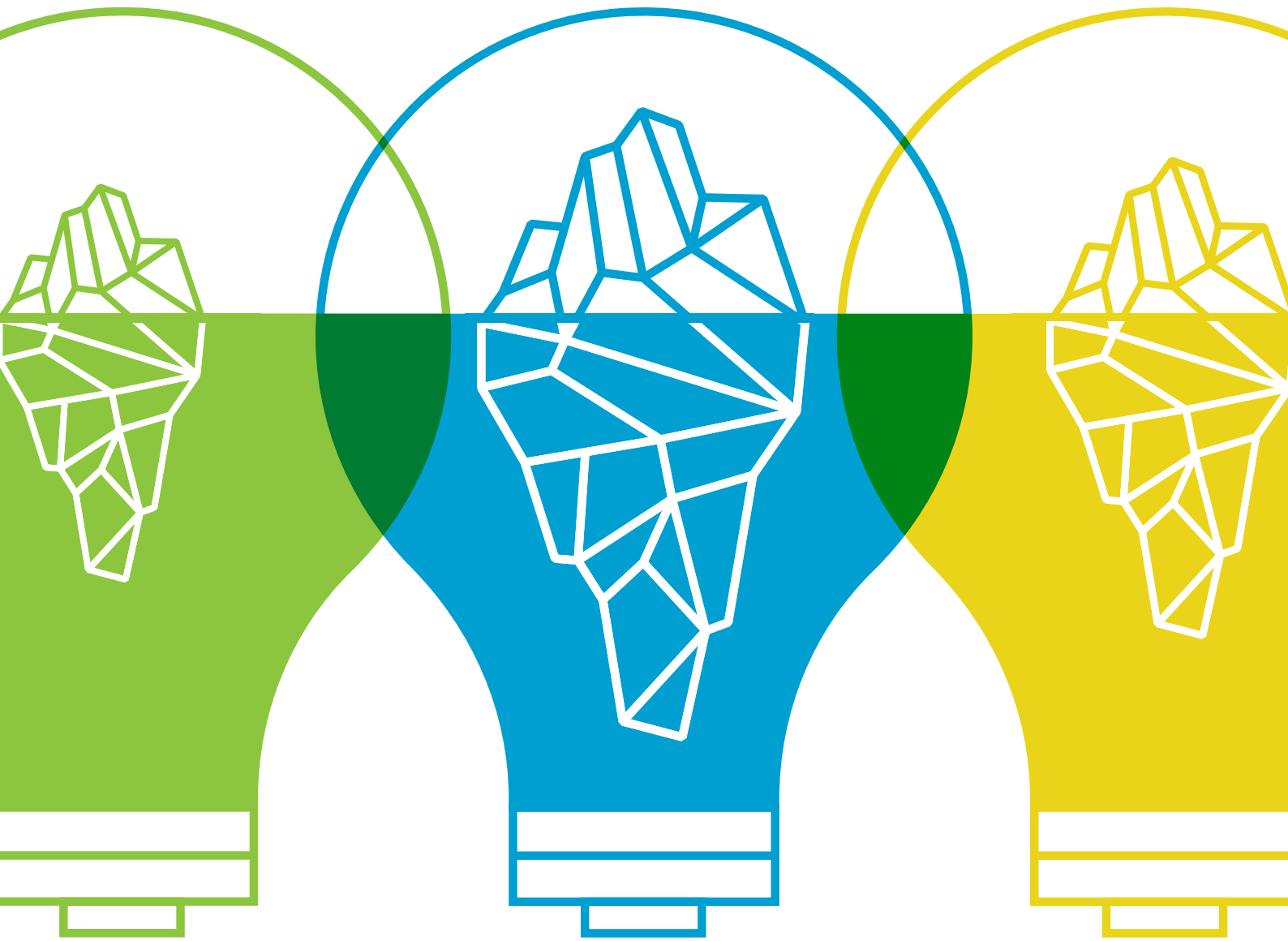
# NEW TECHNIQUES FOR IMPROVING CLIMATE MODELS, PREDICTIONS AND PROJECTIONS

EDITED BY: Matthew Collins, Marcelo Barreiro, Roxy Mathew Koll, Sarah M. Kang, Thomas Frölicher, Karumuri Ashok, Guojian Wang and Renata Goncalves Tedeschi

PUBLISHED IN: Frontiers in Climate



frontiers Research Topics

# frontiers



ISSN 1664-8714
ISBN 978-2-88974-139-7
DOI 10.3389/978-2-88974-139-7

## 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: frontiersin.org/about/contact

# NEW TECHNIQUES FOR IMPROVING CLIMATE MODELS, PREDICTIONS AND PROJECTIONS

Topic Editors:
**Matthew Collins,** University of Exeter, United Kingdom
**Marcelo Barreiro,** Universidad de la República, Uruguay
**Roxy Mathew Koll,** Indian Institute of Tropical Meteorology (IITM), India
**Sarah M. Kang,** Ulsan National Institute of Science and Technology, South Korea
**Thomas Frölicher,** University of Bern, Switzerland
**Karumuri Ashok,** University of Hyderabad, India
**Guojian Wang,** Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
**Renata Goncalves Tedeschi,** Vale Technological Institute (ITV), Brazil

**Citation:** Collins, M., Barreiro, M., Koll, R. M., Kang, S. M., Frölicher, T., Ashok, K., Wang, G., Tedeschi, R. G., eds. (2022). New techniques for improving climate models, predictions and projections. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-139-7

# Table of Contents

frontiers in Climate

EDITORIAL
published: 07 December 2021
doi: 10.3389/fclim.2021.811205

# Editorial: New Techniques for Improving Climate Models, Predictions and Projections

*Matthew Collins[1]*, Karumuri Ashok[2], Marcelo Barreiro[3], Mathew Koll Roxy[4], Sarah M. Kang[5], Thomas L. Frölicher[6,7], Guojian Wang[8] and Renata Goncalves Tedeschi[9]*

[1] College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom, [2] University of Hyderabad, Hyderabad, India, [3] Department of Atmospheric Sciences, School of Sciences, Universidad de la República, Montevideo, Uruguay, [4] Centre for Climate Change Research, Indian Institute of Tropical Meteorology (IITM), Pune, India, [5] Department of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, South Korea, [6] Climate and Environmental Physics, University of Bern, Bern, Switzerland, [7] Oeschger Center for Climate Change Research, University of Bern, Bern, Switzerland, [8] Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, ACT, Australia, [9] Instituto Tecnológico Vale (ITV), Desenvolvimento Sustentável, Belém, Brazil

**Keywords: climate, predictions, projections, machine learning, data assimilation, uncertainties**

**Editorial on the Research Topic**

**New Techniques for Improving Climate Models, Predictions and Projections**

## INTRODUCTION

Complex climate models are the main tools used to make climate predictions and projections. Despite decades of development, models are still imperfect and generations of models have shown persistent mean-state biases such as the "double intertropical convergence zone (ITCZ)." Model imperfections lead to drift and errors in near-term initialized climate prediction systems and uncertainties in long-term future projections. Techniques such as bias correction and drift removal have been developed to reduce the impact of model imperfection in the case of predictions. Techniques such as emergent constraints and model selection have been used in projection studies. Are these techniques adequate, could they be improved upon, or should the community be investing their efforts into significantly improving the performance of climate models? Will higher resolution bring greater accuracy? Are there new techniques which can significantly improve climate predictions and projections?

The goal of this Research Topic was to explore new techniques for improving climate models, climate predictions, and climate projections. The 11 articles that are appearing in this special issue of Frontiers in Climate Predictions and Projections have together shown new avenues in improving the forecasts and projections, and introduce us to new science and new forecast products.

## DATA AND DATA ASSIMILATION

A major topic in data assimilation is the reduction of the drift of the model away from the estimated observed state of the system. Volpi et al. introduce an innovative initialization technique to reduce the initial drift through a quantile matching between the observed state at the initialization time, and the model state distribution. The adjusted initial state pertains to the model attractor. Volpi et al. find added value of the quantile matching initialization in the North Atlantic subpolar region and over the North Pacific surface temperature as well as for the ocean heat content up to 5 years, and improved predictive skill of the Atlantic Meridional Overturning Circulation and the barotropic stream function in the Labrador Sea throughout the 5 forecast years.

OPEN ACCESS

***Edited and reviewed by:***
*Xander Wang, University of Prince Edward Island, Canada*

****Correspondence:***
*Matthew Collins*
*M.Collins@exeter.ac.uk*

***Specialty section:***
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

***Received:*** *08 November 2021*
***Accepted:*** *17 November 2021*
***Published:*** *07 December 2021*

***Citation:***
*Collins M, Ashok K, Barreiro M, Roxy MK, Kang SM, Frölicher TL, Wang G and Tedeschi RG (2021) Editorial: New Techniques for Improving Climate Models, Predictions and Projections. Front. Clim. 3:811205. doi: 10.3389/fclim.2021.811205*

In satellite-based wind retrievals, accuracy is impaired due to noise, while the maximal observable resolution is bounded by the often sparse distribution of sensors. Schweri et al. applying a neural network from numerical simulations on synthetic data show that data recovery at high resolution and high quality can be learned from simulation of physically realizable fluid flows. The study indicates that the learning-based reconstruction is especially powerful in handling large areas of missing or occluded data, relative to traditional models for data recovery. The authors demonstrate the usefulness of the method a real-world flow data set retrieved from satellite imagery of stratocumulus clouds on Guadalupe Island.

## ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING (AI/ML) APPROACHES

Serifi et al. explore the capabilities of neural networks to reconstruct high-resolution data from low-resolution weather simulations and observations datasets. They investigate supervised machine learning using multiple deep convolutional neural network architectures to test the limits of data reconstruction for various spatial and temporal resolutions, low-frequent and high-frequent input data, and the generalization to numerical and observed data. Climate data produced by the COSMO regional climate model at 2.2 km during 2 months of 2008, and observations from Switzerland in 2004 at 1 km are used. While slowly-changing information, such as temperature can be adequately predicted through an error-predicting network, these networks are far less suitable than deconvolutional neural networks for the analysis of high-frequent fields like precipitation due to poor learning performance.

AI/ML techniques also feature in Mitra who uses a probabilistic graphical model, capable of a binary representation, to identify the spatial distribution in the simulated daily rainfall over the Indian landmass during monsoon in several Coupled Model Intercomparison Project Phase 6 (CMIP6) models, and compares the patterns with those from the observed rainfall for the 2000–2014 period. The study suggests that some of the CMIP6 models simulate the spatial distribution of monsoon rainfall to a reasonable degree, but most models underestimate extreme rainfall events, and are unable to reproduce the homogeneity in rainfall across various regions of the landmass, as observed.

## CLIMATE PREDICTABILITY AND PREDICTIONS

Looking at the predictability of the climate system, Ikuyajolu et al. employ a computationally fast method to look at the potential predictability of sea surface temperatures in the tropical Pacific and Indian Oceans during boreal fall. The predictability of the basins is controlled by two regularly varying non-linear oscillations, the El Niño-Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD). Using historical and RCP8.5 outputs from several CMIP5 models and reanalysis data, the authors do not find robust changes in predictability in future projections, despite the discrepancies between the models and the reanalysis. A brief investigation of the discrepancy in predictability in the basin points to a poor representation of the ocean mean state and inter-basin connectivity at the Indonesian Throughflow.

Also focussing on the predictability of climate variability, Sahai et al. present the newest version of the India Meteorological Department multi-physics multi-model ensemble extended range prediction system. While this system includes older options of coupled climate forecast system version 2 and atmospheric global forecast system forced with real-time bias-corrected sea-surface temperature from NCEP coupled forecast system model version 2, unlike the predecessor, the horizontal resolution is "seamless," i.e., the model forecasts are generated at T574 resolution till 15 days, after which, a coarser T382 resolution is selected. In the newer version, model integrations are performed six times in a month for real-time prediction. Analysis of the 15 year-hindcast generated demonstrate appreciable improvements over its predecessor in predicting the large-scale low variability signal and weekly mean rainfall up to 3 weeks lead, and importantly, better performance at subdivisional scales, especially in the northwest and central parts of India.

On longer time scales, Feba et al. highlight considerable advancement in the multi-year prediction and show, for the first time, that decadal predictions from two general circulation models have significant prediction skills for the IOD for at least 2 years and up to 8–10 years after initialization. This skill is present despite ENSO having a lead prediction skill of only 1 year. The source of this multi-year predictability lies in sub-surface signals that propagate from the Southern Ocean into the Indian Ocean. Prediction skill for a prominent climate driver like the IOD has wide-ranging benefits for climate science and society.

Using the technique of event coincidence analysis, Weidermann et al. decipher differential imprints of the East Pacific (EP) and Central Pacific (CP)/Modoki flavors of ENSOs on very low and very high seasonal precipitation patterns over distinct regions across the globe. The authors find that EP periods exhibit statistically significant event coincidence rates with hydrometeorological anomalies at larger spatial scales, whereas sparser patterns emerge along with CP periods. The study documents, for the first time, distinct impacts such as increased rainfall over Central Asia during CP periods. The authors argue for a thorough distinction of El Niño and La Niña into their two respective flavors for understanding the emergence of strong regional hydrometeorological anomalies and anticipating their associated ecological and socioeconomic impacts.

## CLIMATE PROJECTIONS AND UNCERTAINTIES

Clouds are important for feedbacks in the climate system, yet their representation leads to considerable uncertainty in model projections of climate change. Pathak et al. introduce an efficient uncertainty quantification and Bayesian inference for cloud parameters in order to assess the sensitivity of the outputs of the NCAR Single Column Atmosphere Model to various

parameterization schemes. The method involves using two surrogate models that propagate uncertainty in test parameters to model outputs. Their exercise, for example, shows that ∼40–80% of the total variance of the climate variables can be attributed to auto-conversion size threshold for ice to snow, ∼15–30% to the fall speed parameter for stratiform cloud ice, and so on. The study is valuable in quantifying the source of uncertainties in the model physics.

Feedbacks between climate and the carbon cycle also represent a major uncertainty in projections. Using a simple carbon cycle model, and driving emulators of the temperature responses of 41 Coupled CMIP6 emulators with 127 different emission scenarios for the 21st century, Rypdal et al. find almost a perfect linear relationship between maximum global surface air temperature and cumulative carbon emissions, allowing unambiguous estimates of Remaining Carbon Budget (RCB) for each CMIP6 model. The range of these estimates across the model ensemble provides a measure of the uncertainty in the RCB arising from the range in climate sensitivity over this ensemble. Rypdal et al. suggest that observational constraints imposed on the transient climate response in the model ensemble can reduce uncertainty in RCB estimate. They also show that main uncertainty of the transient climate response to cumulative carbon emissions and the associated RCB is not due to the spread of the emission scenarios, but rather the spread of sensitivities over the CMIP6 model ensemble.

Uncertainties may be reduced by the application of constraints on projections. The paper by Hegerl et al. discusses the challenges in understanding the role of observations in skill estimates and constraints, and using them consistently in predicting and projecting changes in European climate. It discusses constraints across prediction and projection methods, their interpretation, and the metrics that drive them. These are illustrated in this paper by examples of state-of-the-art methods for predicting and projecting changes in European climate. The authors study how the skill estimates may vary over time for initialized predictions with different phases of climate variability and climatic conditions, and are influenced by the presence of external forcing. This, the authors state, complicates the systematic use of observational constraints. They also suggest that sub-selecting simulations from large ensembles based on reproduction of the observed evolution of climate variations is a good testbed for combining projections and predictions.

Improving climate predictions and projections is such a big subject area within climate science that this Research Topic could not cover all potential ways in which advances can be made. Nevertheless, some interesting themes emerge. Blending models with observations on climate time scales, and the use of Artificial Intelligence and Machine Learning techniques, are proposed routes for the creation of so-called "digital twins" of the climate system. Such tools are intended to be more powerful than our current models and can be better focussed on policy questions. However, there is still much to be learnt about what is predictable in the near term and how we can quantify and reduce uncertainties in projections of long-term climate change, these efforts are required even if efforts to limit the size of climate change succeed, as there is already a requirement for societies to adapt.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## FUNDING

TLF was supported by the Swiss National Science Foundation (PP00P2_198897), GW was supported by the Centre for Southern Hemisphere Oceans Research and the Climate Systems Hub of the Australian Government's National Environmental Science Program, and MC supported by NE/S004645/1.

**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 Climate

ORIGINAL RESEARCH
published: 12 February 2021
doi: 10.3389/fclim.2021.618548

# Differential Imprints of Distinct ENSO Flavors in Global Patterns of Very Low and High Seasonal Precipitation

*Marc Wiedermann*[1]*, *Jonatan F. Siegmund*[1], *Jonathan F. Donges*[1,2] *and Reik V. Donner*[1,3]

[1] *Potsdam Institute for Climate Impact Research (PIK)—Member of the Leibniz Association, Potsdam, Germany,* [2] *Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden,* [3] *Department of Water, Environment, Construction and Safety, Magdeburg–Stendal University of Applied Sciences, Magdeburg, Germany*

OPEN ACCESS

**Edited by:**
*Marcelo Barreiro, Universidad de la República, Uruguay*

**Reviewed by:**
*Cristina Masoller, Universitat Politecnica de Catalunya, Spain*
*Annalisa Cherchi, Institute of Atmospheric Sciences and Climate (CNR-ISAC), Italy*

***Correspondence:**
*Marc Wiedermann*
*marcwie@pik-potsdam.de*

**Specialty section:**
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

**Received:** *17 October 2020*
**Accepted:** *20 January 2021*
**Published:** *12 February 2021*

**Citation:**
*Wiedermann M, Siegmund JF, Donges JF and Donner RV (2021) Differential Imprints of Distinct ENSO Flavors in Global Patterns of Very Low and High Seasonal Precipitation. Front. Clim. 3:618548. doi: 10.3389/fclim.2021.618548*

The effects of El Niño's two distinct flavors, East Pacific (EP) and Central Pacific (CP)/Modoki El Niño, on global climate variability have been studied intensively in recent years. Most of these studies have made use of linear multivariate statistics or composite analysis. Especially the former assumes the same type of linear statistical dependency to apply across different phases of the El Niño–Southern Oscillation (ENSO), which appears not necessarily a justified assumption. Here, we statistically evaluate the likelihood of co-occurrences between very high or very low seasonal precipitation sums over vast parts of the global land surface and the presence of the respective EP and CP types of both, El Niño and La Niña. By employing event coincidence analysis, we uncover differential imprints of both flavors on very low and very high seasonal precipitation patterns over distinct regions across the globe, which may severely affect, among others, agricultural and biomass production or public health. We particularly find that EP periods exhibit statistically significant event coincidence rates with hydrometeorological anomalies at larger spatial scales, whereas sparser patterns emerge along with CP periods. Our statistical analysis confirms previously reported interrelations for EP periods and uncovers additional distinct regional patterns of very high/low seasonal precipitation, such as increased rainfall over Central Asia alongside CP periods that have to our knowledge not been reported so far. Our results demonstrate that a thorough distinction of El Niño and La Niña into their two respective flavors could be crucial for understanding the emergence of strong regional hydrometeorological anomalies and anticipating their associated ecological and socioeconomic impacts.

**Keywords: El Niño Southern Oscillation, precipitation, Central Pacific El Niño (El Niño Modoki), Central Pacific La Niña, event coincidence analysis**

## 1. INTRODUCTION

The positive (El Niño) and negative (La Niña) *phases* of the El Niño–Southern Oscillation (ENSO) are known to be associated with wide-spread anomalies in the mean hydrometeorological conditions at various distant parts of the Earth. These long-ranged interactions are often referred to as teleconnections (Trenberth, 1997; Neelin et al., 2003; Domeisen et al., 2019). In this context, recent findings indicate that there exist two distinct *types* or *flavors* of El Niño phases, usually referred to as the East Pacific (EP) or canonical El Niño and the Central Pacific (CP) El Niño or El Niño Modoki (Ashok et al., 2007; Kao and Yu, 2009; Marathe et al., 2015), respectively. It has been shown that these two flavors are possibly associated with distinct hydrometeorological responses in certain regions (Taschetto and England, 2009). Examples include reduced rainfall over eastern Australia (Chiew et al., 1998) or Southern Africa (Ratnam et al., 2014) only during EP periods as well as increased precipitation over the tropical regions of Africa (Preethi et al., 2015) or the western Indo-Pacific Oceans (Weng et al., 2011; Feng and Chen, 2014) during CP periods. The main reason for such differential responses are different longitudinal positions of the strongest ENSO related sea surface temperature (SST) anomalies in the tropical Pacific Ocean alongside the different flavors, which results in different modifications of large-scale atmospheric circulation patterns (Ashok and Yamagata, 2009; Domeisen et al., 2019).

A similar discrimination into two types has been suggested for La Niña phases as well (Kug and Ham, 2011), even though their respective imprints on SST patterns seem less distinct than for El Niño (Kao and Yu, 2009; Ren and Jin, 2011). It therefore remains an open problem to provide further statistical and/or dynamical evidence either in favor or against such a distinction of different La Niña flavors in analogy to El Niño (Chen et al., 2015; Zhang et al., 2015; Levine et al., 2016). Still, differential hydrometeorological responses alongside specific La Niña phases have recently been identified especially across the Pacific (Shinoda et al., 2011; Magee et al., 2017; Song et al., 2017; Hidayat et al., 2018), which provides good reason for discriminating La Niña into two types as well.

Most previous studies (including those mentioned above) on the teleconnective impacts of different ENSO phases and flavors have either applied linear statistical tools, such as (partial) correlation analysis (Diaz et al., 2001; Weng et al., 2011; Preethi et al., 2015; Magee et al., 2017; Hidayat et al., 2018) or empirical orthogonal function (EOF) analysis (Taschetto and England, 2009), or investigated composites (i.e., mean spatial patterns for a specific type of ENSO period) of the corresponding climate observable of interest (Feng and Chen, 2014; Hoell et al., 2014; Song et al., 2017). Specifically, the former methods share the common limitation of focusing on linear or average interdependencies between ENSO and possible response variables. At the same time, global climate change has been projected to lead to an increase in the strength and frequency of both, climate extremes (Easterling et al., 2000; Karl and Trenberth, 2003) as well as extreme ENSO phases (Cai et al., 2014, 2015). This calls for more systematically assessing possible statistical as well as dynamical/mechanistic linkages between these two findings (Allan and Soden, 2008). Accordingly, the present work aims to identify and quantitatively characterize spatial patterns of markedly wet/dry seasons that have an elevated probability to co-occur with certain types (EP or CP) of ENSO phases, following upon previous findings that ENSO can have large-scale effects on rainfall patterns at both, global and regional scales (Ropelewski and Halpert, 1987; Dai and Wigley, 2000).

Several strategies to distinguish East Pacific (EP) from Central Pacific (CP) ENSO events have been proposed in the recent past (Hendon et al., 2009). One prominent example is the ENSO Modoki Index that is computed as weighted average SST anomalies over three specific regions (165°E to 140°W and 10°S to 10°N, 110°W to 70°W, and 15°S to 5°N, as well as 125°E to 145°E and 10°S to 20°N) in the equatorial Pacific (Ashok et al., 2007). Other approaches have used empirical orthogonal functions (Kao and Yu, 2009; Graf and Zanchettin, 2012) or combinations of the Nino3 and Nino4 indices (Hu et al., 2011; Kim et al., 2011) to provide the desired categorization.

Wiedermann et al. (2016) recently compiled a synthesis of eight previous studies that used either of the aforementioned frameworks and identified several ENSO periods during which these techniques proposed either ambiguous, mutually inconsistent, or incomplete classifications, with certain ENSO periods remaining entirely unassigned. In order to fill those gaps, Wiedermann et al. (2016) proposed a new index based on spatial correlation structures among global surface air temperature anomalies, which can be conveniently studied in terms of the transitivity of so-called *functional climate networks* (Tsonis et al., 2006; Donner et al., 2017). This transitivity index confirmed the flavors of all El Niños between 1953 and 2013 that had been classified in mutual agreement across the existing literature (Wiedermann et al., 2016). It also allowed to assign types to those cases where former work yielded incomplete or ambiguous categorizations and thereby provided a consistent, comprehensive, and complete classification of the respective flavors. Moreover, Wiedermann et al. (2016) showed that the transitivity index also naturally distinguishes La Niña episodes into two corresponding types, thereby providing a unique advantage over other classification schemes that mainly focus on El Niño phases alone. Even though the existing literature on a discrimination of different La Niña periods is comparatively scarce, the transitivity index confirmed the results of at least two recent studies (Tedeschi et al., 2013; Yuan and Yan, 2013) and again provides a consistent classification for years that previously had types ambiguously or not all assigned (Wiedermann et al., 2016). For both reasons, i.e., the comprehensiveness of the classification and the ability to distinguish La Niña into two types as well, we will directly use the classified periods from Wiedermann et al. (2016) (see **Table 1** for an overview) for the purpose of our present study.

Based on this categorization of El Niño and La Niña phases, we quantify the likelihood of simultaneous or time-delayed co-occurrences of strong seasonal wet/dry anomalies at a local scale across vast parts of the global land surface with a certain type of ENSO phase. Specifically, we consider seasonal precipitation sums for boreal fall (September–November, SON),

**TABLE 1 |** Classification of ENSO periods according to Wiedermann et al. (2016) on the basis of a comprehensive literature synthesis and a consistent classification using the network-based transitivity index.

| El Niño | | La Niña | |
|---|---|---|---|
| **East pacific** | **Central pacific** | **East pacific** | **Central pacific** |
| 1957 | 1951 | 1964 | 1954 |
| 1965 | 1953 | 1970 | 1955 |
| 1972 | 1958 | 1973 | 1967 |
| 1976 | 1963 | 1988 | 1971 |
| 1982 | 1968 | 1998 | 1974 |
| 1997 | 1969 | 2007 | 1975 |
| 2015* | 1977 | 2010 | 1984 |
| | 1979 | | 1995 |
| | 1986 | | 2000 |
| | 1987 | | 2001 |
| | 1991 | | 2011 |
| | 1994 | | |
| | 2002 | | |
| | 2004 | | |
| | 2006 | | |
| | 2009 | | |
| | 2014* | | |

*The given years correspond to the onset-year of each ENSO event such that, e.g., 1951 indicates the 1951/1952 El Niño event. Years marked with an asterisk are classified using the same methodology as in Wiedermann et al. (2016) to extend the data to the period of study considered in this work.*

winter (December–February, DJF), and spring (March–May, MAM) as key seasons of the developmental cycle of El Niño and/or La Niña conditions. By contrast, we omit the boreal summer season, even though large scale hydrometeorological conditions during this season could still be distinctively affected by different (approaching or withdrawing) ENSO phases. The reason for this choice is that there are various cases where El Niño and La Niña conditions occur in subsequent years so that a unique attribution of anomalies to any of these two phases would be hardly possible. In our present analysis, strong wet (dry) anomalies will be defined as seasonal precipitation sums exceeding the empirical 80th (falling below the empirical 20th) percentile of the distribution of all values on record for a given season and location.

To statistically quantify co-occurrences between different types of ENSO phases and very wet/dry seasons, we employ *event coincidence analysis* (ECA) (Donges et al., 2011, 2016). Put in simple terms, ECA counts the fraction of events of one type (here, some particularly wet/dry season at a certain location) that coincide with those of another type (here, some type of ENSO period) while, in contrast to other conceptually related approaches like event synchronization (Quian Quiroga et al., 2002; Malik et al., 2010; Boers et al., 2014), allowing for a precise control of the relative timing between them (Wolf et al., 2020). This statistical framework has already been successfully applied to quantify the likelihood of climatic (extreme) events possibly triggering certain ecological or socioeconomic responses, such as extreme annual (Rammig et al., 2015) and daily (Siegmund et al., 2016a) tree growth or flowering dates (Siegmund et al., 2016b), the outbreak of epidemics (Donges et al., 2016) or armed conflicts (Schleussner et al., 2016; Ide et al., 2020).

In the course of this work, we first evaluate significant event coincidence rates for the canonical (EP) El Niño. This allows us to demonstrate the consistency of our approach by comparing the obtained spatial patterns with a variety of previously reported results. We continue by studying co-occurrences between strong seasonal precipitation anomalies and the so far less intensively studied CP El Niño periods to highlight differences in their teleconnectivity patterns as compared to their EP counterparts. Ultimately, we also study La Niña periods and demonstrate that the most remarkable large-scale spatial patterns of strong seasonal rainfall anomalies non-randomly co-occurring with this negative ENSO phase are associated with the corresponding EP flavor. This finding suggests that in the light of recent discussions on the existence of two distinct types of La Niña periods (Kao and Yu, 2009; Kug and Ham, 2011; Ren and Jin, 2011; Zhang et al., 2015), it is indeed meaningful to consider a global impact-based distinction between one type that significantly affects seasonal wet/dry patterns globally and another that exhibits less spatially coherent impacts. Wherever appropriate, we also briefly discuss possible ecological and socioeconomic consequences of the identified seasonal precipitation anomalies.

## 2. DATA AND METHODS

### 2.1. GPCC Rainfall Data

We utilize gridded monthly precipitation data provided by the Global Precipitation Climatology Centre (GPCC) at a spatial resolution of $2.5^\circ \times 2.5^\circ$ (Schneider et al., 2018). Since reliable discriminations of El Niño and La Niña periods into their respective EP and CP flavors are so far mainly available for the second half of the twentieth century (Graf and Zanchettin, 2012; Yuan and Yan, 2013; Wiedermann et al., 2016; Freund et al., 2019), we restrict our analysis to the period from 1951 to 2016 (the most recent year in the GPCC data set). We derive separate records for three seasons $s$ by aggregating the precipitation sums of the corresponding 3-month periods from September to November (SON, 1951–2015), December to February (DJF, 1951/52–2015/16), and March to May (MAM, 1952–2016) to cover all seasons associated with the 1951/52 El Niño and all following ENSO phases while ensuring the same length of all series. This results in three time series $P_{i,s}(t)$ per grid cell $i$ with $M = 65$ annual values each. Note that the density of stations from which the GPCC data has been derived varies between 0 and more than 100 per grid cell and year (Lorenz and Kunstmann, 2012), which generally results in a lower accuracy and reliability of the data for those areas with only few stations (Rudolf et al., 1994). We therefore consider only grid cells with at least one station present for 95% of the study period in a specific season $s$ (SON, DJF, or MAM). In addition, we exclude those grid cells where the average precipitation sum in a specific season is below 3 cm (i.e., 1 cm per month) since this particular choice of threshold ensures the exclusion of deserts from our analysis (Chatterton et al., 1971; Houston, 2006; Thomas and Nigam, 2018). Both preprocessing steps yield a total number of

$N_{SON} = 1763$, $N_{DJF} = 1,610$, and $N_{MAM} = 1,736$ valid grid cells in SON, DJF, and MAM (see **Supplementary Figure 5** for details on the spatial distribution of valid grid cells).

We note that in the context of the present work, using 3-month precipitation sums provides information fully equivalent to that obtained when using the 3-month standardized precipitation index SPI-3 (Guttman, 1999; Svoboda et al., 2012), which is a solely precipitation-based characteristic commonly employed in drought-related studies. Specifically, SPI-3 values can be derived by a sophisticated monotonic transformation of the 3-month precipitation sums, which does not change the rank order of the values and, hence, the timing of the events considered in this study.

## 2.2. Classification of ENSO Periods

We use the classification of East Pacific (EP) and Central Pacific (CP) flavors of recent El Niño and La Niña episodes from Wiedermann et al. (2016) (**Table 1**). This classification has been based on a comprehensive literature review of eight (two) studies that distinguished certain El Niños (La Niñas) between 1953 and 2010 into their two respective types. Based on this compilation, Wiedermann et al. (2016) identified 11 ENSO periods for which previous works yielded mutually consistent results and another 10 periods for which (a subset) of previous studies yielded incomplete, ambiguous, or mutually inconsistent classifications.

In order to achieve a consistent and comprehensive classification of both, El Niño and La Niña flavors, including also the ENSO periods without previous consensus, Wiedermann et al. (2016) introduced the so-called *transitivity index*, which reflects distinct characteristics (in terms of abundance and localization) of ENSO's teleconnections during EP and CP periods (Radebach et al., 2013; Wiedermann et al., 2016). In brief, this index is obtained from 1-year sliding-window lag-zero absolute correlation matrices between time series at all pairs of grid cells in the global daily surface air temperature anomaly field (Kistler et al., 2001). The most relevant information of these matrices is contained in the 0.5% strongest absolute correlations (Donges et al., 2009; Radebach et al., 2013; Wiedermann et al., 2016), yielding thresholds for each considered time window below which the matrix coefficients are put to zero. The thus obtained sparse matrices are then considered as weighted adjacency matrices of so-called *functional climate networks* (Donges et al., 2009; Radebach et al., 2013; Wiedermann et al., 2016) with positive entries indicating a strong statistical relationship between climate variability in two grid cells. The transitivity index (Newman, 2003; Antoniou and Tsompa, 2008) then describes the degree to which such strong relationships among triples of grid cells are transitive, i.e., the fraction of cases in which connections between two pairs $(i, j)$ and $(i, k)$ of grid cells sharing a common member are accompanied by a third connection between the remaining pair $(j, k)$. In order to provide a representative property, this fraction is further weighted by the respective cell-sizes and strengths of the involved links, i.e., the corresponding values of the absolute cross-correlation (Saramäki et al., 2007). The transitivity index quantifies the (dispersed or localized) spatiotemporal distribution of links in the climate network and directly reflects distinct characteristics in the temporal evolution of spatial autocorrelations and global teleconnections that are unique to specific ENSO flavors (Radebach et al., 2013; Wiedermann et al., 2016).

Given the presence of either El Niño or La Niña conditions as indicated by the Oceanic Nino Index (ONI), the transitivity index then indicates EP phases by a strong peak co-occurring with the respective ENSO period. By contrast, CP phases can be identified by the absence of such a transitivity peak with values of the index close to its baseline (see Wiedermann et al., 2016 for a comprehensive description and interpretation of the framework as well as all necessary mathematical details).

The transitivity index confirmed the classification of all 11 consistently reported El Niño periods from earlier works and additionally provided a comprehensive classification for those periods that were previously classified inconsistently or incompletely. While most other recent approaches for classifying ENSO flavors were mostly tailored to El Niño events, the transitivity index also provides a consistent classification of La Niña periods (Wiedermann et al., 2016), thereby making it particularly useful for our present study.

In total, Wiedermann et al. (2016) identified six EP El Niño and sixteen CP El Niños between 1951 and 2014, which is largely consistent with a recent study by Freund et al. (2019) that used objective supervised machine learning to obtain a similar discrimination. Similarly, Wiedermann et al. (2016) found seven EP La Niñas and 11 CP La Niñas. Since Wiedermann et al. (2016) only provide classifications for all ENSO periods prior to the year 2014, we apply their methodology to extend the event categorization to our present study period. We specifically identify the 2014/2015 and 2015/2016 El Niño periods as CP and EP types, respectively, which is again consistent with the recently proposed classification by Freund et al. (2019). An overview of all ENSO periods and types that are ultimately used in this study is given in **Table 1**.

Based on the classification in **Table 1**, we create binary indicator time series for the four different ENSO phases (see **Figure 1**), e.g., an EP El Niño series $X_{\mathrm{EPEN}}(t)$ with $X_{\mathrm{EPEN}}(t) = 1$ if $t$ marks the onset-year of an EP El Niño and $X_{\mathrm{EPEN}}(t) = 0$ otherwise (solid lines in **Figure 1A**). Correspondingly, we obtain the event series $X_{\mathrm{CPEN}}(t)$ for CP El Niños (dashed lines in **Figure 1A**). The same procedure is applied to La Niña periods, resulting in two event series $X_{\mathrm{EPLN}}(t)$ and $X_{\mathrm{CPLN}}(t)$, respectively (**Figure 1B**).

## 2.3. Data Preprocessing

We identify years with seasons $s$ (DJF, SON, or MAM, see above) exhibiting extraordinary high or low precipitation amounts from the corresponding time series $P_{s,i}(t)$ for each grid cell $i$, individually. Specifically, we consider values above (below) the 80th (20th) percentile $p^{+}_{s,i}$ ($p^{-}_{s,i}$) in each of the time series $P_{s,i}(t)$ as extraordinary high (low) seasonal precipitation sums (**Figure 2A**). We choose these particular thresholds to ensure the presence of a sufficient number of particularly dry and wet seasons that is comparable with the number of different types of ENSO periods in the considered study period. It has been

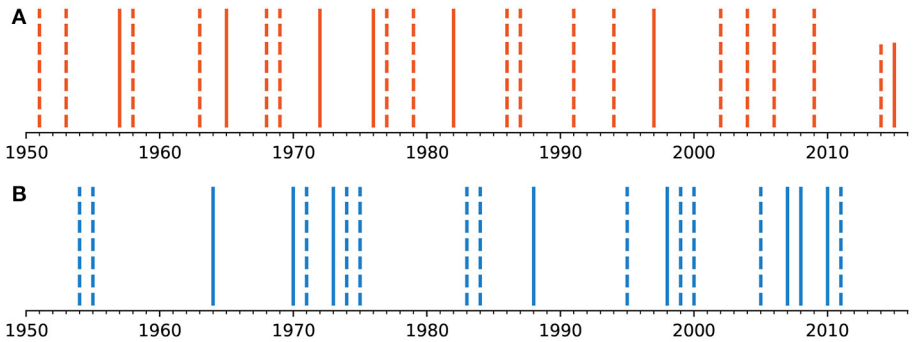



**FIGURE 1 | (A)** Event series $X_{EPEN}$ of East Pacific (solid lines) and $X_{CPEN}$ of Central Pacific (dashed lines) El Niños according to the classification in Wiedermann et al. (2016) and **Table 1**. **(B)** The same for La Niña periods.

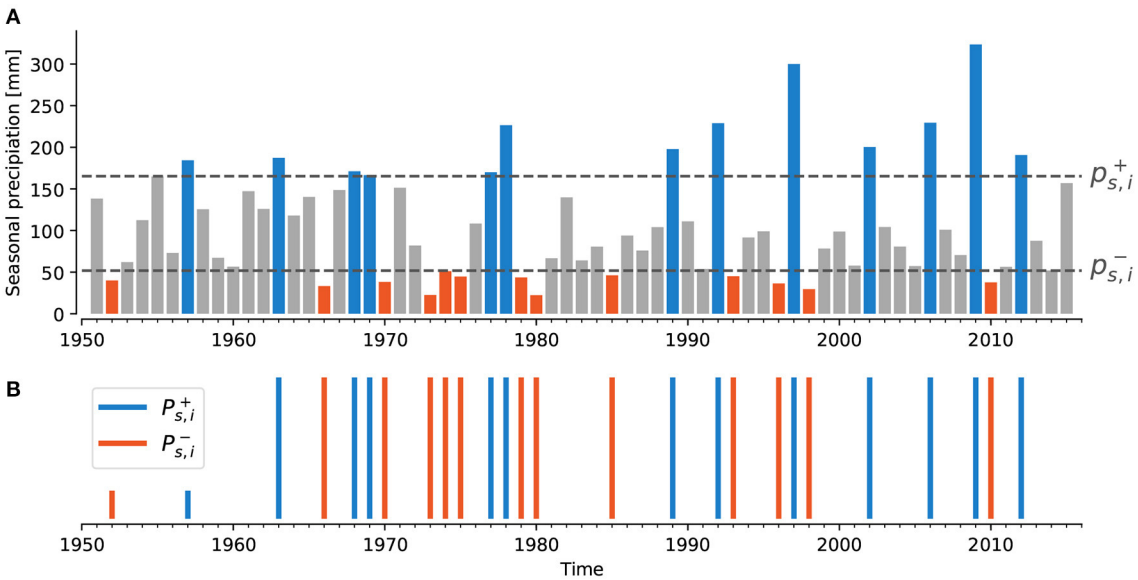



**FIGURE 2 | (A)** Exemplary seasonal precipitation time series $P_{s,i}(t)$ (colored bars) as well as corresponding thresholds $p^+_{s_i}$ and $p^-_{s_i}$ computed as the 80th and 20th percentile of $P_{s,i}(t)$, respectively. Precipitation signals above $p^+_{s_i}$ (blue bars) are considered very wet while those below $p^-_{s,i}$ (red bars) are considered very dry for the subsequent analysis. **(B)** The two corresponding event series $P^+_{s,i}(t)$ and $P^-_{s,i}(t)$ with $P^\pm_{s,i}(t) = 1$ if a corresponding event occurs at time $t$ and $P^\pm_{s,i}(t) = 0$ otherwise.

checked that the results obtained in this work do not change qualitatively if more restrictive or loose thresholds are applied (see **Supplementary Material** for details).

According to these considerations, we obtain six binary (indicator) time series $P^\pm_{s,i}(t)$ for each GPCC grid cell $i$,

$$P^\pm_{s,i}(t) = \Theta(\pm P_{s,i}(t) \mp p^\pm_{s,i}), \tag{1}$$

where $P^+_{s,i}(t) = 1$ ($P^-_{s,i}(t) = 1$) indicates the presence of a very high (very low) seasonal precipitation sum at grid cell $i$ during season $s$ in year $t$ (**Figure 2**). By following the above procedure, no further deseasonalization of the precipitation data is necessary, since the grid cell specific seasonality of precipitation is already taken into account. Furthermore, the events are defined for each grid cell independent of all the others, which means that the specific characteristics of rainfall variability and strength at one location do not influence the definition of events at other locations as it would be the case when considering, e.g., a global (location-independent) threshold for seasonal precipitation.

## 2.4. Event Coincidence Analysis

Event coincidence analysis (ECA) is a statistical tool that quantifies the empirical likelihood of co-occurrences of events in two series (Donges et al., 2011, 2016; Rammig et al., 2015). To complement other conceptually related approaches like event synchronization (Quian Quiroga et al., 2002; Malik et al., 2010; Boers et al., 2014), ECA has been developed based on analytical considerations on paired point processes and combines a precise control of the relative timing of (instantaneous or mutually lagged) events that are considered synchronous with analytical confidence bounds on the obtained event coincidence rates (Donges et al., 2011, 2016). Both aforementioned features

provide some considerable benefit of ECA in the specific context of the present work not shared by other similar methods. Particularly the precise control of the relative timing between events (along with the specific ability to chose a global coincidence interval; Wolf et al., 2020) is crucial for our analysis. It ensures that the we only consider precipitation anomalies that occur simultaneously (for SON or DJF) or exactly a year subsequent (for MAM) to the onset of a respective ENSO phase.

In the context of our work, ECA provides for each grid cell the fraction (called *event coincidence rate*) of EP/CP ENSO periods that co-occur with very high or low precipitation sums in SON or DJF of the same year or MAM of the year following the onset of an El Niño or La Niña phase. Hence, the event coincidence rate $ECR^{\pm}_{s,i,\bullet}$ for one pair of ENSO and precipitation event series is given by

$$ECR^{\pm}_{s,i,\bullet} = \frac{\sum_t X_\bullet(t) P^{\pm}_{s,i}(t-\tau)}{\sum_t X_\bullet(t)}. \quad (2)$$

Here, $X_\bullet(t)$ represents one of the four time series indicating the presence of EP and CP flavors of El Niño and La Niña (see above). While interpreting $t$ as the calendar year, the offset $\tau$ reads $\tau = 0$ for SON and DJF and $\tau = 1$ for MAM.

Note, that our present analysis studies so-called *trigger coincidence rates* (Donges et al., 2016) that quantify the likelihood of a given ENSO period and phase to be *followed* by a specific strong/wet precipitation signal. In contrast, a complementary definition of ECR (denoted *precursor coincidence rate*; Donges et al., 2016) would address the inverse problem of quantifying likelihoods that a given precipitation event (that could arise through a variety of conditions and drivers) is *preceded* by a specific ENSO period and phase. However, since our present work only considers instantaneous coincidences (see Donges et al., 2016 for details), i.e., coincidences between events in the same ENSO period, the two types of coincidence rates only differ in their normalization [i.e., the denominator of Equation (2)]. They therefore provide essentially similar information, except for possible differences in the associated statistical significance resulting from different numbers of events (i.e., different sample sizes).

To assess the statistical significance of the empirical event coincidence rates, we assume both involved event sequences to be distributed randomly, independently and uniformly (Donges et al., 2011, 2016). A corresponding $p$-value is derived analytically from the probability distribution of event coincidence rates that would occur by chance only. We consider an empirical event coincidence rate as statistically significant if its associated $p$-value is smaller than a confidence level of $\alpha = 0.05$ (Donges et al., 2016).

# 3. RESULTS

## 3.1. Seasonal Wet/Dry Patterns and EP El Niño

We first investigate co-occurrences of EP El Niño periods and very wet/dry seasons. **Figures 3A,C,E** highlight areas with significant event coincidence rates between EP El Niños and very dry (red squares) and wet seasons (blue squares) in SON, DJF, and MAM, respectively.

During those SON seasons that correspond to the developing stages of EP El Niños, we find an elevated probability of very dry conditions over Indonesia, the Philippines, and the southwestern Pacific Islands as well as over northern South America and the northern Amazon Basin (**Figure 3A**). Droughts in the latter region have been previously linked to an increased risk of biomass loss in the Amazon which normally serves as a long-term carbon sink (Phillips et al., 2009; Lewis et al., 2011).

For the same season (SON), we also observe an increased likelihood of very wet conditions along the west coast of North America (**Figure 3A**). Similarly, unusually wet conditions frequently emerge over Ecuador and southeastern South America in SON and DJF (**Figures 3A,C**). We further observe more spatially confined regions with wet conditions over parts of the Chilean Andes in SON (**Figure 3A**), which may result in an increased risk for the occurrence of floods in this area (Bookhagen and Strecker, 2012; Boers et al., 2014). All these observations agree well with previous studies (Diaz et al., 2001).

Coinciding with EP El Niños, we also observe more frequent wet conditions over the Mediterranean region and East Africa during SON (**Figure 3A**), both of which have been previously reported in local case studies (Camberlin et al., 2001; Shaman and Tziperman, 2010). The observed tendency toward very dry conditions in southwestern Africa during DJF (**Figure 3C**) has also been reported recently (Hoell et al., 2014).

For MAM seasons, we observe pronounced large-scale patterns of significant event coincidence rates between EP El Niños and low precipitation sums in Northeast Brazil (**Figure 3E**), which is consistent with previous studies (Kane, 1997). Furthermore, the dry conditions over the Philippines that are observed during SON (see above) and DJF also persist into the MAM season (**Figure 3E**). In addition, strong MAM rainfall occurring alongside EP El Niños is most prominently observed in the southeastern United States, which is again consistent with previous works that used composite analysis to determine North American weather patterns associated with El Niño conditions (Ropelewski and Halpert, 1986).

Taken together, the results of ECA are overall in good agreement with previously reported interrelations between El Niño and global precipitation patterns, which have mostly been identified using linear statistical tools such as correlation analysis (Camberlin et al., 2001; Phillips et al., 2009), principal component analysis (Diaz et al., 2001), or composites based on seasonal averages (Shaman and Tziperman, 2010; Hoell et al., 2014). Thus, we conclude that the application of ECA to unveil strong ENSO related hydrometeorological anomalies provides consistent results when compared to previous studies. Our results also imply that strong responses of seasonal precipitation to canonical (EP) El Niño conditions show similar spatial patterns as the average statistical interdependency between ENSO-related indices and the corresponding hydrometeorological observables. However, we note that most previous studies have not discriminated between the two El Niño flavors. Thus, the agreement between our results for EP El Niños and the existing literature suggests that the

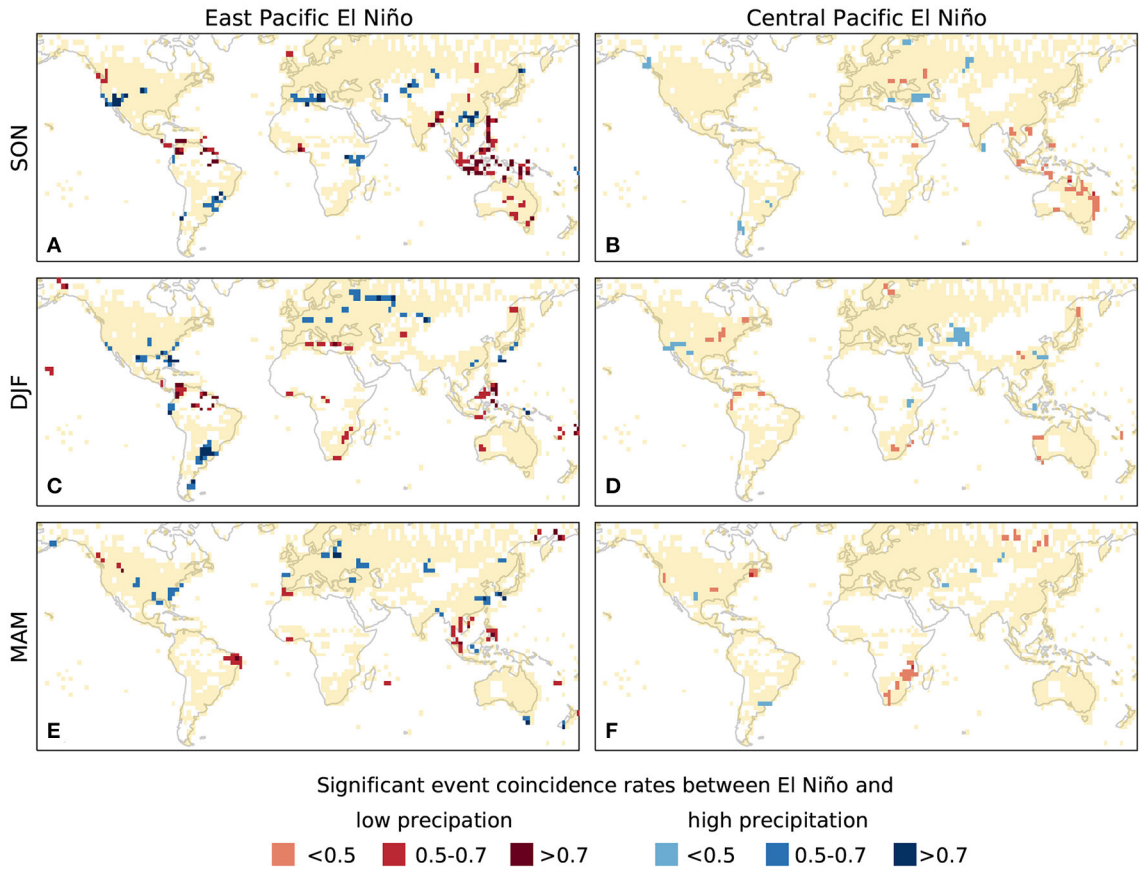



**FIGURE 3 |** Statistically significant event coincidence rates (ECR) between EP **(A,C,E)** and CP **(B,D,F)** El Niños and very dry (red squares) or very wet (blue squares) conditions for the three seasons SON, DJF, and MAM. Dry/wet periods are defined by seasonal precipitation sums below/above the 20th/80th percentile of all years from 1951 to 2016. Only significant clusters of at least two adjacent grid-cells are shown. Different shades of red/blue indicate increasing magnitudes of significant ECR between the respective El Niño events and strong/weak seasonal precipitation. Yellow areas indicate grid cells with non-significant event coincidence rates. White areas over land indicate insufficient quality of the GPCC dataset (i.e., grid cells excluded from our analysis).

previously observed (linear) effects might be dominated by the (on average stronger; Kug et al., 2009; Huang et al., 2016) EP events.

## 3.2. Seasonal Wet/Dry Patterns and CP El Niño

Next, we focus on very high and low seasonal precipitation along with the so far less intensively studied CP El Niño (cf. **Figures 3B,D,F**). We first discuss those regions that display significant event coincidence rates for EP El Niños (see above) but *not* for the corresponding CP periods. Notably, very dry seasons over tropical South America that frequently co-occur together with EP El Niños are markedly less prominent for CP El Niños during DJF (**Figure 3D**) and do not display any significant event coincidence rates at all in SON (**Figure 3B**). The latter also holds true for very wet seasons along the western coast of Central and North America that have been observed for EP El Niños. In the same manner, the wet SON patterns over southern China, the Mediterranean, and East Africa frequently co-occurring with EP El Niños in SON do not exhibit significant coincidence rates with CP El Niños (**Figure 3B**). We further note, that the large-scale dry events over Indonesia observed along with EP El Niños during SON become less spatially coherent for CP El Niños (**Figure 3B**). For MAM, the wet patterns over the northern Iberian peninsula that are observed for EP El Niños cannot be identified for CP El Niños (**Figure 3F**).

While the aforementioned observations indicate decreased or weakened impacts of CP El Niños in comparison with the EP flavor, we also observe new additional patterns of significant event coincidence rates that are not present during EP El Niños but emerge only along with CP periods. Most notably, very dry conditions become more likely along Australia's east coast during SON (**Figure 3B**). Such hydrometeorological anomalies could thus result in severe impacts on river ecosystems and agriculture in that region (Leblanc et al., 2012). In particular, marked drought phases in Eastern Australia are likely to cause a cascade of low river inflows, general water scarcity and large scale floodplain forest mortality as well as an increase of toxicity in the surrounding lakes (Leblanc et al., 2012). These natural hazards can have substantial effects on agricultural production in terms of a severe reduction in irrigated crop yields (van Dijk et al., 2013).

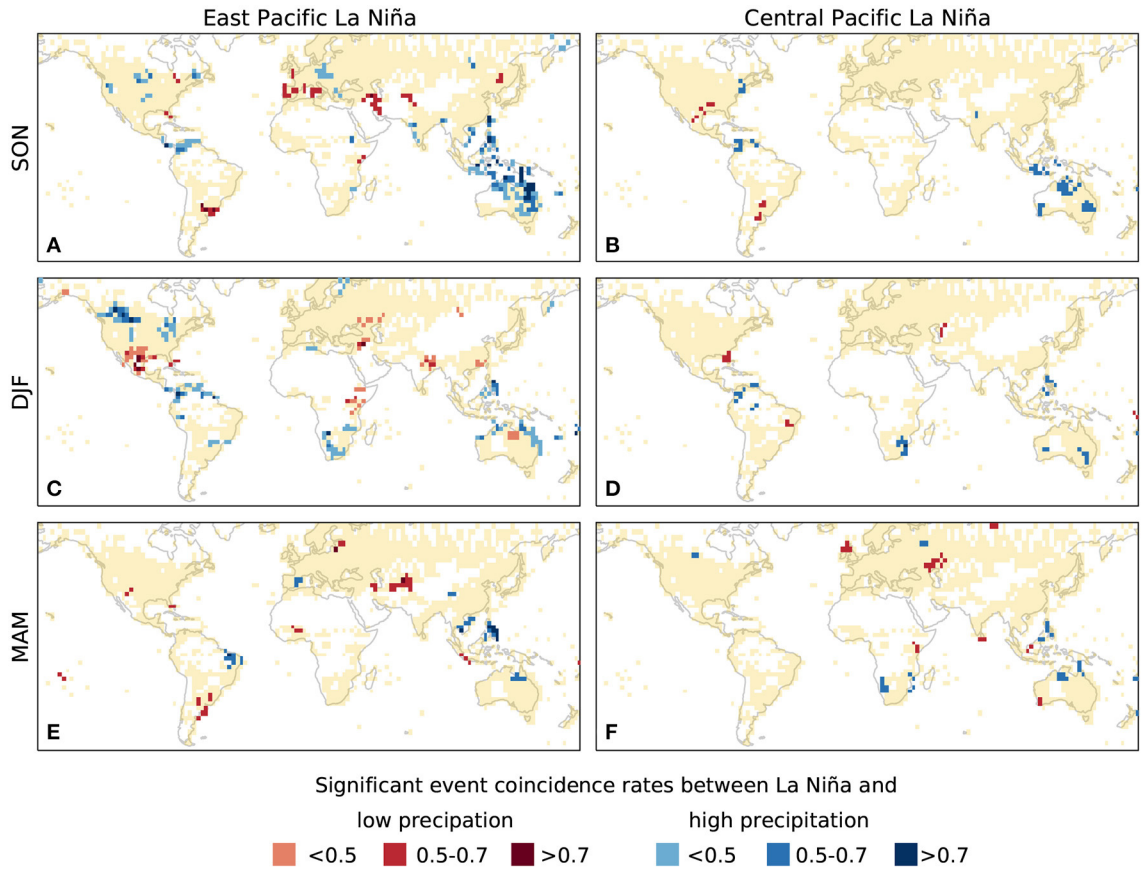



**FIGURE 4 |** Same as **Figure 3**, but for the two types of La Niña.

In addition to reduced rainfall responses, significant event coincidence rates with very wet conditions are found over southern Chile pointing toward increased rainfall during CP El Niños as compared to their canonical counterparts. In DJF months coinciding with CP El Niño periods, we observe the emergence of new wet patterns over Central Asia well as a dry pattern over the north of Peru and Ecuador (**Figure 3D**). Finally, we observe a pronounced dry pattern over Southeast Africa in MAM (**Figure 3F**).

Generally, we note lower event coincidence rates between seasonal wet/dry conditions and CP El Niños as compared to EP periods (compare **Figures 3A,C,E** and **Figures 3B,D,F**). This might be partly explained by the larger of number of 17 CP events as compared to 7 EP events over the considered study period which is consistent with previously reported frequencies of the two flavors (Hendon et al., 2009; Graf and Zanchettin, 2012; Preethi et al., 2015; Wiedermann et al., 2016). In addition, recent studies suggest that CP El Niños might be further discriminated into two subtypes based on their specific impacts on Pacific rainfall and the modes of the Indian Ocean dipole (Wang and Wang, 2013, 2014; Wang et al., 2018). Along those lines, future work should therefore investigate whether the comparatively lower significant event coincidence rates between CP El Niños and strong/weak seasonal precipitation can indeed also be attributed to the presence or absence of any of these two possible subtypes.

## 3.3. Seasonal Wet/Dry Patterns and EP/CP La Niña

Ultimately, we perform the same analysis as above for La Niña periods. For the EP, i.e., canonical, La Niña phases (**Figures 4A,C,E**) we again find various patterns that have already been reported in previous studies. Specifically, during SON coinciding with EP La Niñas (**Figure 4A**), we recapture an increased probability for very wet conditions over Australia and Indonesia (Arblaster et al., 2002) and exceptionally dry conditions in southern Europe (Pozo-Vázquez et al., 2005) and the south of Brazil and Uruguay (Ropelewski and Halpert, 1996). We further observe significant event coincidence rates for dry seasons in the Middle East contrasted by more intense than normal rainfall over central Europe. For DJF, our analysis confirms previous findings of very wet conditions over South Africa and dry episodes over West Africa (Nicholson and Selato, 2000) (**Figure 4C**), the latter having been previously linked to

potential agricultural losses (Karpouzoglou and Barron, 2014) and considerable health risks in that specific area (Rataj et al., 2016). We further observe a prominent seasonal precipitation dipole with dry conditions over Mexico and elevated rainfall over southwestern Canada. The latter has become an important aspect of local water resource management (Lute and Abatzoglou, 2014), but together with increasing air temperature and more frequent storms also poses the threat of landslides in corresponding coastal areas (Guthrie et al., 2010). For MAM seasons associated with EP La Niñas, we observe a tendency towards strong rainfall over the Amazon (Rogers, 1988) and parts of Northern Australia (Arblaster et al., 2002) and the Philippines (**Figure 4E**).

In contrast to the wide-spread spatially coherent wet/dry anomaly patterns observed for EP La Niña periods, we find much fewer spatially extended structures along with CP La Niñas (**Figures 4B,D,F**). Most prominently, we recover previously reported wet conditions over parts of Australia in SON (Arblaster et al., 2002; Cai and Cowan, 2009) (**Figure 4B**). Additionally, we uncover strongly reduced rainfall over Florida in DJF (**Figure 4D**), and over the United Kingdom, Ireland and the west of Kazakhstan in MAM (**Figure 4F**). In summary, we observe that CP La Niñas co-occur with less spatially coherent precipitation responses as compared to their canonical counterparts. We also note that event coincidence rates between La Niña and seasonal wet/dry precipitation signals are quantitatively more similar across EP and CP periods than El Niño (cf. **Figures 4A–F**), which could be partly explained by the comparatively lower difference between the total number of 8 EP and 12 CP La Niña events. In addition, future work should again investigate the potential for further discriminating CP La Niñas into two distinct subtypes in analogy to El Niño (Wang and Wang, 2013, 2014; Wang et al., 2018). Such an approach would possibly allow us to attribute significant event coincidence rates between CP La Niña and strong/weak seasonal precipitation signals to the presence or absence of a certain event type and help explain the different magnitude in significant event coincidence rates observed in **Figure 4**.

## 4. DISCUSSION AND CONCLUSIONS

We have carried out a detailed analysis of ENSO imprints in global patterns of very wet/dry seasons over land. Specifically, we distinguished both, El Niño and La Niña, into two distinct flavors (East Pacific and Central Pacific) by utilizing a classification based on an extensive literature review paired with the assessment of a recently proposed complex network-based index (Wiedermann et al., 2016). From this classification, we obtained event series describing the occurrence-times of the four distinct types of ENSO. Strong seasonal precipitation anomalies have been obtained from the globally gridded GPCC rainfall data set by identifying seasons with precipitation sums above the empirical 80th (below the empirical 20th) percentile of all values for a given grid point as very wet (very dry) periods. This definition follows the spirit of the 3-month aggregate standardized precipitation index (SPI-3), which just provides a monotonic rescaling of seasonal precipitation sums according to the local distributional characteristics of seasonal rainfall. Accordingly, our results can be interpreted in terms of seasonal drought characteristics. Modifying the considered thresholds for defining very wet/dry seasons within reasonable ranges did not qualitatively alter the results reported in this work (see **Supplementary Material**).

We have then employed event coincidence analysis (Donges et al., 2016; Siegmund et al., 2016a) to identify grid points with significant event coincidence rates between different types of ENSO phases and very high or low seasonal precipitation sums. Our analysis confirmed that previously observed interrelationships based on linear correlation or composite analysis in many cases also apply to the timing of events corresponding to the tails of the probability distributions of seasonal precipitation sums. In addition, we identified further patterns of very wet/dry conditions with elevated probabilities alongside different ENSO types, which have to our best knowledge not been described so far. These include increased rainfall over Central Asia during CP El Niños contrasted by rainfall reduction over the same area during CP La Niñas (compare also **Figures 5B,D**) which implies that for such cases standard (linear) statistical methods seem to be insufficient to unveil the underlying interrelations of events. Moreover, our present analysis demonstrates that even though a general linear relationship between ENSO and precipitation might be relatively weak or even absent for a given region, dry periods or very wet seasons can still be possible consequences of the presence of certain ENSO phases. At the same time we observe that several previously reported links between ENSO and global precipitation patterns mostly apply to EP ENSO flavors as far as particularly wet/dry seasons are concerned.

We found that the CP flavor of El Niño (and in parts also that of La Niña) significantly, i.e., non-randomly, co-occurs alongside wet/dry seasonal precipitation signals at lower event coincidence rates which might in parts be explained by the more frequent occurrence (Hendon et al., 2009; Graf and Zanchettin, 2012; Preethi et al., 2015; Wiedermann et al., 2016) compared to its EP counterpart. This finding also suggests that the absolute number of event coincidences [i.e., the numerator in Equation (2)] between EP or CP ENSO periods with wet/dry seasonal precipitation per grid-cell might be of rather comparable size over the period of study, an effect that should be investigated more thoroughly in future research. In addition, recent studies suggest that the CP El Niño can be further discriminated into two types with distinct imprints on precipitation signals especially in the Indo-Pacific area (Wang and Wang, 2013, 2014; Wang et al., 2018). Future work should hence investigate if the comparatively lower significant event coincidence rates between CP El Niño and wet/dry seasonal conditions can potentially be attributed to the presence or absence of a further subtype of CP El Niño. Since El Niño and La Niña show a large degree of symmetry in terms of their frequency and potential for discrimination into different types (Ashok et al., 2007; Kao and Yu, 2009; Hidayat et al., 2018), a similar analysis could also be performed for La Niña as we found lower, yet significant, event coincidence rates for the CP flavor of ENSO's negative phase as well.

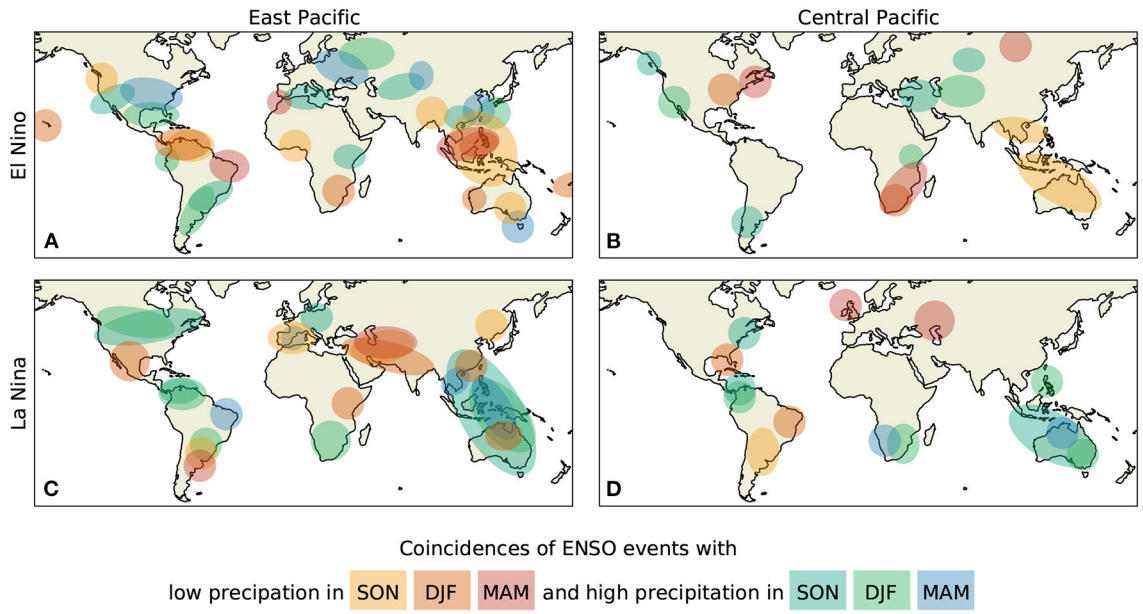



**FIGURE 5 |** Schematic summary of the results presented in this work. Shaded areas indicate major regions in which very high/low seasonal precipitation sums show significant event coincidence rates with EP **(A,C)** or CP **(B,D)** El Niño **(A,B)** or La Niña phases **(C,D)**.

Along those lines, our analysis provides a complementary impact oriented view on the recently raised question whether it is actually (statistically and/or dynamically) meaningful to distinguish La Niña into two types in a similar fashion as El Niño (Kug and Ham, 2011). While some previous studies advocated for such a distinction (Ashok and Yamagata, 2009; Kao and Yu, 2009), others argued that based on correlation analyses between La Niña related SST patterns there is a lack of evidence for the existence of two distinct types (Kug et al., 2009; Ren and Jin, 2011). Contributing to this discussion, Wiedermann et al. (2016) already demonstrated that according to the transitivity index of ENSO's global teleconnections it is indeed meaningful to provide a discrimination of La Niña periods into two flavors in close analogy to El Niño. Complementing recent findings on differential hydrometeorological conditions alongside potentially distinct La Niña flavors (Shinoda et al., 2011; Magee et al., 2017; Song et al., 2017; Hidayat et al., 2018), our results further demonstrate that seasonal wet/dry patterns accompanied by EP La Niñas are generally more likely to arise in a spatially coherent way than such observed along with CP periods (**Figure 4**). The same finding also applies to El Niño periods, which highlights that there exists a certain symmetry between both types of ENSO phases not only in the spatial SST anomaly patterns of El Niño and La Niña themselves, but also with respect to their effects on global precipitation patterns. Thus, from an impact oriented point of view, our work provides further evidence in favor of a distinction between two flavors of La Niña indicated by the presence or absence of very wet/dry regional conditions along with either of the two types of ENSO conditions. In other words, from the viewpoint of event based statistics (and therefore not necessarily in agreement with results based on linear correlations) it appears not only reasonable, but actually relevant to discriminate La Niña into two types in a similar way as common for El Niño periods.

In conclusion, our analysis provides a detailed and global overview on the large-scale differential imprints of different ENSO phases and flavors in the emergence of very wet/dry seasons. All findings reported in this work are ultimately summarized in **Figure 5**, which highlights schematically the main regions where the four different types of ENSO show large scale patterns of significant event coincidence rates with very high/low seasonal precipitation. Especially with respect to the CP flavors we find a variety of regions [such as (northern) Australia or southern Africa] where El Niño and La Niña show opposite impacts in terms of wet/dry seasons (**Figures 5B,D**). In addition, we observe that both flavors of one ENSO phase can also display similar impacts over the same regions, such as reduced precipitation over Australia or south-east Asia for EP *and* CP El Niños (**Figures 5A,B**) or enhanced precipitation over Australia or northern South America for EP *and* CP La Niñas (**Figures 5C,D**). We further uncover unique signatures that are only observed for a single type, e.g., reduced precipitation over northern India and the Middle East for EP (and *not* CP) La Niñas (**Figure 5C**). The general trend toward less significant clusters of large-scale coincidences between ENSO and very wet/dry seasons along with CP periods is observable for both, El Niño and La Niña, as well (c.f. **Figures 5A,B** and **Figures 5C,D**). Our results thus demonstrate that a thorough discrimination of ENSO can be crucial for properly anticipating strong regional and seasonal hydrometeorological anomalies since its differential impacts may not only vary depending on the presence of El Niño and La Niña

conditions but are additionally modulated by their respective EP and CP flavors.

## 5. PERSPECTIVES AND OUTLOOK

Future work should further apply the statistical concepts used in this work to also study ENSO related imprints on other climate variables (e.g., surface air temperature) as well as corresponding effects on socioeconomically and ecologically relevant observables like agricultural yields or water availability. This also implies that if the specific flavor of a developing El Niño or La Niña was to be detected early enough, possible threats like droughts or elevated flood risks, as well as their ecological and socio-economic consequences, could be better anticipated.

Since reliable predictions of EP or CP types of ENSO are scarce, the framework surrounding our present analysis has the potential to serve as an indirect early classification scheme of such periods based on observed globally distributed impacts. Specifically, the observation of early strong/weak climate signals either in SON or the presently ignored boreal summer (June–August, JJA) season of the onset year could be employed to assess the likelihood of approaching an ENSO phase with a specific flavor. Here, ECA could be used to analyze probabilities for specific types of seasonal climate anomalies—that could arise through a variety of conditions and drivers—to be *preceded* by certain ENSO periods and flavors (using so-called *precursor event coincidence rates*; Donges et al., 2016, see section 2.4). Such assessments would complement the estimation of *trigger coincidence rates* (Donges et al., 2016) at which a specific ENSO flavor is *followed* by a certain precipitation response as studied in our present work. The corresponding impact-based classification of ENSO periods could then be used to systematically estimate probabilities of potentially upcoming later (DJF, MAM, or even JJA of the year following the onset of an ENSO period) strong/weak precipitation seasons. In this context, we recall that for the case of instantaneous co-occurrences as studied here, precursor and trigger event coincidence rates solely differ by which of the two series (ENSO or precipitation events) is considered as the reference, i.e., the denominator in Equation (2). Since (sea or air) temperature anomalies are known to exhibit considerable persistence, it may therefore be another relevant extension of the present work to study non-instantaneous (i.e., multi-year) statistical linkages between ENSO episodes and wet/dry patterns, in which case trigger and precursor event coincidence rates become distinct statistical concepts.

Taken together, a corresponding analysis for early classification of ENSO flavors would need to be accompanied by a thorough review of recent approaches for ENSO prediction and attribution, a proper selection of climate variables beyond precipitation that potentially exhibit a mechanistic link with ENSO variability (Di Capua et al., 2019), and an associated estimation of appropriate temporal scales. Ultimately, such a predictive analysis would require a reliable validation of the obtained results by means of dividing the considered study period into training and test intervals, potentially accompanied by extending the analysis at least over the complete twentieth century or even beyond to ensure a sufficient number of events. Even though such an endeavor is clearly beyond the scope of the present work, it remains as a proposal for a potentially important subject of future research with the potential for an early attribution of EP and CP flavors to developing El Niños and La Niñas.

Finally, a more detailed intercomparison between the results obtained from traditional linear and complementary event based statistics could prove useful in assessing which regions are most affected in terms of strong climate responses to the presence of any combination of ENSO phase and flavor. Specifically, and in order to not only assess timing and regional extent of coincidences between ENSO events and precipitation signals, future work should systematically combine ECA with complementary techniques such as composite analysis to also study quantitatively the (relative) changes in magnitude of enhanced/reduced precipitation occurring alongside EP or CP flavours of ENSO. While the current work has solely focused on observational data, it should also be investigated if the reported co-occurrence patterns can be equally observed in historical simulations of state of the art coupled climate models [e.g., those contributing to the Coupled Model Intercomparison Project Phase 6 (CMIP 6)], with any difference pointing to potentially insufficiently represented key processes in the models, thereby contributing to the identification of such processes and, hence, future model improvements. Ultimately, given that the frequency and magnitude of different ENSO phases might be markedly affected by global climate change (Yeh and Kirtman, 2007; Stevenson, 2012; Cai et al., 2014, 2015), it appears additionally useful to further apply the presented framework to future climate projections in order to assess possible changes in the spatial extent and frequency of ENSO related extreme events.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The GPCC gridded precipitation data can be obtained directly from: https://opendata.dwd.de/climate_environment/GPCC/full_data_2018/full_data_monthly_v2018_25.nc.gz.

## AUTHOR CONTRIBUTIONS

All authors designed the study. MW and JS analyzed the data and wrote the manuscript with input from all authors. JD assisted with the analysis. RD supervised the study. MW, JD, and RD substantively revised the work.

## FUNDING

MW, JS, and RD acknowledge funding by the German Federal Ministry for Education and Research via the BMBF projects CoSy-CC$^2$ (grant no. 01LN1306A), GOTHAM (grant no. 01LP16MA), and ROADMAP (grant no. 01LP2002B). JD has been supported by the Stordalen Foundation. MW and JD thank the Leibniz Society (project DOMINOES) for financial support. RD acknowledges the IRTG 1740 funded by

DFG and FAPESP. JS has been supported by the Evangelisches Studienwerk Villigst e.V.

## ACKNOWLEDGMENTS

This paper was developed within the scope of the IRTG 1740/TRP 2015/50122-0, funded by the DFG/FAPESP. The authors gratefully acknowledge the European Regional Development Fund (ERDF), the German Federal Ministry of Education and Research and the Land Brandenburg for supporting this project by providing resources on the high performance computer system at the Potsdam Institute for Climate Impact Research. This manuscript has been released as a pre-print at http://arxiv.org/abs/1702.00218 (Wiedermann et al., 2020). The content of this manuscript has been partially published as part of the dissertation of Wiedermann (2018).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fclim.2021.618548/full#supplementary-material

## REFERENCES

Allan, R. P., and Soden, B. J. (2008). Atmospheric warming and the amplification of precipitation extremes. *Science* 321, 1481–1484. doi: 10.1126/science.1160787

Antoniou, I. E., and Tsompa, E. T. (2008). Statistical analysis of weighted networks. *Discrete Dyn. Nat. Soc*. 2008:375452. doi: 10.1155/2008/375452

Arblaster, J., Meehl, G., and Moore, A. (2002). Interdecadal modulation of Australian rainfall. *Clim. Dyn*. 18, 519–531. doi: 10.1007/s00382-001-0191-y

Ashok, K., Behera, S. K., Rao, S. A., Weng, H., and Yamagata, T. (2007). El Ni no Modoki and its possible teleconnection. *J. Geophys. Res*. 112:C11007. doi: 10.1029/2006JC003798

Ashok, K., and Yamagata, T. (2009). Climate change: the El Ni no with a difference. *Nature* 461, 481–484. doi: 10.1038/461481a

Boers, N., Bookhagen, B., Barbosa, H., Marwan, N., Kurths, J., and Marengo, J. (2014). Prediction of extreme floods in the eastern Central Andes based on a complex networks approach. *Nat. Commun*. 5:5199. doi: 10.1038/ncomms6199

Bookhagen, B., and Strecker, M. R. (2012). Spatiotemporal trends in erosion rates across a pronounced rainfall gradient: examples from the southern Central Andes. *Earth Planet. Sci. Lett*. 327, 97–110. doi: 10.1016/j.epsl.2012.02.005

Cai, W., Borlace, S., Lengaigne, M., van Rensch, P., Collins, M., Vecchi, G., et al. (2014). Increasing frequency of extreme El Ni no events due to greenhouse warming. *Nat. Clim. Change* 4, 111–116. doi: 10.1038/nclimate2100

Cai, W., and Cowan, T. (2009). La Ni na Modoki impacts Australia autumn rainfall variability. *Geophys. Res. Lett*. 36:L12805. doi: 10.1029/2009GL037885

Cai, W., Wang, G., Santoso, A., McPhaden, M. J., Wu, L., Jin, F.-F., et al. (2015). Increased frequency of extreme La Ni na events under greenhouse warming. *Nat. Clim. Change* 5, 132–137. doi: 10.1038/nclimate2492

Camberlin, P., Janicot, S., and Poccard, I. (2001). Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical sea-surface temperature: Atlantic vs. ENSO. *Int. J. Climatol.* 21, 973–1005. doi: 10.1002/joc.673

Chatterton, N. J., Goodin, J. R., McKell, C. M., Parker, R. V., and Rible, J. M. (1971). Monthly variation in the chemical composition of desert saltbush. *Rangeland Ecol. Manag*. 24, 37–40. doi: 10.2307/3896062

Chen, D., Lian, T., Fu, C., Cane, M. A., Tang, Y., Murtugudde, R., et al. (2015). Strong influence of westerly wind bursts on El Ni no diversity. *Nat. Geosci*. 8, 339–345. doi: 10.1038/ngeo2399

Chiew, F. H. S., Piechota, T. C., Dracup, J. A., and McMahon, T. A. (1998). El Ni no/Southern Oscillation and Australian rainfall, streamflow and drought: links and potential for forecasting. *J. Hydrol*. 204, 138–149. doi: 10.1016/S0022-1694(97)00121-2

Dai, A., and Wigley, T. M. L. (2000). Global patterns of ENSO-induced precipitation. *Geophys. Res. Lett*. 27, 1283–1286. doi: 10.1029/1999GL011140

Di Capua, G., Kretschmer, M., Runge, J., Alessandri, A., Donner, R. V., van den Hurk, B., et al. (2019). Long-lead statistical forecasts of the indian summer monsoon rainfall based on causal precursors. *Weath. Forecast*. 34, 1377–1394. doi: 10.1175/WAF-D-19-0002.1

Diaz, H. F., Hoerling, M. P., and Eischeid, J. K. (2001). ENSO variability, teleconnections and climate change. *Int. J. Climatol.* 21, 1845–1862. doi: 10.1002/joc.631

Domeisen, D. I., Garfinkel, C. I., and Butler, A. H. (2019). The teleconnection of El Ni no Southern Oscillation to the stratosphere. *Rev. Geophys*. 57, 5–47. doi: 10.1029/2018RG000596

Donges, J. F., Donner, R. V., Trauth, M. H., Marwan, N., Schellnhuber, H.-J., and Kurths, J. (2011). Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution. *Proc. Natl. Acad. Sci. U.S.A*. 108, 20422–20427. doi: 10.1073/pnas.1117052108

Donges, J. F., Schleussner, C.-F., Siegmund, J. F., and Donner, R. V. (2016). Event coincidence analysis for quantifying statistical interrelationships between event time series: on the role of flood events as triggers of epidemic outbreaks. *Eur. Phys. J.-Spec. Top*. 225, 471–487. doi: 10.1140/epjst/e2015-50233-y

Donges, J. F., Zou, Y., Marwan, N., and Kurths, J. (2009). Complex networks in climate dynamics. *Eur. Phys. J.-Spec. Top*. 174, 157–179. doi: 10.1140/epjst/e2009-01098-2

Donner, R. V., Wiedermann, M., and Donges, J. F. (2017). "Complex network techniques for climatological data analysis," in *Nonlinear and Stochastic Climate Dynamics*, eds C. L. Franzke and T. J. O'Kane (Cambridge University Press), 159–183. doi: 10.1017/9781316339251.007

Easterling, D. R., Evans, J., Groisman, P. Y., Karl, T., Kunkel, K. E., and Ambenje, P. (2000). Observed variability and trends in extreme climate events: a brief review. *Bull. Am. Meteor. Soc*. 81:417. doi: 10.1175/1520-0477(2000)081<0417:OVATIE>2.3.CO;2

Feng, J., and Chen, W. (2014). Influence of the IOD on the relationship between El Ni no Modoki and the East Asian-western North Pacific summer monsoon. *Int. J. Climatol*. 34, 1729–1736. doi: 10.1002/joc.3790

Freund, M. B., Henley, B. J., Karoly, D. J., McGregor, H. V., Abram, N. J., and Dommenget, D. (2019). Higher frequency of Central Pacific El Ni no events in recent decades relative to past centuries. *Nat. Geosci*. 12, 450–455. doi: 10.1038/s41561-019-0353-3

Graf, H.-F., and Zanchettin, D. (2012). Central Pacific El Ni no, the subtropical bridge, and Eurasian climate. *J. Geophys. Res*. 117:D01102. doi: 10.1029/2011JD016493

Guthrie, R. H., Mitchell, S. J., Lanquaye-Opoku, N., and Evans, S. G. (2010). Extreme weather and landslide initiation in coastal British Columbia. *Quart. J. Eng. Geol. Hydrogeol*. 43, 417–428. doi: 10.1144/1470-9236/08-119

Guttman, N. B. (1999). Accepting the standardized precipitation index: a calculation algorithm 1. *J. Am. Water Res. Assoc*. 35, 311–322. doi: 10.1111/j.1752-1688.1999.tb03592.x

Hendon, H. H., Lim, E., Wang, G., Alves, O., and Hudson, D. (2009). Prospects for predicting two flavors of El Ni no. *Geophys. Res. Lett*. 36:L19713. doi: 10.1029/2009GL040100

Hidayat, R., Juniarti, M., and Maŕufah, U. (2018). Impact of La Ni na and La Ni na Modoki on Indonesia rainfall variability. *IOP Conf. Ser. Earth Environ. Sci*. 149:012046. doi: 10.1088/1755-1315/149/1/012046

Hoell, A., Funk, C., Magadzire, T., Zinke, J., and Husak, G. (2014). El Ni no-Southern Oscillation diversity and Southern Africa teleconnections during Austral Summer. *Clim. Dyn*. 45, 1583–1599. doi: 10.1007/s00382-014-2414-z

Houston, J. (2006). Variability of precipitation in the Atacama Desert: its causes and hydrological impact. Int. *J. Climatol*. 26, 2181–2198. doi: 10.1002/joc.1359

Hu, Z.-Z., Kumar, A., Jha, B., Wang, W., Huang, B., and Huang, B. (2011). An analysis of warm pool and cold tongue El Ni nos: air-sea coupling

processes, global influences, and recent trends. *Clim. Dyn.* 38, 2017–2035. doi: 10.1007/s00382-011-1224-9

Huang, B., L'Heureux, M., Hu, Z.-Z., and Zhang, H.-M. (2016). Ranking the strongest ENSO events while incorporating SST uncertainty. *Geophys. Res. Lett.* 43, 9165–9172. doi: 10.1002/2016GL070888

Ide, T., Brzoska, M., Donges, J. F., and Schleussner, C.-F. (2020). Multi-method evidence for when and how climate-related disasters contribute to armed conflict risk. *Glob. Environ. Chang.* 62:102063. doi: 10.1016/j.gloenvcha.2020.102063

Kane, R. (1997). Prediction of droughts in north-east Brazil: Role of ENSO and use of periodicities. *Int. J. Climatol.* 17, 655–665. doi: 10.1002/(SICI)1097-0088(199705)17:6<655::AID-JOC144>3.0.CO;2-1

Kao, H.-Y., and Yu, J.-Y. (2009). Contrasting Eastern-Pacific and Central-Pacific types of ENSO. *J. Clim.* 22, 615–632. doi: 10.1175/2008JCLI2309.1

Karl, T. R., and Trenberth, K. E. (2003). Modern global climate change. *Science* 302, 1719–1723. doi: 10.1126/science.1090228

Karpouzoglou, T., and Barron, J. (2014). A global and regional perspective of rainwater harvesting in sub-Saharan Africa's rainfed farming. *Phys. Chem. Earth* 72–75, 43–53. doi: 10.1016/j.pce.2014.09.009

Kim, W., Yeh, S.-W., Kim, J.-H., Kug, J.-S., and Kwon, M. (2011). The unique 2009-2010 El Ni no event: a fast phase transition of warm pool El Ni no to La Ni na. *Geophys. Res. Lett.* 38:L15809. doi: 10.1029/2011GL048521

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., et al. (2001). The NCEP-NCAR 50-Year reanalysis: monthly means CD-ROM and documentation. *Bull. Am. Meteor. Soc.* 82, 247–267. doi: 10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2

Kug, J.-S., and Ham, Y.-G. (2011). Are there two types of La Nina? *Geophys. Res. Lett.* 38:L16704. doi: 10.1029/2011GL048237

Kug, J.-S., Jin, F.-F., and An, S.-I. (2009). Two types of El Ni no events: Cold Tongue El Ni no and Warm Pool El Ni no. *J. Clim.* 22, 1499–1515. doi: 10.1175/2008JCLI2624.1

Leblanc, M., Tweed, S., Van Dijk, A., and Timbal, B. (2012). A review of historic and future hydrological changes in the Murray-Darling Basin. *Glob. Planet. Chang.* 80, 226–246. doi: 10.1016/j.gloplacha.2011.10.012

Levine, A., Jin, F. F., and McPhaden, M. J. (2016). Extreme noise-extreme El Ni no: How state-dependent noise forcing creates El Ni no-La Ni na asymmetry. *J. Clim.* 29, 5483–5499. doi: 10.1175/JCLI-D-16-0091.1

Lewis, S. L., Brando, P. M., Phillips, O. L., Heijden, G. M. F. v. d., and Nepstad, D. (2011). The 2010 Amazon drought. *Science* 331, 554–554. doi: 10.1126/science.1200807

Lorenz, C., and Kunstmann, H. (2012). The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *J. Hydrometeorol.* 13, 1397–1420. doi: 10.1175/JHM-D-11-088.1

Lute, A. C., and Abatzoglou, J. T. (2014). Role of extreme snowfall events in interannual variability of snowfall accumulation in the western United States. *Water Resour. Res.* 50, 2874–2888. doi: 10.1002/2013WR014465

Magee, A. D., Verdon-Kidd, D. C., Diamond, H. J., and Kiem, A. S. (2017). Influence of ENSO, ENSO Modoki, and the IPO on tropical cyclogenesis: a spatial analysis of the southwest Pacific region. *Int. J. Climatol.* 37, 1118–1137. doi: 10.1002/joc.5070

Malik, N., Marwan, N., and Kurths, J. (2010). Spatial structures and directionalities in monsoonal precipitation over south Asia. *Nonlin. Proc. Geophys.* 17:371. doi: 10.5194/npg-17-371-2010

Marathe, S., Ashok, K., Swapna, P., and Sabin, T. (2015). Revisiting El Ni no Modokis. *Clim. Dyn.* 45, 3527–3545. doi: 10.1007/s00382-015-2555-8

Neelin, J. D., Chou, C., and Su, H. (2003). Tropical drought regions in global warming and El Ni no teleconnections. *Geophys. Res. Lett.* 30:2275. doi: 10.1029/2003GL018625

Newman, M. (2003). The structure and function of complex networks. *SIAM Rev.* 45, 167–256. doi: 10.1137/S003614450342480

Nicholson, S., and Selato, J. (2000). The influence of La Ni na on African rainfall. *Int. J. Climatol.* 20, 1761–1776. doi: 10.1002/1097-0088(20001130)20:14<1761::AID-JOC580>3.0.CO;2-W

Phillips, O. L., Arag ao, L. E. O. C., Lewis, S. L., Fisher, J. B., Lloyd, J., López-González, G., et al. (2009). Drought sensitivity of the Amazon rainforest. *Science* 323, 1344–1347. doi: 10.1126/science.1164033

Pozo-Vázquez, D., Gámiz-Fortis, S. R., Tovar-Pescador, J., Esteban-Parra, M. J., and Castro-Díez, Y. (2005). El Ni no-Southern Oscillation events and associated European winter precipitation anomalies. *Int. J. Climatol.* 25, 17–31. doi: 10.1002/joc.1097

Preethi, B., Sabin, T., Adedoyin, J., and Ashok, K. (2015). Impacts of the ENSO Modoki and other tropical Indo-Pacific climate-drivers on African rainfall. *Sci. Rep.* 5:16653. doi: 10.1038/srep16653

Quian Quiroga, R., Kreuz, T., and Grassberger, P. (2002). Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Phys. Rev. E* 66:041904. doi: 10.1103/PhysRevE.66.041904

Radebach, A., Donner, R. V., Runge, J., Donges, J. F., and Kurths, J. (2013). Disentangling different types of El Ni no episodes by evolving climate network analysis. *Phys. Rev. E* 88:052807. doi: 10.1103/PhysRevE.88.052807

Rammig, A., Wiedermann, M., Donges, J. F., Babst, F., von Bloh, W., Frank, D., et al. (2015). Coincidences of climate extremes and anomalous vegetation responses: comparing tree ring patterns to simulated productivity. *Biogeosciences* 12, 373–385. doi: 10.5194/bg-12-373-2015

Rataj, E., Kunzweiler, K., and Garthus-Niegel, S. (2016). Extreme weather events in developing countries and related injuries and mental health disorders-a systematic review. *BMC Public Health* 16:1020. doi: 10.1186/s12889-016-3692-7

Ratnam, J., Behera, S., Masumoto, Y., and Yamagata, T. (2014). Remote effects of El Ni no and Modoki events on the austral summer precipitation of southern Africa. *J. Clim.* 27, 3802–3815. doi: 10.1175/JCLI-D-13-00431.1

Ren, H.-L., and Jin, F.-F. (2011). Ni no indices for two types of ENSO. *Geophys. Res. Lett.* 38:L04704. doi: 10.1029/2010GL046031

Rogers, J. C. (1988). Precipitation variability over the Caribbean and Tropical Americas associated with the Southern Oscillation. *J. Clim.* 1, 172–182. doi: 10.1175/1520-0442(1988)001<0172:PVOTCA>2.0.CO;2

Ropelewski, C. F., and Halpert, M. S. (1986). North American precipitation and temperature patterns associated with the El Ni no/Southern Oscillation (ENSO). *Mon. Weather Rev.* 114, 2352–2362. doi: 10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2

Ropelewski, C. F., and Halpert, M. S. (1987). Global and regional scale precipitation patterns associated with the El Ni no/Southern Oscillation. *Mon. Weather Rev.* 115, 1606–1626. doi: 10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2

Ropelewski, C. F., and Halpert, M. S. (1996). Quantifying Southern Oscillation-Precipitation relationships. *J. Clim.* 9, 1043–1059. doi: 10.1175/1520-0442(1996)009<1043:QSOPR>2.0.CO;2

Rudolf, B., Hauschild, H., Rueth, W., and Schneider, U. (1994). "Terrestrial precipitation analysis: Operational method and required density of point measurements," in *Global Precipitations and Climate Change*, Number 26 in NATO ASI Series, eds M. Desbois and F. Désalmand (Berlin; Heidelberg: Springer), 173–186. doi: 10.1007/978-3-642-79268-7_10

Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K., and Kertész, J. (2007). Generalizations of the clustering coefficient to weighted complex networks. *Phys. Rev. E* 75:027105. doi: 10.1103/PhysRevE.75.027105

Schleussner, C.-F., Donges, J. F., Donner, R. V., and Schellnhuber, H. J. (2016). Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9216–9221. doi: 10.1073/pnas.1601611113

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., and Ziese, M. (2018). *GPCC Full Data Monthly Version 2018 at 2.5$^{circ}$: Monthly Land-Surface Precipitation from Rain-Gauges Built on GTS-Based and Historic Data: Gridded Monthly Totals*. Global Precipitation Climatology Centre (GPCC) at Deutscher Wetterdienst.

Shaman, J., and Tziperman, E. (2010). An atmospheric teleconnection linking ENSO and southwestern European precipitation. *J. Clim.* 24, 124–139. doi: 10.1175/2010JCLI3590.1

Shinoda, T., Hurlburt, H. E., and Metzger, E. J. (2011). Anomalous tropical ocean circulation associated with La Ni na Modoki. *J. Geophys. Res.-Oceans* 116:C12001. doi: 10.1029/2011JC007304

Siegmund, J. F., Sanders, T. G. M., Heinrich, I., van der Maaten, E., Simard, S., Helle, G., et al. (2016a). Meteorological drivers of extremes in daily stem radius variations of Beech, Oak, and Pine in Northeastern Germany: an event coincidence analysis. *Front. Plant. Sci.* 7:733. doi: 10.3389/fpls.2016.00733

Siegmund, J. F., Wiedermann, M., Donges, J., and Donner, R. V. (2016b). Impact of temperature and precipitation extremes on the flowering dates of four German wildlife shrub species. *Biogeosciences* 13, 5541–5555. doi: 10.5194/bg-13-5541-2016

Song, L., Chen, S., Chen, W., and Chen, X. (2017). Distinct impacts of two types of La Ni na events on Australian summer rainfall. *Int. J. Climatol.* 37, 2532–2544. doi: 10.1002/joc.4863

Stevenson, S. (2012). Significant changes to ENSO strength and impacts in the twenty-first century: results from CMIP5. *Geophys. Res. Lett.* 39:L17703. doi: 10.1029/2012GL052759

Svoboda, M., Hayes, M., and Wood, D. (2012). *Standardized Precipitation Index User Guide*. World Meteorological Organization Geneva.

Taschetto, A. S., and England, M. H. (2009). El Ni no Modoki impacts on Australian rainfall. *J. Clim.* 22, 3167–3174. doi: 10.1175/2008JCLI2589.1

Tedeschi, R. G., Cavalcanti, I. F., and Grimm, A. M. (2013). Influences of two types of ENSO on South American precipitation. *Int. J. Climatol.* 33, 1382–1400. doi: 10.1002/joc.3519

Thomas, N., and Nigam, S. (2018). Twentieth-century climate change over Africa: seasonal hydroclimate trends and sahara desert expansion. *J. Clim.* 31, 3349–3370. doi: 10.1175/JCLI-D-17-0187.1

Trenberth, K. E. (1997). The definition of El Ni no. *Bull. Am. Meteor. Soc.* 78, 2771–2777. doi: 10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2

Tsonis, A. A., Swanson, K. L., and Roebber, P. J. (2006). What do networks have to do with climate? *Bull. Amer. Meteor. Soc.* 87, 585–595. doi: 10.1175/BAMS-87-5-585

van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y., Podger, G. M., et al. (2013). The millennium drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resour. Res.* 49, 1040–1057. doi: 10.1002/wrcr.20123

Wang, C., and Wang, X. (2013). Classifying El Ni no Modoki I and II by different impacts on rainfall in southern China and typhoon tracks. *J. Climate* 26, 1322–1338. doi: 10.1175/JCLI-D-12-00107.1

Wang, X., Tan, W., and Wang, C. (2018). A new index for identifying different types of El Ni no Modoki events. *Clim. Dyn.* 50, 2753–2765. doi: 10.1007/s00382-017-3769-8

Wang, X., and Wang, C. (2014). Different impacts of various El Ni no events on the Indian Ocean Dipole. *Clim. Dyn.* 42, 991–1005. doi: 10.1007/s00382-013-1711-2

Weng, H., Wu, G., Liu, Y., Behera, S. K., and Yamagata, T. (2011). Anomalous summer climate in China influenced by the tropical Indo-Pacific Oceans. *Clim. Dyn.* 36, 769–782. doi: 10.1007/s00382-009-0658-9

Wiedermann, M. (2018). *Classification of complex networks in spatial, topological and information theoretic domains* (Ph.D. thesis). Humboldt-Universität zu Berlin, Berlin, Germany.

Wiedermann, M., Radebach, A., Donges, J. F., Kurths, J., and Donner, R. V. (2016). A climate network-based index to discriminate different types of El Ni no and La Ni na. *Geophys. Res. Lett.* 43, 7176–7185. doi: 10.1002/2016GL069119

Wiedermann, M., Siegmund, J. F., Donges, J. F., and Donner, R. V. (2020). Differential imprints of distinct enso flavors in global patterns of very low and high seasonal precipitation. *arXiv [Preprint]. 1702.00218*.

Wolf, F., Bauer, J., Boers, N., and Donner, R. V. (2020). Event synchrony measures for functional climate network analysis: a case study on South American rainfall dynamics. *Chaos* 30:033102. doi: 10.1063/1.5134012

Yeh, S.-W., and Kirtman, B. P. (2007). ENSO amplitude changes due to climate change projections in different coupled models. *J. Clim.* 20, 203–217. doi: 10.1175/JCLI4001.1

Yuan, Y. and Yan, H. (2013). Different types of La Ni na events and different responses of the tropical atmosphere. *Chin. Sci. Bull.* 58, 406–415. doi: 10.1007/s11434-012-5423-5

Zhang, W., Wang, L., Xiang, B., Qi, L., and He, J. (2015). Impacts of two types of La Ni na on the NAO during boreal winter. *Clim. Dyn.* 44, 1351–1366. doi: 10.1007/s00382-014-2155-z

**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 Climate

ORIGINAL RESEARCH
published: 09 April 2021
doi: 10.3389/fclim.2021.656505

# A Physics-Aware Neural Network Approach for Flow Data Reconstruction From Satellite Observations

*Luca Schweri[1]*, Sebastien Foucher[1], Jingwei Tang[1], Vinicius C. Azevedo[1], Tobias Günther[2] and Barbara Solenthaler[1]*

[1] Department of Computer Science, ETH Zurich, Zurich, Switzerland, [2] Department of Computer Science, Friedrich-Alexander-University of Erlangen-Nuremberg, Erlangen, Germany

An accurate assessment of physical transport requires high-resolution and high-quality velocity information. In satellite-based wind retrievals, the accuracy is impaired due to noise while the maximal observable resolution is bounded by the sensors. The reconstruction of a continuous velocity field is important to assess transport characteristics and it is very challenging. A major difficulty is ambiguity, since the lack of visible clouds results in missing information and multiple velocity fields will explain the same sparse observations. It is, therefore, necessary to regularize the reconstruction, which would typically be done by hand-crafting priors on the smoothness of the signal or on the divergence of the resulting flow. However, the regularizers can smooth the solution excessively and will not guarantee that possible solutions are truly physically realizable. In this paper, we demonstrate that data recovery can be learned by a neural network from numerical simulations of physically realizable fluid flows, which can be seen as a data-driven regularization. We show that the learning-based reconstruction is especially powerful in handling large areas of missing or occluded data, outperforming traditional models for data recovery. We quantitatively evaluate our method on numerically-simulated flows, and additionally apply it to a Guadalupe Island case study—a real-world flow data set retrieved from satellite imagery of stratocumulus clouds.

**Keywords: deep learning–CNN, Karman vortex street, cloud motion winds, satellite wind data, wind velocity retrieval**

OPEN ACCESS

***Edited by:***
*Matthew Collins,*
*University of Exeter, United Kingdom*

***Reviewed by:***
*Subimal Ghosh,*
*Indian Institute of Technology Bombay, India*
*Juliana Anochi,*
*National Institute of Space Research (INPE), Brazil*

****Correspondence:***
*Luca Schweri*
*luca.schweri@alumni.ethz.ch*

***Specialty section:***
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

***Received:*** *20 January 2021*
***Accepted:*** *11 March 2021*
***Published:*** *09 April 2021*

***Citation:***
*Schweri L, Foucher S, Tang J, Azevedo VC, Günther T and Solenthaler B (2021) A Physics-Aware Neural Network Approach for Flow Data Reconstruction From Satellite Observations. Front. Clim. 3:656505. doi: 10.3389/fclim.2021.656505*

## 1. INTRODUCTION

The formation of observable mesoscale vortex patterns on satellite imagery is driven by atmospheric processes. Certain structures, such as Karman vortex streets forming in the stratocumulus-topped wake of a mountainous islands, bear resemblance to patterns observable in laboratory flows. Such flow structures have been studied based on satellite measurements since the 1960s (Hubert and Krueger, 1962; Chopra and Hubert, 1965; Young and Zawislak, 2006). Recent advances in remote sensing (Geerts et al., 2018) enabled the retrieval of high-resolution wind fields at kilometer-scale (Horváth et al., 2017, 2020), which is a necessary requirement for the analysis of atmospheric processes in turbulent environments. While operational wind products based on the Advanced Baseline Imager (ABI) onboard the Geostationary Operational Enrivonmental Satellite-R (GEOS-R) (Schmit et al., 2017) provide a 7.5 km resolution, Horváth et al. (2020)

utilized the internal wind vectors at 2.5 km resolution in their study of vortex patterns in the wake of Guadalupe Island off Baja California on 9 May 2018, which were combined with MODIS-GEOS wind products offering stereo cloud-top heights and semi-independent wind validation data (Carr et al., 2019). Utilizing such high resolutions is necessary for the analysis of small-scale structures, but comes at the price of an increased level of measurement noise, which is accompanied by general uncertainty in regions without or with only few clouds. Disambiguation of possible flow configurations based on imperfect measurement data is thereby challenging. At present, such spaceborne measurements have been cleaned with median filters, smoothing, and thresholding of unreasonably large velocity components (Horváth et al., 2020). Thus, current approaches to reconstruct missing or uncertain information are based on assumptions about the signal smoothness, and do not yet incorporate that the wind fields are the result of fluid dynamical processes. Such physical regularizations, however, are more difficult to model.

Data-driven regularization and data recovery with neural networks offer great potential for data completion tasks. A main challenge is to include physics knowledge in the network design, such that the reconstruction follows the solution of the fluid dynamic equations. Cloud satellite imagery may contain large areas of missing or uncertain data, and high noise levels. Further, the fluid dynamics are affected by the topography. Therefore, the data inference must be powerful enough to generate accurate results even in such challenging settings. Neural networks have excellent properties for data completion: they are universal approximators, and are able to efficiently combine data-driven and physics-based regularizations. Data reconstructed with neural networks, however, typically lack details due to the convolution operations, which manifests as degraded and smoothed flow details. In this paper, we present a novel neural network architecture for 2-D velocity fields extracted from satellite imagery. Our method uses physically-inspired regularization, leveraging surrogate simulations that are generated in the full three dimensional space, which enables a more precise approximation of the transport phenomena. The inference time of our approach is fast, offering new applications for predicting large scale flows in meteorological settings.

Deep learning approaches are currently studied with great interest in climate science in a number of different topics, including convection (O'Gorman and Dwyer, 2018), forecasting (Weyn et al., 2019), microphysics (Seifert and Rasp, 2020), empirical-statistical downscaling (Baño-Medina et al., 2020), and radiative transfer (Min et al., 2020). Estimating missing flow field data has many similarities with the image inpainting task commonly studied in computer vision, as it is essentially a scene completion process using partial observations. The recent success of learning-based image inpainting algorithms demonstrates the capability of deep neural networks to complete large missing regions in natural images in a plausible fashion. Pathak et al. (2016) used Context Encoders as one of the first attempts for filling missing image data with a deep convolutional neural network (CNN). CNN-based methods are attractive due to their ability to reconstruct complex functions with only few sparse samples while being highly efficient. The follow-up work by Iizuka et al. (2017) proposed a fully convolutional network to complete rectangular missing data regions. The approach, however, still relied on Poisson image blending as a post-processing step. Yu et al. (2018) introduced contextual attention layers to model long-range dependencies in images and a refinement network for post-processing, enabling end-to-end training. Zeng et al. (2019) extended previous work by extracting context attention maps in different layers of the encoder and skip connect attention maps to the decoder. These approaches all include adversarial losses computed from a discriminator (Goodfellow et al., 2014) in order to better reconstruct visually appealing high frequency details. However, high frequency details from adversarial losses can result in mismatches from ground truth data (Huang et al., 2017), which can potentially predict missing data that diverge from physical laws. Liu et al. (2018) designed partial convolution operations for image inpainting, so that the prediction of the missing pixels is only conditioned on the valid pixels in the original image. The operation enables high quality inpainting results without adversarial loss. Inpainting approaches have also been successfully used for scene completion and view path planning using data from sparse input views. Song et al. (2017) used an end-to-end network SSCNet for scene completion and Guo and Tong (2018) a view-volume CNN that extracts geometric features from 2D depth images. Zhang and Funkhouser (2018) presented an end-to-end architecture for depth inpainting, and Han et al. (2019) used multi-view depth completion to predict point cloud representations. A 3D recurrent network has been used to integrate information from only a few input views (Choy et al., 2016), and Xu et al. (2016) used spatial and temporal structure of sequential observations to predict a view sequence. We base our method on previous deep learning architectures for image inpainting, namely a U-Net (Ronneberger et al., 2015) with partial convolutions (Liu et al., 2018). Utilizing that we are dealing with velocity fields instead of images, we further include a physically-inspired loss function to better regularize the predicted flow data. Neural networks have also recently been applied to fluid simulations. Applications include prediction of the entire dynamics (Wiewel et al., 2019), reconstruction of simulations from a set of input parameters (Kim et al., 2019b), interactive shape design (Umetani and Bickel, 2018), inferring hidden physics quantities (Raissi et al., 2018), and artistic control for visual effects (Kim et al., 2019a). A comprehensive overview of machine learning for fluid dynamics can be found in Brunton et al. (2020).

## 2. METHOD

In this paper, we demonstrate that the reconstruction of high-quality wind fields from noisy, uncertain and incomplete satellite-based wind retrievals based on GEOS-R measurements (Schmit et al., 2017) can be learned by a neural network from numerical simulations of realizable fluid flows. The training and evaluation of such a supervised approach is, however, challenging due to the lack of ground

truth data. Thus, in a first step we numerically simulate fluid flows, which we synthetically modify to account for uncertainty and lack of measurements. The flow configurations must thereby be chosen carefully to sample the space of possible fluid configurations uniformly. Simply using off-the-shelf reanalysis simulations (Hersbach et al., 2020) would bias the network to perform best on very common fluid flow configurations, leading to poor results in the exceptional situations that are most interesting to study. For this reason, we generate fluid flow configurations in a controlled manner suitable for supervised machine learning. Afterwards, we analyze the performance of our model on satellite wind retrievals in the wake of Guadalupe island by Horváth et al. (2020) and compare those with linear reconstructions obtained via least-squares minimization. Our neural network has similarities with image inpainting approaches, mainly stemming from both sharing a scene completion process using partial observations. The major difference between flow field and image inpainting is that flow field data inherently follows the solution of the fluid dynamic equations. Hence, existing image inpainting algorithms can easily fail in physics-aware completion tasks as they never aim to capture the underlying physical laws. The presented flow inpainting method therefore considers the mathematical equations that model the fluid phenomena in the design of the network architecture and loss functions. The network is designed such that large areas of missing data with and without obstacles can be inferred. In the next sections, we are going to detail how we designed our network architecture, along with challenges and necessary modifications that were made to support fluid flow data.

## 2.1. Network Architecture

Our goal is to train a network that can fill empty regions of velocity fields. The input scheme is similar to standard image inpainting tasks. For a given 2D velocity field $\vec{u}_{in}$ with missing fluid regions represented by a binary mask $\mathbf{M}$ (0 for empty and 1 for known regions), the network predicts the inpainted velocity field $\hat{\vec{u}}$. Existing consistency checks, such as wind speeds not exceeding 8 m/s can be incorporated directly by the mask. The network for fluid data completion consists of three main parts: an encoder, dense blocks, and a decoder. The encoder-decoder pair follows a U-Net structure (Ronneberger et al., 2015) shown in **Figure 1**. It first encodes the original velocity field by reducing the spatial resolution progressively, later decoding it by increasing the resolution until it matches the original size. This way, features are extracted on all scales. The U-Net includes skip connections that forward these scale-dependent features from the encoding phase to the decoding phase in order retain the locality of high-frequent information. To improve the quality of the results obtained by the U-Net further, we add Dense Blocks (Huang et al., 2017) at the bottleneck to enrich the feature representation.

Each layer of the network is implemented by replacing the standard convolution operations with modified partial

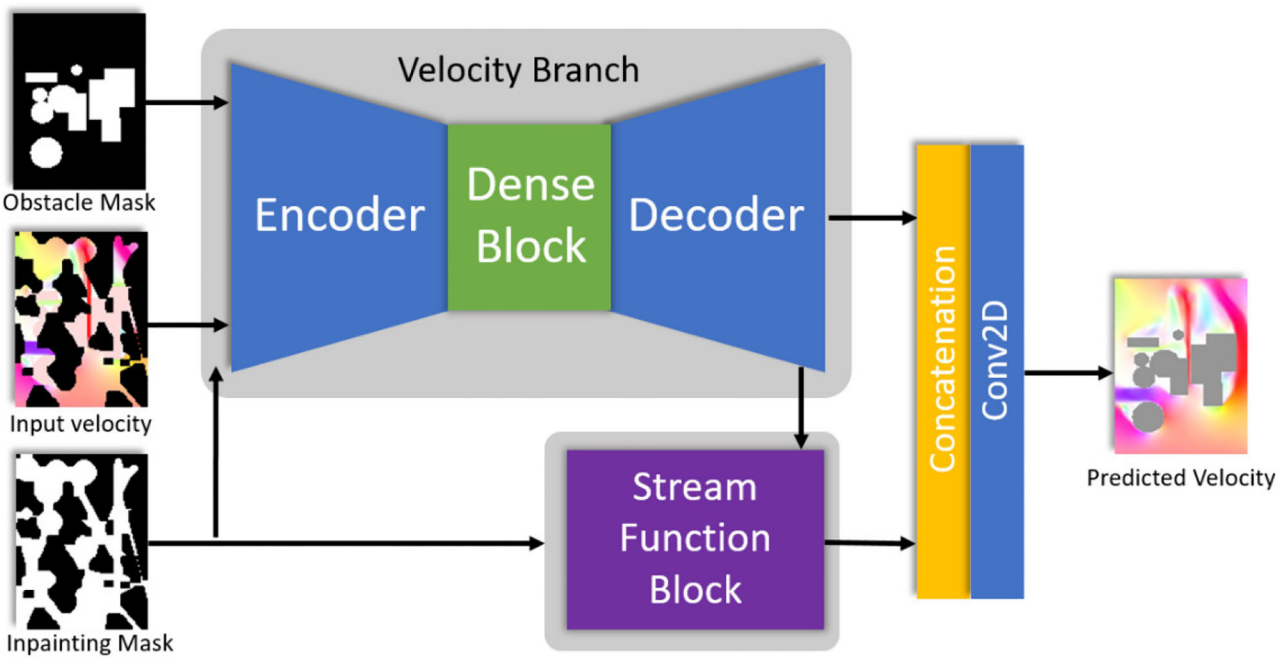



**FIGURE 1 |** Our network architecture uses a U-Net with Dense Blocks to predict velocity fields. A stream function block implemented through another Dense Block can be added for incompressible flow data. Our synthetic data experiments use the stream function block, while the results of the Guadalupe island do not, since they are slices of 3-D incompressible flows, and thus, not divergent free.

convolutions (Liu et al., 2018). The modified partial convolution at every location is defined as

$$x' = \begin{cases} \mathbf{C}^T(\mathbf{X}\cdot\mathbf{M})\frac{sum(\mathbf{1})}{sum(\mathbf{M})} + b, & \text{if } sum(\mathbf{M}) > 1 \\ 0, & \text{otherwise} \end{cases}$$
$$m' = \begin{cases} 1, & \text{if } sum(\mathbf{M}) > 1 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $x'$ and $m'$ are the layer output and updated mask, respectively. $\mathbf{C}$ represents the convolution filter weights and $b$ refers to its corresponding bias. $\mathbf{X}$ are the feature values for the current input window. $\mathbf{M}$ is the corresponding binary mask. $\mathbf{X}\cdot\mathbf{M}$ is an element-wise multiplication, and $\mathbf{1}$ has the same shape as $\mathbf{M}$ with all the elements equal to 1. The key difference between partial convolution and normal convolution is to multiply $\mathbf{X}$ and $\mathbf{M}$ element-wisely. In this way, the output only depends on the unmasked input values. A scaling factor $\frac{sum(\mathbf{1})}{sum(\mathbf{M})}$ adjusts for varying amount of unmasked input values, leading to sharper velocity profiles in the reconstructed field. For incompressible flow fields we can add a stream function block as shown in **Figure 1**. The resulting velocity field can then be reconstructed from the predicted stream function field $\Psi(x, y)$ by $\mathbf{u} = \nabla \times \Psi$. We refer to the **Appendix** for more detailed information about the network architecture and the training setup.

## 2.2. Loss Functions

It is important to define a new set of supervised loss functions to model physical properties and constraints for fluid flow data. Let $\hat{\mathbf{u}}$ be the predicted velocities and $\mathbf{u}$ be the ground truth velocities. We keep the $L^1$ reconstruction loss as it can efficiently reconstruct low-frequency information:

$$L_{vel} = ||(\hat{\mathbf{u}} - \mathbf{u})\cdot\mathbf{M}||_1 + \alpha_{vel}||(\hat{\mathbf{u}} - \mathbf{u})\cdot(\mathbf{1} - \mathbf{M})||_1, \quad (2)$$

where $\alpha_{vel}$ is a scale factor that weights between empty and known regions. We use $\alpha_{vel} > 1$ to emphasize better reconstructions on regions where the flow information is missing. Inspired by Kim et al. (2019b), we additionally minimize the difference of the velocity field Jacobian between ground truth and predicted velocity fields. With a sufficiently smooth flow field data set, high-frequency features of the CNN are potentially on the null space of the $L^1$ distance minimization (Kim et al., 2019b). Thus, matching the Jacobians helps the network to recover high-frequency spectral information, while it also regularizes the reconstructed velocity to match ground truth derivatives. The velocity Jacobian $J(\mathbf{u})$ is defined in 2D as

$$J(\mathbf{u}) = \begin{pmatrix} \frac{\partial \mathbf{u}_x}{\partial x} & \frac{\partial \mathbf{u}_x}{\partial y} \\ \frac{\partial \mathbf{u}_y}{\partial x} & \frac{\partial \mathbf{u}_y}{\partial y} \end{pmatrix}, \quad (3)$$

and the corresponding loss function is simply given as the $L^1$ of vectorized Jacobian between predicted and ground truth velocities:

$$L_{jac} = ||(J(\hat{\mathbf{u}}) - J(\mathbf{u}))\cdot\mathbf{M}||_1 + \alpha_{jac}||(J(\hat{\mathbf{u}}) - J(\mathbf{u}))\cdot(1 - \mathbf{M})||_1. \quad (4)$$

Additionally, we compute a loss function that matches the vorticity of predicted and ground truth velocities. The vorticity field describes the local spinning motion of the velocity field. Similarly to the Jacobian loss, our vorticity loss acts as a directional high-frequency filter that helps to match shearing derivatives of the original data, enhancing the capability of the model to properly match the underlying fluid dynamics. The vorticity loss is defined as:

$$L_{vort} = ||(\nabla\times\hat{\mathbf{u}} - \nabla\times\mathbf{u})\cdot\mathbf{M}||_1 + \alpha_{vort}||(\nabla\times\hat{\mathbf{u}} - \nabla\times\mathbf{u})\cdot(1 - \mathbf{M})||_1. \quad (5)$$

Incompressible flows should have zero divergence, but numerical simulations often produce results that are not strictly divergence-free due to discretization errors. As we inpaint missing fluid regions, we aim to minimize the divergence on the predicted fields by

$$L_{div} = ||\nabla\cdot\hat{\mathbf{u}}||_1. \quad (6)$$

Lastly, all losses modeled by the network are based on the $L_1$ distance function. Distance-based loss functions are known for undershooting magnitude values, creating results that are visibly smoother. This is especially visible for the inpainting task when we substitute original measured values back into the velocity field $\hat{\hat{u}}$ reconstructed by the network. Therefore, we employed a magnitude of the gradient as our last loss function to produce inpainted results that have less discrepancies when using original measured values for known regions:

$$L_{mag} = ||(\nabla||\hat{\mathbf{u}}\cdot(1 - M) + \mathbf{u}\cdot M||_2)\cdot\mathbf{W}_{mag}||_1. \quad (7)$$

The magnitude of the gradient loss needs a special weighting function $\mathbf{W}_{mag}$ that depends on the mask interface, since if used indistinguishably it can also quickly degenerate the convergence of the network. This weighting function is computed based on the morphological gradient of the mask, which is the difference between the mask dilation and its erosion, which yields a 1-ring mask boundary. We expand the mask boundary continuously by dilation, in order to fill the regions of missing information. As the mask is expanded internal weights are assigned based on the iteration of the dilation–initial iterations have higher values that decay as iterations progress. This is similar to compute a level-set distance function from a missing region to the mask boundary, however our implementation is computationally more effective. The mask is bounded from $[1, 3]$, with higher values closer to the mask boundary.

Notice that all our loss functions, excluding the divergence-free one, employ weights to known and unknown regions. This is a common strategy in inpainting works, and we empirically found that for our data sets $\alpha_{vel} = \alpha_{jac} = \alpha_{vort} = 6$ to yield the best results. Other loss functions, such as perceptual loss and style loss (Liu et al., 2018) are not suited for completing flow field data, since they match pre-learned filters from image classification architectures.

## 2.3. Encoding Obstacles

The interaction between fluid and solid obstacles is crucial for fluid dynamics applications, as the interaction creates shear layers

that drive the formation of vortices. To incorporate solid obstacle information as prior knowledge to the network, we concatenate a binary mask **O** indicating whether a solid obstacle occupies a cell (1) or not (0) as an extra input channel. In order to properly propagate the obstacle information to all network layers, **O** is concatenated to previous layers' output as input to the current layer. To account for resolution change between network layers, we downsample and upsample the obstacle map **O** using average pooling.

## 3. RESULTS

We show that the learning-based reconstruction is especially powerful in handling large areas of missing or occluded data, outperforming traditional models for data recovery. We evaluate our method on numerically-simulated flows, and additionally apply it to the Guadalupe Island case study.

### 3.1. Inpainting of Synthetic Data

Due to the lack of publicly available flow data sets captured from real-world experiments, we trained our model on synthetic data. We generated fluid velocity fields with a numerical incompressible flow solver [Mantaflow (Thuerey and Pfaff, 2018)] and used the stream function block in the training. Each data sample consists of a 2-dimensional ground truth vector field $\vec{u}$, as well as the empty regions and obstacles masks **M** and **O**. To use the data sample in both training and testing, we apply empty regions mask **M** on both velocity component through element-wise multiplication to obtain the input velocity to the model $\vec{u}_{in} = \vec{u} \cdot \mathbf{M}$. The model concatenates the input velocity $\vec{u}_{in}$, empty regions mask **M** and obstacle mask **O** as input, and outputs the predicted velocity field $\hat{\vec{u}}$. Our synthetic flow data set for this task completion is computed on a grid resolution of $128 \times 96$. The *wind tunnel* data set implements a scene with transient turbulent flow around obstacles. We define inflow velocities at bottom and top regions of the domain, while the remaining two sides (left and right) are set as free flow (open) boundary conditions. The inflow speed is set to random values, and 12 obstacles varying between spheres, rectangles and ellipses are randomly positioned, yielding a total of 25,500 unique simulation frames. Examples of velocity fields generated by this simulation setup can be seen in **Figure 2**. We split the whole data set into training (78%), validation (10%), and test (12%) data sets. Each split of the data set comes from a different set of simulation runs. Models are trained on the training set and are compared on the validation set. Later, we report visual results on the test set.

During training, different types of empty region masks are generated on the fly with empty to filled area region ratios that vary randomly between 10 and 99%. The gradient magnitude mask $\mathbf{W}_{mag}$ is also automatically generated for all masks used in the training phase. We model three different types of masks for this task: uniform random noise masks mimic possible sampling noise from real-world velocity measurements; scan path masks simulate paths of a velocity probing; and large region masks model large occluded areas that are not reachable by probes or measurement devices. Illustrations of these types of masks can be seen in **Figures 3**, **4**.

Results of our approach can be seen in **Figures 3**, **4** bottom. The results are generated by taking simulations from the test data set, applying an input mask and feeding as the input of the network, along with the obstacle boundary mask. Our results demonstrate that our deep learning approach is able to plausibly reconstruct flows even in regions with large occlusions. In our evaluations we found that the use of the Dense Block and the combination of the proposed losses yield best results in terms of Mean Absolute Error. The effect of the magnitude of the gradient loss (Equation 7) is particularly interesting as it enforces smooth magnitude transitions in the output and hence reduces artifacts. This is demonstrated visually in **Figure 4**, where the top and bottom rows were computed without and with magnitude of the gradient loss, respectively. Without using this loss, there are noticeable differences in the magnitude of the recovered values, and the masks can be seen in the final reconstruction.

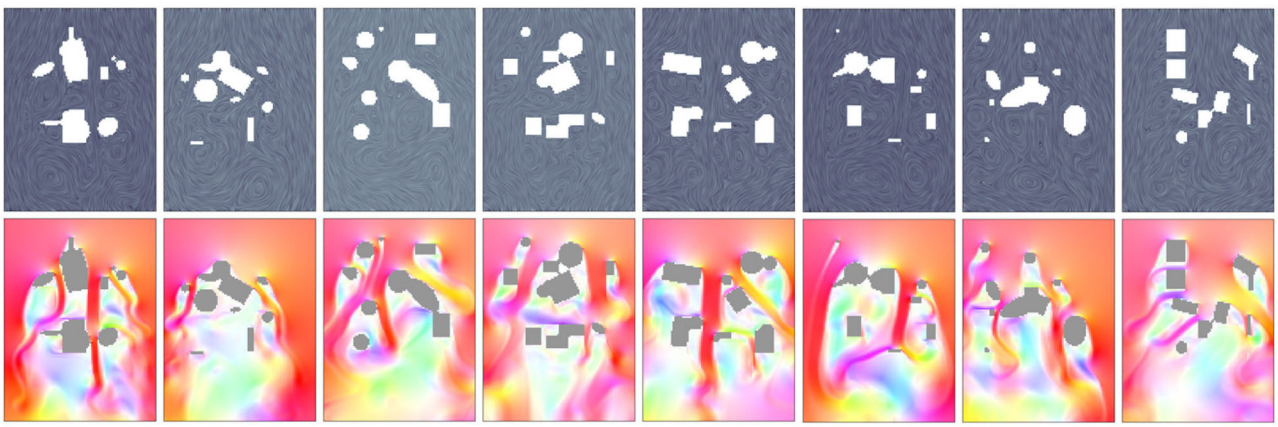


**FIGURE 2 |** Several simulation examples for generating the *wind tunnel* data set. The images above show the line integral convolution (LIC) plots, while images below show HSV color coded velocity fields.

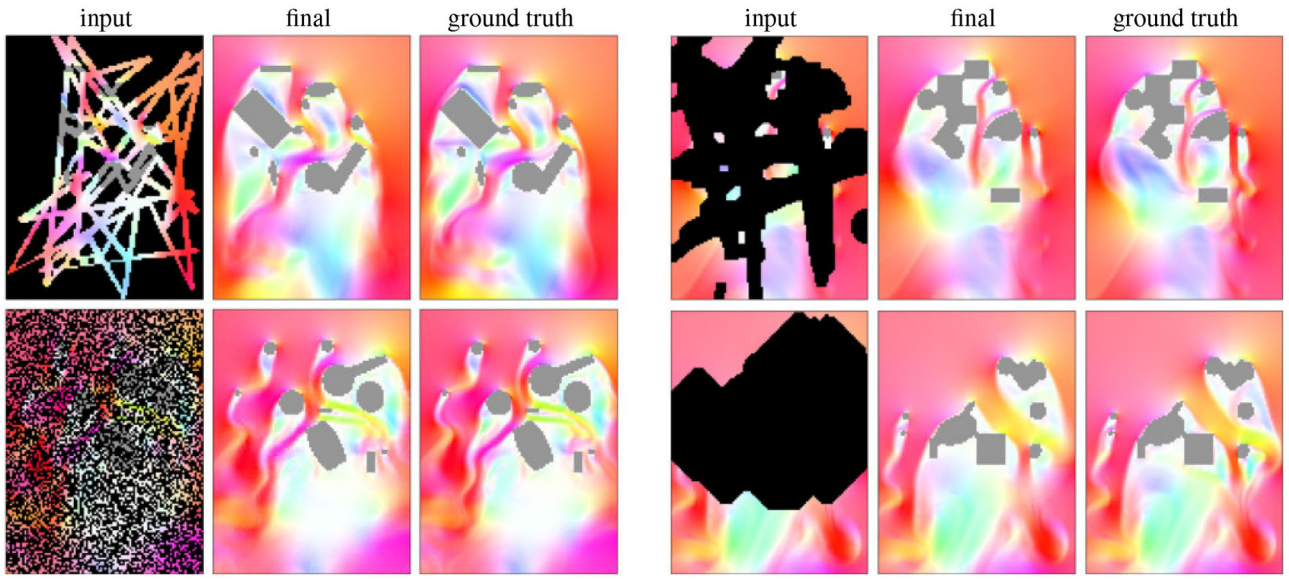



**FIGURE 3 |** Examples of predictions using the *wind tunnel* data set. From left to right of each example: masked velocity input, output of our network, ground truth velocity.

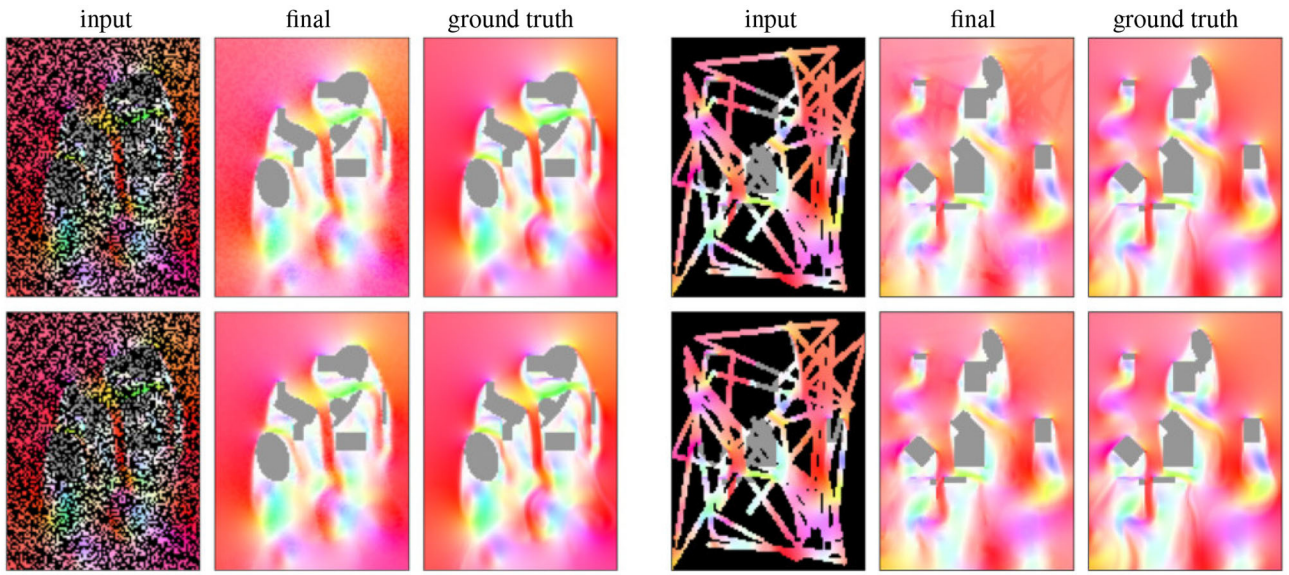



**FIGURE 4 |** Results without (top) and with (bottom) the magnitude of the gradient loss (Equation 7). The latter removes visible artifacts at mask boundaries.

## 3.2. Inpainting of Velocity Measurements Obtained by Optical Flow

We further evaluate the model on a real-world data set consisting of reconstructed velocity fields from satellite imagery of the atmospheric vortex street behind Guadalupe Island (Horváth et al., 2020). **Figure 5** shows the satellite imagery on 9 May 2018 with the vortex generating structure behind the island. Based on 2.5 km GEOS-R observations, patches of 5 × 5 were tracked over time to reconstruct temporal correspondences, resulting in a sequence of 96 time steps with spatial resolution of 6.3 km and 5 min. The reconstructed velocity fields are noisy and lack information outside of the satellite's field of view as well as in areas where the cloud tracking algorithm fails to produce valid results. We show that our model is capable of reconstructing and preserving the vortex generating structure behind Guadalupe Island, whereas conventional methods fail to reconstruct these

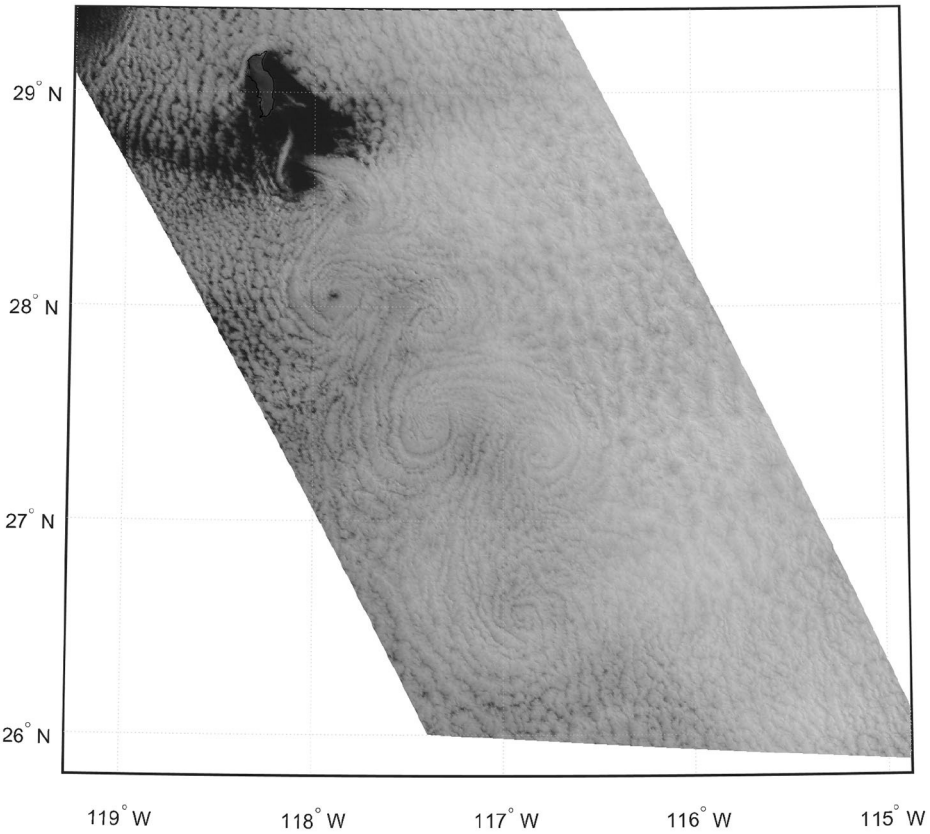



**FIGURE 5 |** Radiance map of Guadalupe island on 9 May 2018 at 14:37 UTC.

structures at such detail. The noise of the real-world data set is successfully removed by the network, generating smooth flow predictions. We demonstrate that physical quantities, such as the vorticity is more accurately captured by our model, especially directly in the wake of the island.

For this task, we generated a three dimensional flow with a voxelized representation of the Guadalupe island geometry immersed in the domain. This setup allows a more realistic flow over the Guadalupe island boundaries, in which the flow can go above and around the island. Since the velocity fields reconstructed from satellite observations are 2-D, we only use the velocity components $u$ and $v$ from the simulated data set at about 800 m above sea level. We also omitted the stream function block of the network and the divergence loss term, since we are evaluating 2-D slices of a 3-D velocity field, which are not guaranteed to be incompressible on the sliced plane. For the scene boundaries, we set the left and bottom with inflow velocities, right and top with outflow conditions. The other two boundaries (above and below) are modeled with free-slip boundary conditions.

The masks used for completing this data set are obtained from evaluating noise patterns that emerge in cloud remote sensing. In **Figure 6**, left, we show how those patterns appear due to errors in the cloud tracking algorithm. We extract many of these samples to generate masks that are similar to the ones that are going to be used to complete the velocity field. This is done by two filtering steps for removing velocities exceeding 8 m/s and noise (**Figure 6**, center and right).

A real-world sample reconstructed with our method is shown in **Figure 7**. The reconstructed vector field is smooth and completes the real-world sample plausibly. We compare our model with a least-squares approximation, minimizing:

$$E = \int_{\mathcal{M}} ||\hat{\mathbf{u}} - \mathbf{u}||^2 \, \mathrm{d}\mathbf{x} + \lambda \int_{\mathcal{D}} ||\hat{\mathbf{u}'}||^2 \, \mathrm{d}\mathbf{x} \rightarrow \min, \qquad (8)$$

where $\mathbf{u}$ is the noisy incomplete data, $\hat{\mathbf{u}}$ is the least-square result, and $\hat{\mathbf{u}'}$ is the gradient of the result. $\lambda = 0.2$ is an empirically chosen weighting term. The first term enforces the preservation of the known data, where the data mask is 1 ($\mathcal{M}$), while the second term enforces a spatially and temporally smooth solution everywhere in the domain ($\mathcal{D}$). Note that $\mathbf{u}$ and $\hat{\mathbf{u}}$ are 3D data with x, y, and time axes, therefore the least squares method can utilize the temporal dependency between the samples, while our model cannot.

**Figure 8** shows the result of our model (center) and the result of the least squares (right) for the same example in **Figure 7**. The curvy pattern of the vortex street is visible in both results,

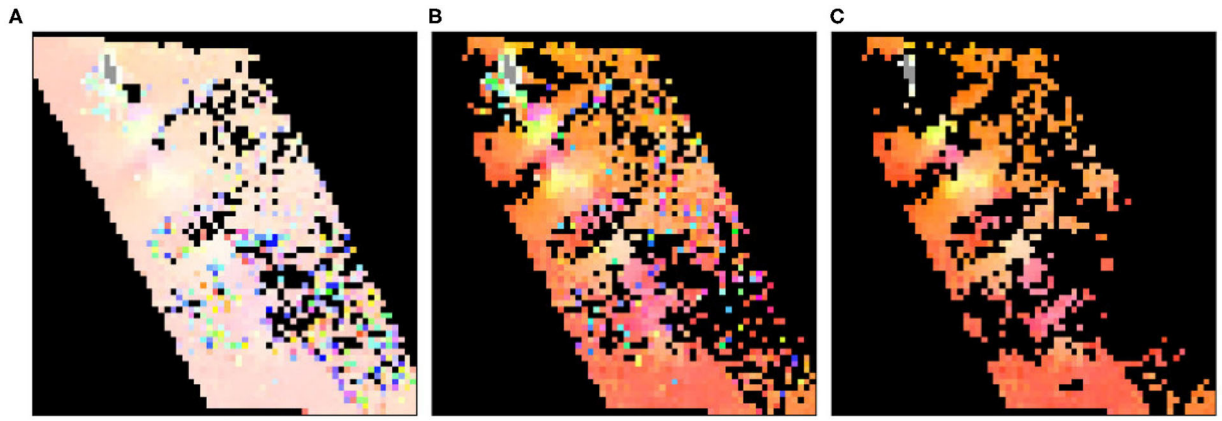



**FIGURE 6 |** A real-world data sample: Original data **(A)**, filtered data to remove velocities exceeding 8 m/s **(B)**, and removal of local noise **(C)**.

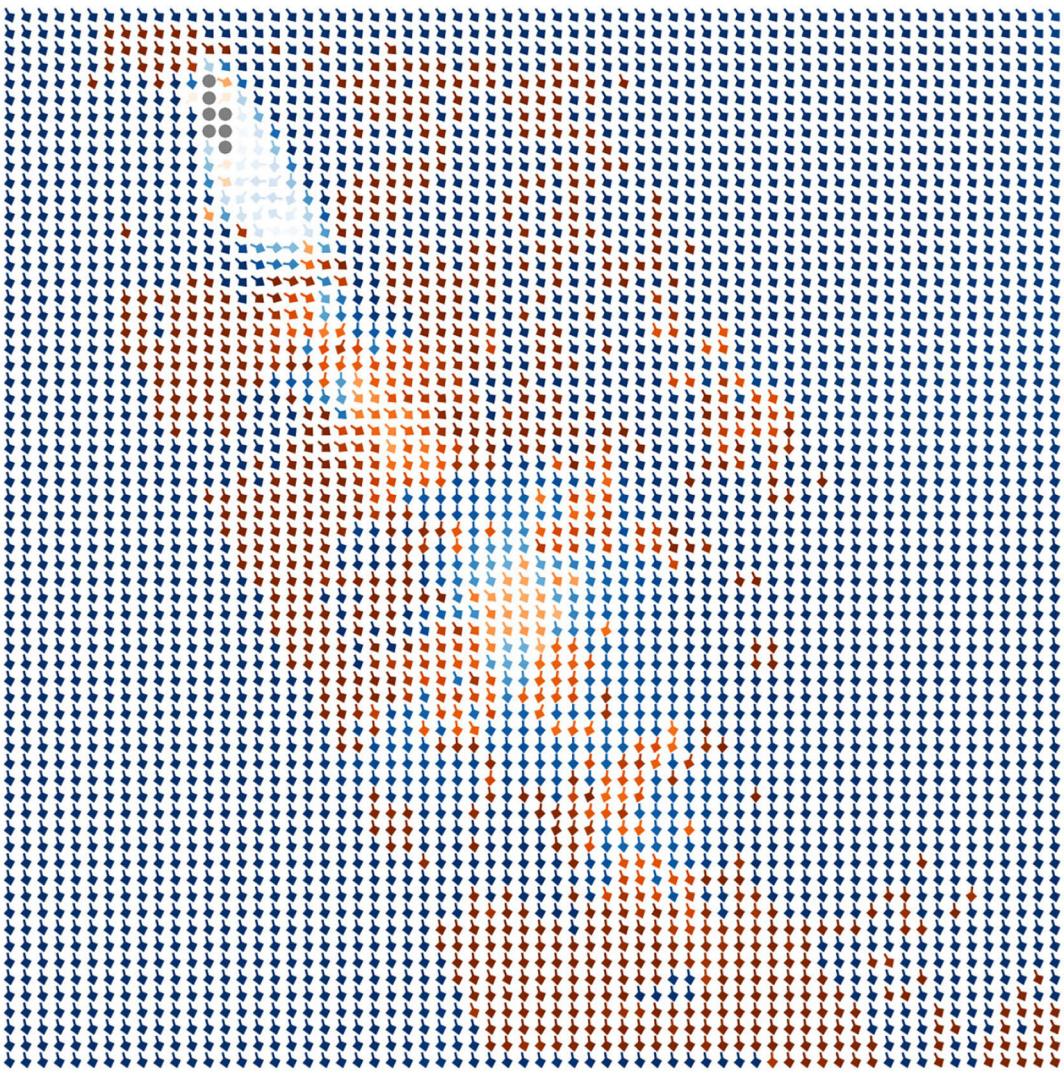


**FIGURE 7 |** Reconstructed real-world sample. The red arrows indicate the known data points, the blue arrows show the reconstructed data points, and the brightness of the arrows encodes the magnitudes.

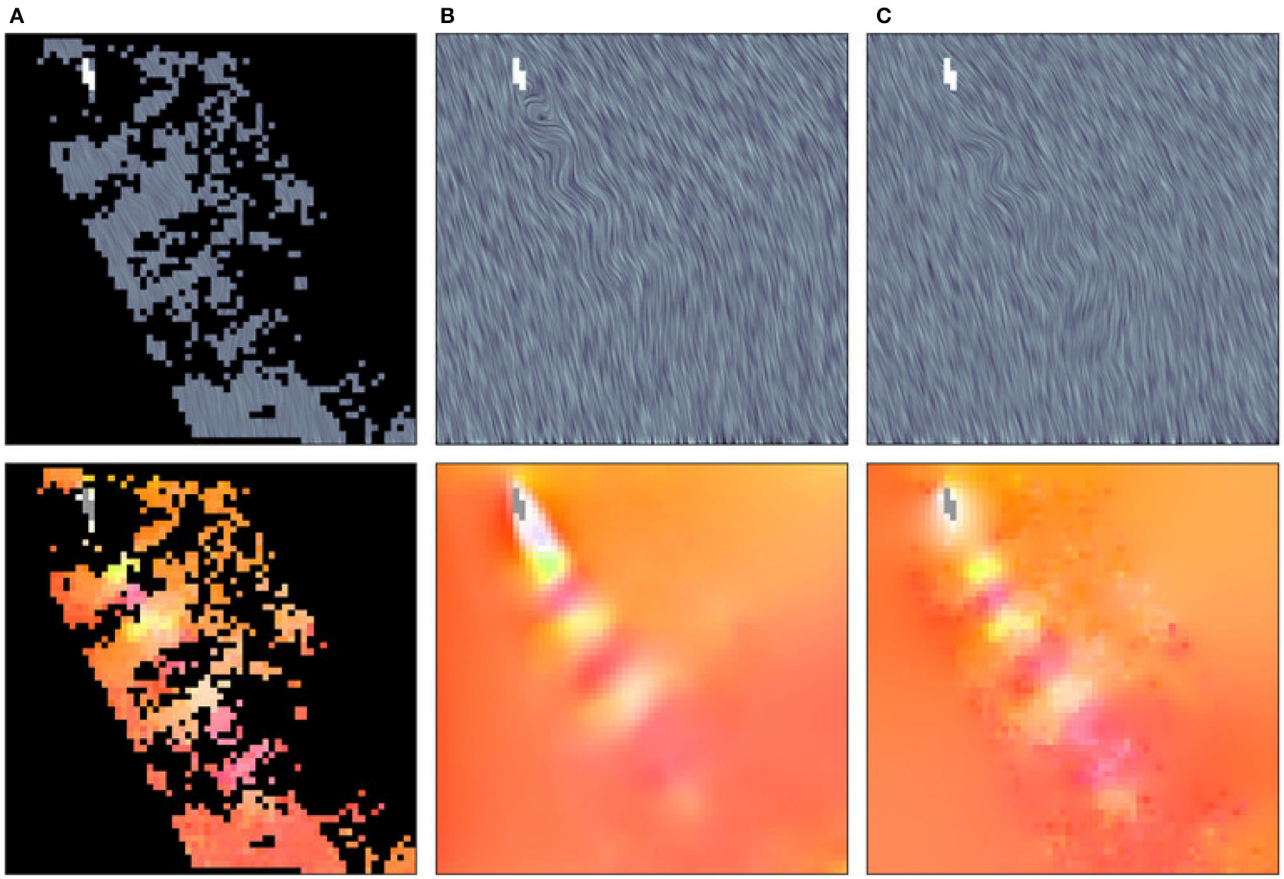



**FIGURE 8 |** LIC and HSV visualization of the results obtained by reconstructing a real-world data sample **(A)** with our model **(B)** and with a least squares method **(C)**.

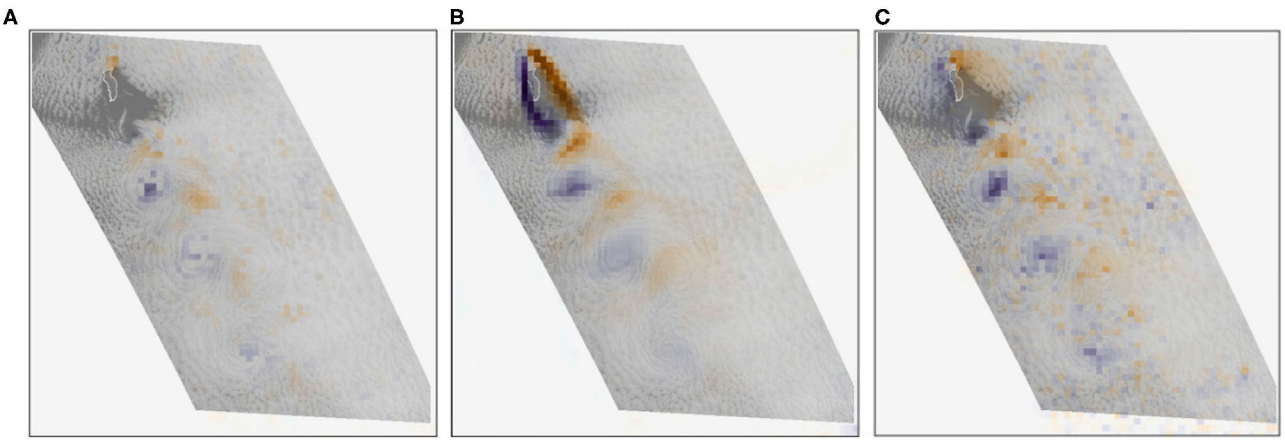



**FIGURE 9 |** Vorticity visualization of the smoothed real-world data sample **(A)**, the reconstruction with our neural method **(B)**, and the least squares method **(C)**.

but unlike the least squares method, our model is also capable of reconstructing the turbulent vortex-generating structure in close vicinity to the island. To evaluate whether the vortices are preserved by our model and the least squares method, we examine the vorticity of the results in **Figure 9**, while showing satellite observations of the clouds in the background to provide context. While there is, unfortunately, no ground truth velocity field available for such observational data, an agreement of vortex

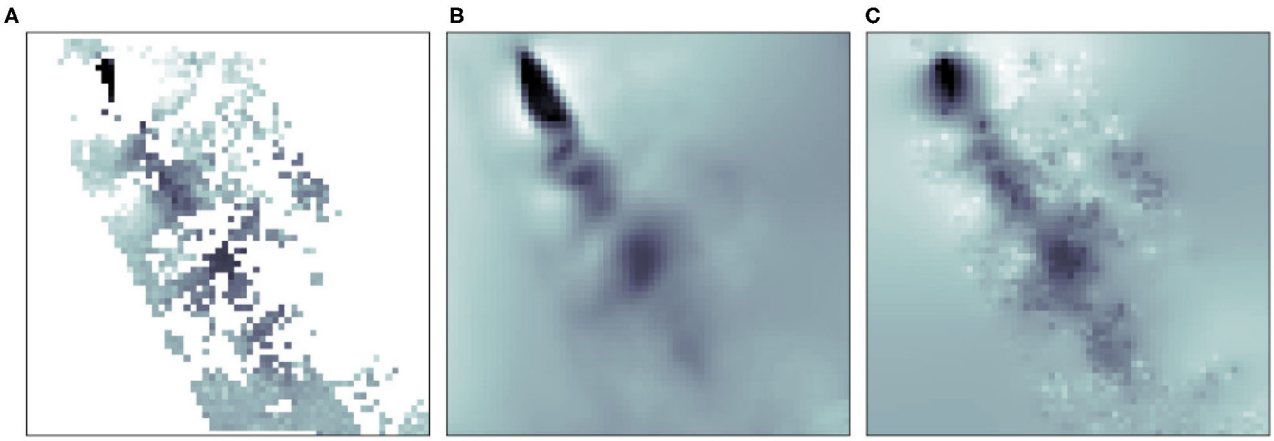



**FIGURE 10 |** Magnitude visualization of the smoothed real-world data sample **(A)**, our reconstruction **(B)**, and the least squares solution **(C)**.

structures with the underlying cloud patterns gives a strong indication for correctness, since the reconstructed flow is able to explain the fluid dynamical processes leading to the observable patterns (Horváth et al., 2020). From left to right we show the vorticity of the original data, our reconstruction with the neural network, and the least squares result. It can be seen that our model gives superior results, preserving the position and size of the different vortices. The visualization of the flow magnitudes in **Figure 10** underlines this, showing that our model (center) results in a more reliable reconstruction of the flow structure than the least squares solution (right).

## 4. CONCLUSIONS

Satellite-based wind retrieval enables a high-resolution view onto atmospheric air flows. This data, however, is intrinsically noisy, which is usually addressed through median filters or spatial smoothing. In this paper, we regularized the reconstruction from partial and noisy data in a data-driven approach using a U-Net based convolutional neural network. Based on model simulations, we generated training data to teach a neural network to disambiguate the partial observations based on physically-realizable flow configurations that have been observed during training. Our results on the synthetic flow data sets demonstrate that the use of neural networks can be successfully applied to flow data recovery tasks. Our case study on the Guadalupe Island data set demonstrates the practical impact of this physics-aware neural network on data post-processing, evaluation, and prediction in atmospheric modeling. The proposed neural-based method is especially powerful in high-occlusion applications, where the least squares method fails to provide reliable results. We further found that dilated convolution increases the number of data points used by the model to predict larger areas of missing data. This leads to better high-resolution results and prevents poor predictions – it can even be helpful in medium-resolution data sets without obstacles.

A neural network can only handle what it has seen during training. For extreme events, such as storms or floods, our currently trained model is likely to fail. The next step would be to study from observational or long-term reanalysis data the different boundary conditions that would occur in extreme events, and to then sample those conditions to initialize our ground truth simulations. It will then be important to create a training data set that is evenly balanced in its different flow configurations, such that extreme events are not too rare in the training data.

For the Guadalupe flow, unfortunately, no ground truth vector field is available, since this is a measured vector field. Instead, plausibility is usually tested by considering the spatial coherence of wind vectors (cf. Horváth et al., 2020). Additionally, Horváth et al. (2020) derived vortex measures assuming that a good agreement of vortex locations and visible cloud patterns is an indicator for a reasonable wind vector field that can explain the fluid dynamical processes that are visible in observations, which is also the approach we took in this manuscript. Determining additional validation heuristics is an interesting avenue in itself, which is beyond the scope of this paper.

Considering subsequent data points in time could potentially improve our model and allow for spatio-temporal reconstructions of flow structures. We found that when reconstructing each frame of the Guadalupe island sequence individually, the neural network prediction is smooth in most areas except the one in close vicinity to the island. This is because turbulent flow structures emerge in those regions, which cannot be captured coherently in time by a single-frame reconstruction technique.

Our model could also be used to determine the data points that give our model the least information (the best data points to mask) by calculating the gradient for the mask and not for the input. It could further inspire new techniques for compressing meteorological data and hence reducing data storage. Beyond meteorological applications, the method could impact related fields as well, such as finding optimal guiding procedures for human-based flow scanning systems, or improved

workflows for digital prototyping where quick flow previews are particularly useful.

When surface altimetry data is available, for example from SWOT (surface water and ocean topography) missions, it could be added as additional channel to the network, allowing the network to pick up information that helps to disambiguate the partial observations further. Studying its effect would be another interesting avenue for future research.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

LS, SF, and VA carried out the implementation and performed the calculations. LS, SF, JT, VA, TG, and BS developed the theoretical formalism and wrote the manuscript. JT, VA, TG, and BS supervised the project and conceived the original idea. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Swiss National Science Foundation under Grant number CRSK-2_190296.

## REFERENCES

Baño-Medina, J. L., García Manzanas, R., and Gutiérrez Llorente, J. M. (2020). Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geosci. Model Dev*. 13, 2109–2124. doi: 10.5194/gmd-2019-278

Brunton, S. L., Noack, B. R., and Koumoutsakos, P. (2020). Machine learning for fluid mechanics. *Annua. Rev. Fluid Mech*. 52, 477–508. doi: 10.1146/annurev-fluid-010719-060214

Carr, J. L., Wu, D. L., Wolfe, R. E., Madani, H., Lin, G. G., and Tan, B. (2019). Joint 3D-wind retrievals with stereoscopic views from MODIS and GOES. *Rem. Sens*. 11:2100. doi: 10.3390/rs11182100

Chopra, K. P., and Hubert, L. F. (1965). Mesoscale eddies in wake of islands. *J. Atmos. Sci*. 22, 652–657. doi: 10.1175/1520-0469(1965)022<0652:MEIWOI>2.0.CO;2

Choy, C. B., Xu, D., Gwak, J., Chen, K., and Savarese, S. (2016). "3D-R2N2: a unified approach for single and multi-view 3D object reconstruction," in *Computer Vision–ECCV 2016* (Amsterdam), 628–644. doi: 10.1007/978-3-319-46484-8_38

Geerts, B., Raymond, D. J., Grubisic, V., Davis, C. A., Barth, M. C., Detwiler, A., et al. (2018). Recommendations for *in situ* and remote sensing capabilities in atmospheric convection and turbulence. *Bull. Am. Meteorol. Soc*. 99, 2463–2470. doi: 10.1175/BAMS-D-17-0310.1

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, eds Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (Montreal, QC: Curran Associates, Inc.), 2672–2680.

Guo, Y., and Tong, X. (2018). "View-volume network for semantic scene completion from a single depth image," in *IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence* (Stockholm), 726–732. doi: 10.24963/ijcai.2018/101

Han, X., Zhang, Z., Du, D., Yang, M., Yu, J., Pan, P., et al. (2019). "Deep reinforcement learning of volume-guided progressive view inpainting for 3D point scene completion from a single depth image," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA). doi: 10.1109/CVPR.2019.00032

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc*. 146, 1999–2049. doi: 10.1002/qj.3803

Horváth, Á., Bresky, W., Daniels, J., Vogelzang, J., Stoffelen, A., Carr, J. L., et al. (2020). Evolution of an atmospheric Kármán vortex street from high-resolution satellite winds: Guadalupe island case study. *J. Geophys. Res. Atmos*. 125:e2019JD032121. doi: 10.1029/2019JD032121

Horváth, Á., Hautecoeur, O., Borde, R., Deneke, H., and Buehler, S. A. (2017). Evaluation of the eumetsat global avhrr wind product. *J. Appl. Meteorol. Climatol*. 56, 2353–2376. doi: 10.1175/JAMC-D-17-0059.1

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings–30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* (Honolulu), 2261–2269. doi: 10.1109/CVPR.2017.243

Hubert, L. F., and Krueger, A. F. (1962). Satellite pictures of mesoscale eddies. *Mon. Weather Rev*. 90, 457–463. doi: 10.1175/1520-0493(1962)090<0457:SPOME>2.0.CO;2

Iizuka, S., Simo-Serra, E., and Ishikawa, H. (2017). Globally and locally consistent image completion. *ACM Trans. Graph*. 36:107. doi: 10.1145/3072959.3073659

Kim, B., Azevedo, V., Gross, M., and Solenthaler, B. (2019a). Transport-based neural style transfer for smoke simulations. *ACM Trans. Graph*. 38, 188:1–188:11. doi: 10.1145/3355089.3356560

Kim, B., Azevedo, V. C., Thuerey, N., Kim, T., Gross, M., and Solenthaler, B. (2019b). Deep fluids: a generative network for parameterized fluid simulations. *Comput. Graph. Forum* 38, 59–70. doi: 10.1111/cgf.13619

Liu, G., Reda, F. A., Shih, K. J., Wang, T. C., Tao, A., and Catanzaro, B. (2018). Image inpainting for irregular holes using partial convolutions. *Lect. Notes Comput. Sci*. 11215, 89–105. doi: 10.1007/978-3-030-01252-6_6

Min, M., Li, J., Wang, F., Liu, Z., and Menzel, W. P. (2020). Retrieval of cloud top properties from advanced geostationary satellite imager measurements based on machine learning algorithms. *Rem. Sens. Environ*. 239:111616. doi: 10.1016/j.rse.2019.111616

O'Gorman, P. A., and Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: potential for modeling of climate, climate change, and extreme events. *J. Adv. Model. Earth Syst*. 10, 2548–2563. doi: 10.1029/2018MS001351

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., and Efros, A. A. (2016). "Context encoders: feature learning by inpainting," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 2536–2544. doi: 10.1109/CVPR.2016.278

Raissi, M., Yazdani, A., and Karniadakis, G. E. (2018). Hidden fluid mechanics: a navier-stokes informed deep learning framework for assimilating flow visualization data. *arXiv [Preprint]*. arXiv:1808.04327.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. *Lect. Notes Comput. Sci*. 9351, 234–241. doi: 10.1007/978-3-319-24574-4_28

Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., and Lebair, W. J. (2017). A closer look at the abi on the goes-R series. *Bull. Am. Meteorol. Soc*. 98, 681–698. doi: 10.1175/BAMS-D-15-00230.1

Seifert, A., and Rasp, S. (2020). Potential and limitations of machine learning for modeling warm-rain cloud microphysical processes. *J. Adv. Model. Earth Syst*. 12:e2020MS002301. doi: 10.1029/2020MS002301

Song, S., Yu, F., Zeng, A., Chang, A., Savva, M., and Funkhouser, T. (2017). "Semantic scene completion from a single depth image," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu) 190–198. doi: 10.1109/CVPR.2017.28

Thuerey, N., and Pfaff, T. (2018). *MantaFlow*. Available online at: http://mantaflow.com

Umetani, N., and Bickel, B. (2018). Learning three-dimensional flow for interactive aerodynamic design. *ACM Trans. Graph*. 37, 89:1–89:10. doi: 10.1145/3197517.3201325

Weyn, J. A., Durran, D. R., and Caruana, R. (2019). Can machines learn to predict weather? Using deep learning to predict gridded 500-HPA geopotential height from historical weather data. *J. Adv. Model. Earth Syst*. 11, 2680–2693. doi: 10.1029/2019MS001705

Wiewel, S., Becher, M., and Thuerey, N. (2019). Latent-space physics: towards learning the temporal evolution of fluid flow. *Comput. Graph. Forum* 38, 71–82. doi: 10.1111/cgf.13620

Xu, K., Shi, Y., Zheng, L., Zhang, J., Liu, M., Huang, H., et al. (2016). 3D attention-driven depth acquisition for object identification. *ACM Trans. Graph.* 35, 238:1–238:14. doi: 10.1145/2980179.2980224

Young, G. S., and Zawislak, J. (2006). An observational study of vortex spacing in island wake vortex streets. *Mon. Weather Rev*. 134, 2285–2294. doi: 10.1175/MWR3186.1

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T. S. (2018). "Free-form image inpainting with gated convolution," in *IEEE/CVF International Conference on Computer Vision (ICCV)* (Istanbul), 4470–4479. doi: 10.1109/ICCV.2019.00457

Zeng, Y., Fu, J., Chao, H., and Guo, B. (2019). "Learning pyramid-context encoder network for high-quality image inpainting," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA), 1486–1494. doi: 10.1109/CVPR.2019.00158

Zhang, Y., and Funkhouser, T. A. (2018). "Deep depth completion of a single RGB-D image," in *IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 175–185. doi: 10.1109/CVPR.2018.00026

**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

## Network Architectures

We specify the network architectures in **Tables A1–A4**. All layers in the encoder, dense block and decoder are Partial Convolution layers described in Equation (1). The layers in the stream function branch are normal convolutional layers. Batch normalization is used for all layers except the output layer and layers in the stream function branch. We use ReLU activation functions for layers in the encoder and dense block, LeakyReLU activation function for layers in the decoder and Swish activation function for the stream function branch. The output of the stream function branch is used to compute the velocity field: $\hat{\tilde{u}}_\psi = (\frac{\partial \psi}{\partial y}, -\frac{\partial \psi}{\partial x})$. Then, $\hat{\tilde{u}}_\psi$ is concatenated with the output from the velocity branch and goes through a normal convolutional layer to produce the final velocity prediction $\hat{\tilde{u}}$.

## Training Details

All models were trained with Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and the learning rate was set to $10^{-5}$. We used a batch size of 8 and train all models for 50 epochs.

**TABLE A1 |** Network layer configurations of the encoder part.

| Layer | E1 | E2 | E3 | E4 | E5 | E6 | E7 |
|---|---|---|---|---|---|---|---|
| In channels | 2 | 64 | 192 | 144 | 432 | 324 | 972 |
| Out channels | 64 | 192 | 144 | 432 | 324 | 972 | 729 |
| Kernel Size | 7 × 7 | 5 × 5 | 5 × 5 | 3 × 3 | 3 × 3 | 3 × 3 | 3 × 3 |

**TABLE A2 |** Network layer configurations of the DenseBlock part.

| Layer | Dense1 | Dense2 | Dense3 | Dense4 | Dense5 | Dense6 | Dense7 | Dense8 |
|---|---|---|---|---|---|---|---|---|
| In channels | 1,701 | 1,733 | 1,765 | 1,797 | 1,829 | 1,861 | 1,893 | 1,925 |
| Out channels | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 |
| Kernel size | 3 × 3 | 3 × 3 | 3 × 3 | 3 × 3 | 3 × 3 | 3 × 3 | 3 × 3 | 3 × 3 |

**TABLE A3 |** Network layer configurations of the decoder part.

| Layer | D1 | D2 | D3 | D4 | D5 | D6 |
|---|---|---|---|---|---|---|
| In channels | 356 | 756 | 576 | 336 | 256 | 66 |
| Out channels | 324 | 432 | 144 | 192 | 64 | 2 |
| Kernel size | 3 × 3 | 3 × 3 | 3 × 3 | 3 × 3 | 3 × 3 | 3 × 3 |

**TABLE A4 |** Network layer configurations of the stream function branch.

| Layer | S1 | S2 | S3 | S4 | S5 |
|---|---|---|---|---|---|
| In channels | 67 | 131 | 163 | 195 | 227 |
| Out channels | 64 | 32 | 32 | 32 | 1 |
| Kernel size | 7 × 7 | 5 × 5 | 5 × 5 | 5 × 5 | 1 × 1 |

frontiers in Climate

ORIGINAL RESEARCH
published: 12 April 2021
doi: 10.3389/fclim.2021.656479

# Spatio-Temporal Downscaling of Climate Data Using Convolutional and Error-Predicting Neural Networks

*Agon Serifi*[1]*, *Tobias Günther*[2] *and Nikolina Ban*[3]

[1] Department of Computer Science, ETH Zurich, Zurich, Switzerland, [2] Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, [3] Department of Atmospheric and Cryospheric Science, University of Innsbruck, Innsbruck, Austria

OPEN ACCESS

**Edited by:**
*Guojian Wang, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia*

**Reviewed by:**
*Rüdiger Westermann, Technical University of Munich, Germany*
*Eduardo Rodrigues, IBM Research, Australia*

***Correspondence:**
*Agon Serifi*
*agon.serifi@alumni.ethz.ch*

**Specialty section:**
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

**Received:** *20 January 2021*
**Accepted:** *15 March 2021*
**Published:** *12 April 2021*

**Citation:**
*Serifi A, Günther T and Ban N (2021) Spatio-Temporal Downscaling of Climate Data Using Convolutional and Error-Predicting Neural Networks. Front. Clim. 3:656479. doi: 10.3389/fclim.2021.656479*

Numerical weather and climate simulations nowadays produce terabytes of data, and the data volume continues to increase rapidly since an increase in resolution greatly benefits the simulation of weather and climate. In practice, however, data is often available at lower resolution only, for which there are many practical reasons, such as data coarsening to meet memory constraints, limited computational resources, favoring multiple low-resolution ensemble simulations over few high-resolution simulations, as well as limits of sensing instruments in observations. In order to enable a more insightful analysis, we investigate the capabilities of neural networks to reconstruct high-resolution data from given low-resolution simulations. For this, we phrase the data reconstruction as a super-resolution problem from multiple data sources, tailored toward meteorological and climatological data. We therefore investigate supervised machine learning using multiple deep convolutional neural network architectures to test the limits of data reconstruction for various spatial and temporal resolutions, low-frequent and high-frequent input data, and the generalization to numerical and observed data. Once such downscaling networks are trained, they serve two purposes: First, legacy low-resolution simulations can be downscaled to reconstruct high-resolution detail. Second, past observations that have been taken at lower resolutions can be increased to higher resolutions, opening new analysis possibilities. For the downscaling of high-frequent fields like precipitation, we show that error-predicting networks are far less suitable than deconvolutional neural networks due to the poor learning performance. We demonstrate that deep convolutional downscaling has the potential to become a building block of modern weather and climate analysis in both research and operational forecasting, and show that the ideal choice of the network architecture depends on the type of data to predict, i.e., there is no single best architecture for all variables.

**Keywords: machine learning, climate data, downscaling, super-resolution, convolutional neural networks**

## 1. INTRODUCTION

A universal challenge of modern scientific computing is the rapid growth of data. For example, numerical weather and climate simulations are nowadays run at kilometer-scale resolution on global and regional domains (Prein et al., 2015), producing a data avalanche of hundreds of terabytes (Schär et al., 2020). In practice, however, data is often available at lower resolution only, for which there are many practical reasons. For example, older archived simulations have been computed on lower resolution or were reduced due to memory capacity constraints. Also, when allocating the computational budget running multiple low-resolution ensemble simulations might be favored over few high-resolution simulations. The loss of high-resolution information is a serious problem that must be addressed for two critical reasons. First, the loss of data limits any form of *post-hoc* data analysis, sacrificing valuable information. Second, an *in-situ* data analysis (Ma, 2009), i.e., the processing of the data on the simulation cluster, is not reproducible by the scientific community, since the original raw data has never been stored. Even if a large amount of computing resources is available for re-running entire simulations and outputting higher frequency data for analysis, it still requires reproducible code, which is cumbersome to maintain due to the changes in super computing architectures (Schär et al., 2020). For these reasons, reconstruction algorithms from partial data are a promising research direction to improve data analysis and reproducibility. Not only is the reconstruction of higher spatial and temporal resolutions valuable for numerical simulations, meteorological observations are also only available at certain temporal resolutions and suffer from sparse observational networks. In many applications, such as in hydrology, higher temporal resolutions are desperately needed, for example to inform urban planners in the design of infrastructures that support future precipitation amounts (Mailhot and Duchesne, 2010).

In climate science, deep learning has recently been applied to a number of different problems, including microphysics (Seifert and Rasp, 2020), radiative transfer (Min et al., 2020), convection (O'Gorman and Dwyer, 2018), forecasting (Roesch and Günther, 2019; Selbesoglu, 2019; Weyn et al., 2019), and empirical-statistical downscaling (Baño-Medina et al., 2020). For example, Yuval et al. (2021) have applied deep learning for parametrization of subgrid scale atmospheric processes like convection. They have trained neural networks on high-resolution data and have applied it as parametrization for coarse resolution simulation. Using deep learning, they demonstrated that they could decrease the computational cost without affecting the quality of simulations.

In computer vision, the problem of increasing the resolution of an image is referred to as the single-image super-resolution problem (Yang et al., 2014). The super-resolution problem is inherently ill-posed, since infinitely many high-resolution images look identical after coarsening. Usually, the recovery of a higher resolution requires assumptions and priors, which are nowadays learned from examples *via* deep learning, which–in the context of climate data–has proven to outperform simple linear baselines (Baño-Medina et al., 2020). For single-image super-resolution, Dong et al. (2015) introduced a convolutional architecture (CNN). Their method receives as input an image that was already downscaled with a conventional method, such as bicubic interpolation, and then predicts an improved result. The CNN is thereby applied to patches of the image, which are combined to result in the final image. The prior selection of an interpolation method is not necessarily optimal, as it places assumptions and alters the data. Thus, both Mao et al. (2016) and Lu and Chen (2019) proposed variants that take the low-resolution image as input. Their architectures build on top of the well known U-Net by Ronneberger et al. (2015). The method learns in a encoder-decoder fashion a sub-pixel convolution filter or deconvolution filter, respectively, which were shown to be equivalent by Shi et al. (2016). A multi-scale reconstruction of multiple resolutions has been proposed by Wang et al. (2019). Further, Wang et al. (2018) explored the usage of generative adversarial networks (GANs). A GAN models the data distribution and samples one potential explanation rather than finding a blurred compromise of multiple explanations. These generative networks hallucinate plausible detail, which is easy to mistake for real information. Despite the suitability of generative methods in the light of perceptual quality metrics, the presence of possibly false information is a problem for scientific data analysis that has not been fully explored yet. For a single-image super-resolution benchmark in computer vision, we refer to Yang et al. (2019).

Next, we revisit the deep learning-based downscaling in meteorology and climate science, cf. Baño-Medina et al. (2020). Rodrigues et al. (2018) took a supervised deep learning approach using CNNs to combine and downscale multiple ensemble runs spatially. Their approach is most promising in situations where the ensemble runs deviate only slightly from each other. In very diverse situations, a standard CNN will give blurry results, since the CNN finds a least-squares compromise of the many possible explanations that fit the statistical variations. In computer vision terms, this approach can be considered a multi-view super-resolution problem, whereas we investigate the more challenging single-image super-resolution. Höhlein et al. (2020) studied multiple architectures for spatial downscaling of wind velocity data, including a U-Net based architecture and deep networks with residual learning. The latter resulted in the best performance on the wind velocity fields that they studied. Following Vandal et al. (2017), they included additional variables, such as geopotential height and forecast surface roughness, as well as static high-resolution fields, such as land sea mask and topography. They demonstrated that the learning overhead of such a network is justified, when considering the computation time difference between a low-resolution and high-resolution simulation. Later, we will show that residual networks will not generally outperform direct convolutional approaches on our data, since the best choice of network is data-dependent, and we also include temporal downscaling in our experiments. Pouliot et al. (2018) studied the super-resolution enhancement of Landsat data. Vandal et al. (2017, 2019) stacked multiple CNNs to learn multiple higher spatial resolutions from a given precipitation field. Cheng et al. (2020) proposed a convolutional

architecture with residual connections to downscale precipitation spatially. In contrast, we also focus on the temporal downscaling of precipitation data, which is a more challenging problem due to motion and temporal variation. Toderici et al. (2017) solved the compression problem of high-resolution data and did not consider the downscaling problem. In principle, it would be imaginable to not store a coarsened version of the high-resolution data (which would be possible in our pipeline), but to store the compressed latent space as encoded by the network (as done by Toderici et al., 2017). The latter requires to keep the encoding/decoding code alongside the data and has the potential downside that many (old) codes have to be maintained, which could turn out impractical for operational scenarios. Instead, we investigate a pipeline in which we start from coarsened data. It is clear, however, that a learnt encoder could provide a better compression than a simple coarsening. CNNs tend to produce oversmoothed results, as they produce a compromise of the possible explanations that satisfy the incomplete data. Different approaches have been tested to improve the spatial detail, including the application of relevance vector machines (Sachindra et al., 2018) and (conditioned) generative neural networks (Singh et al., 2019; Han and Wang, 2020; Stengel et al., 2020). While the latter improves the visual quality, it is not yet clear how much the interpretability of the result is impeded by the inherent hallucination.

When considering the various meteorological variables that are at play, we can observe large differences between the rates at which the structures in the data evolve temporally, how they correlate with spatial locations–for example convection near complex topography, and how much spatial variation they experience. For this reason, we investigate and evaluate meteorological fields from both ends of the spectrum: low-frequent and high-frequent signals. Fundamentally, two different approaches are imaginable. A deep neural network could either predict a high-resolution field directly, or an error-corrector from a strong baseline approach could be learnt, utilizing the strengths of contemporary methods. Thereby, the success of the error-predicting approach depends on the quality of the baseline. We explore both types of architecture in the light of the underlying signal frequency, as we hypothesize that for high-frequent data the baseline might not reach the significant quality needed to be useful for the error-predicting network. In order to avoid over-smoothing of the results, we augment the loss function to enforce the preservation of derivatives. Further, numerically simulated data and measured data have different signal-specific characteristic in terms of smoothness, occurrence of noise and differentiability. As both domains–simulation and observations–profit greatly from highly-resolved data, we investigate the spatial and temporal downscaling on both simulated and observed data.

# 2. METHOD AND DATA

Formally, we aim to downscale a time-dependent meteorological scalar field $s(x, y, t)$ from a low number of grid points $X \times Y \times T$ to a higher number of grid points $\overline{X} \times \overline{Y} \times \overline{T}$, with $\overline{X} = k_x X$, $\overline{Y} = k_y Y$, and $\overline{T} = k_t T$. Thereby, $k_x$, $k_y$, and $k_t$ are called the downscaling factors. We approach the problem through supervised deep learning, i.e., at training time we carefully prepare groundtruth pairs of low-resolution and high-resolution scalar field patches. A patch is a randomly cropped space-time region from the meteorological data. Afterwards, convolutional neural networks are trained to recall the high-resolution patch from a given low-resolution patch. Using patches enables direct control over the batch size, which is an important hyper-parameter during training, as it influences the loss convergence. Since our network architectures are convolutional, the networks can later be applied to full domains, i.e., cropping of patches is not necessary at inference time after training. We follow prior network architectures based on the U-Net by Ronneberger et al. (2015), one called UnetSR by Lu and Chen (2019)–an end-to-end network directly predicting the downscaled output, the other one called REDNet by Mao et al. (2016)–a residual prediction network. Both networks receive trivially downscaled data as input and have an encoder-decoder architecture where skip connections connect the feature maps from the encoder to their mirrored counterpart in the decoder. In the following, we refer to our residual predicting network as RPN and the end-to-end deconvolution approach as DCN. Before explaining the network architectures in detail, we introduce the data and explain the coarsening of high-resolution data to obtain groundtruth pairs for the training process.

## 2.1. Data

Here we describe the two data sets on which we apply and test the method. The data originates from two sources: climate model simulations and observations.

### 2.1.1. Climate Model Data

The described method and approach is tested on the climate data produced by a regional climate model COSMO (Consortium for Small Scale Modeling). It is a non-hydrostatic, limited-area, atmospheric model designed for applications from the meso-$\beta$ to the meso-$\gamma$ scales (Steppeler et al., 2003). The data has been produced by a version of COSMO that is capable of running on GPUs (Fuhrer et al., 2014), and has been presented and evaluated in Hentgen et al. (2019). The climate simulation has been conducted with a horizontal grid spacing of 2.2 km (see Leutwyler et al., 2017; Hentgen et al., 2019). The red box in **Figure 1** shows the domain that we use for the temperature predictions. Since precipitation can be close to zero in many regions of the domain, we expanded the domain to the blue box for the precipitation experiments. We used temperature and precipitation fields available every 5 min for the months June and July in 2008.

### 2.1.2. Observations

The observational data set used in this study is a gridded precipitation dataset for year 2004, covering the area of Switzerland. The horizontal grid spacing of the data is 1 km (Wüest et al., 2010) and it is available at hourly frequency. It is generated using a combination of station data with radar-based disaggregation. The data is often used for climate model evaluation (see e.g., Ban et al., 2014).

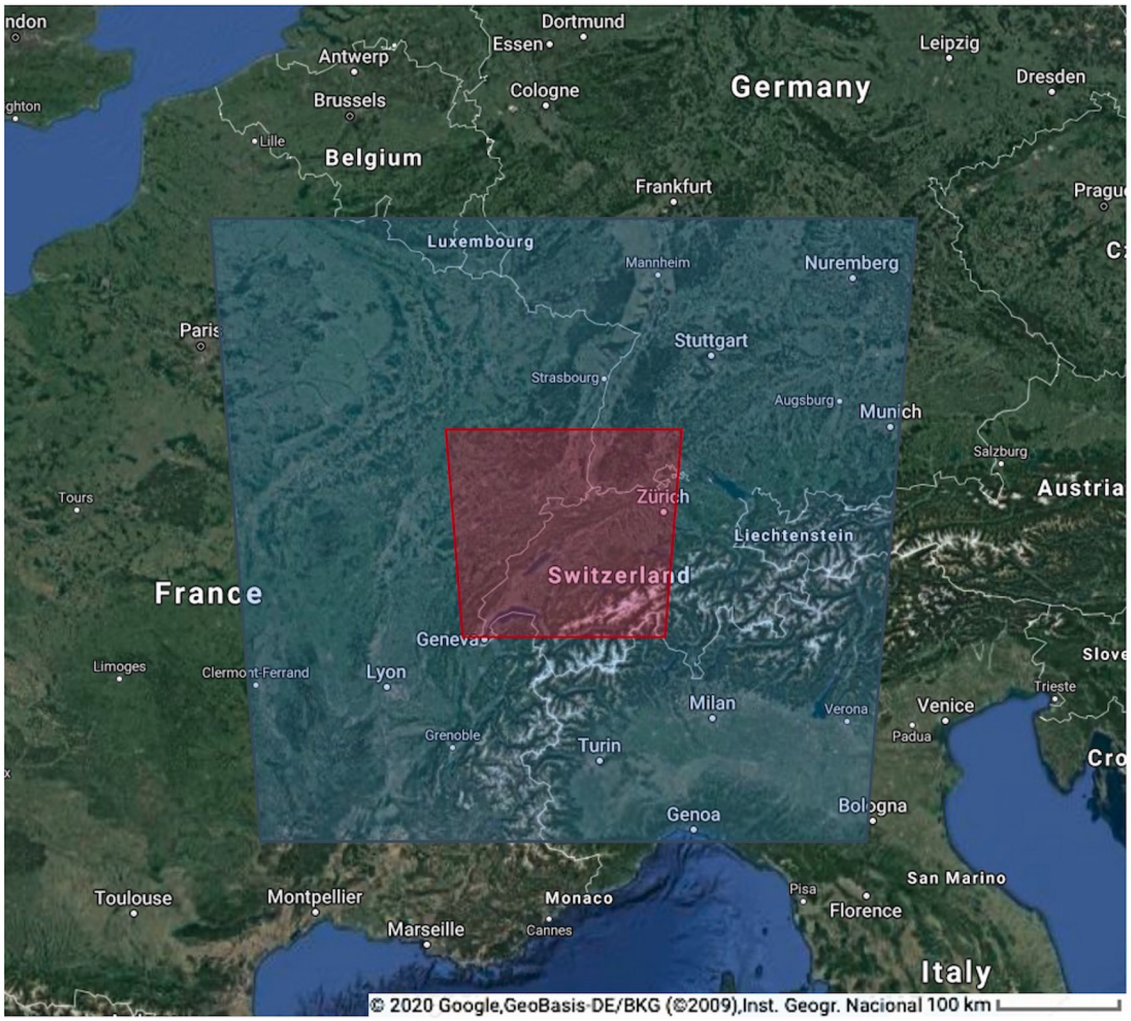



**FIGURE 1 |** The analysis region over central Europe used in this study indicated with red box (temperature) and blue box (precipitation).

## 2.2. Supervised Machine Learning for Downscaling of Meteorological Data

Let $\mathbf{X}$ be a coarse patch with $X \times Y \times T$ regular grid points, and let $\mathbf{Y}$ be the corresponding downscaled patch with $\overline{X} \times \overline{Y} \times \overline{T}$ grid points. Further, let $f(\mathbf{Y})$ be a map that coarsens a high-resolution patch $\mathbf{Y}$ into its corresponding low-resolution patch $\mathbf{X}$:

$$\underbrace{\mathbf{X} = f(\mathbf{Y})}_{\text{coarsening}}, \qquad \underbrace{\mathbf{Y} = f^{-1}(\mathbf{X})}_{\text{downscaling}} \tag{1}$$

The inverse problem $f^{-1}$, i.e., the downscaling problem, is usually ill-posed, since the map $f$ is not bijective. While any high-resolution patch can be turned into a unique low-resolution patch *via* coarsening, the reverse will have multiple possible solutions, i.e., $f$ is surjective, but not injective.

However, not every possible solution to $\mathbf{Y} = f^{-1}(\mathbf{X})$ is physically meaningful and realizable in real-world data. It therefore makes sense to construct the inverse map $f^{-1}$ in a data-driven manner from real-world data to only include mappings that have actually been seen during training, which is the key idea behind supervised machine learning. The inverse map is thereby parameterized by a concatenation of multiple weighted sums of inputs that each go through a non-linear mapping. The composition–a deep neural network–thereby becomes a differentiable, highly non-linear mapping between the input and output space, and can be iteratively trained *via* gradient descent.

The success of a deep neural network thereby hinges on three key criteria:

1. the architecture of the neural network combines low-frequent and high-frequent information, and the gradients $d\mathbf{Y}/d\mathbf{X}$ are well defined to facilitate the training process,
2. the training data is of high quality and expressive, i.e., we explore the space of possible mappings sufficiently and the mappings are sufficiently distinct.

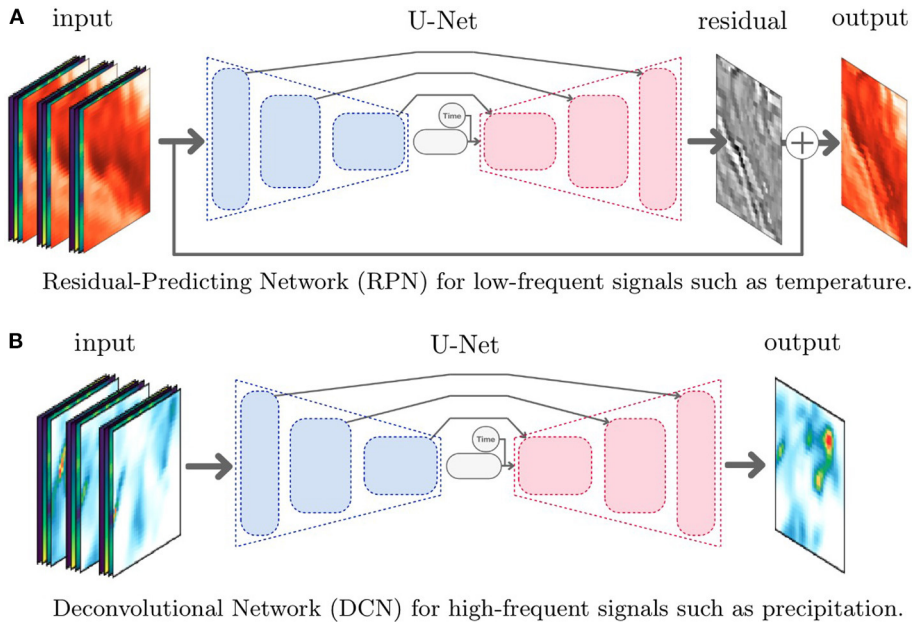



**FIGURE 2 |** Illustrations of the convolutional neural network architectures for downscaling of low-frequent and high-frequent meteorological variables. Both architectures receive three time steps of the scalar field to predict (temperature or precipitation) and additional fields (longitude, latitude, surface height) as input. For RPN, the input data is downscaled conventionally in both space and time. For DCN, the input data is downscaled conventionally only in space. In both networks, the time variable is appended in the latent space and indicates at which relative time between the input frames the output should be downscaled at. **(A)** Residual-predicting network (RPN) for low-frequent signals, such as temperature. **(B)** Deconvolutional network (DCN) for high-frequent signals, such as precipitation.

3. the loss function expresses the desired goal well and the energy manifold is well-behaved to allow a stable (stochastic) gradient descent.

In the following subsections, we elaborate on the network architectures in section 2.2.1, the training data generation in section 2.2.2, and the training procedure and loss function in section 2.2.3.

### 2.2.1. Network Architecture

When observing meteorological variables, such as temperature and precipitation, we can see vast differences in their spatial and temporal variability. While temperature varies slowly in space and time, i.e., it is a comparatively low-frequent signal, precipitation is far more localized and varies faster, i.e., it is a high-frequent signal that is harder to predict with conventional downscaling techniques. To leverage the data characteristics, we design two separate convolutional neural networks to represent the inverse mapping $f^{-1}$.

#### *2.2.1.1. Low-Frequent Data: Residual-Predicting Network (RPN)*

In case, the data has low spatial and temporal variability, a conventional downscaling technique might already take us close to the desired solution. Rather than learning the entire downscaling process, it will then be an easier task to correct the conventional downscaling method, which is the core concept of residual learning (cf. Dong et al., 2015). Let $\widehat{f}^{-1}$ be an existing downscaling technique, such as trilinear interpolation in space-time. Then, the inverse $f^{-1}(\mathbf{X})$ can be formed by:

$$f^{-1}(\mathbf{X}) = \underbrace{\widehat{f}^{-1}(\mathbf{X})}_{\text{trilinear downscaling}} + \underbrace{r(\widehat{f}^{-1}(\mathbf{X}))}_{\text{residual}} \tag{2}$$

where our neural network only learns to predict the residual $r(\widehat{f}^{-1}(\mathbf{X}))$ of the trilinear downscaling method. For this, we follow the architecture of Mao et al. (2016), who applied an encoder-decoder architecture, which is detailed further below. The advantage of this approach is that it is comparatively easier to improve over the existing trilinear baseline method in contrast to learning a downscaling method from scratch. If $\widehat{f}^{-1}(\mathbf{X})$ performs poorly, for example since the scalar field exhibits too much temporal variability, then the next approach will perform better.

#### *2.2.1.2. High-Frequent Data: Deconvolutional Network (DCN)*

Consider a case in which too much motion occurred between time steps, e.g., a cloud got transported to a new location not overlapping with its previous location. Then, the trilinear downscaling method might interpolate two small clouds in the time step in-between at the original and the final location, rather than obtaining a single translating cloud in the middle. Other than before, the linear downscaling in time might not be close enough to benefit from residual prediction. In such cases where a a conventional temporal downscaling method is not helpful, we learn the partial mapping $p(\widetilde{f}^{-1}(\mathbf{X}))$ from spatially-downscaled

**TABLE 1 |** Temperature (°C) downscaling mean-squared errors (MSE), coloring the best (●), intermediate (●) and worst (●) result.

| MSE (°C) | Baseline | | | RPN | | | DCN | | |
|---|---|---|---|---|---|---|---|---|---|
| $k_t \backslash k_{x,y}$ | 1 | 2 | 4 | 1 | 2 | 4 | 1 | 2 | 4 |
| 1 | – | 0.398 | 0.669 | – | 0.168 | 0.216 | – | 0.177 | 0.461 |
| 2 | 0.016 | 0.399 | 0.669 | 0.016 | 0.169 | 0.253 | 0.031 | 0.177 | 0.308 |
| 4 | 0.017 | 0.411 | 0.677 | 0.017 | 0.150 | 0.232 | 0.054 | 0.190 | 0.421 |
| 12 | 0.102 | 0.416 | 0.678 | 0.099 | 0.194 | 0.259 | 0.113 | 0.216 | 0.393 |

*Columns show spatial scaling factors $k_x = k_y \in \{1, 2, 4\}$ and rows show temporal scaling factors $k_t \in \{1, 2, 4, 12\}$. Note that the residual-predicting network (RPN) outperformed the baseline and DCN in all cases. Temporal downscaling introduces marginal errors, since the field is varying slowly in time.*

data to the high resolution:

$$f^{-1}(\mathbf{X}) = \underbrace{p(\widetilde{f}^{-1}(\mathbf{X}))}_{\text{partial downscaling}} \tag{3}$$

where $\widetilde{f}^{-1}(\mathbf{X})$ performs only spatial downscaling using bilinear interpolation, but not temporal downscaling and where $p(\widetilde{f}^{-1}(\mathbf{X}))$ performs both the temporal downscaling and improves over the result of $\widetilde{f}^{-1}$. Since $\widetilde{f}^{-1}(\mathbf{X})$ does not interpolate information in time, a residual prediction is no longer applicable. Hence, the high-resolution data is predicted directly. For the network architecture, we follow a typical U-Net architecture (Ronneberger et al., 2015), which is a general design not limited to downscaling problems. In our downscaling setting, the input data is spatially downscaled with a bilinear baseline method, as was proposed by Lu and Chen (2019) for image super-resolution. In the following, we explain how the networks are structured and which modifications improved the performance for meteorological downscaling problems.

#### *2.2.1.3. Layers and Skip Connections*

The neural network architectures are illustrated in **Figure 2**. In both architectures, the network consists of convolutional layers only. Among the most recent convolutional neural network architectures, U-Nets by Ronneberger et al. (2015) are often the most promising approach. A U-Net extracts both low-frequency and high-frequency features from the data by repeatedly performing feature extraction and coarsening. In the so-called contraction phase, we apply successively two convolutional layers followed by a MaxPooling layer to extract features and then reduce the resolution. To handle convolutions on image boundaries, we use zero-padding and apply the convolutions with a stride of 1, i.e., every pixel of the input data will once be the center of a convolution kernel. We repeat this structure four times where the last time we omit the pooling layer. Within each layer, we extract a certain number of features. Starting with 64 features maps, we double the size until 512 feature maps are reached in the last layer. This is the amount of information available in the subsequent step: the synthesis of the output in the expansion phase. In the expansion phase, the goal is to reconstruct a high resolution image from all previously extracted features by iteratively increasing the number of grid points until the target resolution is reached. We do this by using so-called UpSampling layers, which repeat the values to a larger data block, followed by three convolutional layers. The key to success is to provide in each level of the U-Net the feature maps that have been extracted previously on the same resolution during contraction. This is achieved by skip connections from the contraction to the expansion phase. Adding the skip connections as in the U-Net by Ronneberger et al. (2015) has two positive effects. First, it was shown to smooth the loss landscape (c.f., Li et al., 2018), which makes it easier to perform gradient descent during training. Second, the skip connections give access to the high-frequency information of earlier layers, which greatly helps to construct high-frequent outputs.

#### *2.2.1.4. Inputs and Outputs*

Since we intend to downscale data both in space and time, we provide the network with both spatial and temporal information. Thus, the input to the model is a 4D data block, one dimension is used for the time steps, two for the spatial information, and the last one holds the so-called channels. A difference to the conventional U-Net is that we experimented with additional data channels that provide more information to resolve ambiguities in the inverse map $f^{-1}$. The effectiveness of additional data channels was already demonstrated by Vandal et al. (2017) and Höhlein et al. (2020) for downscaling. These additional channels include latitude and longitude such that the network can learn regional weather patterns, and altitude to include dependencies on the topography. For example, we observed that adding these additional fields improved the residual by 7.3%, for a precipitation downscaling with $k_x = k_y = k_t = 4$. In addition, we provide temporal information to the network, which allows us to give the model information about the relative time between the next known adjacent time steps. Since the time variable is a constant and not a spatially-varying map–unlike all other previously mentioned additional information, we append the time to the latent space, i.e., the feature map at the end of the feature extraction phase. Other options to include the time are imaginable. Including time as a constant separate slice in the input would increase the network size, which is why we opted for appending it to the latent space. Our data concentrated on a specific season. Including the day of the year as additional variable in order to learn seasonal effects would be straightforward to add.

The output of the network depends on the chosen architecture. As described above, we predict the error residual for the low-frequent data in the RPN, e.g., for the temperature field. In the case of high-frequent data, such as precipitation, we directly predict the high-resolution outputs in the DCN. In both cases, the networks are convolutional, thus the network can be applied at inference time to the full input domain at once.

### 2.2.2. Training Data Generation

Supervised machine learning requires groundtruth pairs of low-resolution and corresponding high-resolution patches. In the following, we describe how these groundtruth pairs are generated from the given high-resolution meteorological data.

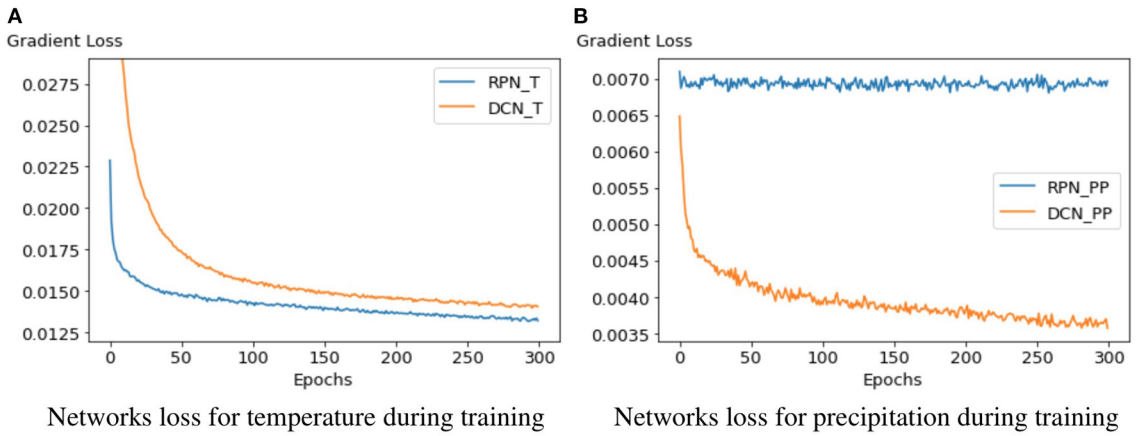



**FIGURE 3 |** Validation loss plots of both architectures on temperature and precipitation. The left plot shows that RPN converges faster and achieves a lower loss during training for the temperature field than DCN. On the other hand, we see that the same RPN architecture is unable to learn when applied on precipitation data. **(A)** Network loss for temperature during training. **(B)** Network loss for precipitation during training.

The coarsening operation depends on the units of the data. When the units remain the same (e.g., absolute temperature in K), then we use an average operation only. When the units change (e.g., total precipitation depends on the time step), then we apply averaging and convert the units afterwards. In case of precipitation, the coarsening in time is equal to an accumulation of the precipitation values. Generally, we recommend to use an averaging operation to do the coarsening, since a max operation or a simple subsampling would cause aliasing artifacts that would not be present if the data was simulated or measured on lower resolution. For the residual predicting network (RPN), we downscale the low-resolution data with a conventional trilinear interpolation method, and feed the downscaled data to the network in order to predict the residual (c.f., section Low-frequent data: residual-predicting network (RPN)). In this work, we applied linear interpolation to avoid extrapolation of minima and maxima. Any other existing downscaling method, such as cubic interpolation, would conceptually also be imaginable. For DCN, the network receives spatially-downscaled input, similar to RPN. In the temporal direction, we input the coarse resolution, since a linear interpolation would cause blending and ghosting artifacts that the network would have to learn to undo. During training, we randomly crop patches with a resolution of $32 \times 32$ from the high-resolution and (conventionally downscaled) low-resolution data. We thereby separate the time sequence into a training period and a testing period to assure that the training and testing sets are independent. For this, we used the last 10% of the time range for testing.

Since the input fields (temperature or precipitation, and longitude, latitude, and surface height) would have different value ranges, we normalize all fields globally across the entire data set to the unit interval [0, 1], which is a common preprocess with neural networks. The scaling factors are stored, such that the results of the network can be scaled back to the physical units later.

### 2.2.3. Training Procedure and Loss

As loss function, we measure the difference between the predicted result $\mathbf{Y}$ and the groundtruth $\overline{\mathbf{Y}}$. Convolutional neural networks are known to oversmooth the output. Hence, we assess the difference with an $L_1$ norm that is combined with a gradient loss to not only penalize differences in the values but also in the derivatives, which aids in the reconstruction of higher-frequency details. We refer to Kim and Günther (2019) and Kim et al. (2019) for a discussion of the norms and weights of the gradient loss.

$$\mathcal{L}(\mathbf{Y}, \overline{\mathbf{Y}}) = \underbrace{|\mathbf{Y} - \overline{\mathbf{Y}}|_1}_{\text{data loss}} + \lambda \underbrace{|\nabla\mathbf{Y} - \nabla\overline{\mathbf{Y}}|_1}_{\text{gradient loss}} \tag{4}$$

Here, $\lambda$ is a weight indicating how much the focus should lie on the difference of gradients. We explored the residual for different choices of $\lambda$ in a precipitation downscaling experiment with scaling factors 2 in temporal and spatial dimension. The baseline obtains a residual of 6.601 MSE $[g/m^2]$. Setting $\lambda = 0$, i.e., not including the gradient loss term, gives the simple L1-norm, which obtains a residual of 8.181 MSE $[g/m^2]$, which is larger than the baseline. Thus, the gradient loss term is required such that the network is able to concentrate on high-frequent details. We empirically set $\lambda = 1$ in our experiments, which result in a residual of 3.882 MSE $[g/m^2]$. Increasing $\lambda$ further, e.g., to $\lambda = 10$, again increased the residual of the network to 5.846 MSE $[g/m^2]$.

As common with neural networks, we performed for both network architectures a hyperparameter optimization, i.e., we empirically adjusted each network parameter, such as the number of layers, the number of features, the batch size, and the activation functions to obtain the best neural network for each problem. Alternatively, automatic hyperparameter optimization frameworks, such as Optuna are

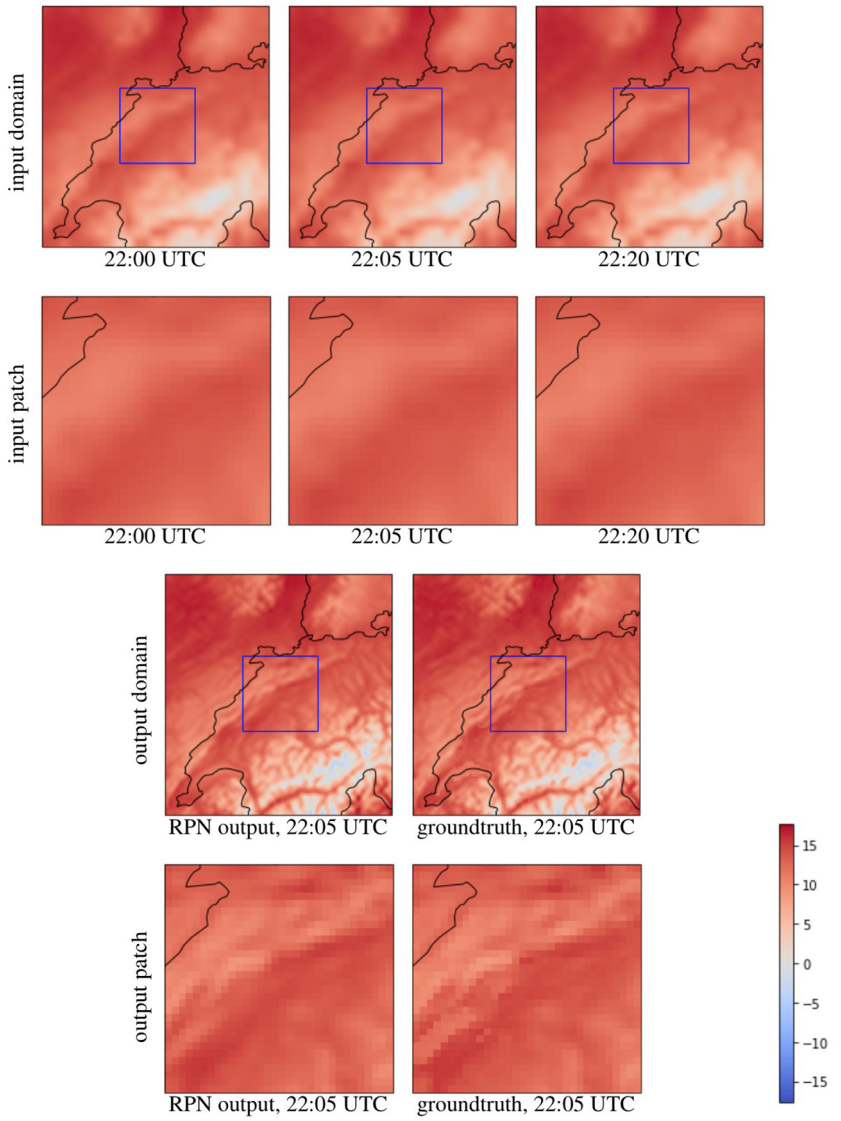



**FIGURE 4 |** Downscaling results for temperature (in °C) by a factor of $k_x = k_y = k_t = 4$ in both the temporal and spatial dimension. The first row shows time steps of the trivially downscaled domain. The second row shows a patch as it is sent into the network. The third row compares the result of the network to the groundtruth on the full domain. The last row shows the groundtruth comparison for the patch that was predicted.

available (Akiba et al., 2019), which could be employed in the future. We choose Adam as optimizer with the default settings (learning rate 0.001) as proposed by Kingma and Ba (2014), and used a batch size of 8 to meet GPU memory constraints. Both networks were trained for 80 h on a single compute node (Intel Xeon CPU E5-2630, Nvidia GeForce 1080Ti). The training time is an important factor in the hyper-parameter optimization. Automatic frameworks, such as Optuna (Akiba et al., 2019), explore many different hyper-parameter combinations, each requiring a training run. For such an automatic hyper-parameter optimization, the total training time would scale linearly in the number of tested parameter configurations.

## 2.3. Analysis

To evaluate the neural networks, we performed a number of experiments, which are detailed in the following sections. To quantify the improvement over the trilinear downscaling in space-time, we calculate the mean squared error (MSE). Let $i \in \{1, \dots, n\}$ be the index of the $n$ grid points of a space-time patch, then MSE is defined as:

$$\mathrm{MSE}(\mathbf{Y}, \overline{\mathbf{Y}}) = \frac{1}{n} \sum_{i}^{n} |\mathbf{Y}_i - \overline{\mathbf{Y}}_i|^2 \tag{5}$$

where $\mathbf{Y}$ is the downscaled result and $\overline{\mathbf{Y}}$ is the groundtruth. Along with the quantitative measures, we visualize the downscaled fields to show the amount of detail that is reconstructed visually.

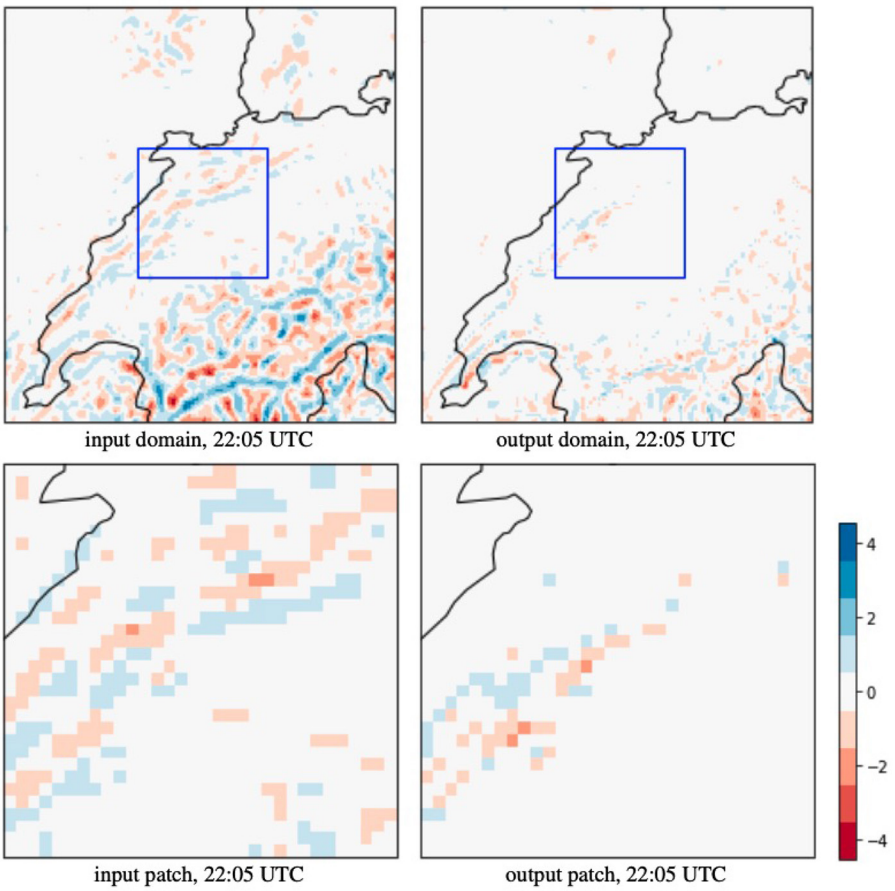



**FIGURE 5 |** The error map of **Figure 4** in MSE (°C) shows a comparison between the trilinear interpolated input and our predicted output relative to the groundtruth. On the full domain, RPN reduces the MSE from 1.438 to 0.143°C which is a 10× improvement, and on the zoom-in (blue box), the MSE reduced from 0.533 to 0.077°C.

To assess how well the network is able to downscale in space and in time, we vary the downscaling factors $k_x$, $k_y$, and $k_t$ in an ablation study in sections 3.1.1 and 3.2.1, and train a network for each case separately. We can expect that small factors will perform better, since less information is missing. The networks were designed for low-frequent input data (RPN) and high-frequent input data (DCN). Therefore, we evaluate both networks on their respective data type, namely temperature fields for RPN, and the precipitation for DCN. To justify the need for DCN, we apply the RPN network to high-frequent precipitation data, as well. Likewise, we apply the DCN network to low-frequent temperature data.

Finally, we train neural networks for observational data in section 3.2.2. Compared to numerical data, observations exhibit very different data characteristics in terms of resolution, noise, and spatial and temporal variation.

## 3. RESULTS

In this section, we report and discuss results of our experiments. We begin with experiments on low-frequent data (temperature), which is followed by reporting results for high-frequent data (precipitation). For all shown metrics, we compare with the high-resolution ground truth, which is equivalent to the result obtained by a full resimulation. A resimulation is prohibitively expensive, taking a full day on Piz-Daint (supercomputer at the Swiss National Supercomputing Center (CSCS) in Switzerland) utilizing 100 GPU nodes (Nvidia Tesla K20X).

### 3.1. Temperature

First, we investigate the downscaling capabilities for both network architectures by reporting the residual errors for different downscaling factors.

#### 3.1.1. Network Comparison

We reduced the number of spatial grid points by a factor of 2 and 4, and the time steps by a factor of 2, 4, and 12. For all scaling factors, we perform downscaling with the baseline method and our two network architectures, and report the MSE (in °C) in **Table 1**. With only temporal downscaling ($k_x = k_y = 1$), RPN and the baseline give similar results, while DCN is about 80% worse. Across all spatial downscaling factors, varying the temporal downscaling does not significantly change the result, since the temporal variation of temperature was low. Compared

**TABLE 2 |** Precipitation (g/m$^2$) downscaling mean-squared errors (MSE), coloring the best (●), intermediate (●) and worst (●) result.

| MSE (g/m$^2$) | Baseline | | | RPN | | | DCN | | |
|---|---|---|---|---|---|---|---|---|---|
| $k_t$ \ $k_{x,y}$ | 1 | 2 | 4 | 1 | 2 | 4 | 1 | 2 | 4 |
| 1 | – | 1.074 | 3.069 | – | 6.233 | 3.631 | – | 0.491 | 1.575 |
| 2 | 1.020 | 1.740 | 3.441 | 6.034 | 5.569 | 3.769 | 0.258 | 0.649 | 1.717 |
| 4 | 2.775 | 3.224 | 4.353 | 5.080 | 4.341 | 3.108 | 1.045 | 1.393 | 2.178 |
| 12 | 5.974 | 6.153 | 6.601 | 4.762 | 12.11 | 27.65 | 3.672 | 3.445 | 3.882 |

*Columns show spatial scaling factors $k_x = k_y \in \{1, 2, 4\}$ and rows show temporal scaling factors $k_t \in \{1, 2, 4, 12\}$. The deconvolution approach (DCN) consistently outperforms the baseline and RPN. Note that both spatial and temporal downscaling affect the error significantly. For such high-frequent data, RPN is rarely able to improve over the baseline.*

to the baseline, RPN is able to reduce the error for $k_x = k_y = 2$ by about 58%, while DCN achieves 53%. A more significant difference occurs for spatial downscaling with $k_x = k_y = 4$, for which RPN achieves 64% and DCN only 41% reduction compared to the baseline (cf. **Table 1**). In **Figure 3A**, we see at the example of $k_x = k_y = 2$, $k_t = 4$ that both networks achieve a reasonable reduction of the loss. RPN improves over the DCN architecture in both the convergence rate and the obtained residual. We can observe that for a low-frequent signal, such as temperature, the residual predicting network (RPN) consistently outperforms the baseline and the deconvolutional approach (DCN). The only exception occurred for $k_x = k_y = 1$ (no spatial down-scaling) and $k_t = 2$ (temporal down-scaling by factor 2). Since temperature varies very slowly in time, the baseline already obtains a very small error. In that case, RPN is on average $0.001°C$ worse than the baseline (only yellow square for RPN in **Table 1**), which is a negligible difference. We can also see that a reconstruction from a high temporal coarsening ($k_t = 12, k_x = k_y = 2$) is better than the reconstruction from larger spatial coarsening ($k_t = 1, k_x = k_y = 4$), which would both reconstruct from the same number of low-resolution grid points. This is because temperature changes slower over time, therefore downscaling in this dimension is easier for the neural network to learn.

In addition to the quantitative measures, we provide a qualitative view onto the reconstructed temperature field. **Figure 4** shows a sample of the testing set with a spatial and temporal downscaling factor of $k_x = k_y = k_t = 4$. The RPN model is able to recover detailed structures, increasing the quality not only quantitatively but also visually. The corresponding error map in **Figure 5** shows that the remaining errors remain highest in regions with complex topography due the high spatial variability. The MSE reduced by a factor of 10.

The reconstruction of temperature data can be done in parallel and takes 1 min on a single Intel i7 4770 HQ (2.2 GHz) per timestep, while the network requires about 125 MB of storage.

## 3.2. Precipitation

In this section, we study a more challenging task: the downscaling of high-frequent precipitation fields.

### 3.2.1. Network Comparison

The numerical precipitation data was given at 5 min intervals. For temporal downscaling, we test the reconstruction from 10, 20 min, and hourly data. In **Table 2**, we report the MSE for the baseline, RPN and DCN for multiple combinations of downscaling factors. While the low-frequent temperature field was best reconstructed by residual learning using RPN, the technique fails on the high-frequent precipitation field, increasing the error on average by a factor of two. Using the DCN architecture instead, consistently leads to better results. For $k_x = k_y = 1$, the DCN improved over the baseline on average by about 43%, for $k_x = k_y = 2$ by about 54%, and for $k_x = k_y = 4$ by about 47%. The less the spatial dimension was downscaled, the higher the improvement when increasing the temporal downscaling. Thus, other than for RPN and low-frequent fields, here, the temporal factor is more important. For example with DCN, we induce more error when reconstructing from a coarsening with a temporal factor to 12 ($k_t = 12, k_x = k_y = 2$) than when reconstructing from a coarsening with a spatial factor of four ($k_t = 4, k_x = k_y = 4$), although the total number of grid points to start from was larger for $k_t = 12, k_x = k_y = 2$. In **Figure 3B**, we see at the example of $k_x = k_y = 2$, $k_t = 4$ that only the DCN network was able to learn for precipitation fields, and that the same RPN architecture that was used before on temperature was not able to reduce the loss, which explains the higher errors of RPN compared to the baseline.

**Figure 6** shows an example of downscaling from 20 to 5 min. The time steps that are sent into the network shown, in which a cloud movement from the top left to the bottom right is apparent, as well as how precipitation decreases over time. Using this information, the DCN network is able to estimate the position and the amount of precipitation at a specific intermediate time. **Figure 7** shows the error map of a conventional linear downscaling and our neural network prediction, where we can see that the DCN output is closer to the ground truth.

### 3.2.2. Application of Deep Learning to Observational Data

Given the experiments on simulated data, another interesting question is to see if the model is able to learn how to downscale observational data. For this, we run eight instances of our model training on the observational data and performed downscaling between different pairs of resolutions. We checked how the model can downscale from 2, 4, 6, 12, and 24 hourly data to 1 h intervals. Additionally, we evaluated the downscaling from 12 and 24 h data to 6 h data, and from 24 h data to 12 h. The results are summarized in **Table 3**. We observe that for small downscaling factors like from 2 to 1 h data, our model is able to reduce the error compared to the baseline by 24.65%. Increasing the downscaling factor decreases the performance and gets worse for high factors like 12 or 24–1 h data. For such extreme downscaling, not enough information is present to disambiguate the correct hourly information. When downscaling smaller factors but on coarser resolution, i.e., with a downscaling factor of 2 but from 12 to 6 h data, the model is able to improve significantly over the baseline and for the extreme case of downscaling from daily data to 12 h data it achieved an

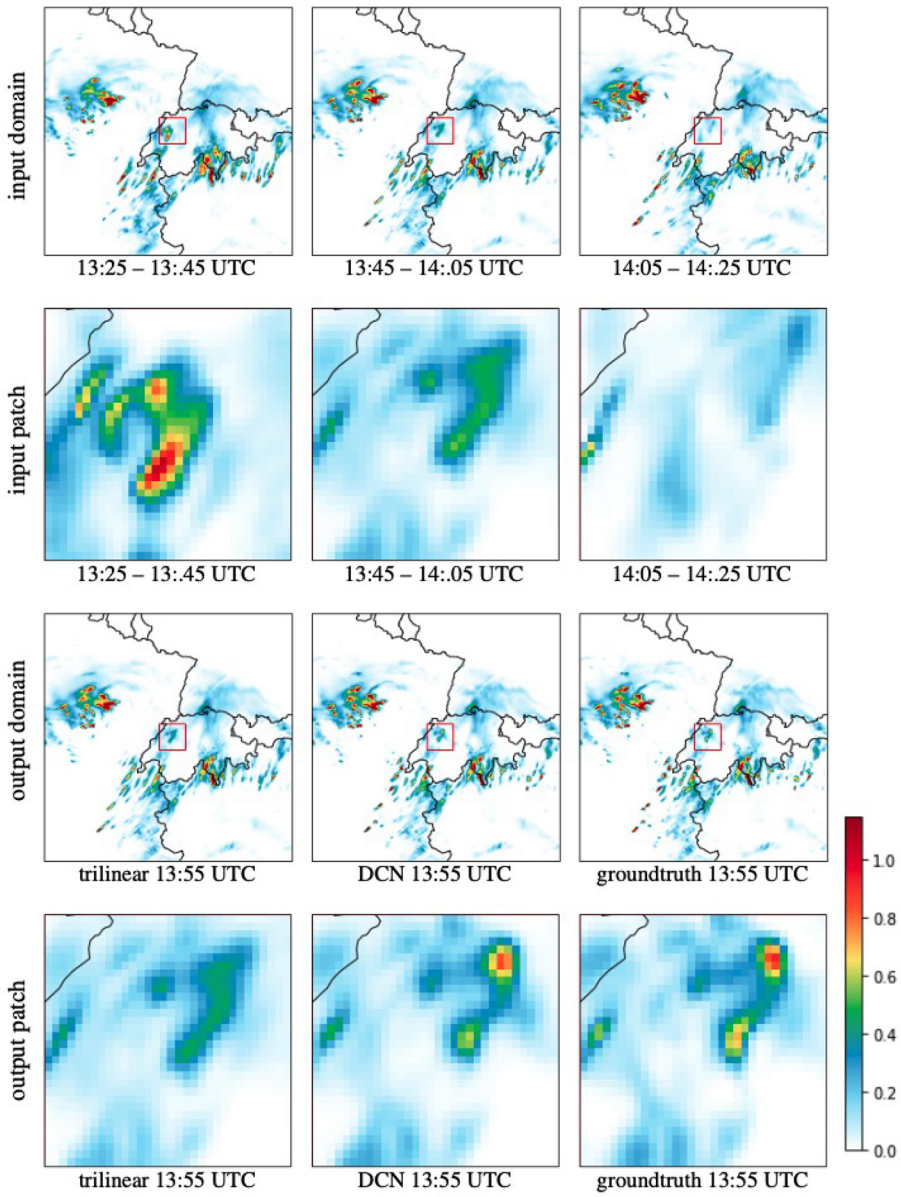



**FIGURE 6 |** Downscaling results for precipitation (g/m$^2$) by a factor of $k_t = 4$ from 20 to 5 min resolution. The first row shows time steps of the trivially downscaled domain (spatially). The second row shows a patch as it is sent into the network. The third row compares the result of the network to the groundtruth on the full domain. The last row shows the groundtruth comparison for the patch that was predicted.

error reduction of up to 70%. **Figure 8** shows an example of this downscaling scenario.

## 4. CONCLUSIONS

In this paper, we investigated the suitability of deep learning for the reconstruction of numerically-simulated and observed low-resolution data in both space and time. We thereby concentrated on two meteorological variables—temperature and precipitation—for which we develop suitable network architectures that utilize additional time-invariant information, such as the topography. We decided on temperature and precipitation to assess the performance of a neural network on both low-frequent and very high-frequent fields in order to test the limits of the architectures. While we observed that slowly-changing information, such as temperature can be adequately predicted through an error-predicting network, we found that fields with larger variations in both space and time, such as precipitation, require a different approach and cannot profit from residual learning, as there is no straight-forward downscaling method to leverage which achieves

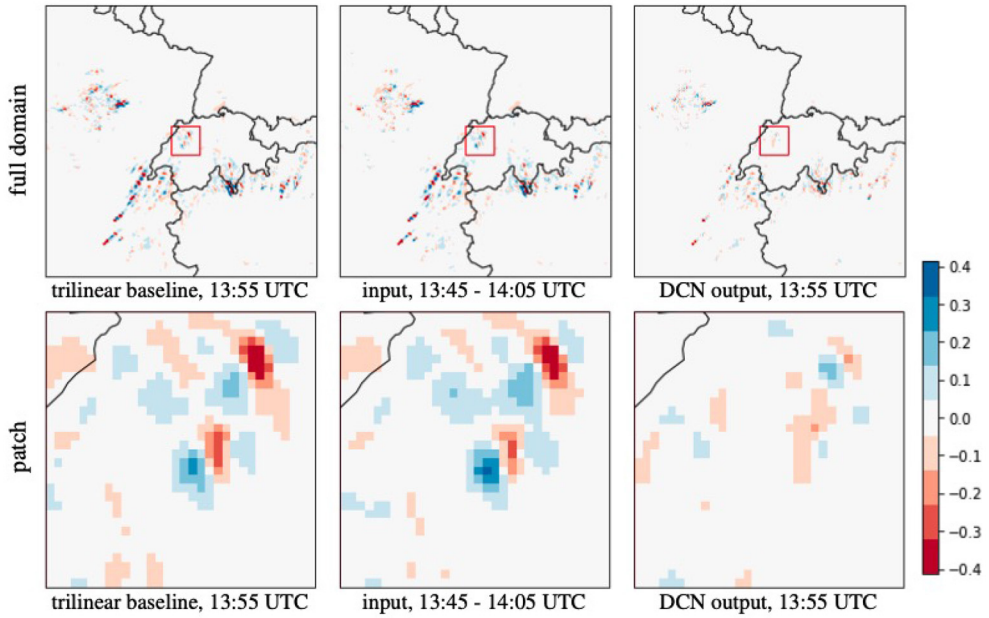



**FIGURE 7 |** The error map of **Figure 6** compares the trilinear interpolated baseline, the time-integrated network input covering the time to predict, and our predicted output. DCN reduces the MSE from 2.269 to 0.745 g/m$^2$ on the full domain, and from 4.531 to 0.910 g/m$^2$ in the shown patch.

**TABLE 3 |** Observation downscaling results in temporal dimension, using the baseline and our DCN.

| MSE (kg/m$^2$) factors | Baseline | DCN | Reduction (%) |
|---|---|---|---|
| 2–1 h | 0.229 | 0.173 | −24.65 |
| 4–1 h | 0.571 | 0.551 | −3.522 |
| 6–1 h | 0.557 | 0.534 | −4.205 |
| 12–1 h | 0.647 | 0.649 | 0.283 |
| 24–1 h | 0.697 | 0.762 | 9.348 |
| 12–6h | 6.254 | 4.531 | −27.558 |
| 24–6 h | 8.762 | 5.836 | −33.395 |
| 24–12 h | 12.925 | 3.802 | −70.586 |

*When targeting hourly outputs, the DCN is beneficial only for small scaling factors. For predictions of 6 and 12 h intervals, the network achieved a 30 or 70% error reduction, respectively.*

close enough baselines. Learning to suppress unnecessary or wrong structures is more difficult, then letting the network directly predict the high-resolution output by itself from the extracted features. For both cases, we developed a convolutional architecture with residual skip connections in order to extract features at different scales and to combine them in the subsequent deconvolution, leading us to a high-resolution prediction.

One possible reason why data is available at lower resolution only is that it has been coarsened for storage. If storage alone was the concern, it would be more effective to apply lossy compression approaches directly to the high-resolution data, especially if the data has low-frequent regions that could be sampled more sparsely than the uniformly chosen coarse resolution used throughout this manuscript for coarsening. That said, a limitation of the presented downscaling approach is that it is not able to compete with lossy compressions that were able to work from the high-resolution data. Instead, we focused on what can be recovered once the damage is done, i.e., once the data has been coarsened. Future work could follow up on the compression, for which an information theoretic approach would be instructive (MacKay, 2003; Yeung, 2010). In the future, it would be interesting to study if there are ways to predict the optimal downscaling parameters. This will be quite challenging, since the best network and the best parameter choice is strongly dependent on the data characteristics, which vary not only spatially but also temporally.

At present, we assumed that the meteorological data is available on regular grids. In such a case, convolutional layers proved useful for feature map extraction in the hidden layers. In the future, it would be interesting to study convolutional approaches for unstructured or irregularly structured data. Possible approaches would include PointNet-like convolutions (Qi et al., 2017) that waive connectivity information by introducing order-independent aggregations, or graph convolutional networks (Kipf and Welling, 2016) that operate on arbitrary grid topologies.

CNNs and GANs similarly share the problem that their interpretation is difficult, since both involve nonlinear mappings. For example, both of our CNN approaches RPN and DCN obtain an error that is theoretically unbounded. It would be imaginable to bound the reconstruction heuristically using the coarse input data, for example by only allowing a certain

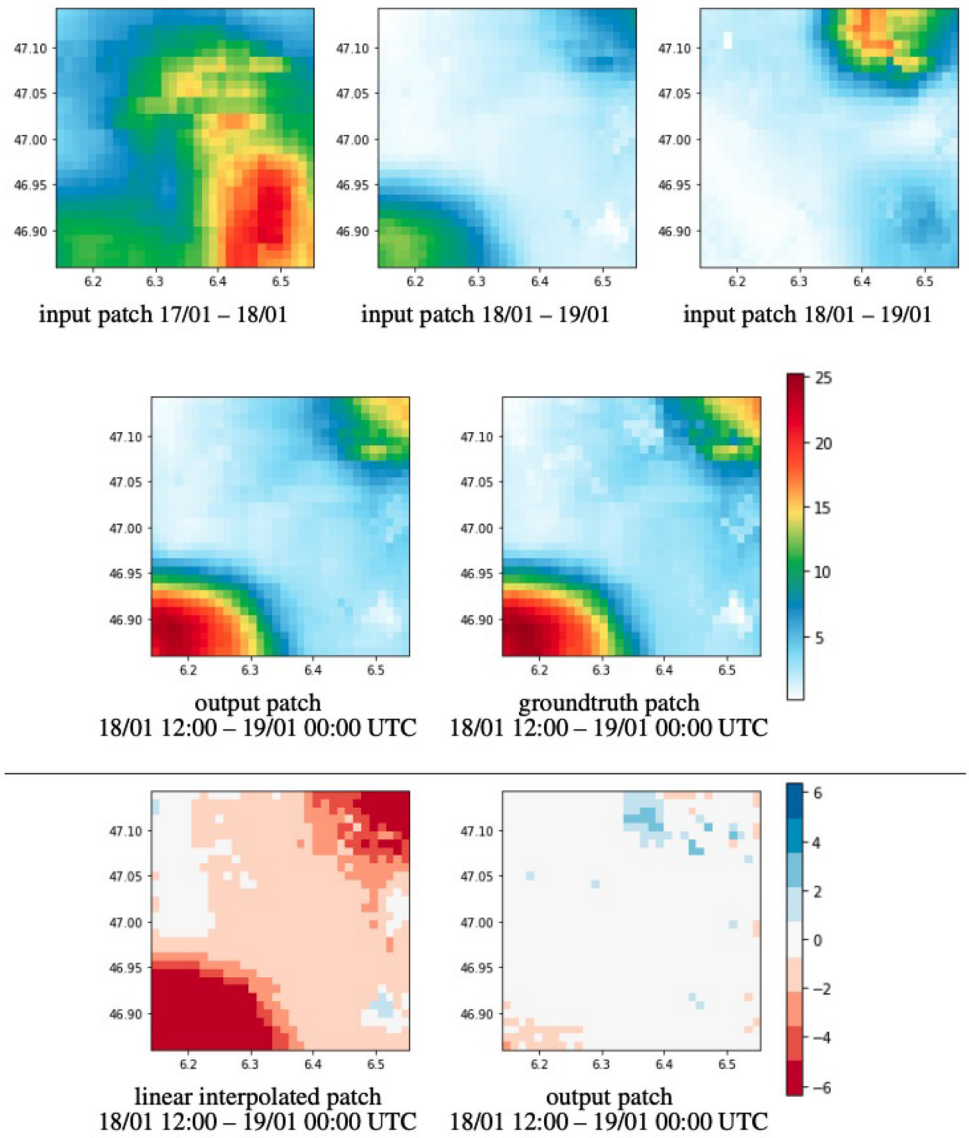



**FIGURE 8 |** Downscaling results of observational precipitation from 24 h resolution to 12 h. Top row shows input patches ranging from 1 day 12:00 UTC to the other 12:00 UTC. Middle row shows the predicted output patch and the groundtruth. Last row compares the baseline and predicted error maps relative to the groundtruth (MSE kg/m$^2$). Our method learns to scale the amount of precipitation dependent on the time we estimate. Our method can reduce the MSE from 14.251 to 0.252.

deviation away from the input signal, but this would of course be rather heuristic. Extreme weather events could be smoothed out since the frequency of their occurrence was not accounted for in the training data. Weighting individual training samples is an interesting direction for future work, which would require more data and an identification of the extreme events.

Neural networks can learn to disambiguate the reconstruction from low-resolution to high-resolution data in a data-driven way. In the future, it would be interesting to include additional regularizers into the loss function to utilize physical conservation laws that needed to hold during simulation. Further, it would be interesting to apply residual predictions to dynamical downscaling models, as this would build up on the meteorological knowledge that went into the design of dynamical models. While running the dynamical models also imposes a computational cost, there is great potential in including more physics into the learning process.

The work presented here shows a proof of concept how neural networks can be used to reconstruct data that has been coarsened, and how this could serve for development/reconstruction of high-resolution model data and observations. For example, trained networks can be

used for disaggregation of daily observational values into subdaily instead of using functions that can introduce statistical artifacts. It still remains to expand the current study to different domains and to longer time periods and it still remains an open problem to investigate if and how the hallucinations of generative neural networks (Singh et al., 2019; Han and Wang, 2020; Stengel et al., 2020) might impede the data analysis.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: gridded precipitation observations are obtained through MeteoSwiss data service. The data is available for research upon signing a license agreement. Due to the high data volume, the COSMO model data is available on request from the authors. The source code for the construction of the DCN and RPN, as well as the gradient loss calculation are available on GitHub at http://github.com/aserifi/convolutional-downscaling. Requests to access these datasets should be directed to the authors.

## AUTHOR CONTRIBUTIONS

AS conducted the analysis and wrote the manuscript. TG conceived the idea and NB provided the data. TG and NB have supervised the work with regular inputs and have contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Swiss National Science Foundation (SNSF) Ambizione grant no. PZ00P2_180114.

## ACKNOWLEDGMENTS

We acknowledge PRACE for awarding us access to Piz Daint at Swiss National Supercomputing Center (CSCS, Switzerland). We also acknowledge the Federal Office for Meteorology and Climatology MeteoSwiss, the Swiss National Supercomputing Centre (CSCS), and ETH Zürich for their contributions to the development of the GPU-accelerated version of COSMO.

## REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). "Optuna: a next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK), 2623–2631. doi: 10.1145/3292500.3330701

Ban, N., Schmidli, J., and Schär, C. (2014). Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *J. Geophys. Res. Atmos.* 119, 7889–7907. doi: 10.1002/2014JD021478

Baño-Medina, J. L., García Manzanas, R., and Gutiérrez Llorente, J. M. (2020). Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geosci. Model Dev.* 13, 2109–2124. doi: 10.5194/gmd-2019-278

Cheng, J., Kuang, Q., Shen, C., Liu, J., Tan, X., and Liu, W. (2020). Reslap: generating high-resolution climate prediction through image super-resolution. *IEEE Access* 8, 39623–39634. doi: 10.1109/ACCESS.2020.2974785

Dong, C., Loy, C. C., He, K., and Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 295–307. doi: 10.1109/TPAMI.2015.2439281

Fuhrer, O., Osuna, C., Lapillonne, X., Gysi, T., Cumming, B., Bianco, M., et al. (2014). Towards a performance portable, architecture agnostic implementation strategy for weather and climate models. *Supercomput. Front. Innov.* 1, 45–62. doi: 10.14529/jsfi140103

Han, J., and Wang, C. (2020). SSR-TVD: spatial super-resolution for time-varying data analysis and visualization. *IEEE Trans. Vis. Comput. Graph.* doi: 10.1109/TVCG.2020.3032123. [Epub ahead of print].

Hentgen, L., Ban, N., and Schär, C. (2019). Clouds in convection resolving climate simulations over Europe. *J. Geophys. Res. Atmos.* 124, 3849–3870. doi: 10.1029/2018JD030150

Höhlein, K., Kern, M., Hewson, T., and Westermann, R. (2020). A comparative study of convolutional neural network models for wind field downscaling. *Meteorol. Appl.* 27:e1961. doi: 10.1002/met.1961

Kim, B., Azevedo, V. C., Thuerey, N., Kim, T., Gross, M., and Solenthaler, B. (2019). Deep fluids: a generative network for parameterized fluid simulations. *Comput. Graph. Forum* 38, 59–70. doi: 10.1111/cgf.13619

Kim, B., and Günther, T. (2019). Robust reference frame extraction from unsteady 2D vector fields with convolutional neural networks. *Comput. Graph. Forum* 38, 285–295. doi: 10.1111/cgf.13689

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv* 1412.6980.

Kipf, T. N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv* 1609.02907.

Leutwyler, D., Lüthi, D., Ban, N., Fuhrer, O., and Schär, C. (2017). Evaluation of the convection-resolving climate modeling approach on continental scales. *J. Geophys. Res. Atmos.* 122, 5237–5258. doi: 10.1002/2016JD026013

Li, H., Xu, Z., Taylor, G., Studer, C., and Goldstein, T. (2018). "Visualizing the loss landscape of neural nets," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Red Hook, NY: Curran Associates Inc.), 6391–6401. doi: 10.5555/3327345.3327535

Lu, Z., and Chen, Y. (2019). Single image super resolution based on a modified U-net with mixed gradient loss. *arXiv* 1911.09428.

Ma, K. (2009). In situ visualization at extreme scale: challenges and opportunities. *IEEE Comput. Graph. Appl.* 29, 14–19. doi: 10.1109/MCG.2009.120

MacKay, D. J. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.

Mailhot, A., and Duchesne, S. (2010). Design criteria of urban drainage infrastructures under climate change. *J. Water Resour. Plan. Manage.* 136, 201–208. doi: 10.1061/(ASCE)WR.1943-5452.0000023

Mao, X. J., Shen, C., and Yang, Y. B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *arXiv* 1603.09056.

Min, M., Li, J., Wang, F., Liu, Z., and Menzel, W. P. (2020). Retrieval of cloud top properties from advanced geostationary satellite imager measurements based on machine learning algorithms. *Rem. Sens. Environ.* 239:111616. doi: 10.1016/j.rse.2019.111616

O'Gorman, P. A., and Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: potential for modeling of climate, climate change, and extreme events. *J. Adv. Model. Earth Syst.* 10, 2548–2563. doi: 10.1029/2018MS001351

Pouliot, D., Latifovic, R., Pasher, J., and Duffe, J. (2018). Landsat super-resolution enhancement using convolution neural networks and sentinel-2 for training. *Rem. Sens.* 10:394. doi: 10.3390/rs10030394

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., et al. (2015). A review on regional convection-permitting climate modeling: demonstrations, prospects, and challenges. *Rev. Geophys.* 53, 323–361. doi: 10.1002/2014RG000475

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017). Pointnet++: deep hierarchical feature learning on point sets in a metric space. *arXiv* 1706.02413.

Rodrigues, E. R., Oliveira, I., Cunha, R., and Netto, M. (2018). "Deepdownscale: a deep learning strategy for high-resolution weather forecast," in *2018 IEEE 14th*

*International Conference on e-Science (e-Science)* (Amsterdam: IEEE), 415–422. doi: 10.1109/eScience.2018.00130

Roesch, I., and Günther, T. (2019). Visualization of neural network predictions for weather forecasting. *Comput. Graph. Forum* 38, 209–220. doi: 10.1111/cgf.13453

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Munich: Springer), 234–241. doi: 10.1007/978-3-319-24574-4_28

Sachindra, D., Ahmed, K., Rashid, M. M., Shahid, S., and Perera, B. (2018). Statistical downscaling of precipitation using machine learning techniques. *Atmos. Res.* 212, 240–258. doi: 10.1016/j.atmosres.2018.05.022

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Di Girolamo, S., et al. (2020). Kilometer-scale climate models: prospects and challenges. *Bull. Am. Meteorol. Soc.* 101, E567–E587. doi: 10.1175/BAMS-D-18-0167.2

Seifert, A., and Rasp, S. (2020). Potential and limitations of machine learning for modeling warm-rain cloud microphysical processes. *J. Adv. Model. Earth Syst.* 12:e2020MS002301. doi: 10.1029/2020MS002301

Selbesoglu, M. O. (2019). Prediction of tropospheric wet delay by an artificial neural network model based on meteorological and gnss data. *Eng. Sci. Technol.* 23, 967–972. doi: 10.1016/j.jestch.2019.11.006

Shi, W., Caballero, J., Theis, L., Huszar, F., Aitken, A., Ledig, C., et al. (2016). Is the deconvolution layer the same as a convolutional layer? *arXiv* 1609.07009.

Singh, A., White, B. L., and Albert, A. (2019). "Downscaling numerical weather models with GANs," in *AGU Fall Meeting 2019* (San Francisco, CA: AGU).

Stengel, K., Glaws, A., Hettinger, D., and King, R. N. (2020). Adversarial super-resolution of climatological wind and solar data. *Proc. Natl. Acad. Sci. U.S.A.* 117, 16805–16815. doi: 10.1073/pnas.1918964117

Steppeler, J., Doms, G., Schattler, U., Bitzer, H., Gassmann, A., Damrath, U., et al. (2003). Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteor. Atmos. Phys.* 82, 75–96. doi: 10.1007/s00703-001-0592-9

Toderici, G., Vincent, D., Johnston, N., Jin Hwang, S., Minnen, D., Shor, J., et al. (2017). "Full resolution image compression with recurrent neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 5306–5314. doi: 10.1109/CVPR.2017.577

Vandal, T., Kodra, E., and Ganguly, A. R. (2019). Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. *Theor. Appl. Climatol.* (Halifax, NS) 137, 557–570. doi: 10.1007/s00704-018-2613-3

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R. (2017). "DeepSD: generating high resolution climate change projections through single image super-resolution: an abridged version," in *International Joint Conferences on Artificial Intelligence Organization*. doi: 10.1145/3097983.3098004

Wang, Y., Perazzi, F., McWilliams, B., Sorkine-Hornung, A., Sorkine-Hornung, O., and Schroers, C. (2018). "A fully progressive approach to single-image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (Salt Lake City, UT), 864–873. doi: 10.1109/CVPRW.2018.00131

Wang, Y., Wang, L., Wang, H., and Li, P. (2019). End-to-end image super-resolution via deep and shallow convolutional networks. *IEEE Access* 7, 31959–31970. doi: 10.1109/ACCESS.2019.2903582

Weyn, J. A., Durran, D. R., and Caruana, R. (2019). Can machines learn to predict weather? Using deep learning to predict gridded 500-HPA geopotential height from historical weather data. *J. Adv. Model. Earth Syst.* 11, 2680–2693. doi: 10.1029/2019MS001705

Wüest, M., Frei, C., Altenhoff, A., M.Hagen, Litschi, M., and Schär, C. (2010). A gridded hourly precipitation dataset for Switzerland using rain-gauge analysis and radar-based disaggregation. *Int. J. Climatol.* 30, 1764–1775. doi: 10.1002/joc.2025

Yang, C. Y., Ma, C., and Yang, M. H. (2014). "Single-image super-resolution: a benchmark," in *European Conference on Computer Vision* (Springer), 372–386. doi: 10.1007/978-3-319-10593-2_25

Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., and Liao, Q. (2019). Deep learning for single image super-resolution: a brief review. *IEEE Trans. Multimed.* 21, 3106–3121. doi: 10.1109/TMM.2019.2919431

Yeung, R. W. (2010). *). Information Theory and Network Coding, 1st Edn*. Springer Publishing Company, Incorporated.

Yuval, J., O'Gorman, P. A., and Hill, C. N. (2021). Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. *Geophys. Res. Lett.* 48:e2020GL091363. doi: 10.1029/2020GL091363

**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 Climate

ORIGINAL RESEARCH
published: 07 May 2021
doi: 10.3389/fclim.2021.654763

# A Comparative Study on the Skill of CMIP6 Models to Preserve Daily Spatial Patterns of Monsoon Rainfall Over India

*Adway Mitra**

*Centre of Excellence in Artificial Intelligence, Indian Institute of Technology Kharagpur, Kharagpur, India*

South Asian monsoon is a phenomena that plays out during June-September every year, due to the northward shift of the ITCZ which causes heavy rainfall over many countries of South Asia, including India. These rains are directly related to the lives and economic well-being of over a billion people. Indian monsoon is highly heterogeneous, due to the vast physiographic variations across the country. There is considerable interest among scientists and other stake-holders about possible future changes to Indian monsoon due to worldwide climate change. Simulations of future climate by global climate models under various scenarios can provide important clues for this. However, simulations of Indian monsoon in the historical period by global climate models under the CMIP5 family were found to be inaccurate in several aspects. Simulations by the new global climate models from the CMIP6 family are now available, and scientists are evaluating their ability to simulate Indian monsoon. In this work, we focus on one particular aspect of simulations by these models: the spatial distribution over daily rainfall over the Indian landmass during monsoon. We use a Machine Learning based probabilistic graphical model that can identify frequent spatial patterns of rainfall after creating a binary representation of rainfall. This model also helps us to identify spatial clusters, i.e., homogeneous regions within the Indian landmass with similar temporal characteristics of rainfall. We identify such frequent spatial patterns and spatial clusters from observed monsoon rainfall data, and also from simulations of monsoon rainfall by different CMIP6 models during the period 2000–2014. We evaluate the models by comparing the patterns and clusters identified from their simulations with those identified from observed data. We find that some of the CMIP6 models can simulate the spatial distribution of monsoon rