. To further select the best algorithm the analysis of variance (ANOVA) was used.

TABLE 2 | Summary of statistics on the *in situ* samples used in this study.

Parameter	Unit	Mean	Min	Max	St Dev
Chl *a*	mg/m^3	5.44	1.01	10.91	2.87
Secchi	m	5.68	3.65	8.00	0.94

Source: UVG, AMSCLAE.

Inference of Best Fit and Algorithm Evaluation

This explains the methodology used to select and validate final algorithm. To validate the resulting algorithm a cross-validation resampling procedure was employed. With a data set of n data stations, the data will be randomly resampled n times, leaving out one station each time, following Chernick (2012). The predictive power of the algorithm will be evaluated using different statistical parameters. The best results obtained from the linear and polynomial regression are evaluated further with other statistical parameters such as the analysis of variance (ANOVA). In the ANOVA analysis the *F*-test (or F-ratio) and *p*-value (or p(F), significance of F) were evaluated. The *F*-test is used to evaluate the hypothesis that all predictor variables under consideration have no explanatory power and that all regression coefficients are zero (Chatterjee et al., 2000).

RESULTS AND DISCUSSION

Field Measurements

Chl *a* measurements collected *in situ* were relatively low, in a range of $1.01 - 10.91\,mg/m^3$ (see **Table 2**). Secchi disk transparency and Chl *a* display modes values of 6 m and 7 mg/m^3, respectively. The datasets represent low optically complex waters, as expected for the dry season, when these datasets were acquired on the field. Forty samples were collected for Chl *a* concentration and 60 samples for Secchi disk transparency. These samples were acquired on January 16, 24, 29; February 22 and April 05, 2013.

Linear Regression

The coefficient of determination, R^2 and the Standard error of estimate were used to assess the results of the linear regression analysis based on band ratios and spectral shape models (FLH and MCI) described in **Table 1**. First, simple surface reflectance (SR) band ratios were assessed and then log-transformed ratios were used. Log-transformation datasets are used in the development of ocean color algorithms (O'Reilly et al., 1998). Extremely low R^2 were obtained using linear regressions for both approaches, the highest R^2 was 0.302, representing not good fit. However, given the multiple bands assessed in this study, this step was used to pre-select band ratios that will be further tested in polynomial regressions. The hyperion bands assessed were in the blue and red-edge part of the spectrum.

Polynomial Regression

Polynomial regressions using a third order polynomial fit following the methods of the ocean color algorithms (O'Reilly et al., 1998; Werdell and Bailey, 2005) were tested. See OC_x model in **Table 1**.

TABLE 3 | Analysis of variance (ANOVA) for the best fit.

Algorithm	Dataset	R^2	Tabulated $F_{(0.01,3,26)}$	F-ratio	ρ-value
Polynomial Log(R467/R559)	Median values	0.7027	4.637	20.486	5.04×10^{-7}
Polynomial Log(R467/R548)	Median values	0.7073	4.637	20.945	4.13×10^{-7}
Polynomial Log(R467/R559)	Mean values	0.7054	4.637	20.752	4.49×10^{-7}
Polynomial Log(R467/R548)	Mean values	0.6999	4.637	20.215	5.69×10^{-7}

Tabulated F-value from ***Table A1*** *in Chatterjee et al. (2000).*

O'Reilly et al. (1998) and Werdell and Bailey (2005) reported the SR443/SR555 ratio maximal to be at Chl $a < 0.3\ mg/m^3$, SR490/SR555 ratio was maximal between 0.3 and 2.0 mg/m^3 and the SR510/SR555 was maximal above 2.0 mg/m^3 and below 30 mg/m^3. Given the Chl *a* concentration values obtained in the study (1–11 mg/m^3), in theory the analogous band ratios used to represent SR490/SR555 and SR510/SR555 (SR487, SR498/SR559 and SR508/SR559, respectively) should have performed better to simulate the Chl *a*. Nonetheless, our best fit was obtained with SR467/SR555, using all datasets. The regression was tested using different combinations of bands. Two datasets, one with all stations containing the whole range of Chl *a* concentration (^{a}), and another with Chl *a* concentrations $< 9\ mg/m^3$ (^{b}) were assessed. Overall, group b had better fits, this is due to the nature of the data, since most of the observations were in this range. This also implies that the algorithm resolves better real Chl *a* conditions that are lower than 9 mg/m^3. Band ratios SR487/SR559 and SR498/SR559 provided good R^2, between 0.65 and 0.66, using mean reflectance values, but were outperformed by SR467/SR555 with an R^2 of 0.7. The 3-band algorithm, FLH, did not improve significantly in the polynomial regression, R^2 of 0.52. Given the overall good performance of the polynomial regression, the band ratios with the best R^2 in this step were selected for further evaluation.

Inference of Best Fit

A summary of the results for the analysis of variance (ANOVA) for the polynomial regressions that have the higher R^2 are shown in **Table 3**. All the algorithms have larger F-ratios than the one tabulated for the significant level of 0.01. Therefore, it can be stated that the results are significant at level 0.01. Consequently the hypothesis that the predictor variables have no explanatory power can be rejected. This means that the Chl *a* concentration can be modeled using a blue to green band ratio. A blue to green band ratio satisfactorily explains the behavior of Chl *a* in the waters of Lake Atitlán at a significant level of 0.01. Evaluating the results from the mean and median values it can be deducted that the third order polynomial of the ratios 467/559 and 467/548 have the best results. The F-ratio is larger than the tabulated one and the ρ-values are significantly smaller than 0.01, see **Table 3**. **Figures 3**, **4** show the polynomial regression graphs for band ratio 467/559, both using the median and mean values datasets respectively.

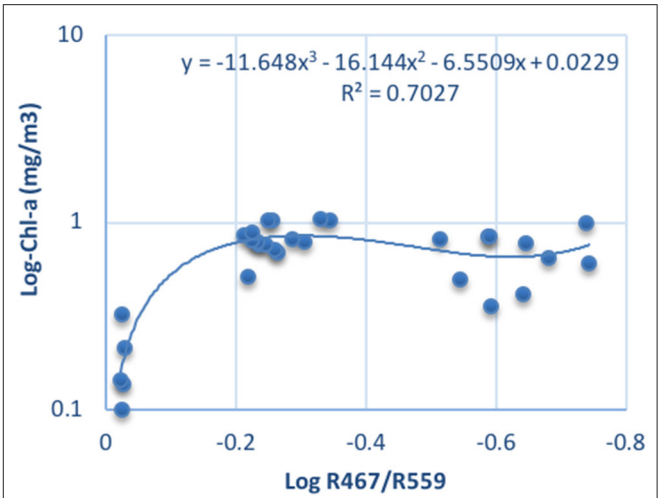


FIGURE 3 | Logarithm plot using median values dataset.

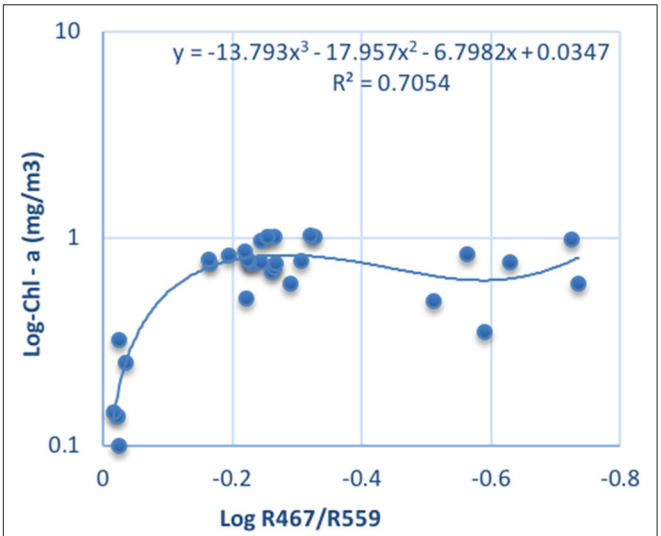


FIGURE 4 | Logarithm plot using mean values dataset.

This is the algorithm selected in this study for chlorophyll a concentration estimation in Lake Atitlán.

Evaluation of the Algorithm

A cross-validation resampling technique was performed to generate a data set that will be used to assess the adequacy of the model. The leave-one-out method was used for this purpose (Chernick, 2012). In this method the complete data set (n = 39) conformed by the data points used for the algorithm development (30) together with the other points that were left out of the algorithm development (9) are randomly resampled n times leaving one point out (n−1). Then the accuracy of the model is tested on the data point left out. This method was used since the number of samples for validation was too small (9 points) which can result in a very large bias (Chernick, 2012). Multiple statistical parameters were used to evaluate the

TABLE 4 | Statistical results for the model evaluation.

Dataset	MRE (%)	RMSE mg/m^3	BIAS mg/m^3	PBIAS (%)	RSR (dimensionsless)	NSE (dimensionsless)
Cross validation	35.12	2.468	±0.2067	−10.21	0.913	0.325
Validation	31.58	1.542	±0.3189	−8.91	0.629	0.603
Average	33.35	2.005	±0.2628	−5.15	0.771	0.464

Cross validation data set, n = 39. Validation data set, n = 9.

model performance. The Mean relative Error (MRE) is the ratio of the absolute error to the real observed measurement, which is assumed to be error free. The MRE provides an estimate of how relevant is the absolute error. MRE is presented as a percentage. The Root mean square error (RMSE) is a common measure to evaluate model performance (Willmott et al., 1985; Moriasi et al., 2007). RMSE has the advantage of indicating an error in the units of the constituent of interest, in this case mg/m^3. However, its disadvantage lies in that large errors are weighted heavily, producing a large RMSE even if there are small errors in a good portion of the data. RMSE values of 0 indicate a perfect fit and the general interpretation is that the smaller the RMSE the better the model performance. Given the uncertainty of what is considered a low RMSE another model evaluation statistic is used to aid the interpretation of RMSE, the RMSE-observations standard deviation ratio (RSR). RSR standardizes RMSE using the observations standard deviation and is calculated as the ratio of the RMSE and standard deviation of observations (Moriasi et al., 2007). The closer the RSR is to zero (0) the better the model performance. The bias error (BIAS) indicates an average model "bias"; that is average over- or under prediction (Willmott and Matsuura, 2005). The percent bias (PBIAS) measures the average tendency of the simulated data to be larger or smaller than their observed counterparts (Gupta et al., 1999). The optimal value of PBIAS is 0.0. Positive values indicate model underestimation bias, and negative values indicate model overestimation (Gupta et al., 1999; Moriasi et al., 2007). PBIAS is the deviation of the data being evaluated, expressed as a percentage (Moriasi et al., 2007). The Nash-Sutcliffe efficiency is a normalized statistic that determines the relative magnitude of the residual variance ("noise") compared to the measured data variance ("information") (Nash and Sutcliffe, 1970; Moriasi et al., 2007).

Table 4 displays the results of the model evaluation statistics. In general the results obtained from the validation and cross-validation data sets agree. It is important to mention that in this analysis the algorithm development and evaluation is being done with satellite data. Meanwhile, the majority of the algorithms developed for water quality parameters had been generated using *in situ* measured reflectance (O'Reilly et al., 1998; Schalles, 2006). The evaluation of these algorithms performances is also commonly done using *in situ* measured reflectance (Le et al., 2013). Our analysis portrays a different approach for algorithm development and evaluation, for an understudied area where, *in situ* reflectance is not available, which can be more easily replicated using a variety of satellite remote sensing data in future research projects.

The algorithm assessed had a MRE error of about 33%. Le et al. (2013) reported its lowest MRE value of 25.66% for a 2-band algorithm using *in situ* reflectance Rrs(λ) in the Tampa Bay area. Le et al. (2013) also reported higher MRE values for 3-band and 4-band algorithms with 50.68% and 48.05%, respectively, using *in situ* reflectance as well. Le et al. (2013) evaluated algorithm performance using real satellite data from MODIS and MERIS. There was no meaningful statistical relationship between *in situ* measurements of Chl-a and coincident MODIS reflectance. This was attributed to atmospheric correction errors in MODIS data and abscence of appropiate spectral bands. The algorithm evaluation performed by Le et al. (2013) using real satellite data from MERIS provided better statistical results, with a MRE of 35.33% for the 2-band algorithm and 46.93%, and 69.15% for the 3-band and 4-band algorithm, respectively. The 2-band algorithm evaluated by Le et al. (2013) was a NIR-red band ratio generated for optically complex waters with Chl *a* concentrations that ranged between 2 and 80 mg/m^3.

However, the MRE was degraded when using the MERIS satellite data (35.33%) compared to using the *in situ* reflectance (25.66%) due to spatial patchiness, differences in Chl *a* concentrations between *in situ* measurement and satellite over pass (Chen et al., 2010; Le et al., 2013) and even atmospheric correction errors (Le et al., 2013). All of these sources of error apply for the analysis performed in Lake Atitlán, and the MRE obtained for the Hyperion data (33%) is very similar to that observed by Le et al. (2013) with MERIS data (35.33%). The algorithm performance results obtained by Le et al. (2013) provide a source of reference for our analaysis, since they are reported for both *in situ* reflectance measurements and satellite-derived reflectance measurements. The RMSE obtained for the algorithm evaluated in this study had an approximate value of 2.0 mg/m^3, which is very similar to the optimal obtained by Le et al. (2013) using MERIS-derived data, 2.01 mg/m^3. To further support the interpretation of the RMSE, the RSR statistic was calculated. An optimal value of RSR would be zero, which indicates zero residual variation, and therefore perfect model simulation (Moriasi et al., 2007). The average RSR value obtained for the algorithm performance evaluation is of 0.77 (see **Table 4**) which is close to zero. The BIAS and PBIAS measurements indicate that overall the model evaluated is overestimating by 0.26 mg/m^3 the simulated Chl *a* concentration values. Finally, the NSE statistic value of 0.46 (see **Table 4**) falls into the range to determine that the model has an acceptable performance.

Discussion

The data set utilized for algorithm development represented a small range of Chl-a concentration (1–10 mg/m^3), which limits the application of the algorithm generated within this concentration range. The final algorithm selected to simulate Chl *a* concentration follows the form of the Ocean color algorithms (OCx), of a polynomial regression fit using blue and green bands. Band ratios of blue and green bands are used for low Chl *a* concentrations, such as the measured in Lake Atitlán in which the major constituent driving the color of water is chlorophyll. The latter was also true for the conditions under which the algorithm was developed in Lake Atitlán. Even though

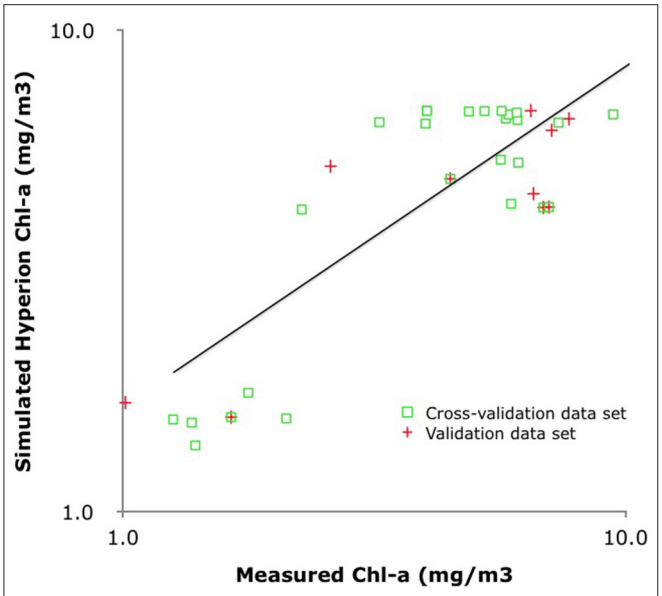


FIGURE 5 | Scatter plot of *in situ* Chl-a from the data sets used for model evaluation and algorithm using satellite reflectance Rrs(λ).

the overall performance of the algorithm selected provided an overestimation of the simulated Chl *a*, the algorithm also presents an under estimation of Chl *a* concentrations when *in situ* Chl *a* was higher than 9 mg/m^3, **Figure 5** portrays this. It is recommended to use a larger calibration and validation data set to fit and test the algorithm to a broader Chl *a* concentration range.

The blue to green ratio selected in this study represents the theoretical behavior of the water dominated by Chl *a*, in which the Chl *a* absorption and scattering characteristics strongly influence the spectral signature of the water. For example, in these waters Chl *a* absorbs in the blue part of the spectrum (400–500 nm) and scatter in the green part of the spectrum (500–600 nm) with an evident reflectance peak at about 550 nm (Schalles, 2006). The band ratio selected in this study SR467/SR559 represents these spectral regions. As recorded by Schalles (2006) this absorption in the blue and scatter in the green is consistent from low (3.3 mg/m^3) to high (more than 60 mg/m^3) Chl *a* concentrations. In theory these blue to green ratios are applicable for water dominated by Chl *a* in concentrations that range from low to high Chl *a* above 60 mg/m^3. However, the evaluation performance of the algorithm in this analysis exhibited larger errors especially in low and high Chl *a* concentrations, for our case the low was defined $<$ 2–4 mg/m^3 and high $>$ 9 mg/m^3, see **Table 2**. This is due to the low representation of these values for the algorithm development and it can be deduced that the majority of data sets are between 6 and 8 mg/m^3.

CONCLUSIONS AND FUTURE PERSPECTIVES

After evaluating the different wavelengths and band ratios used in previous studies to model Chl *a* in Lake Atitlán, the best result was obtained by a blue to green band ratio. The final algorithm obtained had relative error of 33% and assumes that the color of the water in Lake Atitlán is mainly driven by phytoplankton. Given that the data sets used to generate and evaluate this algorithm reflect a small range of Chl *a* concentration (1–10 mg/m^3) it is expected that the relative error will increase when the algorithm is applied in Chl *a* concentrations $>$ 10 mg/m^3. In addition, the analyses performed confirmed the suitability of Hyperion satellite images to model Chl *a* concentrations in Lake Atitlán. The developed algorithm was applied to experimentally estimate Chl *a* in Lake Atitlán during a bloom event, using Landsat OLI data in August 2015. The water quality parameters of Lake Atitlán present a seasonal pattern that is related to the dry and rainy season of the area. At the end of the rainy season lower transparencies and higher Chl *a* concentrations are recorded. Meanwhile, at the end of the dry season higher transparencies and lower Chl *a* concentrations are recorded. Thermal stratification and mixing processes take place seasonally in Lake Atitlàn but given the limited measurements in the lake this seasonality is not well characterized yet (Dix et al., in preparation).

The results of this study represent an early effort, known to the authors, to use satellite images to quantitatively monitor water quality parameters in an inland water body in Central America. The only existing applications of satellite remote sensing for water quality monitoring has been limited to qualitative applications in this region. Therefore, the results of this sutdy demonstrate that Chl *a* can be estimated from hyperspectral imagers like the now-defunct Hyperion, with a relative error of 33%. The algorithm developed is relevant for the current Venμs satellite, DLR Earth Sensing Imaging Spectrometer (DESIS) and the upcoming Hyspiri mission, all hyperspectral, that are or will capture data useful to monitor Chl *a* concentration in Lake Atitlan using the developed algorithm. Given the rapid eutrophication of fresh water bodies world wide (Ho et al., 2019), it is imperative to develop cost-effective methods, such as the one presented in this work, that allow local water managers monitor water quality of their fresh water resources.

The error obtained is slightly below the desired error set by NASA's Ocean Biology and Biogeochemistry Program of 35% (McClain et al., 2006). This confirms the adequacy of the results. Other studies that retrieve Chl *a* concentration in fresh water bodies using multispectral sensors show low degree of certainty (Boucher et al., 2018). However, better performance was obtained when time window of multispectral satellite data acquisition was closer to acquisition of *in situ* data used in correlation. This highlights two issues, (1) the high applicability of hyperspectral sensors and use of narrow spectral bands to retrieve Chl *a* concentration and (2) the relevance of methodologies encompassing same time/date acquisition of satellite data and *in situ* observation to develop algorithms to retrieve water quality parameters from satellite images.

The main sources of error for this algorithm stem from the estimation of surface reflectance obtained from the satellite image and the optical variability presented in the *in situ* data set that was used to generate the algorithm. The latter is related to the *in situ* Chl *a* data set used for the algorithm development, which in this case represented a range of 1–10 mg/m^3. Under higher Chl *a* concentrations ($>$10 mg/m^3) it is expected that the current

algorithm will underperform given the data sets used for its generation. Our validation shows that locations with maximum Chl *a* concentrations (> or = to 10 mg/m^3) retrieved simulated Chl *a* concentration lower than real. Additional testing of applicable band combinations would be required. The significant conclusions of this study can be summarized as follows:

- Hyperion satellite images can be successfully used to model Chl *a* concentrations in Lake Atitlán.
- A semi-empirical algorithm that uses a blue to green band ratio is suitable to model Chl *a* concentrations in Lake Atitlán during the dry season. This algorithm has a relative error of 33%.

The Chl *a* concentrations described in this study are relatively low (1.0 – 10 mg/m^3), but the demonstration of Hyperion's suitability to model such low concentrations (represented by low reflectance values) is promising. It is reasonable to expect that Hyperion can be applied to distinguish Chl *a* in lakes with more variable or higher Chl *a* concentrations. This could be achieved with further *in situ* sampling in more varied conditions and would require fine tuning. The methodology used in this work can also be replicated in other lakes of the region to generate *ad-hoc* algorithms that represent the unique water quality characteristics of those individual water bodies.

Future research should be oriented toward determining the performance of ALI and other multispectral satellite images, such as Landsat MSS, ETM+, and OLI, in estimating Chl *a* concentration re-calibrating this algorithm. This will expand the applicability of the results obtained in this study. The accuracy of the algorithm should be very straightforward to calculate using ALI because images are available for the same dates on which the *in situ* Chl *a* data were collected. The original bands selected from Hyperion will need to be replaced by the closest ALI bands. Landsat 8 and Copernicus, Sentinel 2 is the other logical next step in testing the performance of multispectral imagers, but this will only be possible through the careful coordination of field sampling efforts with satellite overpasses and clear-sky conditions.

Multi-algorithm approaches that combine indices and look-up table approaches are promising (Salem et al., 2017) developments that can expand applicability of locally-tuned algorithms and should be explored when using hyperspectral datasets.

Since the development of this algorithm to the release of the peer-reviewed article, EO-1 has been decommissioned. Hence, it is critical to test and calibrate algorithms using operational sensors, such as Landsat and Sentinel-2.

As a final conclusion we would like to stress the local impact that this research has had in the area of Atitlán and Guatemala. The results of study have provided with new tools to the Lake Authorities (AMSCLAE, CONAP) and academia (UVG, San Carlos) to monitor Chl *a*, such as in the case of the algal bloom in 2015. Satellite remote sensing applications like this will allow for the creation of a systematic record of Chl *a* concentration in the lake that will document the progress of the water quality conditions. At the same time it will be possible to determine the more critical factors that are affecting the water quality of the lake and evaluate the impacts that the implemented policies or conservation actions are having in the lake's water quality.

DATA AVAILABILITY STATEMENT

The datasets analyzed in this article are not publicly available. Requests to access the datasets should be directed to Africa Flores, africa.flores@uah.edu.r.

AUTHOR CONTRIBUTIONS

AF-A developed research idea, performed all remote sensing analysis, and lead article writing. RG provided guidance for analysis and supported article writing. MD collected *in-situ* datasets used in this analysis and supported article writing. Led section for *in situ* Chl a sampling methods. CR-O, JS-A, BP, BH, EC, and FB provided critical feedback and updates related to current state of science, lake conditions and cyanobacteria found in lake. In addition, all co-authors from AMSCLAE were involved in *in situ* data collection, that was used in this analysis.

ACKNOWLEDGMENTS

First and foremost, we would like to thank SERVIR leadership, Dan Irwin, Ashutosh Limaye, and NASA Applied Sciences, Capacity Building Program Manager, Nancy Searby, for their constant support to this work. Thanks also to USAID Climate Change office for their support to SERVIR, Jenny Frankle-Reed, Pete Epanchin, Carry Stokes, and Kevin Coffey. This work was based on the SERVIR thesis research, done by Africa Flores (Flores, 2013) under NASA-UAH Cooperative Agreement NNM11AA01A.

The authors would like to thank Stu Frye from NASA Goddard Space Flight Center Greenbelt, MD, who tasked EO-1 to acquire images at the same time that *in situ* sampling was being carried out. We would also like to recognize the invaluable support provided by researchers in Guatemala, Estuardo Bocel, Juan Carlos Bocel, Isabel Arriola, Cesar Arriola, and the rest of the AMSCLAE team in 2013 who collaborated to collect *in situ* samples for this study. Thank you to Sundar Christopher who provided crucial guidance to perform this analysis. Thanks also to UAH faculty Tom Sever, Udaysankar Nair, Quingyan Han, John Christy, staff Michele Kennedy, Jan Williamson, Linda Berry, David Corredor, and David Cook. Additional support for subsequent research in this area provided by National Geographic and Microsoft AI grant.

REFERENCES

AMSCLAE (2017). *Plan Operativo Anual AL 31 DE MAYO DE 2017*. Technical report.

Bailey, S. W., and Werdell, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sens. Environ.* 102, 12–23. doi: 10.1016/j.rse.2006.01.015

Beck, R. U. O. C. (2003). *EO-1 User Guide v. 2.3*. Technical report, University of Cincinnati.

Boucher, J., Weathers, K. C., Norouzi, H., and Steele, B. (2018). Assessing the effectiveness of Landsat 8 chlorophyll a retrieval algorithms for regional freshwater monitoring. *Ecol. Appl.* 28, 1044–1054. doi: 10.1002/eap.1708

Bukata, R. P. (2013). Retrospection and introspection on remote sensing of inland water quality: "Like Déjà Vu All Over Again". *J. Great Lakes Res.* 39, 2–5. doi: 10.1016/j.jglr.2013.04.001

Chandra, S., Rejmánková, E., Dix, M., Girón, N., Mosquera, V., Komárek, J., J., K., et al. (2013). *Estado del Lago Atitlan, Informe 2013*. Technical report.

Chatterjee, S., Hadi, A. S., and Price, B. (2000). *Regression Analysis by Example, 3rd Edn*. New York, NY: Wiley-Interscience.

Chen, Z., Hu, C., Muller-Karger, F. E., and Luther, M. (2010). Short-term variability of suspended sediment and phytoplankton in Tampa Bay, Florida: observations from coastal oceanographic tower and ocean color satellites. *Estuar. Coast. Shelf Sci.* 89, 62–72. doi: 10.1016/j.ecss.2010.05.014

Chernick, M. R. (2012). Resampling methods. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 2, 255–262. doi: 10.1002/widm.1054

Chesner, C. A., and Halsor, S. P. (2010). "A bathymetric survey of Lake Atitlan, Guatemala," in *American Geophysical Union, Fall Meeting 2010* (San Francisco, CA).

Dix, M., Dix, M., Cabrera, D., Symonds, E., Orozco, E., Toledo, A., et al. (2012a). *Estado Fisico, Quimico y Biologico del Lago Atitlan. Octubre 2009 a Diciembre 2010*. Technical report, Centro de Estudios Atitlan, Laboratorio de Analisis y Monitoreo, Universdiad del Valle de Guatemala.

Dix, M., Orozco, M., Cabrera, D., Bocel, E., Toledo, A., and Symonds, E. (2012b). El Lago de Atitlan, Guatemala: su estado ecologico octubre 2009-diciembre 2011. *Revista de la Universidad del Valle de Guatemala* 24, 35–50.

Eaton, A. D. (2005). *Standard Methods for the Examination of Water and Wastewater, 21st Edn*. Washington, DC: APHA-AWWA-WEF.

Flores, A. (2013). *Hyperspectral remote sensing of water quality in Lake Atitlán, Guatemala*. (M.S. thesis). The University of Alabama in Huntsville, Huntsville, AL, 156.

Gitelson, A. A., Gurlin, D., Moses, W. J., and Barrow, T. (2009). A bio-optical algorithm for the remote estimation of the chlorophyll- a concentration in case 2 waters. *Environ. Res. Lett.* 4:045003. doi: 10.1088/1748-9326/4/4/045003

Gower, J., King, S., Borstad, G., and Brown, L. (2005). Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. *Int. J. Remote Sens.* 26, 2005–2012. doi: 10.1080/01431160500075857

Gower, J. F. R., Doerffer, R., and Borstad, G. A. (1999). Interpretation of the 685 nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS. *Int. J. Geomat. Geosci.* 20, 1771–1786. doi: 10.1080/014311699212470

Gupta, H. V., Sorooshian, S., and Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: comparisson with multilevel expert calibration. *J. Hydrol.* 4, 135–143. doi: 10.1061/(ASCE)1084-0699(1999)4:2(135)

Ho, J. C., Michalak, A. M., and Pahlevan, N. (2019). Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature* 574, 667–670. doi: 10.1038/s41586-019-1648-7

Hu, C., Carder, K. L., and Muller-karger, F. (2001). How precise are SeaWIFS ocean color estimates? Implications of digitization-noise errors. *Remote Sens. Environ.* 76, 239–249. doi: 10.1016/S0034-4257(00)00206-6

Hu, C., Lee, Z., and Franz, B. (2012). Chlorophyll a algorithms for oligotrophic oceans: a novel approach based on three-band reflectance difference. *Geophys. Res.* 117, 1–25. doi: 10.1029/2011JC007395

INE (Instituto Nacional de Estadistica) (n.d.). *Guatemala: Estimaciones de la Población total por municipio. Período 2008–2020*. Retrieved from: http://www.oj.gob.gt/estadisticaj/reportes/poblacion-total-por-municipio(1).pdf

Jordan, T. E., Cornell, J., Miklas, and Weller, D. E. (1991). Long-term trends in estuarine nutrients and chlorophyll, and short term effects of variation in watershed discharge. *Mar. Ecol. Prog. Ser.* 75, 121–132. doi: 10.3354/meps075121

Kirk, J. T. (1994). *Light and Photosynthesis in Aquatic Ecosystems, 2nd Edn*. Cambridge: Cambridge University Press.

Komárek, J., Zapomělová, E., Šmarda, J., Kopecký, J., Rejmánková, E., Woodhouse, J., et al. (2013). Polyphasic evaluation of Limnoraphis robusta, a water-bloom forming cyanobacterium from Lake Atitlán, Guatemala, with a description of Limnoraphis gen. nov. *Fottea* 13, 39–52. doi: 10.5507/fot.2013.004

Le, C., Hu, C., Cannizzaro, J., English, D., Muller-Karger, F., and Lee, Z. (2013). Evaluation of chlorophyll-a remote sensing algorithms for an optically complex estuary. *Remote Sens. Environ.* 129, 75–89. doi: 10.1016/j.rse.2012.11.001

Liao, L., Jarecke, P., Gleichauf, D., and Hedman, T. (2000). "Performance characterization of the Hyperion Imaging spectometer instrument," in *Proceedings of the SPIE 4135, Earth Observing Systems* (San Diego, CA), Vol. 264, 12.

Markert, K. N., Schmidt, C. M., Griffin, R. E., Flores, A. I., Poortinga, A., Saah, D. S., et al. (2018). Historical and operational monitoring of surface sediments in the lower mekong basin using landsat and Google earth engine cloud computing. *Remote Sens.* 10:909. doi: 10.3390/rs10060909

McClain, C., Hooker, S., Feldman, G., and Bontempi, P. (2006). Satellite data for ocean biology, biogeochemestry, and climate research. *EOS Trans.* 87, 337–343. doi: 10.1029/2006EO340002

Mobley, C. D. (1995). "The optical properties of water," in *Handbook of Optics*, Vol. 1, ed M. Bass (New York, NY: McGraw-Hill), 43.3–43.56.

Moriasi, D. N., Arnold, J. G., Liew, M. W. V., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007). Model Evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Am. Soc. Agricult. Biol. Eng.* 50, 885–900. doi: 10.13031/2013.23153

Morrow, J. H., White, B. H., Chimiente, M., and Hubler, S. (2000). A bio-optical approach to reservoir monitoring in Los Angeles. Limnoology and Lake Management. *Adv. Limnol.* 55, 171–191.

Mueller, J. L. (2000). "SeaWiFS algorithm for the diffuse attenuation coefficient, K(490), using water-leaving radiances at 490 and 555 nm," in *Sea-WiFS Postlaunch Calibration and Validation Analyses, Part 3, SeaWiFS Postlaunch Technical Report Series*, Vol. 11, eds S. B. Hooker and E. R. Firestone (Greenbelt, MD: NASA Goddard Space Flight Center), 24–27

NASA, METI, AIST, Japan Space Systems, and U.S./Japan ASTER Science Team (2009). *ASTER Global Digital Elevation Model V003 [Data set]*. NASA EOSDIS Land Processes DAAC. doi: 10.5067/ASTER/ASTGTM.003

Nash, J. E., and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrol.* 10, 282–290. doi: 10.1016/0022-1694(70)90255-6

Ogashawara, I., and Moreno-Madriñán, M. (2014). Improving inland water quality monitoring through remote sensing techniques. *ISPRS Int. J. Geo-Informat.* 3, 1234–1255. doi: 10.3390/ijgi3041234

O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., et al. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.* 103, 24937–24953. doi: 10.1029/98JC02160

Page, B. P., Kumar, A., and Mishra, D. R. (2018). Int J Appl Earth Obs Geoinformation A novel cross-satellite based assessment of the spatio-temporal development of a cyanobacterial harmful algal bloom. *Int. J. Appl. Earth Obs Geoinformat.* 66, 69–81. doi: 10.1016/j.jag.2017.11.003

Palmer, S. C., Kutser, T., and Hunter, P. D. (2015). Remote sensing of inland waters: challenges, progress and future directions. *Remote Sens. Environ.* 157, 1–8. doi: 10.1016/j.rse.2014.09.021

Perez Gudiel, D. B. (2007). *Evaluacion del efecto de la aireacion artificial para mejorar la calidad del agua en el lago de Amatitlan* (Ph.D. thesis). Universidad de San Carlos de Guatemala, Guatemala.

Pérez, L., Bugja, R., Lorenschat, J., Brenner, M., Curtis, J., Hoelzmann, P., et al. (2011). Aquatic ecosystems of the Yucatán Peninsula (Mexico), Belize, and Guatemala. *Hydrobiologia* 661, 407–433. doi: 10.1007/s10750-010-0552-9

Potes, M., Costa, M. J., and Salgado, R. (2012). Satellite remote sensing of water turbidity in Alqueva reservoir and implications on lake modelling. *Hydrol. Earth Syst. Sci.* 16, 1623–1633. doi: 10.5194/hess-16-1623-2012

Rejmánková, E., Komárek, J., Dix, M., Komárková, J., and Girón, N. (2011). Cyanobacterial blooms in Lake Atitlan, Guatemala. *Limnologica* 41, 296–302. doi: 10.1016/j.limno.2010.12.003

Ritchie, R. J. (2006). Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynth. Res.* 89, 27–41. doi: 10.1007/s11120-006-9065-9

Romero Santizo, M. A. (2009). *Valoración económica del lago de Atitlán, Sololá* (M.S. thesis). Universidad de San Carlos de Guatemala, Guatemala.

Romero-Oliva, C. S., Contardo-Jara, V., Block, T., and Pflugmacher, S. (2014). Accumulation of microcystin congeners in different aquatic plants and crops - A case study from lake Amatitlán, Guatemala. *Ecotoxicol. Environ. Safety* 102, 121–128. doi: 10.1016/j.ecoenv.2014.01.031

Rowan, K. S. (1989). *Photosynthetic Pigments of Algae*. New York, NY: Cambridge University Press.

Salem, S. I., Higa, H., Kim, H., Kazuhiro, K., Kobayashi, H., Oki, K., et al. (2017). Multi-algorithm indices and look-up table for chlorophyll-a retrieval in highly turbid water bodies using multispectral data. *Rem. Sens.* 9. doi: 10.3390/rs9060556

Schalles J. (2006). "Optical remote sensing techniques to estimate phytoplankton chlorophyll a concentrations in coastal," in *Remote Sensing of Aquatic Coastal Ecosystem Processes. Remote Sensing and Digital Image Processing,* Vol. 9, eds L. Richardson and E. Ledrew (Dordrecht: Springer), 27–79.

SERVIR (2009). *Pollution in Lake Atitlán, Sololá, Guatemala - Nov 2009*. Retrieved from: https://servirglobal.net/Global/Articles/Article/1082/reappearance-of-pollution-in-lake-atitln-solol-guatemala-may-2011?page=2

SERVIR (2011). *Cambios en la Superficie de Area del Lago Atitlan, Guatemala 1986-2011*. Retrieved from: https://servirglobal.net/Global/Articles/Article/1088/servir-analyzes-changes-in-lake-atitlans-surface-extent-1986-2011

Shang, P., and Shen, F. (2016). Atmospheric correction of satellite GF-1/WFV imagery and quantitative estimation of suspended particulate matter in the yangtze estuary. *Sensors* 16:E1997. doi: 10.3390/s16121997

SIGSA (Sistema de Información Gerencial de Salud) (2019). *Sistema de Información Gerencial de Salud*. Retrieved from: https://sigsa.mspas.gob.gt/datos-de-salud

Stramska, M., and Stramski, D. (2005). Effects of a nonuniform vertical profile of chlorophyll concentration on remote-sensing reflectance of the ocean. *Appl. Opt*. 44, 1735–1747.

U.S. Geological Survey (2018). *Earth Observing 1 (EO-1)*. Retrieved from: https://archive.usgs.gov/archive/sites/eo1.usgs.gov/index.html

UNESCO (2015). *UNESCO Science Report, Towards 2030*. UNESCO.

USGS (2006). *Hyperion Level 1GST (L1GST) Producto Output Files Data Format*. Technical report, Department of Interior, USGS.

Vermote, E., Tanre, D., Deuze, J. L., Herman M., Morcrette J. J., Kotchenova, S. (2006). *Second Simulation of a Satellite Signal in the Solar Spectrum - Vector (6SV)*. Technical report.

Vincent, R. K., Qin, X., McKay, R. L., Miner, J., Czajkowski, K., Savino, J., et al. (2004). Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. *Remote Sens. Environ.* 89, 381–392. doi: 10.1016/j.rse.2003.10.014

Watanabe, F., Alcântara, E., Rodrigues, T., Imai, N., Barbosa, C., and Rotta, L. (2015). Estimation of chlorophyll-a concentration and the trophic state of the barra bonita hydroelectric reservoir using OLI/Landsat-8 images. *Int. J. Environ. Res. Public Health* 12, 10391–10417. doi: 10.3390/ijerph120910391

Weiss, C. (1971). *Water Quality Investigations, Lake Atitlan 1968-1970*. Technical report, University of North Carolina at Chapel Hill.

Werdell, P. J., and Bailey, S. W. (2005). An improved *in-situ* bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote Sens. Environ.* 98, 122–140. doi: 10.1016/j.rse.2005.07.001

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., et al. (1985). Statistics for the evaluation and comparison of models. *J. Geophys. Res.* 90:8995. doi: 10.1029/JC090iC05p08995

Willmott, C. J., and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res.* 30, 79–82. doi: 10.3354/cr030079

Wynne, T. T., Stumpf, R. P., Tomlinson, M. C., Warner, R. A., Tester, P. A., Dyble, J., et al. (2008). Relating spectral shape to cyanobacterial blooms in the laurentian great lakes. *Int. J. Remote Sens.* 29, 3665–3672. doi: 10.1080/01431160802007640

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

TABLE A1 | Comparison of results of curve fitting of data to published semi-empirical algorithms for predicting chlorophyll a (Chl *a*) concentration.

Algorithm	$^{a}R^2$ mean	$^{b}R^2$ mean	$^{a}R^2$ max	$^{b}R^2$ max	$^{a}R^2$ med	$^{b}R^2$ med
SR498/SR548	0.6465	0.7635	0.619	0.693	0.663	0.730
SR467/SR548	0.6999	0.7519	0.549	0.634	0.700	0.738
SR487/SR548	0.5001	0.5371	0.513	0.504	0.576	0.565
SR508/SR548	0.4501	0.4075	0.488	0.520	0.694	0.737
SR498/SR559	0.6618	0.7946	0.562	0.673	0.670	0.750
SR508/SR559	0.5418	0.5954	0.559	0.621	0.693	0.761
SR487/SR559	0.6507	0.766	0.519	0.562	0.675	0.715
SR467/SR559	0.7054	0.760	0.545	0.611	0.703	0.740
SR681/SR711	0.3203	0.397	0.405	0.488	0.214	0.183
FLH	0.5223	0.566	0.336	0.409	0.501	0.533

Regression model: non-linear third polynomial. All Log-transformed data.
[a] all stations; [b] Chl a < 9 mg/m^3; mean, mean reflectance value in a 3 × 3 pixel box; max, maximum value in a 3 × 3 pixel box; med, median value in a 3 × 3 pixel box.
SR, Surface Reflectance; R^2, coefficient of determination.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 19 March 2020
doi: 10.3389/fenvs.2020.00021

Mapping Land Use Land Cover Change in the Lower Mekong Basin From 1997 to 2010

Joseph Spruce[1], John Bolten[2], Ibrahim N. Mohammed[3], Raghavan Srinivasan[4] and Venkat Lakshmi[5]*

[1] Science Systems and Applications, Inc., Consultant, Diamondhead, MS, United States, [2] NASA GSFC Hydrological Sciences Branch, Greenbelt, MD, United States, [3] Science Applications International Corporation, Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, United States, [4] Spatial Sciences Laboratory, Texas A&M University, College Station, TX, United States, [5] Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, United States

The Lower Mekong Basin (LMB) is biologically diverse, economically important, and home to about 65 million people. The region has undergone extensive environmental changes since the 1990s due to such factors as agricultural expansion and intensification, deforestation, more river damming, increased urbanization, growing human populations, expansion of industrial forest plantations, plus frequent natural disasters from flooding and drought. The Mekong river is also heavily used for human transportation, fishing, drinking water, and irrigation. This paper discusses use of pre-existing LULC maps from 1997 and 2010 to derive a LMB regional LULC change map for 9 classes per date using GIS overlay techniques. The change map was derived to aid SWAT hydrologic modeling applications in the LMB, given the 2010 map is currently used in multiple LMB SWAT models, whereas the 1997 map was previously used. The 2010 LULC map was constructed from Landsat and MODIS satellite data, while the 1997 map was from before the MODIS era and therefore based on available Landsat data. The 1997–2010 LULC change map showed multiple trends. Permanent agriculture had expanded in certain sub-basins into previously forested areas. Some agricultural areas were converted to industrial forest plantations. Extensive forest changes also occurred in some locations, such as areas changed to shifting cultivation or permanent crops. Also, the 1997 map under classified some urban areas, whereas the 2010 LULC map showed improved identification of such areas. LULC map accuracy were assessed for 213 randomly sampled locations. The 1997 and 2010 LULC maps showed high overall agreements with reference data exceeding 87%. The LULC change map yielded a moderately high level of overall agreement (78%) that improved to ∼83% once LULC classification scheme specificity was reduced (forests and agriculture were each mapped as singular classes). The change map regionally showed a 4% decrease in agriculture and a 4% increase in deciduous and evergreen forests combined, though deforestation hot spot areas also were evident. The project yielded LULC map data sets that are now available for aiding additional studies that

OPEN ACCESS

Edited by:
Daniel Eric Irwin, Marshall Space Flight Center (NASA), United States

Reviewed by:
Jason A. Tullis, University of Arkansas, United States
Pablo Pacheco, World Wildlife Fund, United States

***Correspondence:**
Joseph Spruce
j_spruce@outlook.com

Specialty section:
This article was submitted to Land Use Dynamics, a section of the journal Frontiers in Environmental Science

Received: *29 May 2019*
Accepted: *17 February 2020*
Published: *19 March 2020*

Citation:
Spruce J, Bolten J, Mohammed IN, Srinivasan R and Lakshmi V (2020) Mapping Land Use Land Cover Change in the Lower Mekong Basin From 1997 to 2010. Front. Environ. Sci. 8:21. doi: 10.3389/fenvs.2020.00021

assess LMB LULC change and the impacts such change may pose to water, agriculture, forestry, and disaster management efforts. More work is needed to map, quantify and assess LULC change since 2010 and to further update the 2010 LULC map currently used in the LMB SWAT models.

Keywords: Lower Mekong Basin, land use land cover change, SWAT hydrologic model, agricultural monitoring, deforestation

INTRODUCTION

The Lower Mekong Basin (LMB) is a region in the southeast Asia known for its agriculture, forests, fisheries, wildlife, and diverse, natural ecosystems [Mekong River Commission MRC (2010)]. About 65 million people live in this region, depending on the Mekong River and its tributaries for transportation, water for drinking and bathing, fishing, and agricultural irrigation. The MRC (2014) reports that agriculture is the most important economic activity in the LMB with rice cultivation occurring on millions of hectares.

The LMB human inhabitants are experiencing multiple environmental threats to their ways of life that include land use land cover (LULC) change (e.g., from river damming, agricultural expansion, and deforestation) and frequent natural disasters (e.g., flooding from severe storms, and drought). The drivers of LULC change in the region include anthropogenic (e.g., economic and demographic) and environmental factors (Tran et al., 2015). Both negative and positive forest change drivers occur, which can collectively affect the amount of agricultural expansion and contraction, well as the amount of deforestation and afforestation (Costenbader et al., 2015a,b; Imai et al., 2018). The causes (i.e., drivers) of LULC change in the Mekong region are multifaceted with the prices of agricultural and forest products, road accessibility, and changes in land titles identified as important drivers of LULC change (Rowcroft, 2008; Xing, 2013). Other drivers of LULC change in the Mekong include anthropogenic drivers such as the construction of large-scale hydropower dams and other abiotic drivers such as climate change (e.g., sea level rise) (Keskinen et al., 2010; Evers and Pathirana, 2018). Subsidence in conjunction with ground water extraction and sea level rise is another driver negatively affecting coastal LMB agriculture (Minderhoud et al., 2018). Climate change in the LMB is regarded as a threat to agriculture (MRC, 2010; MRC, 2014; Pokhrel et al., 2018a) and forests (Estoque et al., 2019).

Such LULC change drivers can affect human settlements and livelihoods, as well as the natural resources that are depended upon by the inhabiting humanity, wildlife, fisheries and other valued wild and domestic life forms (MRC, 2010). The management and sustainability of water resources in the region is particularly of great concern to people in the LMB, given a growing number of hydrologic modifications along the main waterways (e.g., river damming), increased demand on water for irrigation from agricultural intensification, expansion of agriculture into areas formerly occupied by forest, and a growing human population (MRC, 2010; Pokhrel et al., 2018a). In addition, meteorological disasters such as severe droughts and flooding events further threaten regional inhabitants and their economic activities, such as agriculture, forestry, commercial fishing, and transportation (Ribbe et al., 2013).

The LMB and the Mekong region at large has reportedly undergone substantial LULC change in recent decades (Lyon et al., 2017; Yasmi et al., 2017; Pokhrel et al., 2018b), though questions and uncertainties remain as to the location and geospatial extent for specific kinds of LULC change. An assortment of satellite data has been used to map and quantify LULC change in the region, including AVHRR (Giri et al., 2003), Landsat data (Heinimann et al., 2007), SPOT-4 Vegetation (Stibig et al., 2004), and MODIS data (Leinenkugel et al., 2015). Unfortunately, it is difficult to effectively compare results from such change maps given that they can differ in terms of the kind of satellite data used, the observed time span, the methods for generating and validating such products, the LULC change classification scheme employed, spatial resolution of the map, the geographic domain covered by the map, the objectives of the mapping project, and the organizations responsible for making LULC maps. Some of these issues are discussed by Patil and Gumma (2018) with respect to updating south Asia cropland and other land cover types. The challenges arising from the differences in LULC mapping methods may be addressed in part by comparing provenance of geospatial workflows (Tullis et al., 2015). Note that the LULC maps referred to in this paper are digital geospatial data sets as opposed to cartographic hardcopy mapping products.

To help address water, disaster management, and crop security concerns, the Mekong River Commission has been using the Soil and Water Assessment Tool (SWAT) framework (Gassman et al., 2007) for hydrologic modeling in the non-coastal portion of the LMB, including sub-basins 1–8, as discussed by Rossi et al. (2009). The LULC map is a key input parameter to the SWAT framework, given the effect that an area's LULC characteristics can have on surface water flow and runoff. The initial MRC SWAT model employed a LULC map that was based on 1997 Landsat data (Dat, 2013). More recently, MRC SWAT models were updated with a 2010 LULC map for the LMB that was developed using a combination of monthly MODIS NDVI and Landsat data (Spruce et al., 2018). The updated map included more detailed agricultural LULC classes in order to revise and improve SWAT hydrologic modeling products and applications (Mohammed et al., 2018a,b).

The purpose of the project discussed in this paper was to map, quantify and assess spatio-temporal changes in non-coastal LMB LULC that occurred from 1997 to 2010, using pre-existing LULC map data sets as inputs. This work aimed to assess apparent LULC changes during the observed time frame mainly with

respect to forest and agricultural areas. Such information was needed to augment understanding and interpretation of LULC effects on SWAT modeling of runoff using either the 2010 or 1997 data. The LULC patch mosaic can either accelerate or impede the runoff depending on the characteristics of LULC, including the location and extensiveness of LULC types. Although the runoff characteristics of common LULC types are generally known (USDA, 1986; Hong and Adler, 2008), an assessment of LULC change for the updated versus previous LULC map datasets was needed to help resource management organizations (e.g., the MRC and the Asian Disaster Preparedness Center or ADPC) assess SWAT modeled hydrologic response of surface runoff in the LMB, especially in watershed reaches with extensive LULC change. In doing so, such an assessment could help further document and understand the provenance of the LMB SWAT models.

There are several published reviews of raster imagery-based two date change detection methods on the use of remotely sensed data or pre-existing map datasets from remotely sensed data (Mas, 1999; Coppin et al., 2004; Jensen et al., 2012; Hussain et al., 2013). These methods include: (1) image ratioing; (2) image differencing; (3) Principal Components Analysis (PCA); (4) multi-date classification; and (5) Change Vector Analysis (CVA); and (6) post classification comparison. The choice of method can depend on such factors as the input data used for deriving the change map dataset and the kind of change map (e.g., general forest change or more specific LULC change). Excluding the post-classification method, most change detection mapping methods require that remotely sensed imagery from two or more dates as input to a data processing workflow in order to derive a change map.

The post-classification method is the main method used to derive LULC change maps from pre-existing LULC map digital datasets. In doing so, Geographic Information System (GIS) methods (e.g., additive overlays or indexing techniques) are employed to integrate the two LULC map datasets into a digital change map. Jensen et al. (2012) reported that post classification comparison is one of the more commonly used change detection methods. Also termed the delta classification technique, the post classification comparison method has the advantage of not requiring that remote sensed data be acquired and then processed into a digital change map. Post classification change map datasets are also relatively easy to calculate and yield "from to" change information that resource managers can usually understand (Wang et al., 2009). However, the accuracy of this method depends heavily on the accuracy of the two input LULC map datasets. The accuracy of a post classification comparison-derived change map dataset can approximate the product of the accuracies of the 2 input LULC digital maps (e.g., each map scaled on a 0 to 1 floating point scale) (Coppin et al., 2004). This can but not necessarily result in a low accuracy change map dataset. However, the same kind of change detection errors can also occur with change maps derived with methods other than the post classification comparison technique. For example, change maps regardless of method can include thematic mapping errors due to misregistration (i.e., geo-positional or locational) errors in the input data sets used to derive the digital change map. Also, the post classification method can produce superior change map datasets. For example, Mas (1999) conducted a study comparing different change mapping methods, observing that the post classification method yielded the best results. The post classification method is also employed as a method used in part to compute operational, digital LULC mapping products, including the USGS National Land Cover Database (NLCD) (Yang et al., 2018) and NOAA Coastal Change Analysis Program (C-CAP) (McCombs et al., 2016) LULC datasets.

All maps of LULC and LULC change are imperfect and have biases (e.g., classification and positional errors) and uncertainties about the quality of such products (Verberg et al., 2011). The biases (i.e., errors) that occur for a given LULC change map data product can depend on kind and quality of remotely sensed data, how it is processed, minimum mapping unit, classification scheme, characteristics of the study area, and other factors (Jensen et al., 2012; Congalton, 2015). Such imperfections can limit and pose challenges on how digital LULC maps should be used, depending on the application. LULC and LULC change map accuracy assessments can provide information on the quality of such map data set and how to effectively use these products.

MATERIALS AND METHODS

Study Area Description

The LMB study area is in southeast Asia that includes portions of Laos, Myanmar, Thailand, and Vietnam, and Cambodia (**Figure 1**). The area includes 8 sub-basins (SBs) discussed by Rossi et al. (2009) that are used in the MRC SWAT modeling applications. This location includes extensive areas agricultural and forested land uses in addition to locally present land uses (e.g., urban). The study area includes most of the interior LMB, excluding the most coastal portion (e.g., Tonle Sap Basin and Mekong Delta). The study area has a tropical climate with dry and wet seasons, an average maximum temperature of 20.4 to 33.6°C, and a mean annual precipitation of ∼900 to 3300 mm, based on data from International Centre for Environmental Management (ICEM, 2013) study that is hosted on a Greater Mekong Subregion Environment Operations Center (GMS-EOC) Information Portal (GMS-EOC, 2019). The study area mostly has a moderately high or high hydrologic runoff potential, according to a 250 m hydrologic soil group (HSG) map by Ross et al. (2018). Elevations in the study area varies from ∼9 to 2,433 m, based on a 90-m Digital Elevation Model (DEM) data from the Shuttle Radar Topography Mission (SRTM) (NASA and JPL, 2013).

Data Acquisition

The project used pre-existing LULC GIS data layers for the 1997 and 2010 time periods to generate the needed LULC change map. Acquired from the MRC as a vector data layer, the 1997 LULC map was derived from image interpretation, delineation, and digitization of LULC patches from 1:250,000 scale hard copies of Landsat Thematic Mapper imagery (Malyvanh and Feldkotter, 1999). The 2010 LULC map dataset was obtained from a previous project discussed by Spruce et al. (2018) that used a combination

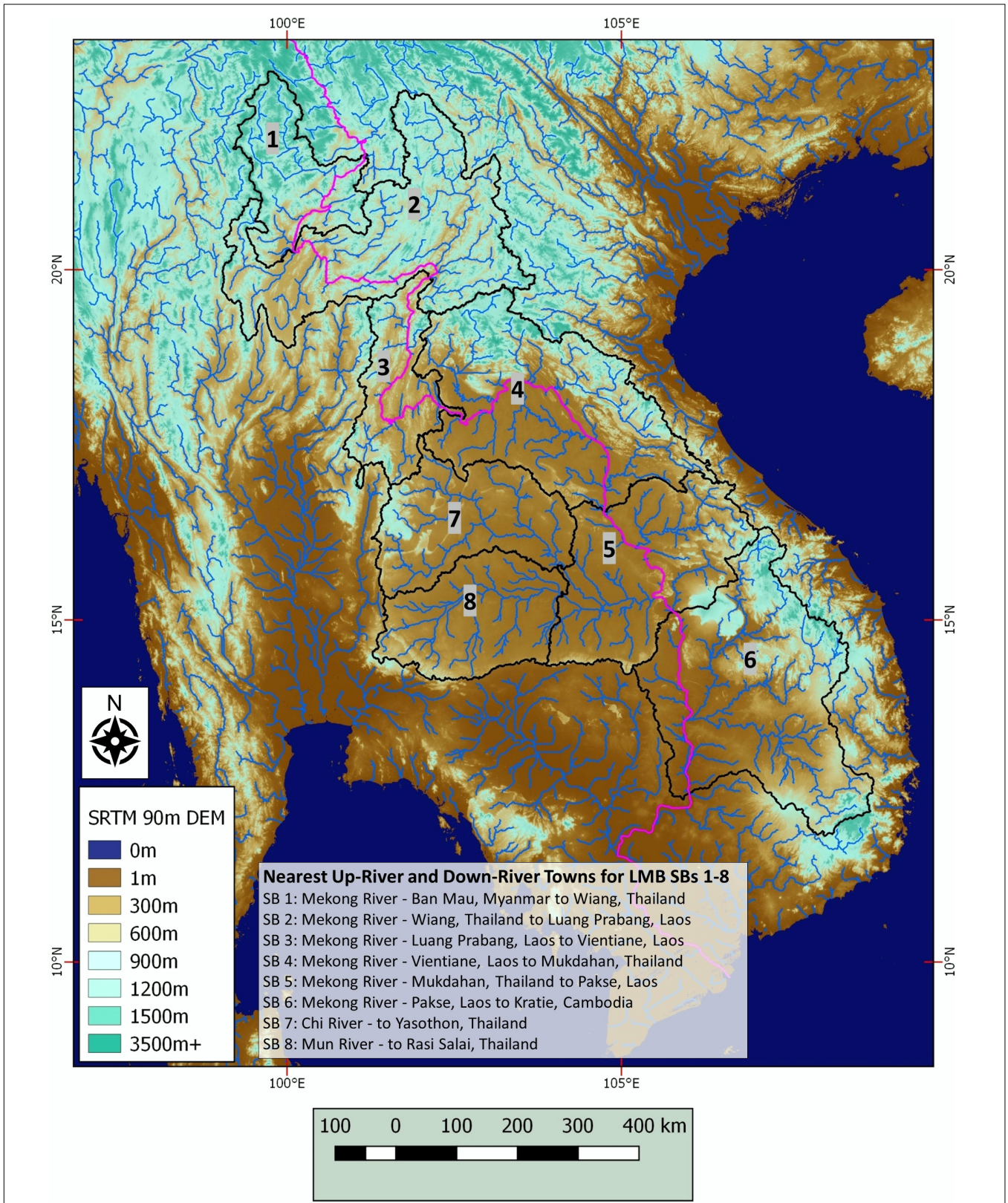


FIGURE 1 | Study area location in Lower Mekong Basin of southeast Asia with Sub-basins (SBs) 1–8 (in black) along with river boundaries (blue) overlain onto a colorized 90 m digital elevation model (DEM) from the NASA Shuttle Radar Topography Mission (SRTM). The main branch of the Mekong River is shown in a magenta tone and the SB #s are included as black text.

of ∼quarter kilometer resolution 2010 MODIS 8-day reflectance data from the LP-DAAC web site[1] and single date circa 2010 dry season Landsat Thematic Mapper reflectance data at 30 m resolution from the USGS GLOVIS web site[2]. For this project, the 1997 and 2010 LULC maps were nominally gridded to a ∼231.66-m resolution.

To aid the project's change detection product validation, additional Level-1 Landsat data for 1997 and 2010 was acquired from the USGS Earth Explorer[3] and GLOVIS web sites. The LULC change map derivation and assessment also made use of other LMB LULC data acquired from the MRC, including circa 2010 field survey data (Nam et al., 2015), an alternative circa 2010 LULC map from Landsat data (Kityuttachai et al., 2016), and crop calendars (Halcrow Group Limited, 2004).

Data Processing and Analysis

To compute the needed LULC change map, the LULC maps for the 1997 and 2010 time periods were recoded into 9 LULC classes per date (**Table 1**). The 1997 map (**Figure 2**) from the MRC were coded into intervals of 10 and the 2010 map (**Figure 3**) were coded into the intervals of 1. This recoding strategy enabled the two input maps to be effectively combined into a change map via the QGIS Raster Calculator (by adding the two LULC maps together). The recoded LULC maps were first aligned with each other in QGIS and then added together to derive the LULC change map (**Figure 4**). In doing so, change was calculated only for pixels that were classified as having terrestrial (non-water) LULC classes on both input dates.

The LULC change map was then assessed for classification accuracy, considering guidance from Congalton (1991, 2015), Congalton and Green (2009), and Olofsson et al. (2013, 2014). The goal of the map validation was to assess the overall accuracy of the LULC change map compared to available reference data. In doing so, the LULC change map dataset was reprojected from geographic coordinates to the UTM map projection and then summarized in terms of pixel frequency for each LULC change class. For this analysis, a stratified random sampling of pixels was then drawn from the LULC change map using the QGIS (QGIS, 2017) and its plugin "AcATaMa" with the number of pixels drawn per class based on the proportion of the LULC class pixel frequency relative to the total number of mapped pixels. The adjacent pixels around each candidate sample pixel were considered in the random selection process. Candidate pixels were selected when they were the center pixel of a 3 by 3 pixel window and that most of the pixels in that window (5 or more) were of the same LULC class as that in the center pixel. In all, a total of 213 pixels were selected for map accuracy assessment. The frequency of occurrence and the number of samples selected per LULC change class is given for the 25 most common classes in **Table 2**.

Each selected sample location was then assessed by a trained, experienced image analyst to identify the LULC class for 1997 and 2010 time periods using assorted reference data that included Landsat false color RGBs and higher spatial resolution imagery from Bing and Google via the QGIS OpenLayers plugin. We considered using the Collect Earth tool (Bey et al., 2016; Saah et al., 2019a) for accuracy assessment, though selected QGIS for this task mainly due to its image enhancement, vector grid overlay, vector editing, and the "AcATaMa" accuracy assessment plugin capabilities. Within QGIS, some of the random sample areas were also reviewed on various dates of high-resolution true color imagery resident to the Google Earth Pro software. The sample locations were also compared and assessed with respect to other digital LULC mapping products primarily from the MRC, along with descriptive information on LULC classes from the MRC. In addition, a vector-based grid map of each pixel's boundaries was superimposed on remote sensing imagery products. Each sample was image interpreted at a 1:6,000 scale which enables viewing individual applicable Landsat pixels of 30 m resolution (∼8 × 8 or 64 pixels) located within a

TABLE 1 | Description of LULC classes used to generate change map.

LULC class	Code	Description
Barren	BA	Barren areas that are not developed, including rock outcrops (e.g., bare karst topography)
Urban/Developed	URB	Urbanized and other developed areas mostly with predominantly bare inert surfaces (e.g., pavement and cement).
Permanent Crops	PC	Current agricultural areas with intensively managed crops on long term basis for commercial purposes (e.g., irrigated and rainfed rice, annual row crops, and pastures).
Shifting Cultivation	SC	Current agricultural areas managed on a rotational basis in which crops are grown temporarily for a few years and then allowed to be in fallow or forested state until the site has sufficiently recovered for a new cropping cycle.
Scrub/Shrub/Herbaceous	SSH	Mixture of low woody (e.g., forest regeneration) and non-woody vegetation that is typically transitional in nature, occurring after agricultural abandonment or forest harvesting.
Deciduous Forest/Scrub	DFS	Deciduous broadleaved forest and/or scrub that can have variable canopy closure and degrees of deciduousness.
Evergreen Forest	EGF	Evergreen forest that is primarily composed of broadleaved tree species and typically has closed canopies.
Industrial Forest Plantation	IFP	Industrial forest plantations that include rubber tree plantations.
Wetland Cover	WET	Wetland areas with woody and/or non-woody vegetation that is not used for agriculture.

Example field photos of these LULC classes can be found in Spruce et al. (2018).

[1]https://lpdaac.usgs.gov/

[2]https://glovis.usgs.gov/

[3]https://earthexplorer.usgs.gov/

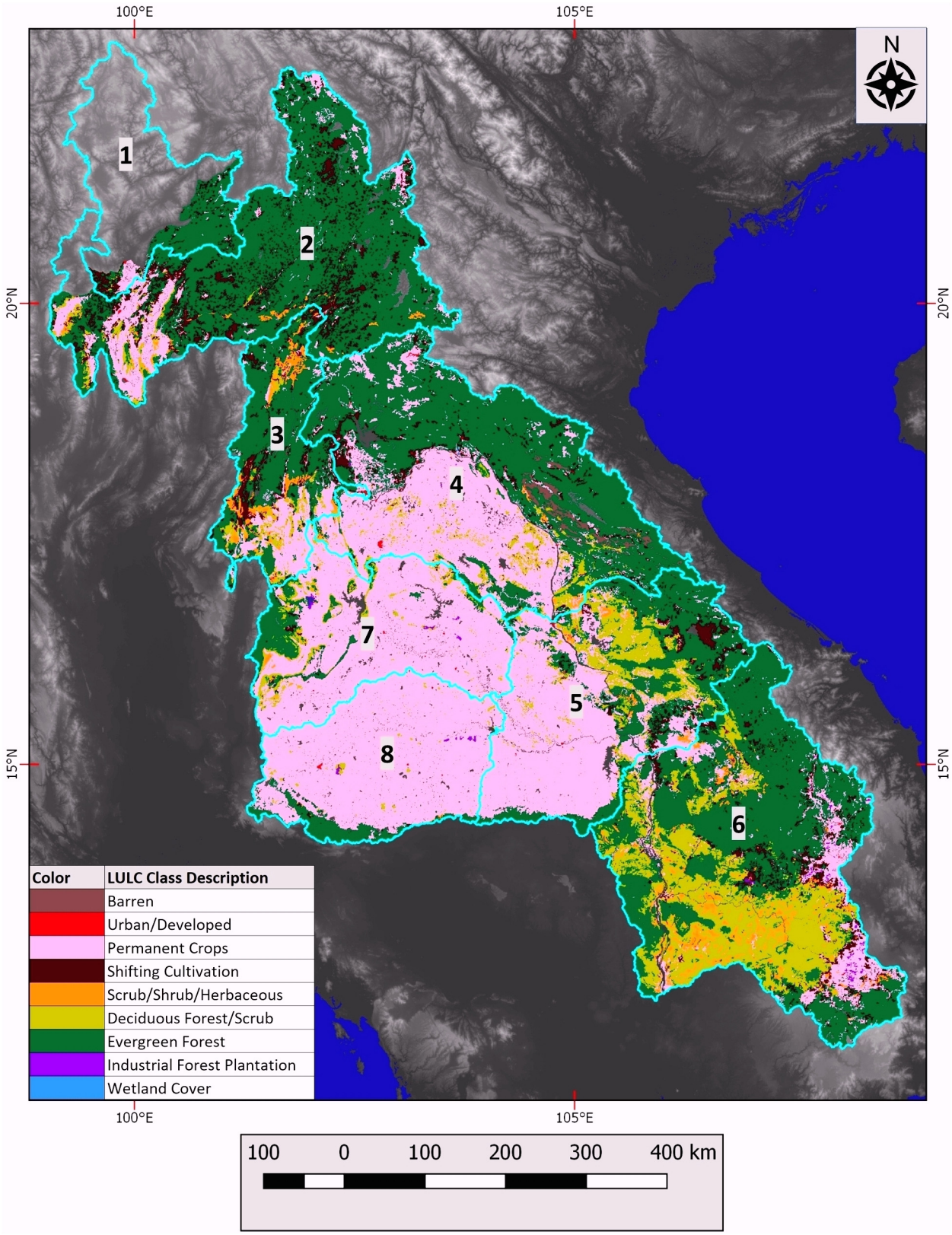


FIGURE 2 | Study area 1997 LULC map with SB boundaries overlain in cyan for SBs 1–8. The SB #s are included as black text. The backdrop image adjacent to the LULC map is a 90 m DEM from SRTM data.

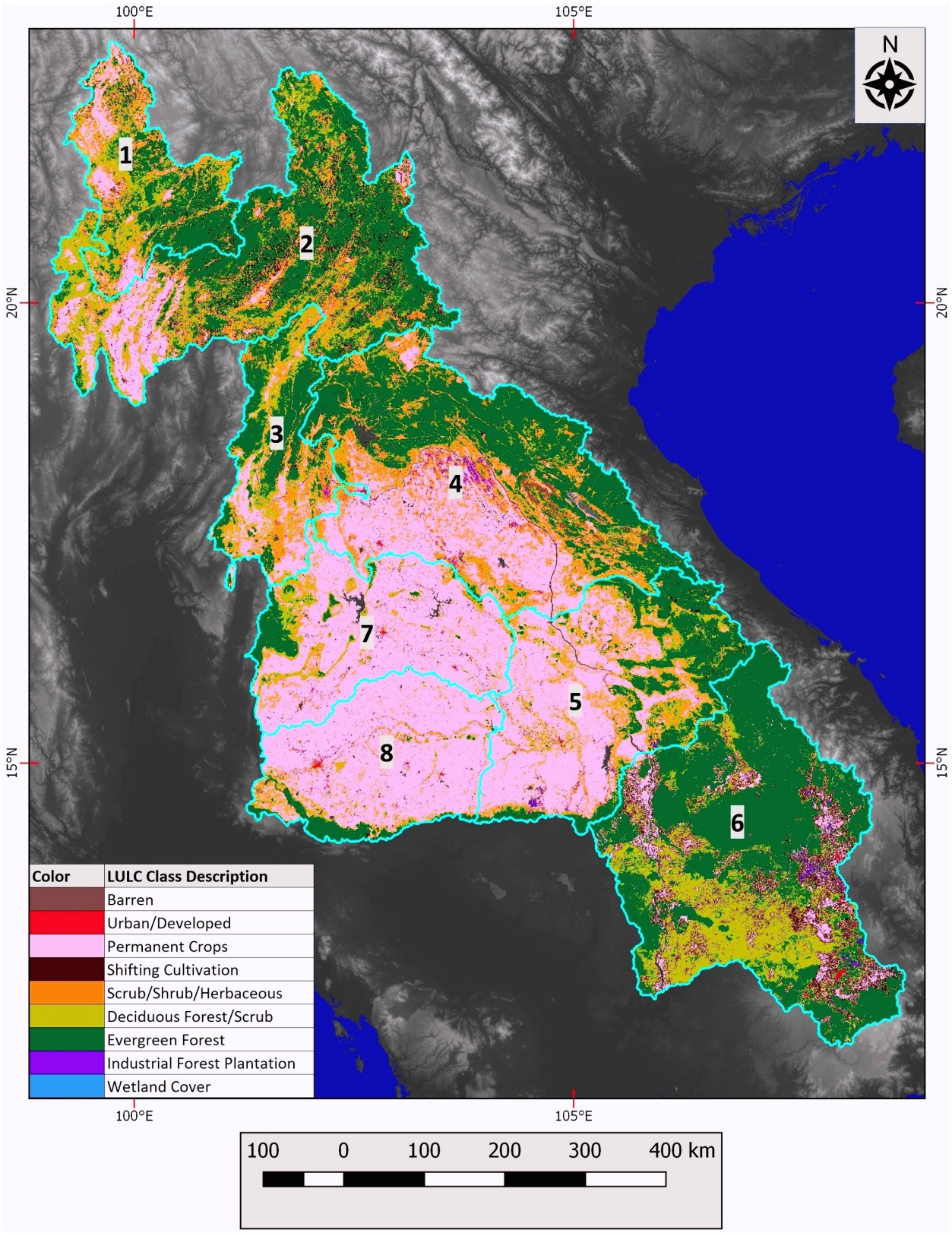
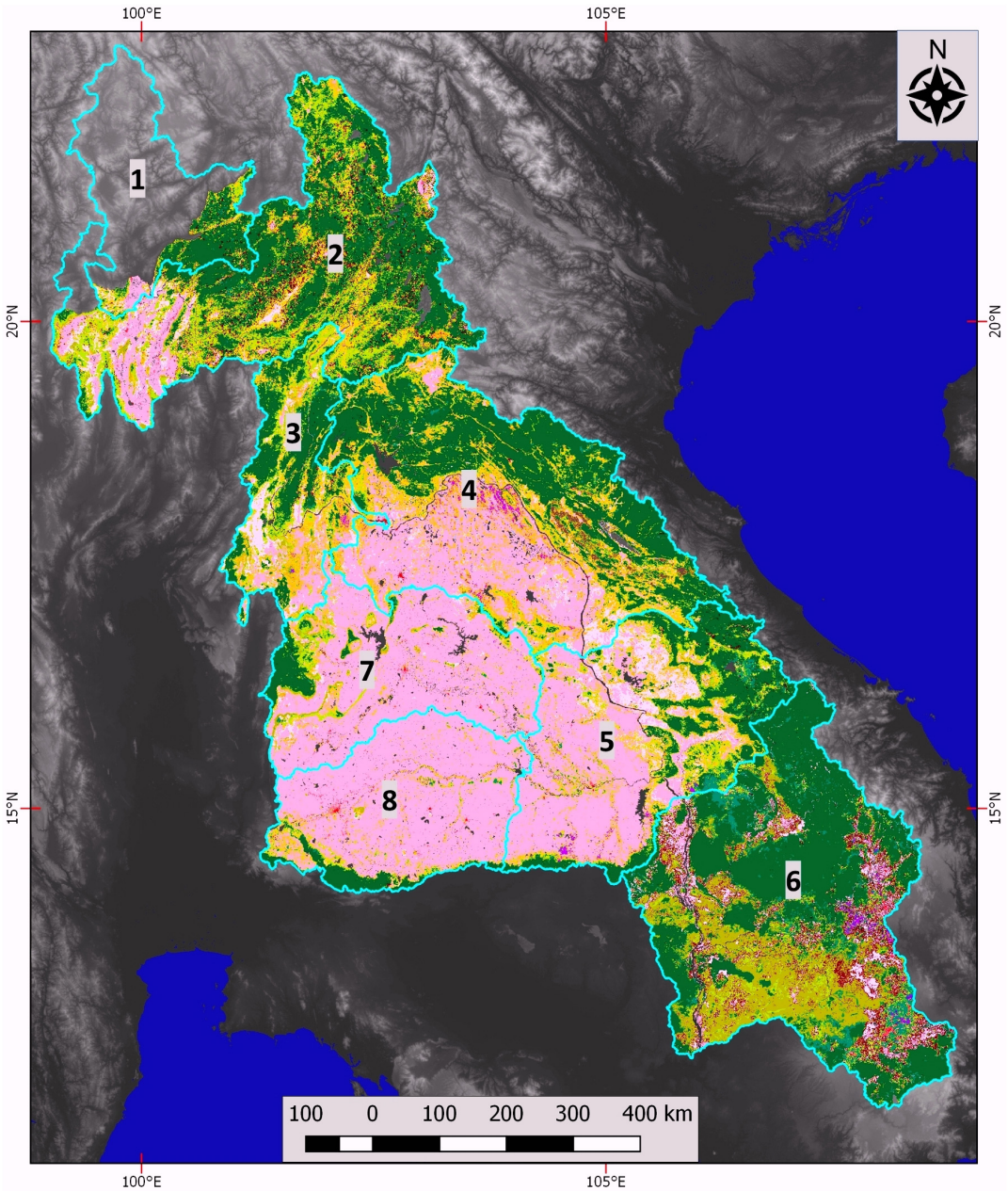


FIGURE 3 | Study area 2010 LULC map with SB boundaries overlain in cyan for SBs 1–8. The SB #s are included as black text. The backdrop image adjacent to the LULC map is a 90 m DEM from SRTM data.



Color	No Change - LULC Class Description	Color	Change - LULC Class Description
	No Change - Barren		Change to Barren in 2010
	No Change - Urban/Developed		Change to Urban/Developed in 2010
	No Change - Permanent Crops		Change to Permanent Crops in 2010
	No Change - Shifting Cultivation		Change to Shifting Cultivation in 2010
	No Change - Scrub/Shrub/Herbaceous		Change to Scrub/Shrub/Herbaceous in 2010
	No Change - Deciduous Forest/Scrub		Change to Deciduous Forest/Scrub in 2010
	No Change - Evergreen Forest		Change to Evergreen Forest in 2010
	No Change - Industrial Forest Plantation		Change to Industrial Forest Plantation in 2010
	No Change - Wetland Cover		Change to Wetland Cover in 2010

FIGURE 4 | 1997–2010 LULC change map for study area with SB boundaries overlain in cyan for SBs 1–8. All LULC change classes are colorized to show the LULC class for the end date of 2010, as well as whether the class was of a "change" or "no-change" type, given the change map consists of both change and "no change" classes. For a given LULC category, the "change" class color was assigned a lighter color than the "no change" class (e.g., barren no change was set to dark brown, while change to barren was set to light brown). The SB #s are included as black text. The backdrop image adjacent to the LULC map is a 90 m DEM from SRTM data.

TABLE 2 | Total area for the 25 most common change classes of the 1997–2010 LULC change map, along with # samples drawn per class.

LULC change class #	1997 LULC class description	2010 LULC class description	Total area (km^2)	% total area	Rank	Total random samples
11	Barren	Barren	1,071	0.2%	25	1
32	Permanent Crops	Urban/Developed	1,653	0.4%	23	1
33	Permanent Crops	Permanent Crops	130,540	28.9%	2	62
34	Permanent Crops	Shifting Cultivation	2,948	0.7%	20	1
35	Permanent Crops	Scrub/Shrub/Herbaceous	15,668	3.5%	5	7
36	Permanent Crops	Deciduous Forest/Scrub	7,809	1.7%	9	4
37	Permanent Crops	Evergreen Forest	4,897	1.1%	15	2
38	Permanent Crops	Industrial Forest Plantation	1,543	0.3%	24	1
43	Shifting Cultivation	Permanent Crops	5,028	1.1%	14	2
44	Shifting Cultivation	Shifting Cultivation	2,675	0.6%	21	1
45	Shifting Cultivation	Scrub/Shrub/Herbaceous	4,268	1.0%	16	2
46	Shifting Cultivation	Deciduous Forest/Scrub	7,129	1.6%	10	3
47	Shifting Cultivation	Evergreen Forest	5,295	1.2%	12	3
53	Scrub/Shrub/Herbaceous	Permanent Crops	2,104	0.5%	22	1
56	Scrub/Shrub/Herbaceous	Deciduous Forest/Scrub	3,906	0.9%	18	2
63	Deciduous Forest/Scrub	Permanent Crops	10,373	2.3%	7	5
64	Deciduous Forest/Scrub	Shifting Cultivation	3,887	0.9%	19	2
65	Deciduous Forest/Scrub	Scrub/Shrub/Herbaceous	5,119	1.1%	13	2
66	Deciduous Forest/Scrub	Deciduous Forest/Scrub	22,882	5.1%	4	11
67	Deciduous Forest/Scrub	Evergreen Forest	3,956	0.9%	17	2
73	Evergreen Forest	Permanent Crops	6,099	1.4%	11	3
74	Evergreen Forest	Shifting Cultivation	8,412	1.9%	8	4
75	Evergreen Forest	Scrub/Shrub/Herbaceous	12,357	2.7%	6	6
76	Evergreen Forest	Deciduous Forest/Scrub	36,928	8.2%	3	18
77	Evergreen Forest	Evergreen Forest	138,738	30.7%	1	67
Total						213

*See **Supplementary Table S1** for comparable summary areas for all mapped LULC change classes.*

given LULC change map pixel of a nominally 231.66 m spatial resolution. The apparent LULC class was assessed separately for the 1997 and 2010 time periods. For each sampled location, the reference LULC for each date was then additively combined using MS Excel into a change map score that was then compared to the test LULC change map to determine agreement or disagreement.

The assessment results were then used to summarize the percent overall agreement between the 1997–2010 LULC change map and reference datasets. An error matrix was constructed (**Supplementary Table S2**) to compute the percent overall accuracy for the LULC change map and the two LULC maps. This matrix given in basic form differs from guidelines of Olofsson et al. (2014), who suggested also including information on the percent area that a given LULC class represents in relation to the entire mapped area. This information on the percent of total area per LULC class is alternatively reported in **Supplementary Table S1**.

Percent overall agreement with reference data was computed for the LULC change map as well as the 1997 and 2010 LULC maps. In addition, individual class accuracies were computed in terms of user and producer agreement for the permanent crops and evergreen forest LULC classes, given that each of these two classes each were allotted 50 or more randomly sampled locations in the stratified random sample. The results of mentioned overall and individual classes agreement statistics are shown in **Figure 5**. In addition, percent overall agreement of the LULC change map was also computed for more simplified classification schemes in which included mature forest (grouping evergreen and deciduous forest types) and agriculture (grouping permanent crops and shifting cultivation classes).

The LULC class frequencies for the 1997 and 2010 maps were summarized in terms percent of total area for each mapped LULC class (**Figure 6**) with additional information provided in **Supplementary Table S3**. In addition, a Sankey diagram was computed to illustrate and further assess LULC change transitions between 1997 and 2010 (**Figure 7**). Summary areas for select LULC change classes were then merged into more general categories to further assess the main LULC changes across the study area (**Figure 8** and **Supplementary Table S4**).

RESULTS AND DISCUSSION

LULC Change Map Validation

Compared to available reference data, the LULC change map dataset produced an estimated overall agreement of 78% with 6.7% margin of error at the 95% confidence interval, based on 166 sample locations in agreement out of a total of 213 sample locations (**Figure 5**). However, the overall agreement level for the LULC change map with reference data increases to ∼82% (174/213 sample locations agreed) if one groups the evergreen and deciduous forest into 1 class. The overall agreement further

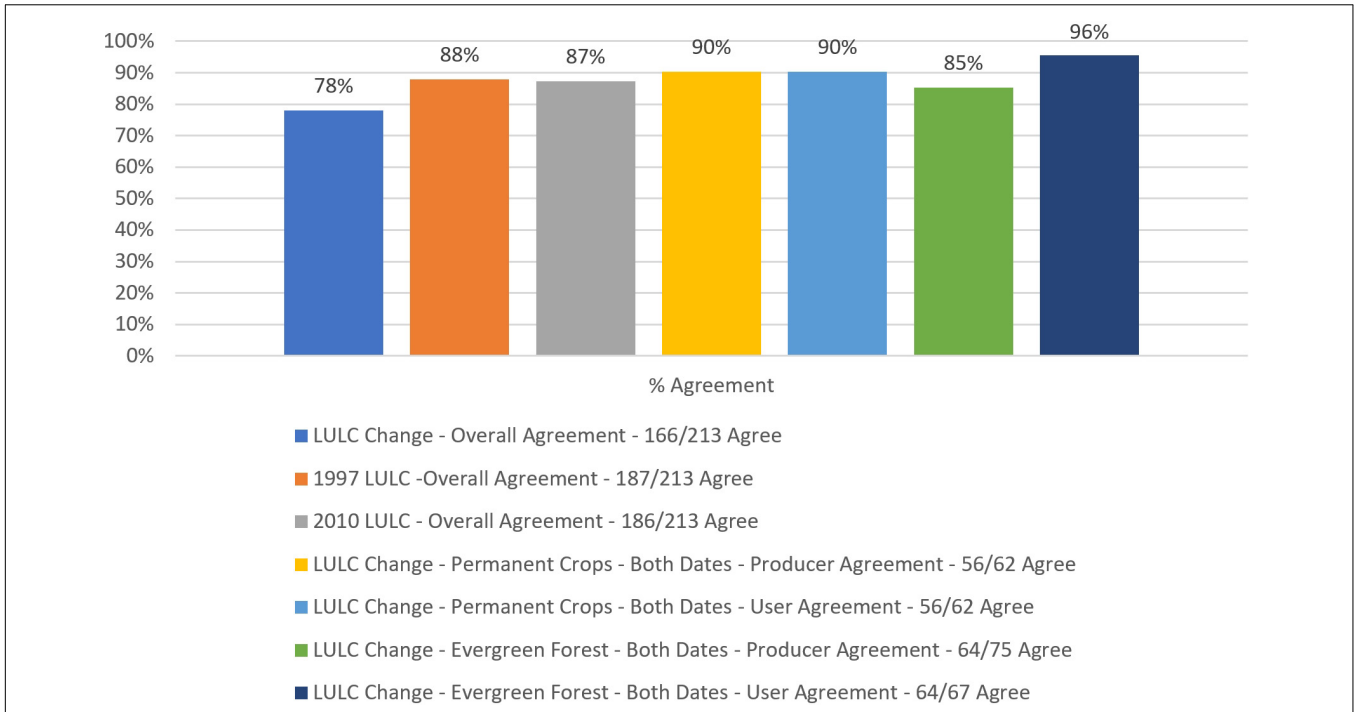


FIGURE 5 | Results of LULC change map accuracy assessment, including percent overall agreement of LULC change map, 1997 LULC map, and 2010 LULC map was well as percent user and producer agreement for the no change permanent crops and evergreen forest LULC change classes.

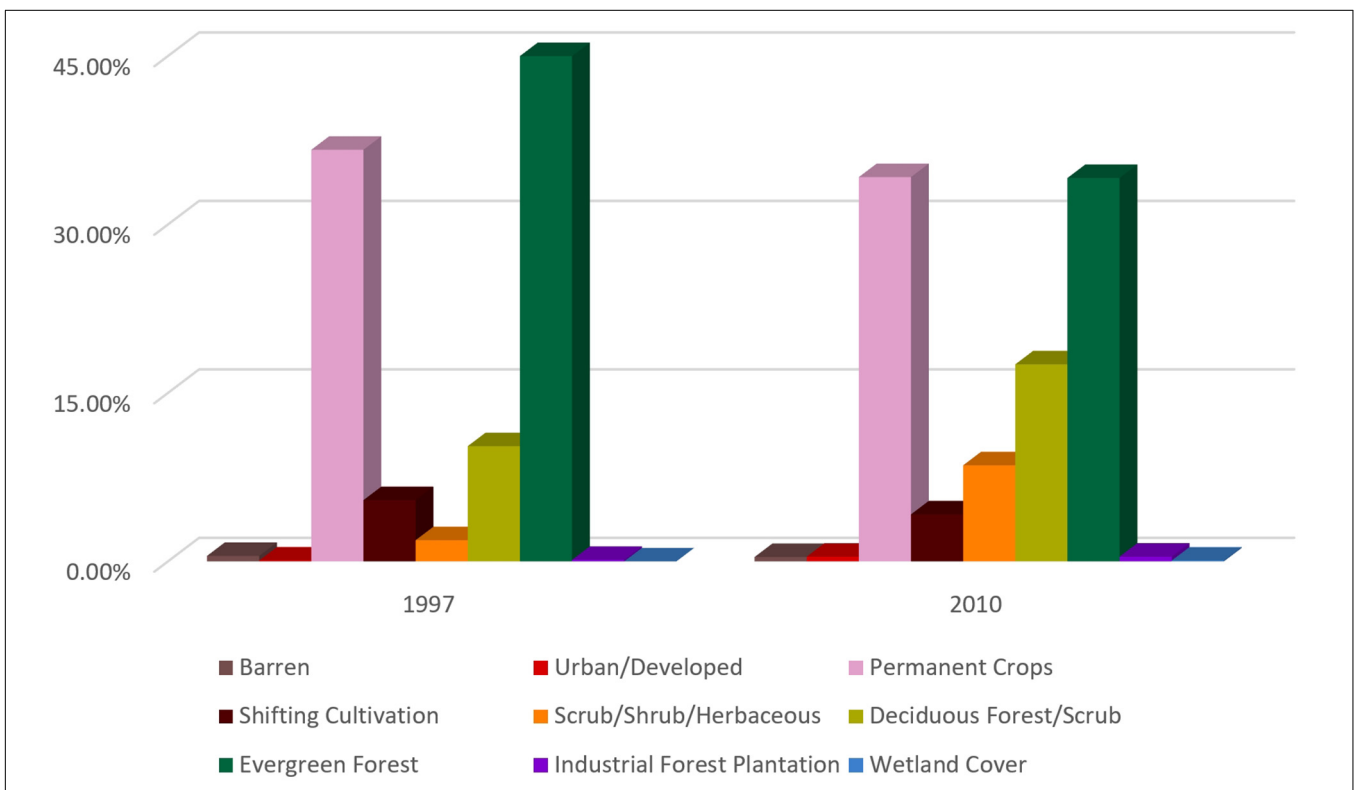


FIGURE 6 | Percent class frequency for the 1997 and 2010 LULC maps.

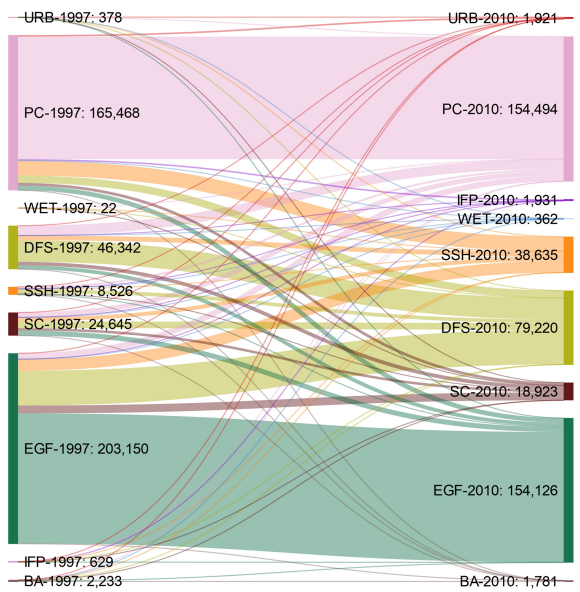


FIGURE 7 | Sankey diagram showing LULC change transitions for 1997–2010. LULC class frequency is reported in terms of km^2. LULC classes include: (1) BA (Barren); (2) URB (Urban/Developed); (3) PC (Permanent Crops); (4) SC (Shifting Cultivation); (5) SSH (Scrub/Shrub/Herbaceous); (6) DFS (Deciduous Forest/Scrub); (7) EGF (Evergreen Forest); (8) IFP (Industrial Forest Plantation); and (9) WET (Wetland Cover). For additional descriptions of these LULC classes, see **Table 1**.

improved to ∼83% (177/213 total) by grouping the forest classes into a mature forest category and grouping the agricultural classes into a single agriculture class. The level of overall agreement for the LULC change map with reference data was influenced in part by the complexity of the map in terms of total mapped 78 change classes out of possible maximum of 81 (**Supplementary Table S1**).

The validation of the change map also enabled estimates of percent overall agreement for each input LULC map compared to reference data (**Figure 5**). Both the 1997 and 2010 LULC maps (each with 9 total classes) showed similar levels of overall agreement to available reference data. The 1997 map produced an 87.8% overall agreement and the 2010 map yielded an 87.3% overall agreement. The observed disagreement between the change map and reference data (i.e., classification errors) were caused by disagreement on either the 1997 and/or 2010 LULC map compared to reference data, including 6 sample locations with errors disagreement on both dates of LULC maps, 20 locations with apparent errors only on the 1997 LULC map, and 21 locations with apparent errors only on the 2010 LULC map. The observed overall agreement for the 1997 and 2010 LULC maps to reference data is regarded as good for general LULC classification schemes and single dates of LULC maps, given that the overall agreement for both dates equaled or exceeded 85% (Anderson et al., 1976). These accuracy assessment results for the individual dates of LULC maps are also comparable to what was produced in another multi-date LULC mapping and change detection study discussed by Ellis et al. (2011) and Spruce et al. (2014).

In addition, the two most frequent LULC change classes, no change classes for permanent crops and evergreen forest, both yielded user and producer class agreements that were high, ranging from 85 to 96% (**Figure 5**). The frequency of occurrence for other mapped LULC change classes were much lower (**Table 2**) and as a result the number of samples for other LULC change classes were much less than 50 sample locations per class (e.g., the third most frequent LULC change class was allotted 18 samples), which is too small to viably estimate individual class producer and user agreement statistics. Given available resources, the main goal of the LULC map validation was to estimate the overall accuracy of the LULC maps. More analysis is needed to further assess accuracy of the low frequency, regionally scarce to rare LULC change classes.

The LMB SWAT models use single dates of LULC map data as an input and not the LULC change map dataset. Although the LULC change map is not to be directly utilized by LMB SWAT models, it provides information that is complementary to the 1997 and 2010 LULC maps.

LMB LULC Change Trends

The LULC maps for 1997 and 2010 are shown in **Figures 2**, **3** and the LULC change map for 1997–2010 is given in **Figure 4**. These maps collectively show areas with both LULC stability and change with the main LULC changes occurring in LMB sub-basins that have more extensive forests and are less agriculturally dominated. For this region, some conversion of forest to agriculture (i.e., deforestation) was observed on the LULC change map. Increases in permanent crops were apparent in multiple SBs, including SBs 3, 5, and 6. Some of such new permanent agriculture areas for 2010 were mapped as forest in 1997. In addition, some of the new permanent crops areas in 2010 were mapped as shifting cultivation in 1997. Agriculture dominated SBs (SBs 7 and 8) on the 1997 LULC map showed less overall change than other SBs that were not agriculturally dominated as of 1997. The SBs that were mixed agriculture and forest in 1997 tended to have more frequent, visibly apparent forest to agriculture conversion by 2010. Some areas adjacent to contiguous forests in 1997 appeared to convert from evergreen (broadleaved) forest or shifting cultivation to deciduous forest/scrub. SBs with extensive forest in 1997 tended to retain extensive forests as of 2010, especially with respect to more remote, higher elevations. As of 1997, much of the lowland forests (e.g., in the Khorat Plateau) were already largely converted to agriculture, including areas along waterways that were presumably woody wetlands. An expansion of industrial forest plantations (IFPs) was observed primarily in SBs 4 and 6. Some expansion of urban/developed areas was also observed from 1997 to 2010, though some of the increase in urban areas were due to commission errors on the 1997 map (compared to Landsat false color imagery) in addition to urban expansion. In SB 6, some increases in evergreen forests were observed on the change map, along with increases in permanent and shifting cultivation that occurred mostly adjacent to extensive deciduous forests.

Of the 25 most frequent LULC change classes, only 5 were for "no change" categories where the LULC class stayed the same for both dates (**Table 2**). These 5 most frequent non-change classes represented 65.6% of the total mapped area. In

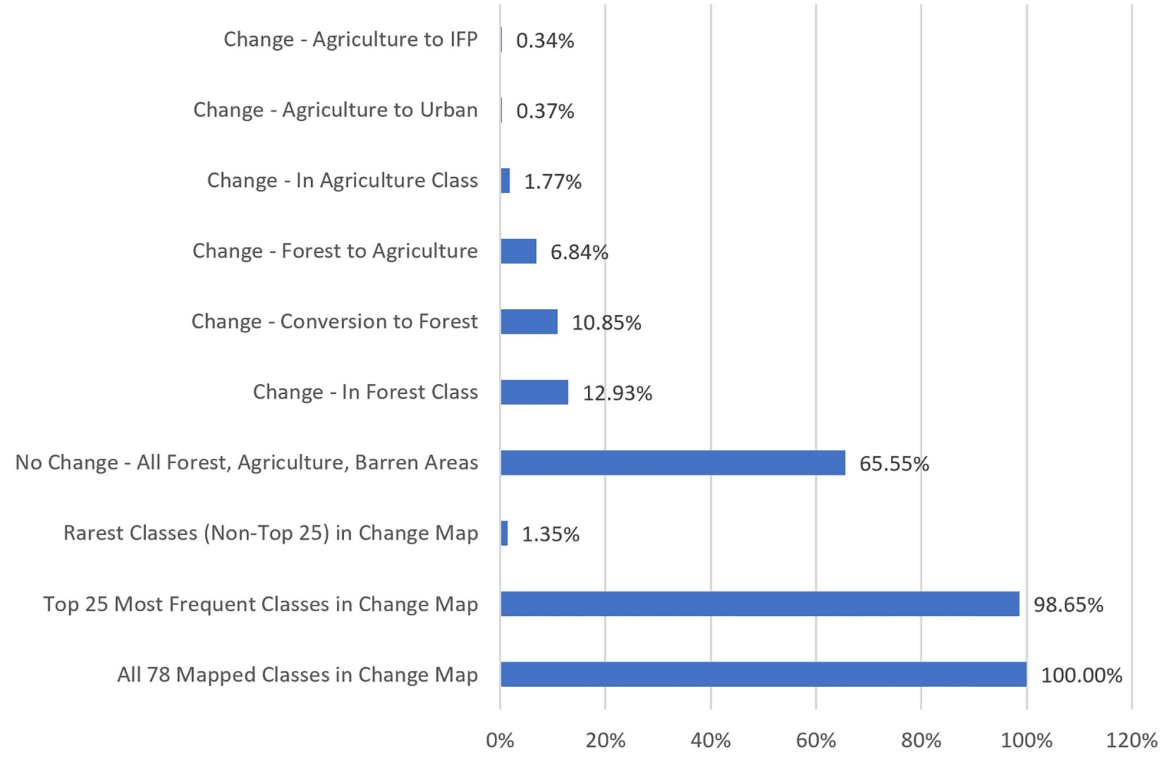


FIGURE 8 | Percent frequency of grouped LULC change categories compared to total mapped area. For this figure: (1) forests pertain to semi-natural deciduous forests, evergreen broadleaved forests, and scrub/shrub/herbaceous areas; (2) agriculture includes permanent crops and shifting cultivation areas; (3) IFP = Industrial Forest Plantations. For additional information, see **Supplementary Table S4**.

addition, 3 of the top 5 most frequent LULC change classes were for non-change categories. The latter categories regarded evergreen forest, permanent crops and deciduous forest/scrub and collectively represented ∼64.7% of the total mapped area. The no change classes for evergreen forest and permanent crops covered 30.7 and 28.9% of the total area, respectively. Collectively these 2 classes occurred on ∼59.7% of the total mapped area. Even for the more stable LULC no change classes (e.g., evergreen forest and permanent crops) we also observed losses (to other LULC classes) and gains from other LULC class changes to a given LULC type. The more transitional LULC classes were as expected particularly dynamic, including the shifting cultivation, scrub/shrub/herbaceous, and deciduous forest/scrub classes that can be related to swidden cultivation practices. Of the 25 most frequent LULC change classes, 7 of these regarded permanent crops (including 1 no change class). In addition, there was a 5 total of LULC change classes each for shifting cultivation, deciduous forest/scrub, and evergreen forest LULC classes (including 1 no change class per mentioned LULC category). The scrub/shrub/herbaceous class mapped in 1997 was not one of the top 25 most frequent LULC change classes, which may be in part due to it being a highly transitional or forest successional LULC class as opposed to a more permanent (i.e., stable) LULC class. The 25 most frequent LULC change classes all together comprised 98.65% of the total mapped area. In contrast, the highly rare LULC change classes collectively represented 1.35% of the total area and consisted of 53 total classes with each class ranging from 0.00 to less than 0.24% of the total area. **Supplementary Table S1** includes summary area estimates for all the LULC change classes, including these rarer classes not reported in **Table 2**.

Figure 6 reports percent of total area of individual LULC classes recorded on the 1997 and 2010 maps and it, along with **Supplementary Table S3**, were used to identify multiple trends: Barren areas were reduced in 2010 slightly from 0.49 to 0.40% of the total mapped area. Although infrequent in occurrence on both dates, urban areas increased by ∼5 times, which was due to part to increased urbanization and the apparent under classification of urban areas on the 1997 LULC map. The permanent agriculture class decreased by ∼−2.5% of the total mapped area, which could be due to multiple factors such as abandonment of permanent crops or the conversion of permanent crops to forests and other LULC classes. The shifting cultivation class decreased as well by ∼−1.3%, which was expected due to a reported decreased use of this agricultural practice in the region (FAO/IWGIA/AIPP, 2015). In contrast, the scrub/shrub/herbaceous category showed a ∼6.7% increase in total mapped area that could be due to abandoned agricultural areas reverting to forest. The latter class also commonly occurs amongst shifting cultivation and permanent crops areas. The deciduous forest class also appeared to increase by ∼7.3% with some of this occurring in areas that

were mapped as evergreen broadleaved forest in 1997. This trend could be due in part to the methods and data used to classify forest deciduousness in 1997 compared to 2010. In addition, some agricultural areas in 1997 had changed to deciduous forest/scrub in 2010.

The evergreen broadleaved forests showed the largest change in area estimated of any LULC class with a 10.8% decrease in the total mapped area from 2010 versus 1997 (**Figure 6**). The decrease in evergreen forest is probably due to multiple factors. We think it is in part related to the remotely sensed data and mapping method used to derive the 1997 LULC map. The forest deciduousness levels for the 1997 LULC map did not include Landsat image data collected at monthly intervals across the calendar year, while the 2010 LULC map used the monthly MODIS NDVI data to help classify forest deciduousness according to vegetation greenness phenology (Spruce et al., 2018). Given that the 2010 map used higher temporal resolution MODIS data to classify forest deciduousness, we think that the 2010 LULC map probably shows forest deciduousness classes more realistically compared to the 1997 map dataset. The production of the 2010 LULC map leveraged the vegetation greenness phenology to classify forest types based on deciduousness and multiple agricultural types with comparatively unique phenology (Spruce et al., 2018). Other potential factors include conversion to agriculture or transition to deciduous forest as a result of timber harvesting according to Rundel (2009). In addition, the drought of 2010 (Zhang et al., 2013; Guo et al., 2017) may be a factor why there was less evergreen forest in 2010, given that drought can cause deciduousness in otherwise evergreen forest (Rundel, 1999).

In addition, the IFP was mapped in 2010 was ∼3 times more than in 1997 but still a very small percentage of the total mapped area (0.3%) (**Figure 6**). Although rare in occurrence (0.1% of total area), about 15 times more wetlands were mapped in 2010, perhaps due to: (1) certain permanent agricultural areas on wetland sites in 1997 had apparently been abandoned and reverted to wetland vegetation in 2010; and (2) occasional commission errors observed on the 2010 LULC map.

The Sankey diagram (**Figure 7**) shows LULC class change transitions occurring from 1997 to 2010. The transitions are colorized on this diagram so that one can trace back from 2010 to 1997 to assess how LULC changed. Most of the permanent crops and evergreen forest areas in 2010 were also the same class in 1997. Some of the permanent crops in 2010 was previously either a forest class or shifting cultivation in 1997. Also, certain areas mapped as agriculture in 1997 transitioned to the evergreen forest class by 2010. The deciduous forest class in 2010 was previously mapped as a mixture of forest and agricultural classes in 1997. Also, some areas with forest and agricultural classes in 1997 had transitioned to the transient scrub/shrub/herbaceous class in 2010. Only a few scrub/shrub/herbaceous areas in 1997 were retained as the same class in 2010, which is related to the transitional nature of this LULC class. Most of the shifting cultivation areas in 2010 were previously mapped as evergreen forest in 1997. Some deciduous forest and agricultural areas in 1997 also transitioned to shifting cultivation in 2010. The remaining 4 of 9 mapped LULC classes in 2010 were comparatively rare and tended to be mapped as a wide variety of LULC classes in 1997. Taken as a whole, the Sankey diagram indicates that extensive forest and agricultural areas of the LMB region had the same LULC class for both dates, though some of the forest changed from evergreen forest to some other forest class or agriculture. Also, conversely, certain agricultural areas had transitioned to various forest classes. The mapped area at large includes mixture of forest and agriculture areas for both dates with occasional area specific fluctuations between forest and agricultural types occurring between the two dates. Also, there appears to be a diminished frequency of shifting cultivation by 2010, which is possibly related to afforestation or else conversion to permanent agriculture. The shifting cultivation type is visibly dynamic for 1997 versus 2010 as well.

More generalized groupings of the 25 most frequent LULC change classes are summarized as a percent of total mapped in **Figure 8** (see **Supplementary Table S4** for additional information). The most frequent group was for no change LULC classes, representing 65.55% of the total mapped area. Changes in forest class was 2nd most frequent group estimated at 12.93% of the total mapped area. In this analysis, forest classes are regarded as being either evergreen forest, deciduous forest, or scrub/shrub/herbaceous (i.e., forest regeneration). The 3rd most frequent LULC change group was for non-forest (mostly agricultural classes) changing to some form of forest. The latter change group represented 10.85% of the total area mapped. Forest to agriculture conversion occurred on ∼6.84% of the total area. The remaining 3 LULC class groups shown in **Figure 8** were much less frequent in occurrence, including 0.34% (of total area) for changes from agriculture to IFP, 0.37% for change from agriculture to urban, and 1.77% for changes in agricultural class (e.g., from shifting cultivation to permanent crops).

Some of the observed 1997–2010 LULC change areas (**Figure 4**) were for locally common yet regionally scarce to rare classes, including LULC classes of concern such as those related to the expansion of IFPs and urban areas (**Supplementary Table S1**). Such infrequent yet important LULC change classes were not sampled in the accuracy assessment, due in part to the comparatively low class frequency, the adopted stratified random sampling approach, plus the time and resources available to do the accuracy assessment. In addition, other scarce classes (that were more frequent than rare but less frequent than common classes) were only minimally sampled in the accuracy assessment. The utilized stratified random sampling approach did provide a means to assess the overall agreement between the LULC change map and reference data. It also yielded some insight into the ability of the two most predominant LULC change classes (evergreen forest and permanent crops no change categories) to be mapped. However, more work is needed to further assess map accuracy of other individual LULC change classes, such as the comparatively scarce to rare LULC change classes.

More effort is also be needed to map and assess LULC change according to a more specific LULC scheme. To construct the change map, the adopted LULC classification scheme for agriculture was more general than desired, given that the 1997

LULC only included one general permanent crops class and that it did not distinguish the number of crops per year and between row crops and rice paddies (Spruce et al., 2018). As a result, the LULC scheme used for both 1997 and 2010 in this study included only 1 all-encompassing permanent crops class, due to the 1997 LULC map only containing 1 permanent crops class. It would be useful from a SWAT modeling perspective to know more about the changes in irrigated rice cultivation from 1997 to 2010. Although 1 permanent crops class in the overall LULC scheme is sub-optimal desired for MRC SWAT modeling applications, this general category was still useful in our study for assessing the change status of permanent cropped areas within the study area and observation period.

The assessment of the LULC change map also revealed other certain limitations. For example, the 1997 LULC map did not have complete coverage for all 8 SBs that were used in the SWAT modeling framework, perhaps due to clouds on the utilized Landsat data and much less area mapped for SB 1. In contrast, the 2010 LULC map included full coverage for SB 1 and did not have no-data areas due to clouds.

Implications of Results

The 1997–2010 LULC change map indicates a mix of forest and agricultural LULC change classes occurred, including about 2/3 of the total area staying the same LULC class for both dates. Conversely, about 1/3 of the mapped area underwent some form of LULC change, according to the LULC change map. Although some of the study area has undergone noteworthy LULC change over the 13-year interval from 1997 to 2010, extensive areas of evergreen forest and permanent crops were not converted to some other land use. Some of the apparent LULC change during the observation period appeared to be of an ephemeral or transitional nature (e.g., change related to agricultural and forestry practices). Other observed LULC change was apparently more permanent in nature (e.g., forest cover to permanent crops). In addition, the LULC change map included areas with deforestation and afforestation, which help to illustrate the dynamic nature of LULC in the region during the observed 13-year time frame.

Some new IFP areas were observed with ∼3 times more IFP mapped on the 2010 vs. 1997 LULC map (**Figure 6**). The amount of mapped IFP in 2010 is a very low percentage of the total mapped area. However, the amount of IFP mapped on at least the 2010 date is somewhat conservatively estimated, given IFP mapping for 2010 was based on 1 date of dry season Landsat data in which some of the IFPs were highly green and other IFP patches were senesced. Improved classifications of IFP may be produced using multiple dates of Landsat data from different times of year when such is available in cloud-free form (Fan et al., 2015; Hurni et al., 2017). Based on spot checks on more recent high-resolution aerial and satellite data on Google Earth Pro and recent publications (Fox and Castella, 2013; Fox et al., 2014), it seems likely that there has been some expansion of IFP areas since 2010 in the SBs 1–8 of the LMB. More work is needed to produce a more up to date LULC map for assessing expansion of IFP areas as well as other LULC changes in the LMB since 2010. When doing so, it might be possible to improve maps of IFP areas using a combination of harmonized Landsat and Sentinel-2 data so that classification of phenology of rubber trees can be better leveraged in deriving more effective IFP maps. Such a strategy could increase the availably of cloud free data in the cloud prone tropical portion of the LMB for both the leaf-on and leaf-off states of IFP, though additional data processing may be needed to effectively combine use of the Landsat and Sentinel-2 data (Flood, 2017; Claverie et al., 2018). Use of time series data processing methods (Robinson et al., 2017; Jönsson et al., 2018) could further improve the Landsat/Sentinel-2 reflectance data products needed to map IFP areas across a broad region like the LMB.

We think there are multiple drivers affecting the patterns evident on the LULC change map with the importance of specific drivers somewhat dependent on the SB. LULC change related to permanent agricultural expansion and intensification, reduced shifting cultivation patch sizes, and forest practices could be related to economic and policy drivers. Part of the observed agricultural intensification may be related to the increased cultivation of cash crops. Also, the observed increase in industrial forest plantations mostly pertains to a woody cash crop (i.e., rubber plantations). We observed increased urbanization which in part could be due to demographic drivers (i.e., human population growth). Based on reviews of 1997 and 2010 Landsat imagery, we could see some forest change (e.g., deforestation and conversion to permanent crops) in areas where roads had expanded into forested areas. Road access was also identified as a LULC change driver by Rowcroft (2008) and Xing (2013). Based on summary areas per class for the 1997 and 2010 LULC maps (**Figure 6** and **Supplementary Table S3**), mature broadleaved forests (deciduous and evergreen forest combined) increased by ∼4% of the entire mapped area, while all agriculture (permanent and shifting cultivation) decreased by ∼4% of the total mapped area. The decreased extent of agricultural areas could possibly be explained by a combination of agricultural intensification, technical advances in agriculture that required less but more intensively cropped land, and agricultural conversion to other LULC types, such as semi-natural forest, industrial forest, and urbanized areas. We further believe that the water and land use planning and policy making in the region may have influenced forest and agricultural change occurring during the observed time frame. Although afforestation appears to have occurred in the LMB at large from 1997 to 2010, the change map also showed a decrease in evergreen broadleaved forest that is in part related to recovery from forest disturbances. Also, on the LULC change map, there are hot spot areas in certain SBs where deforestation is quite prevalent. Afforestation at the country scale was also recently noted in a 1990 to 2015 forest change map produced for the FAO 2015 Global Forest Resources Assessment for Laos, Thailand, and Vietnam (FAO, 2015). The same study noted that forest loss occurred in Myanmar and Cambodia, though very little area in Myanmar is mapped in LMB area covered by our maps. Also, our study area only partially covers Cambodia. Netzer et al. (2019) noted that considerable deforestation occurs is in the Tonle Sap sub-basin of Cambodia, which is outside the study area for our LULC change map.

The LULC change map showed a moderate level of overall agreement (∼78%) with reference data, which improved to ∼83%

by reducing the total number of LULC classes per date from 9 to 5 as follows: (1) merging the deciduous and evergreen broadleaved forest classes into 1 category; and (2) merging the permanent crops and shifting cultivation types into 1 class. At 9 classes per date, the 1997 and 2010 LULC maps used to make the change map each showed ∼87% level of overall agreement. Although imperfect, the LULC mapping products provided previously unavailable information on LULC change from 1997 to 2010. We acknowledge that the sample intensity used in the accuracy assessment of LULC mapping products in this study is lower than some studies, such as those by McCombs et al. (2016) and Yang et al. (2018). However, our study, as is, enabled quantification of the percent overall accuracy of the LULC change map and the individual LULC maps used to compute the change map. Such information, along with corresponding LULC maps is now available to aid SWAT modeling efforts in the non-coastal LMB, which helps given the lack of comparable LULC mapping studies in the region for the same time frame and purpose. The total sampling size used in the LULC change map accuracy assessment was constrained by available resources and time, though was greater than or comparable to the sample intensities used in LULC mapping studies by Ellis et al. (2011) and Spruce et al. (2014). However, the margin of error for our LMB LULC change map accuracy estimate could be reduced by increasing the sampling size.

We think it is also possible to improve LULC change map accuracy for the region and time period, though doing so would likely require different methods and data than was used in this project. The 1997 LULC map dataset that we used was based on Landsat data hard copies that were photo interpreted, delineated, labeled, and digitized. There also isn't any MODIS data available for 1997, given that the first MODIS sensor wasn't deployed until 2000. In contrast, the 2010 LULC map was based on both MODIS and Landsat data in conjunction with image classification and GIS techniques (Spruce et al., 2018). While it is unclear if the 2010 versus 1997 LULC change map data from our project can be improved with additional processing, a more viable 2010 versus 2003 LULC change map dataset could be derived using MODIS Terra/Aqua and Landsat data for both dates of LULC. In doing so, the begin date could be processed at a higher classification scheme specificity than what was feasible with the 1997 LULC map data set. In particular, the 1997 LULC map data only included one type of permanent crops, which is less desirable for SWAT modeling. The LULC change map for this project was generated to aid interpretation of SWAT modeling output using either the 2010 or 1997 LULC map data as inputs. However, the SWAT model requires a single date of LULC data as opposed to a LULC change map. Given this fact, the overall accuracy of each date of LULC map used in LMB SWAT models is probably a more important consideration to SWAT modelers than the overall accuracy of the LULC change map for application at hand (e.g., SWAT hydrologic modeling).

LULC map data is an important input to the SWAT hydrologic model in addition to data on soils, terrain, and precipitation (Gassman et al., 2007; Rossi et al., 2009). Based on the soils data alone (e.g., on hydrologic soil group), most of the LMB is regarded as having either moderately high or high runoff potential (Ross et al., 2018). The runoff risk is further acerbated at least seasonally by the high amounts of rainfall that can occur in this tropical region during the rainy season (MRC, 2010). LULC classes can have unique runoff potential which are described in part via Soil Conservation Service (SCS) curve numbers (USDA, 1986; Hong and Adler, 2008). For example, evergreen broadleaved forests have a lower runoff potential (i.e., SCS curve number) than agricultural cropland for given hydrologic soil group and hydrologic condition class (Hong and Adler, 2008). Extensive conversion of forests to permanent crops or other non-forests can change the runoff characteristics within a sub-basin (Netzer et al., 2019). Consequently, forests are a preferred LULC type by the MRC for catchments being used for water supplies (MRC, 2010).

The LULC change map data product from our study enabled more insight into the location, extent and kind of LMB LULC change in the region and within SBs (#1–8) used in SWAT hydrologic modeling applications, such as SWAT models of runoff for watersheds within a given SB. The project results increased understanding of LULC change in the region during the observed time frame, which could help those using and further developing LMB SWAT modeling products. For example, we think the change map from this project can be employed to aid use and development of LMB SWAT modeling products to support NASA/USAID SERVIR water and disaster management applications in the LMB (Mohammed et al., 2018a,b,c; McDonald et al., 2019). The LULC change map produced through this project (**Supplementary Data File 1**) can be used to view and assess potential 1997–2010 LULC change impacts for watersheds within LMB SBs 1–8. The data products from this project provide a baseline on LULC change for 1997–2010 that could be used to aid management of water, agriculture, and natural disasters in the region, as well as more in-depth follow-on studies on the LULC change occurring in the LMB since 2010. Such LULC change data sets could also possibly benefit the SERVIR Mekong Hub's Regional Land Cover Monitoring System (Saah et al., 2019b), which was recently applied by Ingalls et al. (2018) in LULC change assessment and land resource planning.

CONCLUSION

A LULC change mapping study for the LMB was conducted using GIS techniques along with 1997 and 2010 LULC map data sets to produce a LULC change map. Both dates of LULC maps have been used as inputs to the MRC SWAT hydrologic model framework for the SBs 1–8 of the LMB. The change map was derived to: (1) map, quantify, and assess 1997–2010 LULC change to aid assessment of the LMB SWAT modeling results; and (2) to provide previously unavailable information on LMB LULC change for 1997–2010 that could be used for supporting water and land resource management and planning. While much (almost 2/3rds) of the LULC change map showed no change, there were other observed areas with apparent LULC change. This was as expected given that the area includes a dynamic mix of semi-natural and human-dominated LULC. In addition, not all of the observed change appeared to be of a permanent nature, but

instead some form of ephemeral or transitional change associated with agricultural and/or forestry practices.

The study provided needed supplemental geospatial information on 1997–2010 LULC change to modelers for assessing LMB SWAT modeling results that are based on either the currently used 2010 or the previously used 1997 LULC map. The study also helped to identify LULC change hotspot locations within specific SBs where LULC change may be affecting hydrologic parameters of sub-watersheds that are modeled with the SWAT modeling framework. The project's LULC change map dataset (**Supplementary Data File 1**) and corresponding analysis discussed in this paper enabled a baseline record of the LULC change for 1997–2010 that is available for aiding follow-on studies in support of water, disaster, forest, and agricultural management efforts in the LMB. The 1997–2010 LULC change map from the project could possibly be refined with additional data processing techniques. More work is also needed to further update the 2010 LULC map now used in LMB SWAT modeling for the current time frame and to map LMB LULC change since 2010.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

The project design was developed by JS and JB. The project was implemented by JS with input from JB. The article was drafted by JS with input from JB. The article was reviewed by JB, IM, RS, and VL, revised and submitted by JS.

FUNDING

This work was funded from a 2011 NASA ROSES A.33 Earth Science Applications: Disasters grant (NNH11ZDA001N-DISASTER) in conjunction with a NASA Goddard Space Flight Center (GSFC) Applied Sciences grant (NNG15HQ01C) to Science Systems and Applications, Inc. (SSAI) for supporting work by the Hydrospheric and Biospheric Sciences (HBS) Laboratory at NASA GSFC.

ACKNOWLEDGMENTS

The project was benefitted by collaboration, documentation, and data from the Mekong River Commission provided by Dr. Sopheap Lim (who works with LMB SWAT models). The 2010 LULC map from the project employed MODIS monthly NDVI data products that were originally compiled for a project conducted at NASA Stennis Space Center to aid the USGS Forecast Mekong project. This study also aided by cooperation with the SERVIR Mekong Hub and the Asian Disaster Preparedness Center. Opinions, findings, and conclusions or recommendations found in this work are those of the author(s) and do not necessarily represent the views of NASA, SSAI, and or other participating organizations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2020.00021/full#supplementary-material

REFERENCES

Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E. (1976). *A Land Use and Land Cover Classification System for Use with Remote Sensor Data. Geological Survey Professional Paper 964*. Reston, VA: U.S. Geological Survey.

Bey, A., Sanchez-Paus Diaz, A., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., et al. (2016). Collect earth: land use and land cover assessment through augmented visual interpretation. *Rem. Sens.* 8:807. doi: 10.3390/rs8100807

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J. C., et al. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens. Environ.* 219, 145–161. doi: 10.1016/j.rse.2018.09.002

Congalton, R., and Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd Edn. Boca Raton, FL: CRC/Taylor & Francis, 183.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37, 35–46. doi: 10.1016/0034-4257(91)90048-b

Congalton, R. G. (2015). "Assessing positional and thematic accuracies of maps generated from remotely sensed data," in *Remotely Sensed Data Characterization, Classification, and Accuracies*, ed. P. Thenkabail, (Boca Raton, FL: CRC Press), 583–601.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., and Lambin, E. (2004). Review article digital change detection methods in ecosystem monitoring: a review. *Int. J. Remote Sens.* 25, 1565–1596. doi: 10.1080/0143116031000101675

Costenbader, J., Broadhead, J. S., Yasmi, Y., and Durst, P. B. (2015a). *Drivers Affecting Forest Change in the GMS: an Overview*. Rome: FAO/USAID LEAF. doi: 10.1080/0143116031000101675

Costenbader, J., Varns, T., Vidal, A., Stanley, L., and Broadhead, J. S. (2015b). *Drivers of Forest Change in the Greater Mekong Subregion. Regional Report*. Rome: FAO/USAID LEAF.

Dat, N. D. (2013). "Model setup in Mekong Basin," in *Presentation given by MRC Modeling Team From the Information and Knowledge Management Programme*, (Washington, DC: NASA Headquarters).

Ellis, J. T., Spruce, J. P., Swann, R. A., Smoot, J. C., and Hilbert, K. W. (2011). An assessment of coastal land-use and land-cover change from 1974–2008 in the vicinity of Mobile Bay. *Alabama. J. of Coastal Conservation* 15, 139–149. doi: 10.1007/s11852-010-0127-y

Estoque, R. C., Ooba, M., Avitabile, V., Hijioka, Y., DasGupta, R., Togawa, T., et al. (2019). The future of Southeast Asia's forests. *Nat. Commun.* 10:1829. doi: 10.1038/s41467-019-09646-4

Evers, J., and Pathirana, P. (2018). Adaptation to climate change in the Mekong River Basin: introduction to the special issue. *Climate Change* 149, 1–11. doi: 10.1007/s10584-018-2242-y

Fan, H., Fu, X., Zhang, Z., and Wu, Q. (2015). Phenology-Based Vegetation Index Differencing for Mapping of Rubber Plantations Using Landsat OLI Data. *Remote Sens.* 7, 6041–6058. doi: 10.3390/rs70506041

FAO (2015). *Global Forest Resources Assessment 2015 - Desk Reference, Report, Food and Agriculture Organization of the United Nation*. Rome: FAO, 245.

FAO/IWGIA/AIPP (2015). "Shifting cultivation, livelihood and food security - new and old challenges for indigenous peoples in Asia," in *Report Published Jointly by the Food and Agriculture Organization of the United Nations, International Work Group for Indigenous Affairs (IWGIA), and Asia Indigenous Peoples Pact (AIPP), Christian Erni Editor*, (Rome: FAO/USAID LEAF), 434.

Flood, N. (2017). Comparing Sentinel-2A and Landsat 7 and 8 using surface reflectance over Australia. *Remote Sens* 9, 659. doi: 10.3390/rs9070659

Fox, J., and Castella, J. C. (2013). Expansion of rubber (*Hevea brasiliensis*) in mainland Southeast Asia: what are the prospects for smallholders? *J. Peasant Stud.* 40, 155–170. doi: 10.1080/03066150.2012.750605

Fox, J., Castella, J. C., Ziegler, A., and Westley, S. (2014). Rubber plantations expand in mountainous Southeast Asia: What are the consequences for the environment? *Asia Pacific Issues* 114, 1–8.

Gassman, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G. (2007). The soil and water assessment tool: Historical development, applications, and future research directions. *Trans. of the ASABE* 50, 1211–1250. doi: 10.13031/2013.23637

Giri, C. P., Defourny, P., and Shrestha, S. (2003). Land cover characterization and mapping of continental Southeast Asia using multi-resolution satellite sensor data. *Int. J. Remote Sens.* 24, 4181–4196. doi: 10.1080/0143116031000139827

GMS-EOC (2019). *Climate Change in Lower Mekong Basin, Web Site for GIS Data Layers Used in USAID Mekong Adaptation and Resilience to Climate Change (ARCC) Project Conducted by ICEM.* Avaliable online at: http://portal.gms-eoc.org/maps?cmbIndicatorMapType=data&cmbIndicatorTheme=9&cmbIndicatorMap=30 (last accessed 5/10/19).

Guo, H., Bao, A., Liu, T., Ndayisaba, F., He, D., Kurban, A., et al. (2017). Meteorological drought analysis in the Lower Mekong Basin using satellite-based long-term CHIRPS product. *Sustainability* 9:901. doi: 10.3390/su9060901

Halcrow Group Limited (2004). "Tables of crop calendars, crop factors, and irrigation efficiency and return factors," in *Annex B.2 to: Decision Support Framework. Water Utilization Project Component A: Final Report*, (Phnom Penh: MRC Technical Reference Report, DSF 620 SWAT and IQQM Models), 11.

Heinimann, A., Messerli, P., Schmidt-Vogt, D., and Wiesmann, U. (2007). The dynamics of secondary forest landscapes in the Lower Mekong Basin. *Mt. Res. Dev.* 27, 232–241. doi: 10.1659/mrd.0875

Hong, Y., and Adler, R. F. (2008). Estimation of global SCS curve numbers using satellite remote sensing and geospatial data. *Int. J. Remote Sens.* 29, 471–477. doi: 10.1080/01431160701264292

Hurni, K., Schneider, A., Heinimann, A., Nong, D. H., and Fox, J. (2017). Mapping the expansion of boom crops in mainland Southeast Asia using dense time stacks of Landsat data. *Remote Sens.* 9:320. doi: 10.3390/rs9040320

Hussain, M., Chen, D., Cheng, A., Wei, H., and Stanley, D. (2013). Change detection from remotely sensed images: from pixel-based to object-based approaches. *ISPRS J. Photogramm. Remote Sens.* 80, 91–106. doi: 10.1016/j.isprsjprs.2013.03.006

ICEM (2013). *USAID Mekong ARCC Climate Change Impact and Adaptation Study for the Lower Mekong Basin: Main Report*. Bangkok: USAID Mekong ARCC Project.

Imai, N., Furukawa, T., Tsujino, R., Kitamura, S., and Yumoto, T. (2018). Factors affecting forest area change in Southeast Asia during 1980-2010. *PLoS One* 13:e0197391.

Ingalls, M. L., Diepart, J. C., Truong, N., Hayward, D., Neil, T., Phomphakdy, C., et al. (2018). "State of Land in the Mekong Region," in *Centre for Development and Environment, University of Bern and Mekong Region Land Governance*, (Bern: With Bern Open Publishing).

Jensen, J., Guptil, S., and Cowen, D. (2012). "*Change Detection Technology Evaluation*": *U.S. Census Bureau Geography Division, Deliverable #9 from Syneren Technologies Contract Task T007*, Report No. 240. Suitland, MA: Census Bureau Geography Division.

Jönsson, P., Cai, Z., Melaas, E., Friedl, M. A., and Eklundh, L. (2018). A method for robust estimation of vegetation seasonality from Landsat and Sentinel-2 time series data. *Remote Sens* 10:635. doi: 10.3390/rs10040635

Keskinen, M., Chinvanno, S., Kummu, M., Nuorteva, P., Snidvongs, A., Varies, O., et al. (2010). Climate change and water resources in the Lower Mekong River Basin: putting adaptation into the context. *J. Water Clim. Chang.* 1, 103–117. doi: 10.2166/wcc.2010.009

Kityuttachai, K., Heng, S., and Sou, V. (2016). "Land cover map of the Lower Mekong Basin," in *MRC, Technical Paper No. 59, Information and Knowledge Management Programme, Mekong River Commission*, Vol. 82, (Phnom Penh), 1683–1489.

Leinenkugel, P., Wolters, M. L., Oppelt, N., and Kuenzer, C. (2015). Tree cover and forest cover dynamics in the Mekong Basin from 2001 to 2011. *Remote Sens. Environ* 158, 376–392. doi: 10.1016/j.rse.2014.10.021

Lyon, S. W., King, K., and Polpanich, O. U. (2017). Lacombe G. assessing hydrologic changes across the Lower Mekong Basin. *J. Hydrol. Reg. Stud.* 12, 303–314. doi: 10.1016/j.ejrh.2017.06.007

Malyvanh, M., and Feldkotter, C. (1999). "Application of remote sensing and GIS for forest cover monitoring in LAO P.D.R," in *Conference Paper, Proceedings of Application of Resource Information Technologies (GIS/GPS/RS) in Forest Land & Resources Management, October 18 - 20, 1999*, Hanoi, 48–61.

Mas, J. F. (1999). Monitoring land-cover changes: a comparison of change detection techniques. *Int. J. Remote Sens.* 20, 139–152. doi: 10.1080/014311699213659

McCombs, J. W., Herold, N. D., Burkhalter, S. G., Christopher, J., and Robinson, C. J. (2016). Accuracy assessment of NOAA coastal change analysis program 2006-2010 land cover and land cover change data. *Photogramm. Eng. Remote Sens.* 82, 711–718. doi: 10.14358/pers.82.9.711

McDonald, S., Mohammed, I. N., Bolten, J. D., Pulla, S., Meechaiya, C., Markert, A., et al. (2019). Web-based decision support system tools: the soil and water assessment tool online visualization and analyses (SWATOnline) and NASA earth observation data downloading and reformatting tool (NASAaccess). *Environ. Modell. Software* 120, doi: 10.1016/j.envsoft.2019.104499

Minderhoud, P., Coumou, L., Erban, L., Middelkoop, H., Stouthamer, E., and Addink, E. (2018). The relation between land use and subsidence in the Vietnamese Mekong Delta. *Sci. Total Environ.* 634, 715–726. doi: 10.1016/j.scitotenv.2018.03.372

Mohammed, I. N., Bolten, J. D., Srinivasan, R., and Lakshmi, V. (2018a). Improved hydrological decision support system for the Lower Mekong River Basin using satellite-based earth observations. *Remote Sens.* 10:885. doi: 10.3390/rs10060885

Mohammed, I. N., Bolten, J. D., Srinivasan, R., and Lakshmi, V. (2018b). Satellite observations and modeling to understand the Lower Mekong River Basin streamflow variability. *J. Hydrol.* 564, 559–573. doi: 10.1016/j.jhydrol.2018.07.030

Mohammed, I. N., Bolten, J. D., Srinivasan, R., Meechaiya, C., Spruce, J. P., and Lakshmi, V. (2018c). Ground and satellite based observation datasets for the Lower Mekong River Basin. *Data in Brief* 21, 2020–2027. doi: 10.1016/j.dib.2018.11.038

MRC (2010). *State of the Basin Report 2010*. Vientiane: Mekong River Commission.

MRC (2014). *Crop Production for Food Security and Rural Poverty - Baseline and Pilot Modeling*. Vientiane: Mekong River Commission.

Nam, V. T. B., Haase, M., Kityuttachai, K., and Virak, S. (2015). "Land cover information catalogue of the Lower Mekong Basin," in *MRC Technical Paper, Information and Knowledge Management Programme*, (Phnom Penh: Mekong River Commission), 18.

NASA and JPL (2013). *NASA Shuttle Radar Topography Mission Global 1 arc second, distributed by NASA EOSDIS Land Processes DAAC.* Avaliable online at: doi: 10.5067/MEaSUREs/SRTM/SRTMGL1.003 (last accessed 5/11/19).

Netzer, M. S., Sidman, G., Pearson, T. R. H., Walker, S. M. W., and Srinivasan, R. (2019). Combining global remote sensing products with hydrological modeling to measure the impact of tropical forest loss on water-based ecosystem services. *Forests* 10, 413. doi: 10.3390/f10050413

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* 148, 42–57. doi: 10.1016/j.rse.2014.02.015

Olofsson, P., Foody, G. M., Stehman, S. V., and Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* 129, 122–131. doi: 10.1016/j.rse.2012.10.031

Patil, P., and Gumma, M. K. (2018). A review of the available land cover and cropland maps for South Asia. *Agriculture* 8, 111. doi: 10.3390/agriculture8070111

Pokhrel, Y., Burbano, M., Roush, J., Kang, H., Sridhar, V., and Hyndman, D. W. (2018a). A review of the integrated effects of changing climate, land use, and dams on Mekong River. *Hydrology* 10:266. doi: 10.3390/w10030266

Pokhrel, Y., Shin, S., Lin, Z., Yamazaki, D., and Qi, J. (2018b). Potential disruption of flood dynamics in the Lower Mekong River Basin due to upstream flow regulation. *Sci. Rep.* 8:17767. doi: 10.1038/s41598-018-35823-4

QGIS (2017). *QGIS Training Manual, Release 2.14*, 418. Available online at: https://docs.qgis.org/2.14/pdf/en/QGIS-2.14-UserGuide-en.pdf (accessed September 03, 2020).

Ribbe, L., Nauditt, A., Meinardi, D., Morbach, M., Becker, R., and Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) (2013). *Comparison of Key Drivers Regarding their Significance for Hydro-meteorological Extremes and their Impacts on Selected Hotspots Within the Mekong River Basin*. GIZ Report, No. 128. Cologne: University of Applied Sciences.

Robinson, N. P., Allred, B. W., Jones, M. O., Moreno, A., Kimball, J. S., Naugle, D. E., et al. (2017). A dynamic Landsat derived Normalized Difference Vegetation Index (NDVI) product for the conterminous United States. *Remote Sens.* 9, 863. doi: 10.3390/rs9080863

Ross, C. W., Prihodko, L., Anchang, J., Kumar, S., Ji, W., and Hanan, N. P. (2018). HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling. *Sci. Data* 5:180091. doi: 10.1038/sdata.2018.91

Rossi, C. G., Srinivasan, R., Jirayoot, K., Le Duc, T., Souvannabouth, P., Binh, N., et al. (2009). Hydrologic evaluation of the Lower Mekong River Basin with the soil and water assessment tool model. *Int. Agric. Eng. J.* 18, 1–13.

Rowcroft, P. (2008). Frontiers of change: the reasons behind land-use change in the Mekong Basin. *Ambio* 37, 213–218. doi: 10.1579/0044-7447(2008)37%5B213:foctrb%5D2.0.co;2

Rundel, P. W. (1999). "Forest habitats and flora in Lao PDR," in *Cambodia, and Vietnam. World Wildlife Fund Indochina Programme*, Hanoi, 171.

Rundel, P. W. (2009). "Vegetation of the Mekong Basin," in *The Mekong: Biophysical Environment of an International River Basin*, ed. I. C. Campbell, (Amsterdam: Academic Press), 143–160. doi: 10.1016/b978-0-12-374026-7.00007-3

Saah, D., Johnson, G., Ashmall, B., Tondapu, G., Tenneson, K., Patterson, M., et al. (2019a). Collect earth: an online tool for systematic reference data collection in land cover and use applications. *Environ. Modell. Software* 118, 166–171. doi: 10.1016/j.envsoft.2019.05.004

Saah, D., Tenneson, K., Matin, M., Uddin, K., Cutter, P., Poortinga, A., et al. (2019b). Land cover mapping in data scarce environments: challenges and opportunities. *Front. Environ. Sci.* 7:150. doi: 10.3389/fenvs.2019.00150

Spruce, J., Bolten, J., Srinivasan, R., and Lakshmi, V. (2018). Developing land use land cover maps for the Lower Mekong Basin to aid hydrologic modeling and basin planning. *Remote Sens.* 10, 1910. doi: 10.3390/rs10121910

Spruce, J. P., Smoot, J. C., Ellis, J. T., Hilbert, H., and Swann, R. (2014). Geospatial method for computing supplemental multi-decadal US coastal land use and land cover classification products, using Landsat data and C-CAP products. *Geocarto Int.* 29, 470–485. doi: 10.1080/10106049.2013.798357

Stibig, H. J., Achard, F., and Fritz, S. (2004). A new forest cover map of continental Southeast Asia derived from satellite imagery of coarse spatial resolution. *Appl. Vegetation Sci.* 7, 153–162. doi: 10.1111/j.1654-109x.2004.tb00606.x

Tran, H., Tran, T., and Kervyn, M. (2015). Dynamics of Land Cover/Land Use Changes in the Mekong Delta, 1973–2011: A Remote Sensing Analysis of the Tran Van Thoi District, Ca Mau Province, Vietnam. *Remote Sens.* 7, 2899–2925. doi: 10.3390/rs70302899

Tullis, J. A., Cothren, J. D., Lanter, D. P., Shi, X., Limp, W. F., Linck, R. F., et al. (2015). "Geoprocessing, workflows, and provenance," in *Remotely Sensed Data Characterization, Classification, and Accuracies*, ed. P. Thenkabail, (Boca Raton, FL: CRC Press), 401–421.

USDA (1986). *Urban Hydrology for Small Watersheds; Technical Release 55, Version 2.31*. Washington, DC: USDA Natural Resource Conservation Service, Engineering Division, 164.

Verberg, P. H., Neumann, K., and Nol, L. (2011). Challenges in using land use and land cover data for global change studies. *Global Change Biol.* 17, 974–989. doi: 10.1111/j.1365-2486.2010.02307.x

Wang, Y., Mitchell, B., Nugranad-Marzilli, J., and Bonynge, G. (2009). Remote sensing of land-cover change and landscape context of the National Parks: a case study of the Northeast Temperate Network. *Remote Sens. Environ.* 113, 1453–1461. doi: 10.1016/j.rse.2008.09.017

Xing, L. (2013). "'Land-Use change in the Mekong region,'" in *The Water–Food–Energy Nexus in the Mekong Region: Assessing Development Strategies Considering Cross-Sectoral and Transboundary Impacts*, eds A. Smajgl, and J. Ward, (New York: Springer), 179–190.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Case, A., et al. (2018). A new generation of the United States national land cover database: requirements, research priorities, design, and implementation Strategies. *ISPRS J. Photogramm. Remote Sens.* 146, 108–123. doi: 10.1016/j.isprsjprs.2018.09.006

Yasmi, Y., Durst, P., Haq, R. U., and Broadhead, J. (2017). *Forest Change in the Greater Mekong Subregion (GMS): An overview of Negative and Positive drivers. FAO Report.* Bangkok.

Zhang, B., Zhang, L., Guo, H., Leinenkugel, P., Zhou, Y., and Shen, Q. (2013). Climate and drought impacts on vegetation productivity in the Lower Mekong Basin. *Int. J. Remote Sens.* 35, 2835–2856. doi: 10.1080/01431161.2014.890298

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 11 June 2020
doi: 10.3389/fenvs.2020.00069

Modeling Invasive Plant Species in Kenya's Northern Rangelands

Edward Ouko[1], Steve Omondi[1], Robinson Mugo[1], Anastasia Wahome[1], Kenneth Kasera[1], Emmanuel Nkurunziza[1], John Kiema[1], Africa Flores[2,3], Emily C. Adams[2,3], Samson Kuraru[4]* and Margaret Wambua[5]*

[1] Regional Centre for Mapping of Resources for Development, Nairobi, Kenya, [2] SERVIR Science Coordination Office, NASA Marshall Space Flight Center, Huntsville, AL, United States, [3] Earth System Science Center, The University of Alabama in Huntsville, Huntsville, AL, United States, [4] Northern Rangelands Trust, Isiolo, Kenya, [5] Laikipia Wildlife Forum, Nanyuki, Kenya

Kenya is composed of diverse geographic regions and is heavily impacted by climatic variability. Habitat heterogeneity has led to a diverse number of plants and animals. Invasive species, however, threaten this biodiversity. This study mapped the current distribution of *Acacia reficiens* and *Opuntia* spp. using occurrence data, then applied a species distribution model to identify where suitable habitats occur under current and projected climatic scenarios under Representative Climate Pathways (R) 2.6 and 8.5. Occurrences of the two invasive plant species were sampled using an android-based application and a GPS (Global Positioning System) device. Predictor variables included: elevation, distance to streams and rivers, human population density, and vegetation indices (monthly Normalized Difference Vegetation Indices (NDVI) and Enhanced Vegetation Indices (EVI) derived from MODIS products 1-km spatial resolution). The mean of 25 replicates was used in identifying suitable habitats. We evaluated model performance using the average test AUC, mean testing omission rate metrics, and mean regularized training gain. The predictive models for both species performed better than random chance ($p < 0.05$). Mean test AUC values of 0.96 and 0.97 for *A. reficiens* and *Opuntia* spp. respectively, were achieved and their associated 95% confidence intervals showed the fitted models realized the high discriminative ability to differentiate optimal conditions for invasive plant species from random pseudo-absence points. The mean test AUC results for *A. reficiens* (0.97 ± 0.02) and *Opuntia* spp. (0.985 ± 0.01) were regarded as high. The models yielded moderate test gain values of 2.4 and 2.7, respectively. The model predictions show the distributions of *A. reficiens* and *Opuntia* spp. may increase under future climatic scenarios; with current extents estimated at 339,000 and 183,000 ha, respectively, with projected future spread reaching 732,800 and 206,900 ha, respectively, by 2070. Data on mapping, monitoring, and assessment of the invasive species can provide governments with insight into how the poor and vulnerable people are affected by the loss and degradation of biodiversity and ecosystems due to the spread of such species. This information is key in achieving the Sustainable Development Goals 15 (SDG) of the UN, aimed at the protection, restoration, and promotion of sustainable use of terrestrial ecosystems.

Keywords: invasive species, maxent, modeling, prediction, species distribution modeling, climate

OPEN ACCESS

Edited by:
Niall Patrick Hanan, New Mexico State University, United States

Reviewed by:
Jamil Hasan Kazmi, University of Karachi, Pakistan
Wenjie Ji, New Mexico State University, United States
David W. Kimiti, Other, Isiolo, Kenya

****Correspondence:***
Edward Ouko
eouko@rcmrd.org
Samson Kuraru
Samson.kuraru@nrt-kenya.org

Specialty section:
This article was submitted to Land Use Dynamics, a section of the journal Frontiers in Environmental Science

Received: *01 May 2019*
Accepted: *11 May 2020*
Published: *11 June 2020*

Citation:
Ouko E, Omondi S, Mugo R, Wahome A, Kasera K, Nkurunziza E, Kiema J, Flores A, Adams EC, Kuraru S and Wambua M (2020) Modeling Invasive Plant Species in Kenya's Northern Rangelands. Front. Environ. Sci. 8:69. doi: 10.3389/fenvs.2020.00069

INTRODUCTION

Invasive species are a major threat to global human well-being, biodiversity, and economies. The threat of incursion is directly linked to the rate at which the invasive species propagules are introduced and the level of disturbances that encourages their establishment (Hernández et al., 2006). Invasive plant species are plants that produce large numbers of offspring, with greater chances of spreading widely (Ratnayake, 2015). Deforestation, climate change, and habitat degradation have led to the loss of biodiversity and have allowed for the proliferation of invasive species (Thomas and Thomas, 2013). There are situations when native species are regarded as "invasive." This occurs when conditions responsible for controlling a species are weakened or absent, for instance, due to climate change. Suddenly, the species realizes an enormous and abnormal increase. *Acacia reficiens,* native to Kenya, have generally shown an aggressive invader tendency similar to those of invasive species. Though such species are important components of their habitats and ecosystems at large, some suggest they should not be referred to as "invasive" but "expansive" or "super-dominant" (Ratnayake, 2015).

A. reficiens is a bush, or small tree, about 3–4 m tall, the top is usually flattened, the branches are reddish-brown and it's a native of Ethiopia, Somalia, Kenya, Namibia, Sudan, Uganda, and Angola (Witt, 2017). *A. reficiens* are regarded as a very aggressive invader in many places, especially, but not necessarily, in disturbed areas. *Opuntia* spp. of the family *Cactaceae* is found in many arid and semi-arid parts of Kenya. They are drought tolerant with flattened succulent stems that keep moisture most of the year. The spines cause injury to the intestines and mouths of wildlife and livestock (Witt, 2017). *Opuntia* spp. is among the most spread and naturalized invasives in the arid and semi-arid areas of Northern Kenya, hampering rural livelihoods and the ecosystems. The invasion has been linked to changes in land-use practices which leads to degradation. They were introduced mostly intentionally for ornamental purposes (Obiri, 2017; Githae, 2019).

There is no known record or quantification of either current or potential distribution of *A. reficiens* or *Opuntia* spp. in Kenya. The use of species distribution models can help identify habitat suitability for the occurrence and potential distribution based on climate data. Correlative models allow for the prediction of species' potential niches by associating their occurrences to environmental proxies like climate, topography, vegetation indices, and then apply a relationship to identify areas of similar environmental conditions over which the species are likely to establish (Mitchell et al., 2016). The objectives of this research were to (1) use citizen science derived *A. reficiens* and *Opuntia* spp. occurrence data to map their current distributions, (2) map the current distribution of *A. reficiens* and *Opuntia* spp. in Samburu – Laikipia region using a time-series of MODIS vegetation indices and topographic environmental variables, and (3) predict the potential distribution under different climate change scenarios using bio-climatic variables.

MATERIALS AND METHODS

Study Area

The larger Laikipia-Samburu is a diverse ecosystem, consisting of different habitats and land use practices (Wittemyer et al., 2010). Laikipia is predominantly made up of large-scale ranches with resident wildlife species. Conversely, Samburu is a lower-elevation pastoralist grazing region composed of forested ranges (Omondi et al., 2002). The region is in a transition area for the three major vegetation types; semi-desert grassland, shrubland, and *Acacia*. The vegetation is mainly grassland, woodland, bushland, and dry forest with a scattered declining riparian forest that is important for the maintenance of the region's biodiversity. Forests cover 6% of Laikipia, and the region's soils are mainly black cotton, with significant areas also characterized by red sandy soils' (Jong et al., 2015). The rainfall is generally bimodal, where the long rains occur during April-May period, often accounting for 80% of the total yearly rainfall. The short rains fall later in October-November. The long-term annual precipitation mean between 1990 and 2010 was recorded at approximately 630 mm (Bergmann et al., 2016). The Laikipia drainage constitutes the upper Ewaso Ng'iro River catchment, which is the only major source of water. In dry spells, water flowing through perennial rivers are fed by the Mt. Kenya and Ndarua Range catchments (Government of Kenya, 2007; **Figure 1**).

Species Occurrence Data

The study focused mainly on five counties (Laikipia, Samburu, Isiolo, Marsabit, and parts of Meru) which have been heavily affected by *A. reficiens* and *Opuntia* spp. reducing forage spaces for livestock and wildlife. Both *Opuntia* spp. and *A. reficiens* occur in the same geographical space, so we used the same sets of environmental predictors.

Citizen Science Data

Because reliable spatial information on the invasive plant species was not available, we collected point locations from the study area using a custom-built electronic application installed in android phones, christened the Invasive Species Mapper (ISM) (see Appendix A) currently available on the Google Play Store. The ISM is customized to include any list of local invasive plant species, take photos, and work offline in remote areas with limited internet access.

We applied the ISM Android App, in fixed plot sizes of 1,000 m^2. This ensured that the data collected were consistent with assessments of the relative cover of invasive species in a particular location. In addition to fixed plot sizes, plots were stratified by the conservancy and randomized to ensure that the landscape was sampled consistently and that the plots without *Opuntia* spp. or other invasive species were also included. In addition to estimating the infested area within each plot, the total number (count) of individual plants were counted for the invasive species detected.

Local field assistants and rangeland coordinators that had functional knowledge of local vegetation distribution, especially

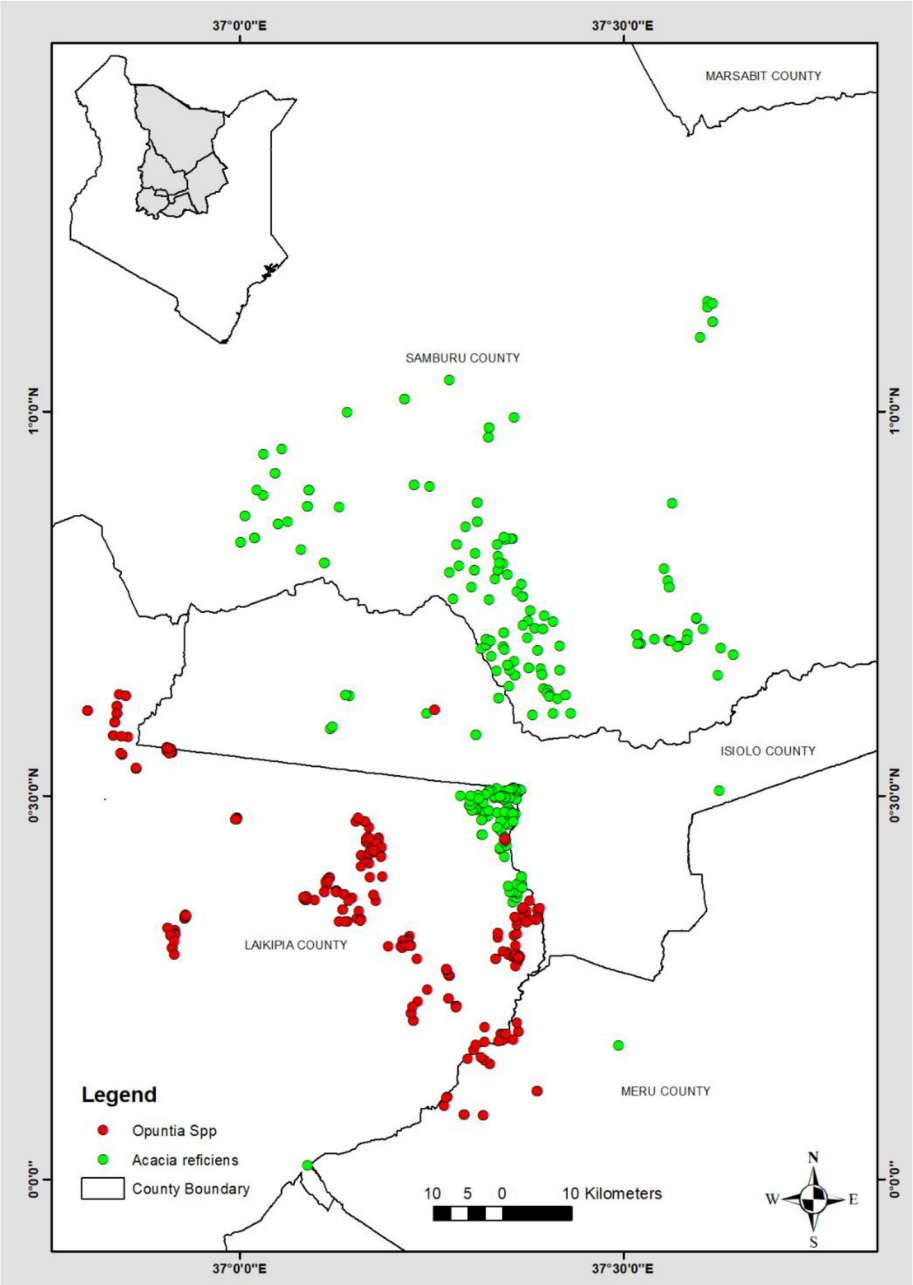


FIGURE 1 | Study area and locations of records used for modeling.

of the invasive species were trained on how to use the ISM in data collection. The data collected are remotely archived on an online platform for visualization and sharing[1] (see Appendix A). To reduce the spatial bias/autocorrelation, the data collection on *A. reficiens* and *Opuntia* spp. occurrences were random within a minimum distance of 200 m between the occurrence points. A total of 362 *A. reficiens* and 338 *Opuntia* spp. geo-tagged presence observations were randomly collected from 2016. Another set was collected from May to June 2017 and finally from September to November 2019. Five (5) counties of Meru, Samburu, Laikipia, Marsabit, and Isiolo were targeted. These counties comprise several conservancies with most occurrences recorded in Samburu and Laikipia counties (**Figure 1**).

Environmental Predictors

At the sub-national scale, we assumed climate is an important parameter for plant growth and survival. Correlative modeling of species ranges needs environmental data that have a direct

[1]http://mobiledata.rcmrd.org/invspec/

or indirect link to the species' spatial distribution. The proximal variables (direct) have an immediate effect on species while the distal variables (indirect) are dependent on the former with some varying degrees (Koh, 2008). Elevation, slope angle, and slope aspect are indirect variables and only correlate with organisms through their interactions with parameters like temperature and precipitation (Austin, 2007). Air temperature, soil water levels, and solar radiation directly define plant niches (Dymond and Johnson, 2002). The current and projected future climate data used were obtained at the highest-available resolution of (30 arc-seconds (∼1 km). The current climate data were derived through interpolation of observed data representing 1960 through 1990 (v.1.4). Future climate projections relied on predictions of the GISS – E2 – R model being part of CMIP5 (IPPC 5th Assessment) for climate average 2041–2060 for RCPs 2.6 and 8.5 being the lowest and highest emission climate scenarios, respectively. For comparison of the single-sourced derived current and future climate data, the WorldClim's current data (version 1.4) were used in the downscaling and calibration process[2]. MODIS products (MOD 13Q1) derived at 1-km^2 resolution comprising Enhanced Vegetation Index (EVI) and monthly Normalized Difference Vegetation Index (NDVI) were used as measures of vegetation productivity for the year 2017. Non-climatic variables were resampled to 0.00833 degrees (∼1 km^2 at the equator) using the nearest neighborhood algorithm to coincide with WorldClim climate derived predictors. Data sets used in modeling were derived from different sources at different resolutions, so scale conversion was done for consistent analyses (Park, 2011). Slope and elevation were derived from Shuttle Radar Topography Mission (30 m SRTM).

Modeling Approach

The MaxEnt program uses a maximum-likelihood algorithm to produce a probability distribution. The MaxEnt algorithm applies pixels of known species occurrence data and randomly generated pseudo-absence background data to form sample points (Hernández et al., 2006; Young et al., 2013). A collinearity analysis was computed to eliminate highly correlated variables. The correlation coefficient and the variance inflation factor (VIF) were calculated. The VIF helps measure inflation of variance of the coefficient estimate due to multicollinearity. In this study, the analysis of collinearity was done within the full list of original variables. Any variable with a VIF value above five was flagged off as collinearity concern (Dormann et al., 2013).

The distal variables are normally avoided because such indirect correlation tends to propagate errors when models predict species' potential niches under different climate scenarios (Baldwin, 2009; Pearson, 2010). A species distribution model (SDM) hierarchical process was applied. First, the invasive species were modeled specifically using the 19 bioclimatic variables[3]. The subsequent models run included the 24 MODIS variables representing monthly EVI and NDVI, elevation, soils types, distance to rivers, and population density. The relative importance of the predictors was determined by the percent contribution and permutation importance derived from MaxEnt results. The variables of less importance were removed systematically, leaving variables with a significant contribution to model performance, resulting in an AUC $>$ 0.8. The initial analyses reduced the number of variables to 8 non-correlated vegetation indices and topographic predictors for both *A. reficiens* and *Opuntia* spp., respectively, and seven non-correlated climatic variables for mapping the occurrence and projecting potential niches, respectively.

The resulting model was tested for variable correlation, because correlated variables may mislead interpretations. We ran a pairwise Pearson correlation in R Studio, and only variables with $r \geq \pm\ 0.8$ were used in the final model prediction. In the mapping of current *A. reficiens* and *Opuntia* spp., distribution, NDVI for August and May, and EVI for August and January were used. Other remote sensing proxies included: altitude, distance to rivers, soil drainage, and population density. The current climate data derived through interpolation of observed data representing 1960 through 1990 (v.1.4) were also used in mapping the current extent of the invasive species. These were Isothermality, mean diurnal range, temperature seasonality, precipitation seasonality, annual precipitation, precipitation of wettest quarter, and precipitation of wettest month. For predicting the potential habitats of the two species, similar current climate variables were used under RCPs 2.6 and 8.5, respectively (**Table 1**).

The model parameters were set as follows: replication type was set to sub-sample, 30% random test, the number of iterations was set to 5,000 with replicates of 25 and a regularization value of 1. The MaxEnt model allows one to run a model multiple times and then averages the results from all the model runs. A default setting of 10,000 background points was used and over 700 sites across the study area were used to ensure good representation of all environments (Elith et al., 2011). The model performance was evaluated based on the mean test AUC, testing omission rate, and mean training gain. The threshold used in converting MaxEnt probability outputs into binary maps have effects on the extent of predicted distribution (Baldwin, 2009). The minimum training presence logistic considers suitable all sites that are at least suitable within the training set. It is a conservative approach preferred in modeling invasive species. The last models used to identify suitable niches were built based on the means of 25 replicates grids in MaxEnt. The average logistic threshold was used in estimating optimal niches for the invasive plant species. Continuous output binary is created by choosing a value of the relative occurrence rate under which a given species being modeled is considered present (Merow et al., 2013). Determining biologically accurate thresholds may depend on a species population density or prevalence, for example in this study this information is not fully known, therefore arbitral threshold values are not recommended (Hernández et al., 2006). MaxEnt uses threshold-dependent and independent tests to evaluate a model output. For the threshold-dependent tests, it uses a specific threshold to divide a response as either suitable or unsuitable. A variety of threshold-dependent values are generated, and it's at the user's discretion to choose a value based on their objectives (Young et al., 2013). For instance, in the case of research or management reasons, an accurate

[2]http://www.worldclim.org/version1

[3]https://www.worldclim.org/bioclim

TABLE 1 | Permutation and percent importance of remote sensing and topo-climatic predictor variables of the two MaxEnt models for *A. reficiens* and *Opuntia* spp. under the current and potential distribution.

Variable	% contribution	Permutation importance	% contribution	Permutation importance
	Acacia reficiens		*Opuntia* spp.	
Altitude	48	60	53	70
Population density	29	19	N/A	N/A
May NDVI	11	9	4	5
Distance to rivers	7	2	N/A	N/A
August NDVI	6	1	8	5
Soil drainage	3	4	0	1
August EVI	2	2	19	8
January EVI	1	5	9	8
Isothermality	45	1	32	10
Temperature Seasonality	19	45	N/A	N/A
Mean Diurnal Range	19	6	N/A	N/A
Precipitation of the Wettest Month	7	4	N/A	N/A
Precipitation Seasonality	N/A	N/A	22	48
Annual Precipitation	NA	NA	14	29
Precipitation of the Wettest Quarter	NA	NA	10	5
Precipitation of Coldest Quarter	NA	NA	8	1

prediction of species presence (sensitivity) rather than species absence (specificity) would be of greater emphasis. A threshold weighted to achieve sensitivity would be ideal for this study. The models' significance against random chance is determined in MaxEnt using the threshold – dependent metrics which applies a one-tailed binomial test (Phillips, 2008). The output binary was converted to raster format and the distribution maps were generated by classifying images into two classes, 0.00 to 0.5 and 0.5 to 1.0. Pixels in the lower range were considered as areas of less than 50% chances of species occurrence and the higher ranges to depict areas with at least 50% probability of species existence, hence a highly suitable habitat. The areas which are projected to have changed under different climate scenarios were estimated by calculating the difference in future distribution and present distribution. The negative values denoted range expansion; the zero values denoted no change while positive values were areas of range reduction (**Figures 2**, **3**). In most cases selecting suitable modeling algorithms and associated datasets are a challenge in ecological modeling, so limitations of overestimation of the presence of species are inherent. The SDMs assume random sampling of the presence of species within a grid cell, which leads to a high probability of presence in each grid cell, and overestimation (Thapa et al., 2018).

RESULTS

The models generated $p < 0.005$, performing better than random prediction. The high mean test AUC values (0.97 and 0.98 for *A. reficiens* and *Opuntia* spp., respectively), is an indication that the models fitted could easily discriminate optimal conditions for the invasive species from the randomly generated background points. The highest test AUC value of 0.986 was derived from the models built for *Opuntia* spp. under future climatic projections (see Appendix B). An analysis of the relative contributions of individual remote sensing variables used in mapping the current distribution of both *A. reficiens* and *Opuntia* spp. showed that elevation was the greatest predictor of presence points of both *A. reficiens* and *Opuntia* spp. with percent importance of 48.3 and 53.4, respectively. Besides, climate variables whose contribution was significant in mapping the current extents of both species included temperature seasonality and Isothermality. In predicting the future extents under different climate scenarios, Isothermality was the most important variable predictor for both species (**Table 1**). We diagnosed the input variables based on the results of the jackknife procedure. This helps to identify the loss or gain in the predictor power as each variable is omitted from the models or used independently. A jackknife test of variable importance indicated that the variables which decreased model test AUC most when removed were August EVI and Soil drainage for *A. reficiens* and *Opuntia* spp., respectively, for the current distribution. Under future projection Precipitation of Wettest Month and Mean Diurnal Range reduced the test AUC value the most for *A. reficiens* and *Opuntia* spp., respectively, and contain information that is not present in other variables. Suitable habitats for both species under study were predicted widely in most conservancies. From the areas calculated based on the model results, the current extents of *A. reficiens* and *Opuntia* spp. were 339,000 and 183,000 ha, respectively, (**Figures 2**, **3**). The future distribution is predicted to expand to other conservancies within the projected climate scenarios within RCP 2.6 and RCP 8.5 by the year 2050 and 2070. Quantitatively, reduction of the suitable habitats is also expected to increase marginally especially for both species. This will be highly pronounced by

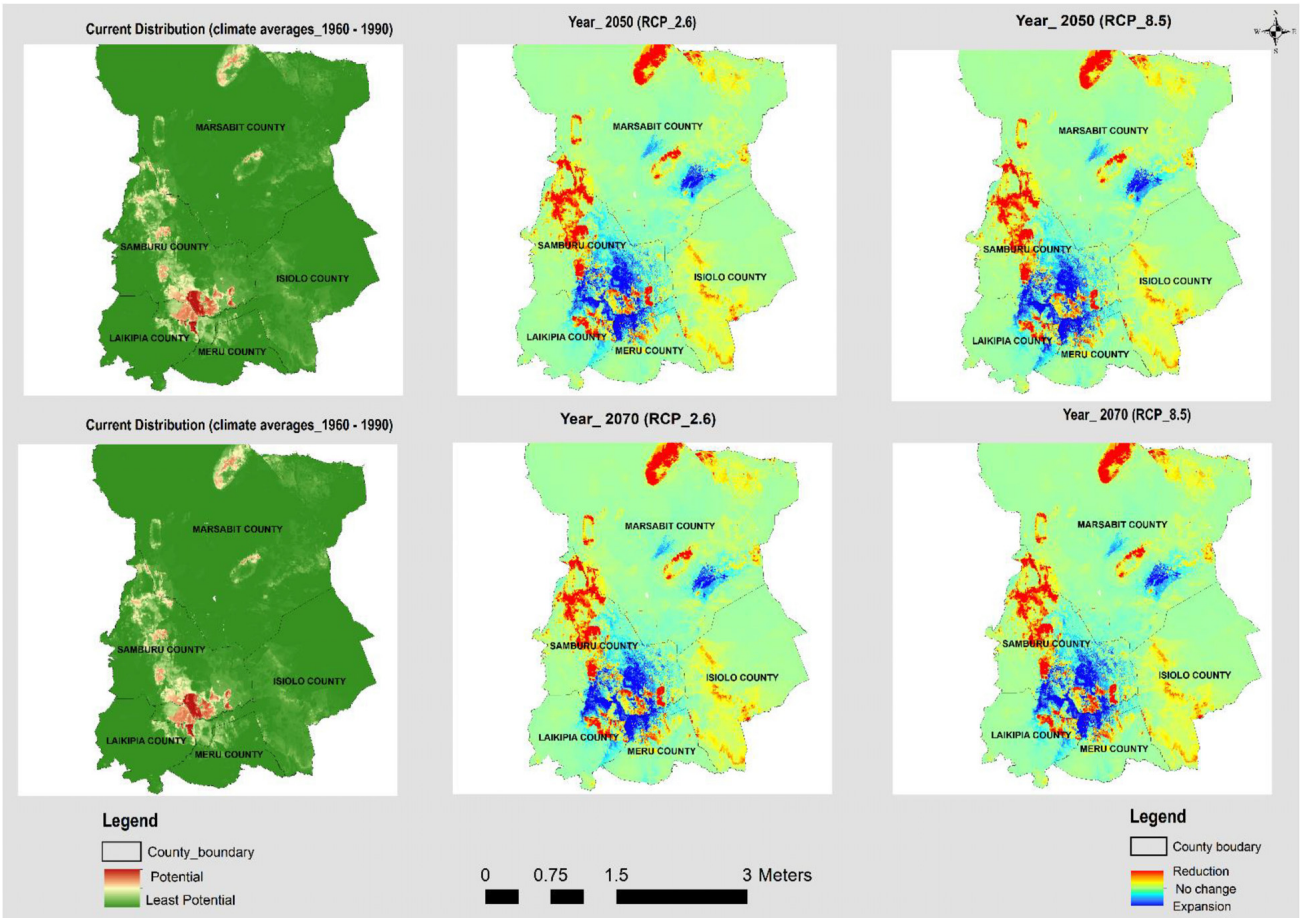


FIGURE 2 | The current and potential distribution under RCP 2.6 and RCP 8.5 for the years 2050 and 2070 for *A. reficiens*.

2070 (**Figures 2**, **3**). The future rates of expansion and reduction of suitable habitats are projected to shift consistently within the conservancies over different climate scenarios. The average proportion of suitable habitats for both species under reduction is slightly higher than areas under expansion over the same RCPs (see Appendix C).

DISCUSSION

In this study, altitude, population density, distance to rivers, NDVI of May and August, EVI of January and August were important variables in identifying the current extents of the two invasive species whose distributions were modeled. Minimal seasonal variations in temperature and rainfall were important predictors in mapping the current extents and in predicting suitable areas of invasion in 2050 and 2070. Generally, rainfall, human interventions, distance to rivers, soil drainage, precipitation seasonality, and temperature seasonality explained the distributions of *A. reficiens* and *Opuntia* spp. (see Appendix D). The model prediction shows the distributions of *A. reficiens* and *Opuntia* spp. are projected to shift (extend and reduce) under future climatic scenarios (2050 and 2070). This could pose an imminent threat to native plants and the well-being of the local communities who are largely pastoralists. The current distribution of *A. reficiens* and *Opuntia* spp. are mainly within the community conservancies within Laikipia and Samburu counties. The areas infested were overlaid on soil type information, and it showed that well-drained soils on the slopes of the hills with a mean elevation of about 1200 m above sea level, provided suitable conditions for the growth of *A. reficiens* and *Opuntia* spp. (see Appendices E, F). Livestock, humans, and wildlife feed on *Opuntia* spp. and are believed to be agents of dispersal (Witt, 2017). This is confirmed by the existence of *Opuntia* spp. on the banks of rivers and in densely populated areas within the conservancies. The relative probability of presence of *A. reficiens and Opuntia* spp. increases with the increase in altitude and population density, though they diminish with the increase of distance from rivers and streams (see Appendices E, F). The results show that the current predicted distributions of *A. reficiens* infestation cover mainly Samburu County, while *Opuntia* spp. is predominant in Laikipia County. Both range reduction and expansion for *A. reficiens* and *Opuntia* spp. are consistent over different climate scenarios.

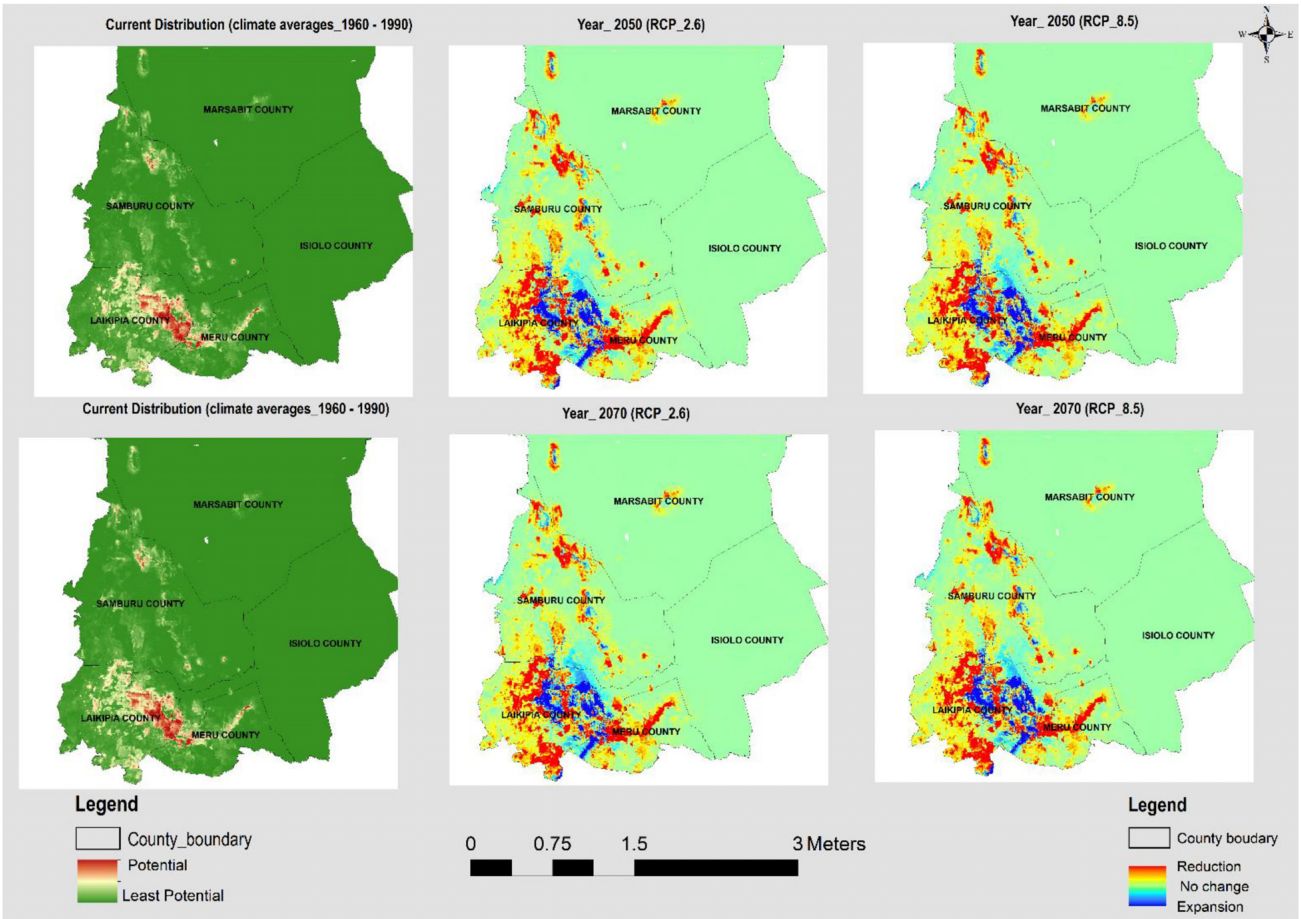


FIGURE 3 | The current and potential distribution under RCP 2.6 and RCP 8.5 for the years 2050 and 2070 for *Opuntia* spp.

The invasive species range expansion is projected to begin as early as 2050. *A. reficiens* and *Opuntia* spp. are projected to expand by 5 and 1%, respectively, relative to the study area, with expansion marginally increasing by 2070. Comparatively, the rate of reduction of suitable niches for both *A. reficiens* and *Opuntia* spp. is slightly higher than the rate expansion over 2050 and 2070 climate scenarios (see Appendix C). When *A. reficiens* encroaches landscapes, it results in an imbalance in bush grass ratios and decreasing biodiversity, lowering the grass productivity (Winowiecki, 2014). Its seeds germinate easily and can displace other species without disturbance. It is found mostly on plains but also grows on hills and dry rivers seldom on the sand. Even though *A. reficiens* causes challenges, several of its parts are utilized by the locals as a source of traditional medicine (Wakshum Shiferaw et al., 2018). Both fire and herbivory independently can affect tree cover by altering demographic height transitions. According to the local communities, *A. reficiens* is preferred by wild animals and livestock and could as well explain the increase in their invasion over-time.

The relative probability of presence of *A. reficiens* increases with increases in variance in temperature parameters (Isothermality, temperature seasonality and mean diurnal range) (see Appendix D). The phenological variations of *A. reficiens* were detected by NDVI of May and August, EVI of January and August. The grasses, crops, and other herbaceous vegetation are likely to turn brown in August and January, while the shrubs and woody vegetation would appear green making it possible to discriminate predominantly *A. reficiens* infested areas. Due to its deep root system, *A. reficiens* is likely to absorb moisture from precipitation of the driest months to remain green and can be easily discriminated in the dry months of August, January, February, and March due to contrast in general vegetation cover. With global climate change, the potential distribution of *A. reficiens* in northern Kenya may increase; thereby expanding the areas at risk of invasion. *A. reficiens* is an aggressive invader and is expected to increase its water use efficiency in the future allowing it to invade xeric habitats[4] (Accessed on the 1st of May, 2020). Its increased probability of presence in wide annual temperature ranges may enable it to out-compete other native species, especially for water resources.

[4]http://www.nbri.org.na/sites/default/files/treeatlas/pdf/TAP_Acacia%20reficiens.pdf

The relative probability presence of *Opuntia* spp. increases steadily with an increase in precipitation of the wettest quarter, precipitation seasonality, and annual precipitation (see Appendix D). They are easily identified in the dry months of August and January by both EVI and NDVI of those months. *Opuntia* spp. displace native species of plants, degrades the pasture, and inhibits the free movement of wildlife, people, and livestock. It also blocks access to water with a high frequency of replication, creating a dense colony over new areas (Githae, 2019). Most of *Opuntia* spp. have shoots comprising of flattened stem portions (cladodes) which are relatively tolerant to lower temperatures of −6°C, though extremely endures high temperatures of up 65°C (Nobel et al., 2003). It can thrive in rainfall regimes of 250–1200 mm per year with very hot summers of over 40°C. *Opuntia* spp. have a superficially extending root system, penetrating in the shallow and rocky substratum. Consequently, precipitation of the driest months reaches to their roots easily and enable them to thrive (Rocchetti et al., 2018).

The accuracy of the models was determined by the AUC value, which can either be ≤0.5 (no-better-than-random) or 1 (perfect). The discriminative ability of species distribution models has been widely tested using AUC statistics. There are concerns about relying on the AUC as the sole measure of model performance (Jiménez-valverde, 2012). Though AUC has been found to perform comparatively better than other measures in scenarios where the species, the target area, and the occurrence data (test and training samples) are uniform across the compared model, same case as this study (Lecours et al., 2016).

Though model overall accuracy is high for current and future maps of the two species, we believe this can be improved by increasing the number of samples within the wider Samburu Laikipia region. Ecological niche models usually suffer some limitations, for instance, over-estimation of the presence of species. MaxEnt uses presence-only data which may lead to high predicted figures for environmental conditions beyond the specified range. Though, this is avoided by applying a threshold value to the output raster (Thapa et al., 2018). The presence-only and climate-based models can potentially contribute to error propagation in the interpretation of the results, for instance, areas visited more often will depict strong geographic bias. The MaxEnt modeling technique employs cross-validation which uses fewer data sets as a way to offset the challenges of data deficiency. Besides, species distribution models can be biased resulting in local biasness and model generalization impacting on model reliability. MaxEnt depends on the jackknife kriging which maintains ordinary kriging simplicity and unbiasedness as well as reducing local-scale bias and over-generalization tendency (Odeny et al., 2019).

Information on current and possible future extent of invasive species would help ecosystem managers focus scarce conservation and restoration resources on areas with a high probability of invasion, narrowing down to areas with high probabilities of suitable environments and regions of lower probabilities values. Habitat suitability thresholds are usually selected subjectively with limited information to guide in choosing appropriate thresholds for presence- dependent modeling (Phillips et al., 2006). The results of this study demonstrated that climate and remotely sensed data can be analyzed to help in the development of predictive models over areas of conservation concern, providing conservationists with vital information in developing current and future eradication and control plans. Early detection methods help in the control or eradication of the invasive species thereby minimizing the control costs (Rejmánek and Pitcairn, 2002). Also, the outcome of this research will help achieve sustainable development goal number 15, which aims at protecting, restoring, and promoting sustainable use of terrestrial ecosystems. This research demonstrated the strengths of citizen science in data collection through a mobile-based application. Through this process, over 1,000 presence points of both *A. reficiens* and *Opuntia* spp. have been collected reducing financial and time constraints. By providing invasive species extent and distribution data to the conservation practitioners, the impacts of invasive alien species on land and water ecosystems can be reduced. Formulating policies on informed data would help in achieving the SDG target number 15.8 by 2020, through controlling and eradicating the priority invasive species.

We are cognizant of some challenges associated with species distribution modeling when applied to near and long-term climate projections (Thuiller and Maa, 2009; Jarnevich et al., 2015). In most cases, biological field data are biased representing an untrue picture of species distribution and abundance. Important areas may have been under-sampled, models may be heavily influenced by sample bias, and there may be mismatches in the resolution of sample plots relative to the resolution and accuracy of predictor variables. Furthermore, climate projections, especially long-term climate projections, have unknown (and often unknowable) accuracies and uncertainty associated with them. We view these model results as "hypotheses" that can only be verified and improved with iterative monitoring and modeling (Jarnevich et al., 2015).

CONCLUSION

Our findings suggest that suitable habitats for *A. reficiens* and *Opuntia* spp. are throughout most parts of Laikipia Samburu regions. The seeds from *A. reficiens* and *Opuntia* spp. can be dispersed by domestic, wild animals and by run-off water. This explains the distributions of these species mostly along the streams and river banks. The predicted expansion of *Opuntia* spp. and *A. reficiens* throughout most of their ranges means that if the seeds continue to be propagated by agents like floods, humans, and animals, increased vigilance is needed to identify and eradicate new invasion with focus on floods, humans, and wildlife dispersal. It will also be important to raise public awareness on the proliferation threats posed by invasive species, identification and, appropriate control measures. For the first time, this research estimated the current and potential extents of *A. reficiens* and *Opuntia* spp. in Northern Kenya. Though there was limited occurrence data, we anticipate further expansion of both species in most parts of Laikipia Samburu region. The study revealed that topo-climatic variables combined with remotely-sensed data (vegetation indices) can be used with the

invasive species occurrence data in a predictive model to quantify the current and potential extents of *A. reficiens* and *Opuntia* spp. The method used is easy and transferable to areas with similar challenges of invasive species. Currently, there is a push for increased evidence-based conservation, with challenges in collecting better site-specific information to prioritize areas for conservation and inform actions on priority areas. Documenting the distribution data on invasive species is important to this end. The predictive maps created from the models are being used as a baseline for current and future monitoring initiatives. Regional governments can use the modeled maps and the distribution data to help conservationists and decision-makers in the formulation of policies to assist in managing and monitoring the ecosystems. For future studies, we propose the inclusion of presence data from the broader Samburu Laikipia area including the species native ranges. High-resolution time-series images and additional variables such as grazing density may result in new findings on the distribution of invasive species in Northern Kenya.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: www.worldclim.com.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

The authors are greatly indebted to RCMRD/SERVIR-Eastern and Southern Africa Project, NASA, and USAID for providing the technical support and funding, respectively. Special thanks to Thomas J. Stohlgren and Paul H. Evangelista whose technical advice made a significant contribution to this publication.

REFERENCES

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecol. Model.* 200, 1–19. doi: 10.1016/j.ecolmodel.2006.07.005

Baldwin, R. A. (2009). Use of maximum entropy modeling in wildlife research. *Entropy* 11, 854–866. doi: 10.3390/e11040854

Bergmann, C., Roden, P., Bergmann, C., Ulrich, A., and Nüsser, M. (2016). Tracing divergent livelihood pathways in the drylands: A perspective on two spatially proximate locations in Laikipia County, Kenya Tracing divergent livelihood pathways in the drylands: A perspective on two spatially proximate locations in Laikipia Co. *J. Arid Environ.* 124, 239–248. doi: 10.1016/j.jaridenv.2015.08.004

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 027–046. doi: 10.1111/j.1600-0587.2012.07348.x

Dymond, C. C., and Johnson, E. A. (2002). Mapping vegetation spatial patterns from modeled water, temperature and solar radiation gradients. *ISPRS J. Photogramm. Remote Sens.* 57, 69–85. doi: 10.1016/S0924-2716(02)00110-7

Elith, J., Phillips, S. J., Hastie, T., and Dudı, M. (2011). A statistical explanation of MaxEnt for. *Divers. Distribut.* 17, 43–57. doi: 10.1111/j.1472-4642.2010.00725.x

Githae, E. W. (2019). *Status of Opuntia Invasions in the Arid and Semi-Arid Lands of Kenya*. Wallingford: CABI, doi: 10.1079/PAVSNNR201813003

Government of Kenya (2007). *Laikipia fact sheet 2005–2007*. Kenya: Government of Kenya.

Hernández, P. A., Catherine, H. G., Master, L. L., and Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29, 773–785. doi: 10.1111/j.0906-7590.2006.04700.x

Jarnevich, C. S., Stohlgren, T. J., Kumar, S., Morisette, J. T., and Holcombe, T. R. (2015). Ecological Informatics Caveats for correlative species distribution modeling. *Ecol. Inform.* 29, 6–15. doi: 10.1016/j.ecoinf.2015.06.007

Jiménez-valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species. 5, 498–507. doi: 10.1111/j.1466-8238.2011.00683.x

Jong, Y. A., De, Africa, E., Diversity, P., Program, C., Butynski, T. M., et al. (2015). *Laikipia County Geography, Environment, and Biodiversity*. Kenya: Eastern Africa Primate Diversity and Conservation Program, doi: 10.13140/RG.2.1.1257.2640

Koh, L. P. (2008). Assessing ecological responses to environmental change using statistical models. *J. Appl. Ecol.* 45, 1321–1329. doi: 10.1111/j.1365-2664.2007.0

Lecours, V., Brown, C. J., Devillers, R., Lucieer, V. L., and Edinger, N. (2016). Comparing selections of environmental variables for ecological studies: a focus on terrain attributes. *PLoS One* 11:e0167128. doi: 10.1371/journal.pone.0167128

Merow, C., Smith, M. J., and Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36, 1058–1069. doi: 10.1111/j.1600-0587.2013.07872.x

Mitchell, P. J., Monk, J., and Laurenson, L. (2016). Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. *Methods Ecol. Evol.* 8, 12–21. doi: 10.1111/2041-210X.12645

Nobel, P. S., Barrera, E., De, Nobel, P. S., et al. (2003). Tolerances and acclimation to low and high temperatures for cladodes, fruits and roots of a widely cultivated cactus, Opuntia ficus-indica. *New Phytol.* 157, 271–279.

Obiri, J. F. (2017). Invasive plant species, and their disaster-effects in dry tropical. (forests) and rangelands of Kenya and Tanzania. *J. Disaster Risk Stud.* 3, 417–428.

Odeny, D., Karanja, F., Mwachala, G., Pellikka, P., and Marchant, R. (2019). Impact of Climate Change on Species Distribution and Carbon Storage of Agroforestry Trees on Isolated East African Mountains. *Am. J. Clim. Change* 364–386. doi: 10.4236/ajcc.2019.83020

Omondi, P., Bitok, E., Kahindi, O., and Mayienda, R. (2002). Total Aerial Count of Elephants in\nSamburu-Laikipia. *Ecosystem*

Park, N.-W. (2011). The effects of spatial patterns in low resolution thematic maps on geostatistical downscaling. *Korean J. Remote Sens.* 27, 625–635. doi: 10.7780/kjrs.2011.27.6.625

Pearson, R. G. (2010). Species' distribution modeling for conservation educators and practitioners. *Lessons Conserv.* 3, 54–89. doi: 10.1016/S0140-6736(10)61462-6

Phillips, A. R. S. (2008). *A Brief Tutorial on Maxent*. Available online at: https://biodiversityinformatics.amnh.org/open_source/maxent/Maxent_tutorial2017.pdf (accessed March 25, 2019).

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Ratnayake, R. M. C. S. (2015). *Why plant species become invasive? Characters Related to Successful Biological Invasion*. Available online at: https://www.researchgate.net/publication/273452847_Why_plant_species_become_invasive (accessed April 20, 2019).

Rejmánek, M., and Pitcairn, M. J. (2002). "When is eradication of exotic pest plants a realistic goal?," in *Turning the Tide: The Eradication of Invasive Species*, eds C. R. Veitch, M. N. Clout (Gland: IUCN), 249–253.

Rocchetti, G., Pellizzoni, M., Montesano, D., and Lucini, L. (2018). Italian Opuntia ficus-indica Cladodes as Rich Source of Bioactive Compounds with Health-Promoting Properties. *Foods* 7:24. doi: 10.3390/foods7020024

Thapa, S., Chitale, V., Rijal, S. J., and Bisht, N. (2018). Understanding the dynamics in distribution of invasive alien plant species under predicted climate change in Western Himalaya Understanding the dynamics in distribution of invasive alien plant species under predicted climate change in Western Himalaya. *PLoS One* 13:e0195752. doi: 10.1371/journal.pone.0195752

Thomas, S., and Thomas, S. M. (2013). *Predicting the Spatial Distribution of an Invasive Plant Species and Modeling Tolerance to Herbivory Using Lythrum salicaria L. as a model system*. Graduate Theses and Dissertations, Iowa State University, Iowa.

Thuiller, X., and Maa, W. (2009). Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology* 90, 1301–1313. doi: 10.1890/08-0134.1

Wakshum Shiferaw, Demissew, S., and Tamrat Bekele. (2018). Invasive alien plant species in Ethiopia: ecological impacts on biodiversity a review paper. *Int. J. Mol. Biol.* 3, 169–176. doi: 10.15406/ijmboa.2018.03.00072

Winowiecki, L. A. (2014). *Baseline Assessment of Rangeland Health - Kalama and Namunyak Conservancies*. Kenya: Northern Rangelands Trust.

Witt, A. (2017). *Guide to the Naturalized and Invasive Plants of Laikipia*. Wallingford: CAB International, doi: 10.1079/9781786392152.0000

Wittemyer, G., Douglas-hamilton, I., Samburu, L., Kahindi, O., Wittemyer, G., King, J., et al. (2010). Employing participatory surveys to monitor the illegal killing of elephants across diverse land uses in Laikipia – Samburu, Kenya Employing participatory surveys to monitor the illegal killing of elephants across diverse land uses. *Afr. J. Ecol.* 48, 972–983. doi: 10.1111/j.1365-2028.2009.01200.x

Young, K. E., Abbott, L. B., Caldwell, C. A., and Schrader, T. S. (2013). Estimating suitable environments for invasive plant species across large landscapes: A remote sensing strategy using Landsat 7 ETM +. *Int. J. Biodivers. Conserv.* 5, 122–134. doi: 10.5897/IJBC12.057

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 19 June 2020
doi: 10.3389/fenvs.2020.00077

Wheat Area Mapping in Afghanistan Based on Optical and SAR Time-Series Images in Google Earth Engine Cloud Environment

Varun Tiwari[1], Mir A. Matin[1], Faisal M. Qamer[1], Walter Lee Ellenburg[2], Birendra Bajracharya[1], Krishna Vadrevu[3], Begum Rabeya Rushi[4] and Waheedullah Yusafi[1]*

[1] International Centre for Integrated Mountain Development, Kathmandu, Nepal, [2] Earth System Science Center, University of Alabama, Huntsville, AL, United States, [3] NASA Marshall Space Flight Center, Hunstville, AL, United States, [4] ENSCO, Inc., Falls Church, VA, United States

OPEN ACCESS

Edited by:
Niall Patrick Hanan,
New Mexico State University,
United States

Reviewed by:
Ahmad Khan,
University of Maryland, College Park,
United States
Qiuyan Yu,
New Mexico State University,
United States

***Correspondence:**
Mir A. Matin
mir.matin@icimod.org

Specialty section:
This article was submitted to Land Use Dynamics, a section of the journal Frontiers in Environmental Science

Received: 31 May 2019
Accepted: 19 May 2020
Published: 19 June 2020

Citation:
Tiwari V, Matin MA, Qamer FM, Ellenburg WL, Bajracharya B, Vadrevu K, Rushi BR and Yusafi W (2020) Wheat Area Mapping in Afghanistan Based on Optical and SAR Time-Series Images in Google Earth Engine Cloud Environment. Front. Environ. Sci. 8:77. doi: 10.3389/fenvs.2020.00077

Wheat is cultivated on more than 2.7 million hectares in Afghanistan annually, yet the country is dependent on imports to meet domestic demand. The timely estimation of domestic wheat production is highly critical to address any potential food security issues and has been identified as a priority by the Ministry of Agriculture Irrigation and Livestock (MAIL). In this study, we developed a system for in-season mapping of wheat crop area based on both optical (Sentinel-2) and synthetic aperture radar (SAR, Sentinel-1) data to support estimation of wheat cultivated area for management and food security planning. Utilizing a 2010 Food and Agriculture Organization (FAO) cropland mask, wheat sown area for 2017 was mapped integrating decision trees and machine learning algorithms in the Google Earth Engine cloud platform. Information from provincial crop calendars in addition to training and validation data from field-based surveys, and high-resolution Digitalglobe and Airbus Pleiades images were used for classification and validation. The total irrigated and rainfed wheat area were estimated as 912,525 and 562,611 ha, respectively for 2017. Province-wise accuracy assessments show the maximum accuracy of irrigated (IR) and rainfed (RF) wheat across provinces was 98.76 and 99%, respectively, whereas the minimum accuracy was found to be 48% (IR) and 73% (RF). The lower accuracy is attributed to the unavailability of reference data, cloud cover in the satellite images and overlap of spectral reflectance of wheat with other crops, especially in the opium poppy growing provinces. While the method is designed to provide estimation at different stages of the growing season, the best accuracy is achieved at the end of harvest using time-series satellite data for the whole season. The approach followed in the study can be used to generate wheat area maps for other years to aid in food security planning and policy decisions.

Keywords: sentinel 1, sentinel 2, GEE, crop type, random forest

INTRODUCTION

The agricultural sector in Afghanistan supports the livelihoods of nearly three-quarters of the total population and contributes nearly 28% to the Gross Domestic Product (GDP) (Muradi and Boz, 2018). Thus, agricultural growth is vital for driving the country's economy and for ensuring national food security (World Bank, 2014). Wheat is the most important crop in Afghanistan, followed by rice, barley, and cotton. Most cereal crops are utilized for self-consumption. Wheat is prominent in all of the major farming systems prevailing in the country and cultivated in every province. It dominates the total cultivated cereal area estimated as 2.7 to 3 million hectares. Despite being the dominant cereal crop in Afghanistan, the production of wheat fails to fulfill the internal demand. About 1 million tons (equivalent to 25% of internal demand) of wheat are imported annually to meet internal requirements (Martínez and Gilabert, 2009). This makes Afghanistan one of the leading importers of wheat in the world. Afghanistan imports wheat mainly from Turkmenistan and Pakistan, two of its neighboring countries. Timely and effective management and estimation of wheat production in Afghanistan are therefore of high importance for overall food security. It can help in managing local food demand and provide stability for social security (Tilman et al., 2011). It can also support decision-makers in national-level planning for formulation and implementation of policies related to food procurement, pricing, import-export, transportation and storage, advance planning, etc. (United Nations, 2013; Pham et al., 2017).

Limited work has been done in the past for wheat area estimation by utilizing a conventional ground-based sampling approach which only provides a qualitative assessment. In 2016, the Food and Agriculture Organization (FAO) carried out rice mapping using sentinel data in a few provinces of Afghanistan (Latham, 2017; Haworth et al., 2018). Similarly, the United Nations Office of Drugs and Crime (UNODC), a pro-active organization working in opium poppy monitoring using high-resolution satellite images, conduct annual assessments of opium poppy sown areas (Simms and Waine, 2016; Avetisyan, 2017). Concerning wheat sown area mapping, some qualitative assessments have been done in the past by the USDA (United States Department of Agriculture) using NDVI (Normalized Difference Vegetation Index) anomalies (Shahriar et al., 2014; Baker, 2015). Currently, the Ministry of Agriculture Irrigation and Livestock (MAIL), Afghanistan is undertaking yearly qualitative assessments of wheat sown area using ground sample data (crop cut survey) and some conventional remote sensing based techniques, i.e., mainly based on visual interpretation of satellite images (UN FAO, 2016). Recently, donor agencies like the United States Agency for International Development (USAID) have shown interest in food security management in Afghanistan. They have started projects, such as the Grain Research and Innovation (GRAIN) and the Kandahar Food Zone (KFZ), funded by USAID, started working in crop area mapping and health monitoring to support livelihoods in Afghanistan (USAID, 2017). Currently, no operational system exists in Afghanistan to provide a rapid assessment of wheat sown area which is essential to support the food security management.

Developing an operational system for wheat sown area assessment for Afghanistan is challenging despite the availability of several methods based on remote sensing. The major challenges include security concerns for collecting the reference data from the ground, small field sizes, cloudy optical imagery, low internet bandwidth for satellite data downloads, and limited computing infrastructure for data processing and analysis. Despite these challenges, through collaborative efforts with MAIL and other organizations in Afghanistan, we present a detailed study that develops a map of wheat sown areas utilizing advanced satellite remote sensing techniques which can be used to address food security planning and management in the country.

Crop type mapping using optical and SAR remote sensing techniques have been attempted by several researchers globally (Inglada et al., 2015). Optical remote sensing approaches use spectral-temporal profiles to identify seasonal thresholds of phenological characteristics to separate different crop types (Foerster et al., 2012). The acquisition time of the image is critical to identify seasonal thresholds and distinguish different crop types. Although spectral-temporal profiles based on seasonal thresholds require less ground sample points and provide good accuracy, they fail to classify crops having similar phenology. Specific to classification algorithms, machine learning classifiers, such as Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), etc. require systematic sampling approaches and a large number of accurate ground data for training the classification model (Camps-Valls et al., 2003; Murmu and Biswas, 2015; Tatsumi et al., 2015). Poor field level data can result in underfitting or over-fitting of the classification model and result in overestimation or underestimation of the classification results (Liakos et al., 2018). Although optical data have shown potential in the identification of crop types, the data is not reliable under cloudy conditions. Alternatively, Synthetic Aperture Radar (SAR) is an emerging technique in crop mapping (Oguro et al., 2001). SAR utilizes the temporal backscatter (physical) response of a crop and, along with machine learning techniques, can be effectively used for crop mapping and monitoring (Sonobe et al., 2014; Tamiminia et al., 2015; Gao et al., 2018). Recent studies utilize coarse to fine resolution satellite imagery for crop type mapping (Wardlow and Egbert, 2010). Some of the well-known approaches for crop type mapping using different sensors and resolutions are listed in **Table 1**.

In this study, we developed a system for in-season wheat sown area mapping by harnessing the power of multisensory remote sensing imagery (optical and SAR) and cloud computing (GEE) techniques (Dong et al., 2016; Gorelick et al., 2017). The system is designed keeping in mind the challenges in Afghanistan and provides the capacity for operationalization. The system can provide independent and evidence-based information on the status of annual crops at the province level. Ingesting field data at regular intervals for different seasons in the system will lead to higher accuracy in crop area estimates at the province level.

TABLE 1 | Crop mapping approaches.

Research topic	Imagery data	Spatial resolution	Temporal coverage	Approach
A comparison of MODIS 250-m EVI and NDVI data for crop mapping: a case study for southwest Kansas (Wardlow and Egbert, 2010)	MODIS	250 m	22 March to 30 September 2001	Spectral-temporal classification
Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery (Inglada et al., 2015)	Sentinel-2, SPOT4, Landsat 8	10–30 m	NA	Temporal-supervised classification
Corn monitoring and crop yield using optical and RADARSAT-2 images (Soria-Ruiz et al., 2007)	RADARSAT-2 Images, SPOT	3–100 m	NA	Temporal back scattered classification and LAI
Crop Classification Using Short-Revisit Multi temporal SAR Data. (Skriver et al., 2011)	Airborne–SAR		April to August 2006	Temporal back scattered classification
Crop identification using harmonic analysis of time-series AVHRR NDVI data (Jakubauskas et al., 2002)	AVHRR	1.1 km	NA	Harmonic Time series NDVI classification
Crop type mapping using spectral-temporal profiles and phenological information (Foerster et al., 2012)	Landsat TM/ETM	30 m	NA	Spectral-temporal profiles and Phenological information
First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. (Immitzer et al., 2016)	Sentinel	10 m	NA	Supervised classification
Toward operational radar-only crop type classification: comparison of a traditional decision tree with a random forest classifier (Deschamps et al., 2012)	RADARSAT 2	3–100 m	NA	Decision tree and random forest classifier
3D Convolutional Neural Networks for Crop Classification. (Ji et al., 2018)	Gafoen 2	15 m	NA	3D Convolutional Neural Networks

METHODS AND MATERIALS

Study Area

The study area (**Figure 1**) covers the whole of Afghanistan (34° 32′ and 38° 1′ 32.16″ N latitude and 69° 9′ and 38° 20′ 49.92″ E. longitude). The Hindu Kush mountain range divides the country into three very different geographic regions: (a) The central highlands, characterized by dry hot summers and very cold winters; (b) the southern plateau consist of sandy deserts with arable lands along the rivers; (c) the northern plains, which are highly fertile and include most of the land under agriculture. The total area of the country is 652,230 sq km with a population of 34.9 million. Agricultural lands represent 58% of the country with most designated as permanent pasture (48%), leaving only 11.8% as arable land (CIA, 2019). Total arable land is 6.5 million hectares of which 3.1 million ha is irrigated and 3.4 million ha is rainfed (FAO, 2010). Wheat, rice, barley, and maize are the main cereal crops grown in the country, with wheat accounting for 80.2% of total cereal production. Thus, wheat is the most important crop for the food security of the country (Ahmad, 2018). However, other than cereals, fruits, vegetables, and opium poppy are also important crops. The average area under different crops are: wheat−2.2 million hectares; rice−0.13 million hectares; barley−0.19 million hectares; maize−0.145 million hectares; pulses−0.102 million hectares; fruits−0.295 million hectares; vegetables−0.104 million hectares; (Rashid, 1997); opium poppy−0.216 million hectares (Avetisyan, 2017).

Dataset Used

Satellite and Other Data

This study used Sentinel 1 SAR and Sentinel 2 multispectral optical satellite images as the main data sources (**Table 2**). For the wheat area classification, bands B4 (Red), B8 (NIR), and B11 (SWIR) from the S2 were used; whereas the VV (Vertically transmit Vertically receive) Polarization band was utilized from the Sentinel 1 data. High-resolution images from Digitalglobe and Airbus Company (Pleiades) were also used for collecting reference data for training and validation.

Apart from satellite datasets, agriculture mask (irrigated and rainfed) from the Afghanistan 2010 land cover (FAO, 2010) was used to aid in crop mapping. The land cover maps (**Figure 1**) have eleven land cover classes viz. irrigated agricultural land, rainfed agricultural land, fruit trees, vineyards, barren land, sand cover, forests and shrubs, rangeland, permanent snow, built up, and water bodies and marshland (FAO, 2010).

Reference Data

Reference data were collected from various sources for training and validation of the classification model. The field survey was conducted by professionals from MAIL to collect samples from the crop field. A random sampling approach was utilized for the

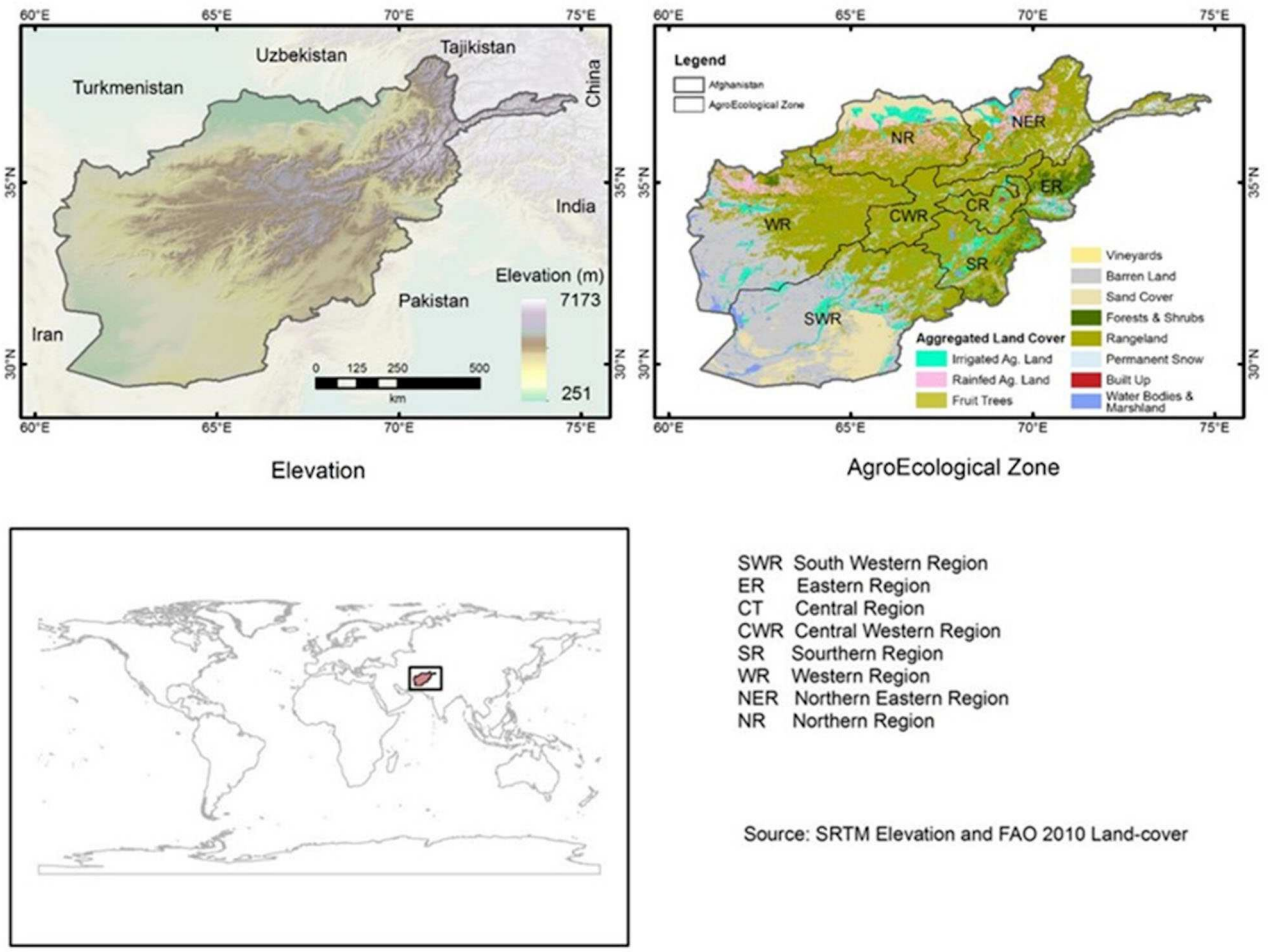


FIGURE 1 | Study area.

collection of the field data/sample points. During the collection of samples, the location of the crop field together with field multi-direction photographs for different crops were collected. The second set of reference data were collected by MAIL during a crop cutting survey that covered 17 provinces. Most of the reference data collected through field and crop cut survey were mainly from wheat fields. Very few samples were collected from non-wheat crops, such as vegetable farms, orchards, and vineyards. Samples for other crops were generated through visual interpretation by relevant experts and using earlier land cover maps and time-series images.

The reference data obtained through field and crop cutting survey covered only 25 out of 34 provinces. For the remaining 10 provinces, samples were generated from high-resolution images through visual interpretation and analysis of NDVI time-series for the current and previous years. Google Earth images and Digital Globe high-resolution images acquired during October 2016–June 2017 were used for the interpretation. Overall 16,383 reference? points from wheat (4,797) and non-wheat (11,586) class were collected. Out of which 70%, i.e., 11,468 samples were used for training and 30%, i.e., 4,915 samples were utilized for validation (discussed in the validation section).

For Helmand province, a set of reference data were received from the United Nations Office of Drug Control (UNODC) for opium poppy fields. Additional reference data were collected through visual interpretation of Airbus Pleiades images. The distribution of reference data and sources are shown in **Figure 2**.

Crop Phenology and Crop Calendar

The goal of the classification algorithm was to distinguish the phenology of wheat from other crop types and land cover. Land surface phenology (LSP) refers to the timing of different life-cycle stages of plants (Martínez and Gilabert, 2009). The study of LSP is important to understand vegetation-growth pattern changes (Myneni et al., 1997; Fisher and Mustard, 2007). Satellite-based analysis of LSP addresses the development patterns in photosynthetic biomass by way of derived vegetation indices (Ahl et al., 2006), such as the normalized difference vegetation index (NDVI), the enhanced vegetation index (EVI), and a two-band enhanced vegetation index (White et al., 1997; Zhang et al., 2003, 2014; Piao et al., 2006). Phenology is measured commonly by (i)

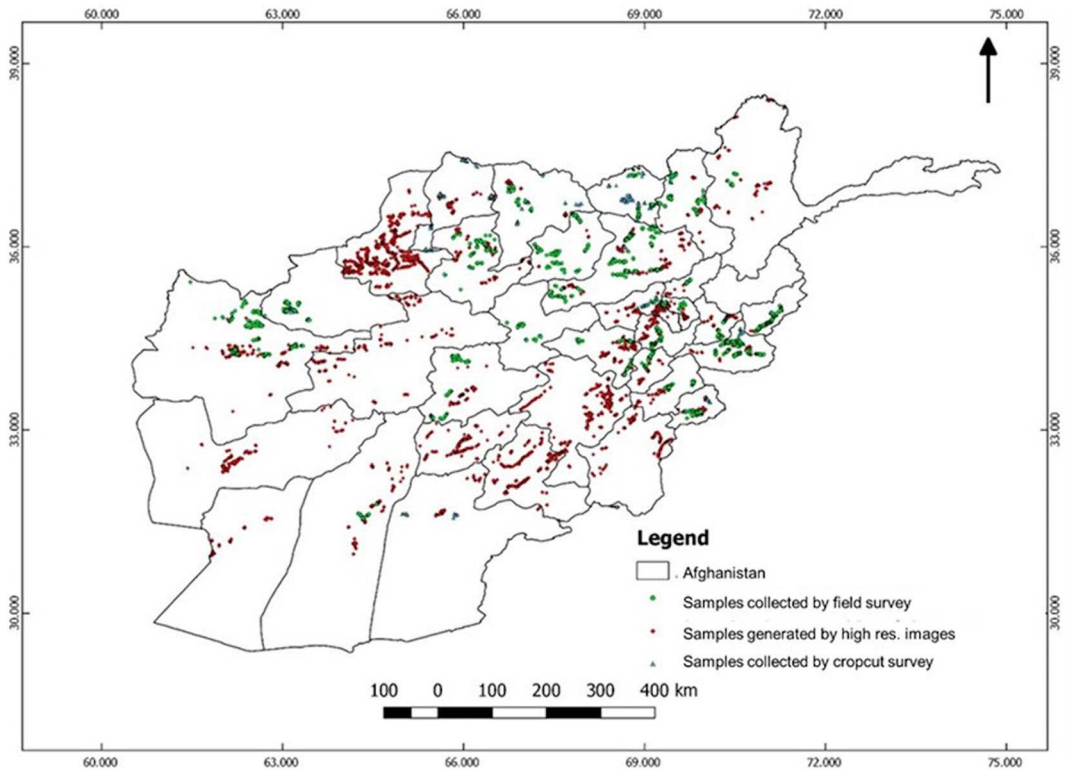


FIGURE 2 | Distribution of reference data.

TABLE 2 | Satellite data specification.

Characteristics	Sentinel 1	Sentinel 2	Airbus Pléiades
Acquisition date	Oct. 2016 to July 2017	Oct. 2016 to July 2017	1, 4, and 27 April 2017
Bands	VV	Red (B4), NIR (B8), SWIR (B12)	Blue Green Red Near-infrared
Wavelength range	5.5 cm	443–2,190 nm	430–950 nm
Spatial resolution (m)	10 m	10, 20, 60 m	0.5 m
Swath	250 km	290 km	20 km
Temporal resolution	12 days	5.5 days	Daily (constellation)

onset of greening, (ii) onset of senescence, (iii) peak development during the growing period, and (iv) the length of the growing season (Hudson and Keatley, 2010). Various methods have been used for the assessment of phenology including threshold, derivative, smoothing, and model-based methods (Hudson and Keatley, 2010). Among these, the threshold-based method is the simplest and is used by many researchers. In the threshold-based method, the values of VI are plotted against time of year and single values are chosen to define different stages of phenology (Karlsen et al., 2006) though the method for specifying the threshold varies. Some authors use single arbitrary thresholds, e.g., 0.17 (Fischer, 1994), 0.09 (Markon et al., 1995), and 0.099 (Lloyd, 1990), whereas some authors use threshold specifiers like the long-term average (Karlsen et al., 2006) or % peak amplitude of VI (Jonsson and Eklundh, 2002). In this study we have used NDVI for training samples to determine the thresholds (see section Wheat Area Mapping Using Optical Images).

Afghanistan has diverse topographic and climatic conditions resulting in wide variability in growing seasons across the entire landscape. Knowledge of the growing season is important for the acquisition of satellite data. The crop calendar is a tool that provides information on the sowing, growing and harvesting stages of crops (in our case, wheat). The crop calendar information can also be used for crop type mapping using the satellite data. Broad crop calendars at a province-level were provided by MAIL; these were compiled in 2012. Because of the variability in climate/weather and other factors, there can be a shift in the timing of sowing and harvest of wheat over the years. The calendars (**Figure 3**) were utilized as a starting point to characterize the timing of phenological stages of wheat.

Study Methodology

In this study, optical and SAR data were utilized in two steps in the process of mapping wheat areas. The flowchart of the detailed methodology is shown in **Figure 4**. The description of the methodology is given in the following sections.

Reference Data Preparation

In the first step, reference data collected from wheat and other crops for 34 provinces of Afghanistan were subjected

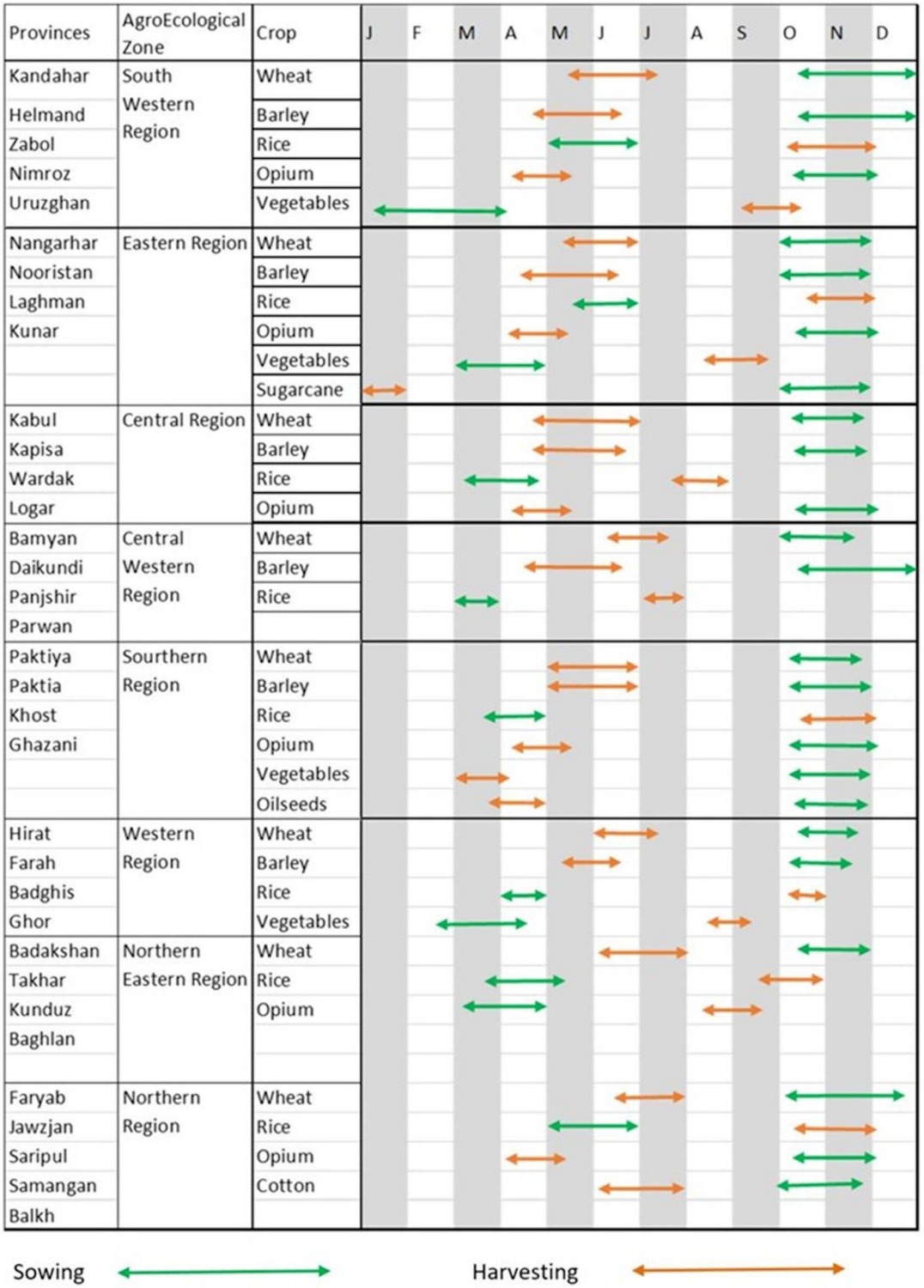


FIGURE 3 | Cropping Calendar by Agro-Ecological Zones for major crops.

to quality check. This is because some of the sample points collected by the field staff were not always inside the crop fields. Accordingly, adjustments were made to correct the location based on three criteria: (a) direction and orientation of the field photographs; (b) phenological characteristics of the crop; and (c) visual interpretation through high-resolution Google Earth

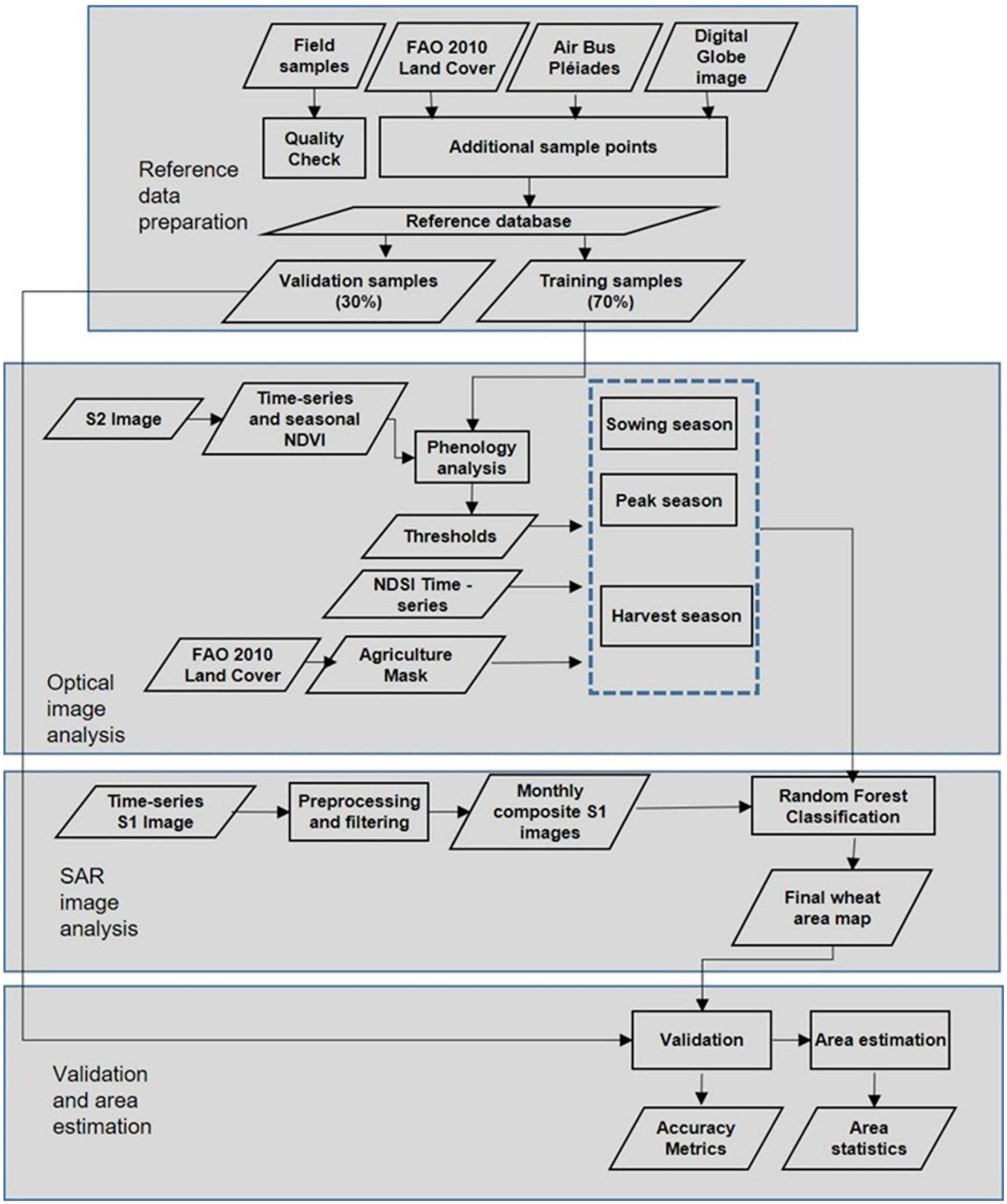


FIGURE 4 | Flow chart for wheat area mapping and estimation.

images. For each province, the reference points were merged and divided randomly into two categories, i.e., training and validation samples; 70% of the samples were used for training and the remaining 30% for validation.

Wheat Area Mapping Using Optical Images

The wheat mapping was done at the provincial level. Optical Sentinel-2A Level 1-C (top-of-atmosphere) satellite images with <30% cloud cover from November 2016–July 2017 were

used for the analysis. The data were preprocessed to remove clouds/bad pixels. The cloud masking utilizes Sentinel-2 Band QA60, a quality flag band, to identify and mask out flagged cloud and cirrus pixels. After that, the median-compositing function was used on the cloud-masked Sentinel-2 images to generate a per-pixel median composite of each of the multi-spectral bands and the indices for every province (Hird et al., 2017).

NDVI is an effective means to characterize these growth patterns during the crop cycle (Menenti et al., 1993). Using randomly collected training samples and the seasonal composite of sentinel 2 images, NDVI thresholds were identified to separate the wheat from other crops during sowing, peak and harvest time at the provincial level. NDVI thresholds were identified for different seasons (sowing, peak and harvest) and were different for each province. The difference in the NDVI thresholds for different provinces is mainly because of the shift in the phenological cycle (early and late sowing) of wheat and other crops.

The Normalized Difference Soil Index (NDSI; Equation 2) values were used as an additional metric to separate the wheat from fallow land during peak season. The NDSI is preferred because it is sensitive to canopy structure (Jin et al., 2016) and is very effective in separating bare soil from other features like water and sparse vegetation.

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

$$NDSI = \frac{SWIR - NIR}{SWIR + NIR} \tag{2}$$

To define the thresholds for separating the wheat from other crops, the minimum and maximum values of NDVI were calculated for wheat using the training samples. The separation was done as below:

$$\text{Minimum of NDVI}_{\text{wheat samples}} < \text{Wheat}_{\text{sowing}} \leq \text{Maximum of NDVI}_{\text{wheat samples}} \tag{3}$$

$$\text{Wheat}_{\text{peak}} \geq (\text{Minimum of NDVI}_{\text{wheat samples}} \text{ and NDSI} < 0) \tag{4}$$

$$\text{Minimum of NDVI}_{\text{wheat samples}} < \text{Wheat}_{\text{harvest}} \leq \text{Maximum of NDVI}_{\text{wheatsamples}} \tag{5}$$

The NDVI threshold derived for 2016–2017 were specifically derived on the basis of collected ground sample points from the field. The NDVI threshold values depend on various factors, such as: (i) whether it is a dry year or wet year; (ii) whether there has been early or late sowing of the crop; and (iii) atmospheric conditions at the time of data acquisition. Hence, the NDVI thresholds are specific to the 2016–2017 growing season. However, Equations (3)–(5) can be utilized for deriving the NDVI thresholds for other years if field sample points for that particular year are available.

Refinement of the Wheat Map Using SAR Data

Compared to the other crops, wheat has a different cropping calendar and growth patterns (**Figure 3**), except for barley, some vegetables and opium poppy have growth patterns similar to wheat. Due to the similarity in cropping season and growth pattern, it was difficult to accurately separate barley, opium poppy and some vegetables using Sentinel 2 data despite having a 5-days temporal resolution. This is because of the limited availability of cloud-free images which makes it difficult to utilize the image of a specific time (where wheat can be separated from other overlapping crops). To overcome this limitation, the wheat area map obtained from the optical image analysis was further refined using Sentinel-1 SAR data. SAR sensors have all-weather capability to acquire images and are sensitive to plant structure; however, to use the SAR (S1) based classification alone would require much more sample data for all the crops. The initial separation of crops using optical data enabled the use of SAR for only separating the wheat from crops with similar crop calendar and phenology. The S1 data has a consistent time-series in terms of incidence angle and has a wide scope in mapping different crops (Inglada et al., 2015). Initially, SAR (S1) datasets were preprocessed with VV polarization imagery. Pre-processing includes orbital file correction, thermal noise removal and terrain correction. For removing speckle noise, the median filter with kernel window size 5 × 5 was used for this research. The selection is based on previous studies that demonstrate that the median filter with window size 5 × 5 produces consistent and satisfactory results than other speckle filters and window sizes (e.g., 3 × 3 or 7 × 7) (Ozdarici and Akyurek, 2010). Monthly median composites were developed for the entire wheat crop cycle (i.e., sowing until harvesting).

The analysis of S1 data shows the difference in response patterns from different crops. However, the variability of responses shows overlap and makes it difficult for the threshold-based separation (**Figure 5**). Thus, a Random Forest (RF) classification technique was utilized using time series S1 data and training points to separate the wheat from other crops. The Random Forest (RF) randomly selects a subset of training sample through replacement to build a single tree, i.e., it uses bagging technique where for every tree, data is sampled from the original complete training set. There are two important user- define parameters in RF, i.e., (i) Number of trees; (ii) Number of variables. The generalization error always converges as the number of trees increases (Breiman, 2001). Therefore, RF classifier doesn't have any issue of overfitting which can also be attributed to the Strong Law of Large Numbers (Bercovici and Pata, 1996). There is no well-defined rule for selection of the number of trees. However, Guan et al. (2013) suggest that number of trees can be as large as possible but beyond a certain point, additional trees will not help in improving the performance of the classifier. Also, the increased number of trees would require high computation. In this study, we have used the number of trees as 100. The selection was based on the hit and trial method. Secondly, the number of variables highly affects the performance of RF classifier, which is usually set to the square root of the number of input variables. In our case, we have used time series of the monthly median of VV polarized sentinel−1 backscatter values. The application of Random Forest was applied within the classified mask generated

FIGURE 5 | Seasonal phenological characteristics using NDVI. **(A)** Laghman, **(B)** Helmand province.

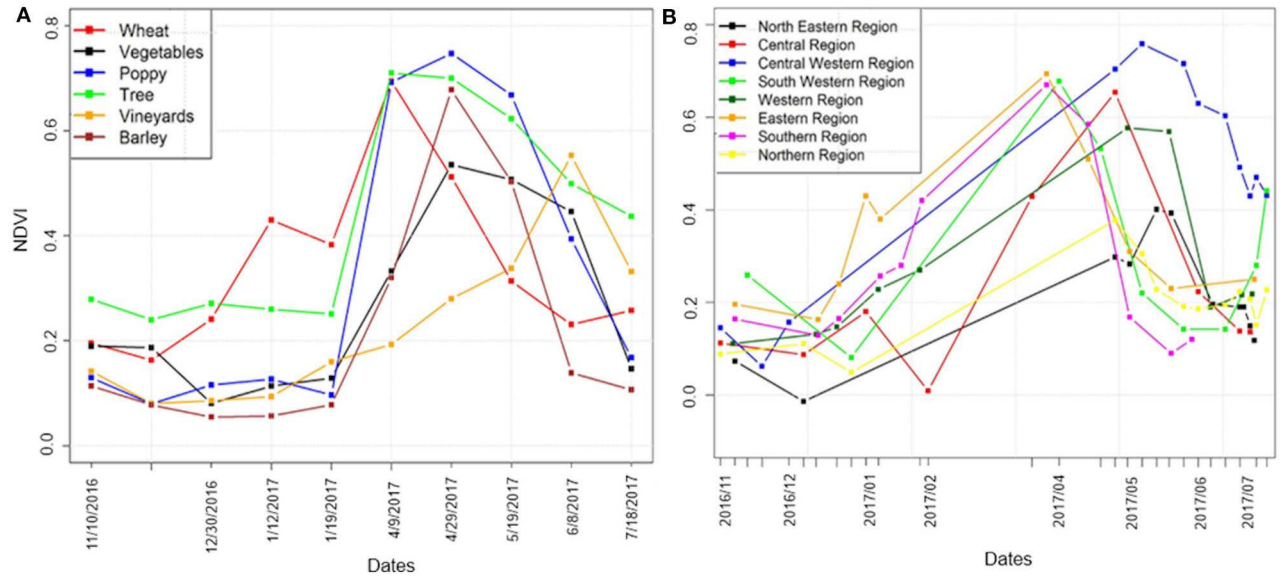


FIGURE 6 | (A) Phenological characteristics of different crops (Laghman Province); **(B)** Phenological characteristics of wheat in different AEZ.

from the optical image analysis. This step was applied only after the harvest season.

Accuracy Assessment

In the context of remote sensing based land cover classification, accuracy assessment can be defined as an agreement between a standard assumed to be correct and a classified image of unknown quality (Grenier et al., 2008). Classification errors occur when a pixel (or feature) belonging to one category is assigned to another category. Accuracy assessments can be done using qualitative methods through visual interpretation and quantitative evaluation based on statistical methods (Cochran, 1997; Olofsson et al., 2014). The accuracy assessment for this study was conducted in two steps. First, the results were checked by comparing with various ancillary data to identify gross errors. Second, the final data were used for quantitative accuracy assessment. Of the total reference samples (ground sample points) collected for different crops, 70% were used for training and the remaining 30% for validation. Error matrices were generated for each province separately. Using these error matrices, the statistical accuracy assessment was done by generating producer's and user's accuracy including Kappa coefficients. Kappa coefficient is an indicator of accuracy of the classified map. It is a measure of how the classification results compare to values assigned by chance. The value of the kappa coefficient ranges from 0 to 1. If kappa coefficient equals to 0, there is no agreement between the classified image and the reference image. If kappa coefficient equals to 1, then the

classified image and the ground truth image are identical. Higher the value of kappa coefficient, more accurate the classified map is.

Tools Used

The workflow for the wheat area mapping is implemented in the Google Earth Engine (GEE) environment using custom Java scripts. The reference data and other required data were loaded into GEE asset storage. The preprocessed images were also loaded into GEE asset to run the classification. The resulting wheat maps were exported as Geocoded rasters and imported to ArcGIS software for visual interpretation and accuracy analysis.

RESULTS AND DISCUSSION

Understanding the Phenological and Temporal Backscatter Characteristics of Wheat and Other Vegetation

Phenological Characteristics Using Optical Data

Cropping calendar information aggregated by province (**Figure 3**) suggests that the sowing season of wheat overlaps with barley, opium poppy, and vegetables. However, some differences can be seen in the length of the season including the start and end of the season, and peak. While analyzing the data, we consulted province-wise crop calendars. However, due to cloud cover, it was not possible to create monthly composites of Sentinel 2 images covering all the provinces to generate monthly phenology. Therefore, seasonal median composites of NDVI were created during the sowing, peak and harvest seasons for wheat. As an example, crop growth patterns for Laghman province are shown in **Figures 5A**, **6A** which depict vineyards having distinct signals compared to wheat areas during the sowing season. The growth pattern is also significantly different. The orchards had higher NDVI both during the peak and harvest time. NDVI response from vegetables varied a lot but the values were lower than the wheat during the peak and the harvest season. The NDVI values for opium poppy showed higher overlap during the sowing period. It has relatively higher NDVI values during the peak and the harvest time. The opium poppy has a shorter cropping season so separation with barley and opium poppy would have been possible if cloud-free monthly images could be obtained. The NDVI characteristics of the opium poppy in Helmand province (**Figure 5B**) showed higher separability from the wheat during the sowing and peak season. Overall, the NDVI seasonal composites were useful to distinguish the wheat from orchards, vineyards and some vegetables. Not much separation between these crops could be achieved using the sowing period data alone; much more improved results can be obtained by integrating sowing, peak and harvest season data. However, significant overlap in NDVI was still observed between the wheat, opium poppy, and barley using the optical image composites.

Figure 6B shows the difference in the phenological characteristics of wheat for different Agro-Ecological Zones (AEZ) in Afghanistan. By examining **Figure 6B**, it was observed that there is not much difference in the phenological characteristics of wheat in the Eastern, Southern and the South Western Region. The Western and the Central Western region have late peak and harvest times. In the Northern and North Eastern Region, the NDVI values at peak season were generally low compared to other regions and also the harvest time of the wheat was late. The shift in the sowing, peak and harvest time of wheat is due to the different altitude, climatic zones and that varies according to the agro-ecological zones. The difference in the growth pattern of the wheat in different AEZs also suggests the use of different NDVI thresholds for sowing, peak and harvest season for different provinces.

Analysis of Temporal Backscatter Characteristics of S1 (SAR Data)

SAR has all-weather capability and good temporal resolution. Therefore, backscatter characteristics of wheat and other overlapping crops/vegetation features were also examined using Sentinel 1 SAR data. The monthly temporal median composite images from November 2016–July 2017 were

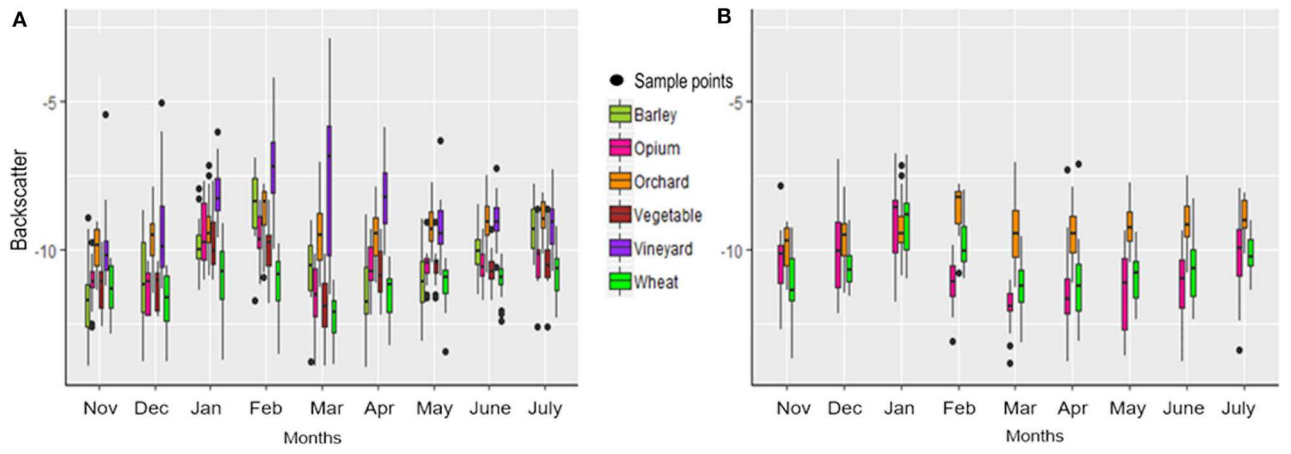


FIGURE 7 | Monthly SAR Backscatter profile by crop. **(A)** Laghman, **(B)** Helmand.

utilized to study the response of backscatter characteristics of overlapping crops at different periods (**Figure 7**). Since SAR backscatter varies by plant structure, distinct signals were observed for different crops. However, due to high overlapping responses from different crops (**Figure 7**), the threshold-based separation was not possible using SAR backscatter datasets. Random Forest classification was used on the monthly composites of backscatter data to constrain the classification within the mask (threshold from optical S2 data) obtained by phenological analysis from S2 images in the earlier stage.

TABLE 3 | Confusion matrix.

Class	Irrigated wheat			
	Non-wheat	Irrigated wheat	Total	User's accuracy (%)
Non-wheat	1,839	282	2,121	86
Irrigated-wheat	341	1,388	1,729	80
Total	2,180	1,670	3,850	
Producer's accuracy (%)	84	83		
	Overall	83.8 (%)		
	AC	0.50		
	Kappa	0.67		

Class	Rainfed wheat			
	Non-wheat	Rainfed wheat	Total	User's accuracy (%)
Non-wheat	710	59	769	92
Rainfed wheat	58	238	296	80
Total	768	297	1,065	
Producer's accuracy (%)	92	80		
	Overall accuracy	89 (%)		
	AC	0.59		
	Kappa	0.77		

The Accuracy of the Wheat Area Estimation

Classification error matrices were generated for each province for wheat and non-wheat areas using validation samples. **Table 3** shows the confusion matrix for Irrigated and Rainfed wheat.

The statistics of accuracy (overall, users, producers, and Kappa) achieved at a provincial level for irrigated and rainfed wheat is depicted in **Figure 8**. The mean overall accuracy for all provinces for irrigated and rainfed wheat areas was 83.8 and 89.0%, respectively. The minimum overall accuracy was 48% for irrigated wheat in the Faryab province. This was exceptional because of three reasons (i) cloud cover, which hampers the quality of the images over that province; (ii) the limited number of sample points, which makes it difficult to identify the thresholds; and (iii) poor quality of ground sample points. The overall accuracy was <75% for irrigated wheat for only six provinces out of 34. For rainfed wheat, only one province had <75% accuracy. The accuracy for the provinces without the reference data was generally lower than those where reference samples were available. The Kappa value forirrigated wheat for provinces with available field data was 0.69 whereas it was 0.54 for provinces with no field data. After evaluating the accuracy of each province, the wheat area was estimatedusing the

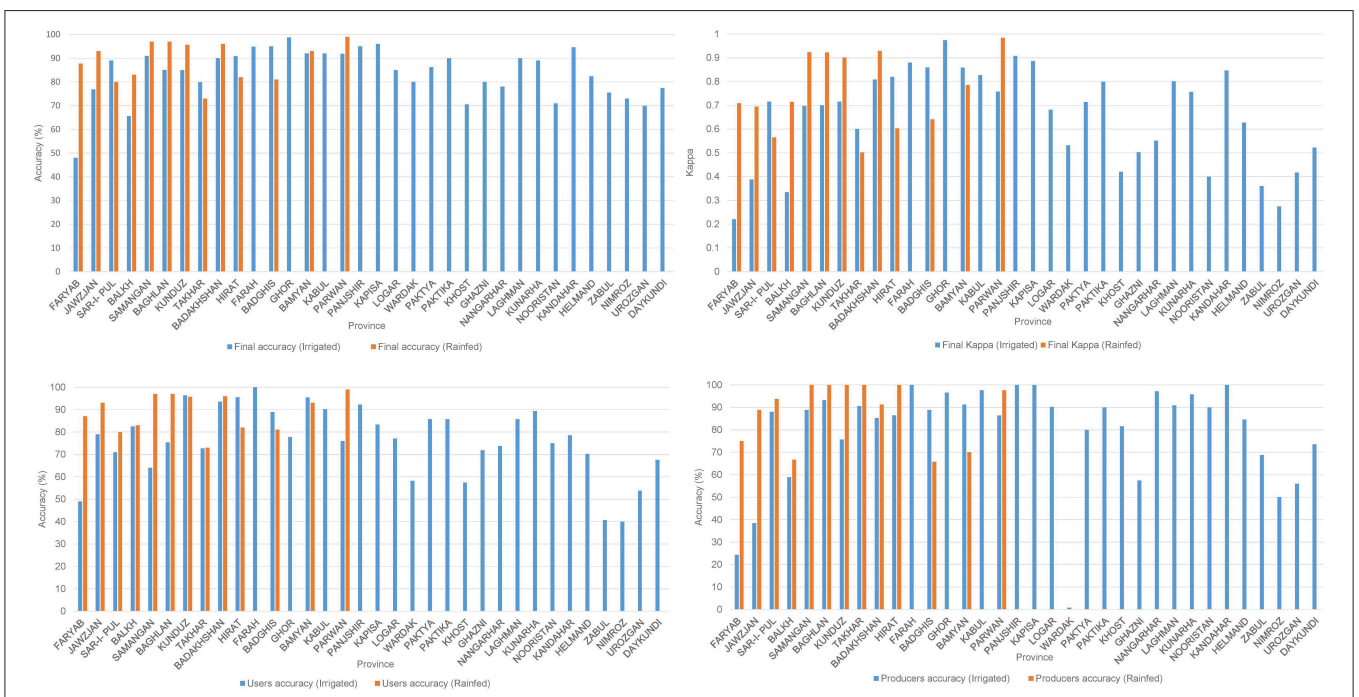


FIGURE 8 | Accuracy analysis results by provinces.

TABLE 4 | Area and accuracy using different combinations of datasets (example Kabul province).

Case	Sensor used		Temporal			Overall accuracy (%)	Area (ha)
	S1	S2	Sowing	Peak	Harvest		
Case 1	–	Yes	Yes	Yes	–	72	31,783
Case 2	Yes	Yes	Yes	Yes	–	86	17,466
Case 3	–	Yes	Yes	Yes	Yes	89	14,780
Case 4	Yes	Yes	Yes	Yes	Yes	93	13935.9

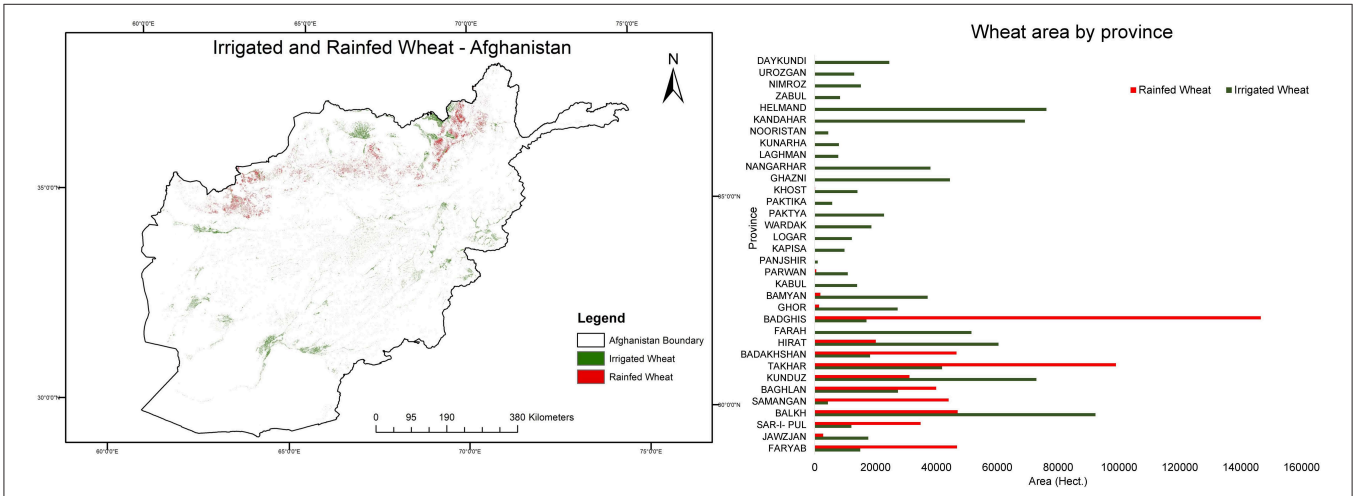


FIGURE 9 | Distribution of Irrigated and Rainfed Wheat area of Afghanistan in 2017.

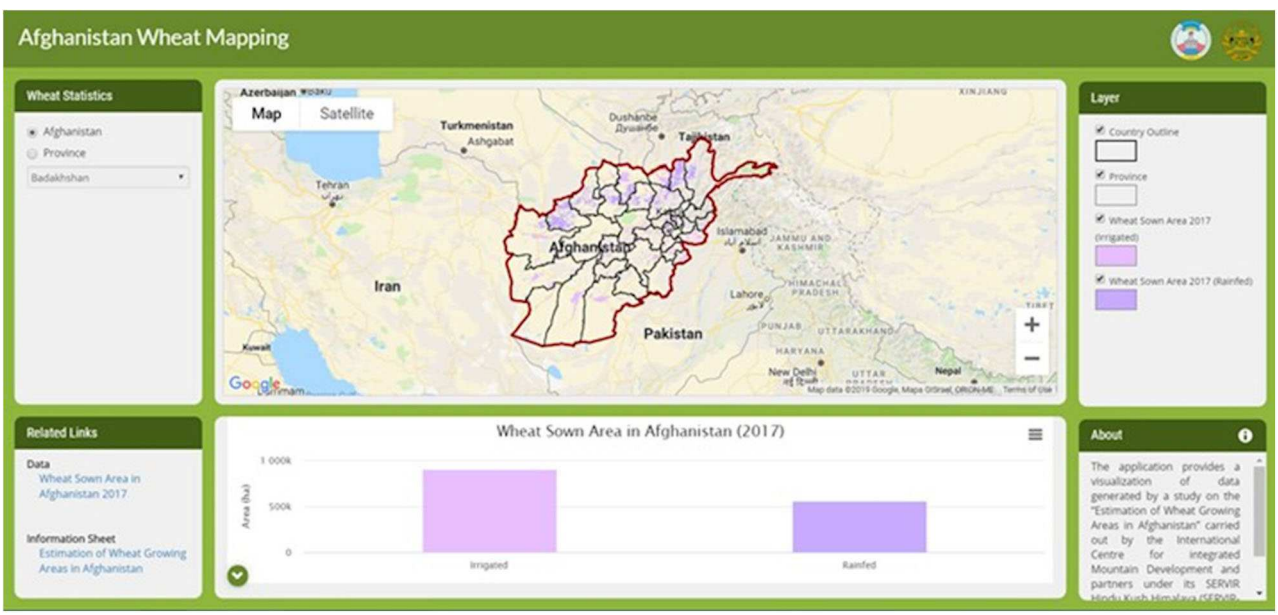


FIGURE 10 | Web mapping system for dissemination and visualization of wheat map.

equation (Equation 6).

$$\text{Wheat area } (ha) = \frac{(\text{Pixel count}) * (\text{resolution of the image}) * (\text{resolution of the image})}{10000} \quad (6)$$

In the study, we also explored the potential accuracy for using S1 and S2 datasets and accuracy achieved for the estimation at different crop growth stages, i.e., sowing, peak, and harvest for Kabul as a case study (**Table 4**). The accuracies and the area estimates for different combinations and periods are presented in **Table 4**. From **Table 4**, it can be observed that the accuracy of the classification is highest when the assessment is done at the end of the season, i.e., utilizing both optical and SAR datasets. However, a decent accuracy can be achieved during peak season using both optical and SAR images. The peak season area estimates can be highly useful for food security management and to address major deviations in the wheat area cultivated during the specific season.

Distribution of Wheat in Afghanistan

Figure 9 shows the distribution of irrigated and rainfed wheat areas for 2017 produced using both optical (S2) and SAR (S1) data and images from sowing until harvest. By observing the map (**Figures 3**, **9**), it can be concluded that the majority of the rainfed areas were located in the northern region of the country, which is mountainous (Hindu Kush range) and experiences a decent amount of precipitation and snowfall during the winter season, whereas irrigated the area is distributed across the entire country. More than 80% of wheat (both irrigated and rainfed) is produced in sixteen provinces of which Badghis, Takhar, Balkh, Kunduz, Herat, Helmand, and Kandahar are major wheat-producing provinces accounting for nearly 50% of the wheat production in the country.

Operationalization and Dissemination

The wheat mapping workflow is implemented in GEE using a customized interface for each module. A systematic capacity building program, including formal and on the job training, was conducted for MAIL professionals to run the workflow. A web-based visualization system was developed (**Figure 10**) to disseminate the final results. The portal can be accessed via the following URL link: http://geoapps.icimod.org/afwheat/.

LIMITATIONS

While the study demonstrated a method for in-season classification of wheat area for food security planning in Afghanistan, there are few limitations of the study. Firstly, the classification system is implemented in the free cloud platform of Google earth Engine assuming that GEE will continue ingesting the Sentinel−2 and Sentinel−1 data in its image collection. If for any reason, GEE discontinues providing access to the GEE cloud platform, the system could not be automated and all the Sentinel data would have to be downloaded and processed offline. Secondly, the threshold-based separation of wheat from other crops has a limitation in the area where wheat is mixed with other crops especially vegetables and opium poppy which have similar phenology. Though SAR images were used to separate these crops from wheat, it requires field samples for each year which is quite challenging in the context of Afghanistan. Alternative source like high-resolution satellite images could be used for the collection of samples from those areas. Third, currently, Sentinel-1 SAR images only capture VV polarized data over Afghanistan. Multiple polarized SAR data could be a better option for classifying wheat using the RF algorithm. Fouth, for operationalization of the system, capacity development has been done for Geospatial staff in MAIL but, many of the MAIL staff have left during the study period. Though most of the system is automated and the system could be implemented with quick training, MAIL has to ensure the availability of staff and knowledge transfer during staff turnover.

CONCLUSION AND WAY FORWARD

In this study, a systematic methodology for wheat area mapping was developed for Afghanistan with the potential for operationalization to support the management of food security in the country. To overcome the issues related to low internet bandwidth, lack of sufficient ground samples and limited availability of cloud-free optical satellite images, a cloud-based system combining phenological characteristics using optical images and temporal backscatter profiles using SAR images was adopted. The system uses a multi-step approach to provide area estimation as the wheat season progresses. The first estimation is provided during the peak season to give an early indication of wheat cultivated area. The more accurate estimation is provided immediately after the harvest season. Considering the low capacity on the use of remote sensing based crop type mapping in Afghanistan, this approach was automated in GEE. Through NASA SERVIR, training activities are underway to enhance the skills of the local staff in government agencies on mapping and monitoring of crop areas using GEE.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

MM: concept development, analysis, and writing. VT: concept development, analysis, implementation, and writing. WE, FQ, BB, and KV: review methodology, review analysis, and writing. BR and WY: data analysis and writing.

FUNDING

This study was partially supported by the SERVIR-HKH program. SERVIR-HKH was a joint development initiative of the National Aeronautics and Space Administration (NASA) and the United States Agency for International Development (USAID). SERVIR works in partnership with leading regional organizations worldwide to help developing countries use the information provided by Earth-observing satellites and geospatial technologies for managing climate risks and land use.

The study was also partially supported by core funds of ICIMOD contributed by the governments of Afghanistan, Australia, Austria, Bangladesh, Bhutan, China, India, and Myanmar, Nepal, Norway, Pakistan, Sweden, and Switzerland.

ACKNOWLEDGMENTS

The authors express their sincere gratitude to the Airbus company for providing the high-resolution Pleiades images at a discounted price.

REFERENCES

Ahl, D. E., Gower, S. T., Burrows, S. N., Shabanov, N. V., Myneni, R. B., and Knyazikhin, Y. (2006). Monitoring spring canopy phenology of a deciduous broadleaf forest using MODIS. *Remote Sens. Environ.* 104, 88–95. doi: 10.1016/j.rse.2006.05.003

Ahmad, J. (2018). *International Conference on Food and Agricultural Economics, 2nd Edn.* Available online at: http://www.ageconalanya.com/ICFAEC%202018%20PROCEEDINGS%20BOOK.pdf

Avetisyan, A. (2017). *Afghanistan Opium Survey 2017 Cultivation and Production*. United Nation Office of Drugs and Crime. Available onlie at: https://www.unodc.org/documents/crop-monitoring/Afghanistan/Afghan_opium_survey_2017_cult_prod_web.pdf

Baker, W. (2015). *Wheat Production Above Average but Down From Last Year*. United States Department of Agriculture. Available online at: https://ipad.fas.usda.gov/highlights/2015/07/Afghanistan/Index.htm

Bercovici, H., and Pata, V. (1996). The law of large numbers for free identically distributed random variables. *Ann. Probab.* 24, 453–465. doi: 10.1214/aop/1042644726

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Camps-Valls, G., Gómez-Chova, L., Calpe-Maravilla, J., Soria-Olivas, E., Martín-Guerrero, J. D., and Moreno, J. (2003). "Support vector machines for crop classification using hyperspectral data," in *Pattern Recognition and Image Analysis*, Vol. 2652, eds F. J. Perales, A. J. C. Campilho, N. P. de la Blanca, and A. Sanfeliu (Berlin; Heidelberg: Springer), 134–141. doi: 10.1007/978-3-540-44871-6_16

CIA (2019). *South Asia: Afghanistan—The World Factbook—Central Intelligence Agency*. The World Fact Book. Available online at: https://www.cia.gov/library/publications/the-world-factbook/geos/af.html

Cochran, W. G. (1997). *Sampling Techniques*, Vol. 3. John Wiley and Sons. Available online at: https://hwbdocuments.env.nm.gov/Los%20Alamos%20National%20Labs/General/14447.pdf

Deschamps, B., McNairn, H., Shang, J., and Jiao, X. (2012). Towards operational radar-only crop type classification: comparison of a traditional decision tree with a random forest classifier. *Can. J. Remote Sens.* 38, 60–68. doi: 10.5589/m12-012

Dong, J., Xiao, X., Menarguez, M. A., Zhang, G., Qin, Y., Thau, D., et al. (2016). Mapping paddy rice planting area in northeastern Asia with landsat 8 images, phenology-based algorithm and google earth engine. *Remote Sens. Environ.* 185, 142–154. doi: 10.1016/j.rse.2016.02.016

FAO (2010). *The Islamic Republic of Afghanistan land cover* atlas. FAO.

Fischer, A. (1994). A model for the seasonal variations of vegetation indices in coarse resolution data and its inversion to extract crop parameters. *Rem. Sens. Environ.* 48, 220–230. doi: 10.1016/0034-4257(94)90143-0

Fisher, J. I., and Mustard, J. F. (2007). Cross-scalar satellite phenology from ground, Landsat, and MODIS data. *Rem. Sens. Environ.* 109, 261–273. doi: 10.1016/j.rse.2007.01.004

Foerster, S., Kaden, K., Foerster, M., and Itzerott, S. (2012). Crop type mapping using spectral–temporal profiles and phenological information. *Comput. Electr. Agric.* 89, 30–40. doi: 10.1016/j.compag.2012.07.015

Gao, Q., Zribi, M., Escorihuela, M., Baghdadi, N., and Segui, P. (2018). Irrigation mapping using sentinel-1 time series at field scale. *Rem. Sens.* 10:1495. doi: 10.3390/rs10091495

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Rem. Sens. Environ.* 202, 18–27. doi: 10.1016/j.rse.2017.06.031

Grenier, M., Labrecque, S., Benoit, M., and Allard, M. (2008). *Accuracy Assessment Method for Wetland Object-Based Classification*. Available online at: https://www.isprs.org/proceedings/xxxviii/4-c1/Sessions/Session1/6739_Grenier_Proc_pos.pdf

Guan, H., Li, J., Chapman, M., Deng, F., Ji, Z., and Yang, X. (2013). Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. *Int. J. Remote Sens.* 34, 5166–5186. doi: 10.1080/01431161.2013.788261

Haworth, B., Biggs, E., Duncan, J., Wales, N., Boruff, B., and Bruce, E. (2018). Geographic information and communication technologies for supporting smallholder agriculture and climate resilience. *Climate* 6:97. doi: 10.3390/cli6040097

Hird, J., DeLancey, E., McDermid, G., and Kariyeva, J. (2017). Google earth engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. *Remote Sens.* 9:1315. doi: 10.3390/rs9121315

Hudson, I. L., and Keatley, M. R. (Eds.). (2010). *Phenological Research: Methods for Environmental and Climate Change Analysis*. Dordrecht; New York, NY: Springer.

Immitzer, M., Vuolo, F., and Atzberger, C. (2016). First experience with sentinel-2 Data for crop and tree species classifications in central europe. *Remote Sens.* 8:166. doi: 10.3390/rs8030166

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., et al. (2015). Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sens.* 7, 12356–12379. doi: 10.3390/rs70912356

Jakubauskas, M. E., Legates, D. R., and Kastens, J. H. (2002). Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Comput. Elect. Agric.* 37, 127–139. doi: 10.1016/S0168-1699(02)00116-3

Ji, S., Zhang, C., Xu, A., Shi, Y., and Duan, Y. (2018). 3D Convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sens.* 10:75. doi: 10.3390/rs10010075

Jin, Y., Sung, S., Lee, D., Biging, G., and Jeong, S. (2016). Mapping deforestation in north Korea using phenology-based multi-index and random forest. *Remote Sens.* 8:997. doi: 10.3390/rs8120997

Jonsson, P., and Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Trans. Geosci. Remote Sens.* 40, 1824–1832. doi: 10.1109/TGRS.2002.802519

Karlsen, S. R., Elvebakk, A., Høgda, K. A., and Johansen, B. (2006). Satellite-based mapping of the growing season and bioclimatic zones in Fennoscandia. *Glob. Ecol. Biogeogr.* 15, 416–430. doi: 10.1111/j.1466-822X.2006.00234.x

Latham, S. J. (2017). *Northern and Eastern Afghanistan: Monitoring of Rice Cultivation Using Satellite Remote Sensing and GIS Technologies*. Food and Agriculture Organization United Nation. Available online at: http://www.fao.org/3/a-i6986e.pdf

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: a review. *Sensors.* 18:2674. doi: 10.3390/s18082674

Lloyd, D. (1990). A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery. *Int. J. Remote Sens.* 11, 2269–2279. doi: 10.1080/01431169008955174

Markon, C. J., Fleming, M. D., and Binnian, E. F. (1995). Characteristics of vegetation phenology over the alaskan landscape using AVHRR time-series data. *Polar Rec.* 31, 179–190. doi: 10.1017/S0032247400013681

Martínez, B., and Gilabert, M. A. (2009). Vegetation dynamics from NDVI time series analysis using the wavelet transform. *Remote Sens. Environ.* 113, 1823–1842. doi: 10.1016/j.rse.2009.04.016

Menenti, M., Azzali, S., Verhoef, W., and van Swol, R. (1993). Mapping agroecological zones and time lag in vegetation growth by means of fourier analysis of time series of NDVI images. *Adv. Space Res.* 13, 233–237. doi: 10.1016/0273-1177(93)90550-U

Muradi, A. J., and Boz, I. (2018). The contribution of agriculture sector in the economy of Afghanistan. *Int. J. Sci. Res. Manage.* 6, EM-2018-750-755. doi: 10.18535/ijsrm/v6i10.em04

Murmu, S., and Biswas, S. (2015). Application of fuzzy logic and neural network in crop classification: a review. *Aquat. Proc.* 4, 1203–1210. doi: 10.1016/j.aqpro.2015.02.153

Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., and Nemani, R. R. (1997). Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* 386, 698–702. doi: 10.1038/386698a0

Oguro, Y., Suga, Y., Takeuchi, S., Ogawa, M., Konishi, T., and Tsuchiya, K. (2001). Comparison of SAR and optical sensor data for monitoring of rice plant around hiroshima. *Adv. Space Res.* 28, 195–200. doi: 10.1016/S0273-1177(01)00345-3

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* 148, 42–57. doi: 10.1016/j.rse.2014.02.015

Ozdarici, A., and Akyurek, Z. (2010). "A comparison of SAR filtering techniques on agricultural area identification," in *ASPRS 2010 Annual Conference* (San Diego, CA), 10.

Pham, H. T. M., Huynh, T. T. T., Duong, T. T., and Kawarazuka, N. (2017). *Food System Policy Baseline Assessment Report From Vietnam*. Hanoi: International Center for Tropical Agriculture (CIAT).

Piao, S., Fang, J., Zhou, L., Ciais, P., and Zhu, B. (2006). Variations in satellite-derived phenology in China's temperate vegetation. *Glob. Change Biol.* 12, 672–685. doi: 10.1111/j.1365-2486.2006.01123.x

Rashid, A. (1997). *FAO/GIEWS Special Report on Afghanistan 08/97*. Available online at: http://www.fao.org/3/w6059e/w6059e00.htm

Shahriar, P., Budde, M., and Rowland, J. (2014). Mapping irrigated areas in Afghanistan over the past decade using MODIS NDVI. *Remote Sens. Environ.* 149, 155–165. doi: 10.1016/j.rse.2014.04.008

Simms, D. M., and Waine, T. W. (2016). "Opium yield estimates in Afghanistan using remote sensing," in *Seventh International Conference on Agricultural Statistics (ICAS VII)*. Available online at: https://www.istat.it/storage/icas2016/b09-simms.pdf

Skriver, H., Mattia, F., Satalino, G., Balenzano, A., Pauwels, V. R. N., Verhoest, N. E. C., et al. (2011). Crop Classification using short-revisit multitemporal SAR data. *IEEE J. Appl. Earth Observ. Remote Sens.* 4, 423–431. doi: 10.1109/JSTARS.2011.2106198

Sonobe, R., Tani, H., Wang, X., Kobayashi, N., and Shimamura, H. (2014). Random forest classification of crop type using multi-temporal TerraSAR-X dual-polarimetric data. *Remote Sens. Lett.* 5, 157–164. doi: 10.1080/2150704X.2014.889863

Soria-Ruiz, J., Fernandez-Ordonez, Y., McNairm, H., and Bugden-Storie, J. (2007). "Corn monitoring and crop yield using optical and RADARSAT-2 images," in *2007 IEEE International Geoscience and Remote Sensing Symposium* (Barcelona: IEEE), 3655–3658. doi: 10.1109/IGARSS.2007.4423638

Tamiminia, H., Homayouni, S., and Safari, A. (2015). Clustering of multi-temporal fully polarimetric l-band sar data for agricultural land cover mapping. ISPRS–International Archives of the Photogrammetry. *Remote Sens. Spatial Inform. Sci.* XL-1-W5, 701–705. doi: 10.5194/isprsarchives-XL-1-W5-701-2015

Tatsumi, K., Yamashiki, Y., Canales Torres, M. A., and Taipe, C. L. R. (2015). Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Comput. Electr. Agric*, 115, 171–179. doi: 10.1016/j.compag.2015.05.001

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20260–20264. doi: 10.1073/pnas.1116437108

UN FAO (2016). *Afghanistan Special Report: Pre-Harvest Assessment*. Available online at: https://reliefweb.int/sites/reliefweb.int/files/resources/2016%20Pre-Harvest_Assessment%20Report.pdf

United Nations (2013). Ensuring food and nutrition security. *United Nations World Econ. Soc. Surv.* 2013, 85–119. doi: 10.18356/0e3c4bbb-en

USAID (2017). *Agriculture Consolidated Project Appraisal Document*. Available online at: https://www.usaid.gov/sites/default/files/documents/1871/USAID_Afghanistan_Project_Appraisal_Document_-_Public_Version_2017-04-03.pdf

Wardlow, B. D., and Egbert, S. L. (2010). A comparison of MODIS 250-m EVI and NDVI data for crop mapping: a case study for southwest Kansas. *Int. J. Remote Sens.* 31, 805–830. doi: 10.1080/01431160902897858

White, M. A., Thornton, P. E., and Running, S. W. (1997). A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Glob. Biogeochem. Cycles* 11, 217–234. doi: 10.1029/97GB00330

World Bank (2014). *Islamic Republic of Afghanistan Agricultural Sector Review (No. AUS9779)*. Available online at: http://documents.worldbank.org/curated/en/245541467973233146/pdf/AUS9779-REVISED-WP-PUBLIC-Box391431B-Final-Afghanistan-ASR-web-October-31-2014.pdf

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., et al. (2003). Monitoring vegetation phenology using MODIS. *Remote Sens. Environ.* 84, 471–475. doi: 10.1016/S0034-4257(02)00135-9

Zhang, X., Tan, B., and Yu, Y. (2014). Interannual variations and trends in global land surface phenology derived from enhanced vegetation index during 1982–2010. *Int. J. Biometeorol.* 58, 547–564. doi: 10.1007/s00484-014-0802-z

Disclaimer: The views and interpretations in this paper are those of the authors and are not necessarily attributable to ICIMOD, USAID, or NASA.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership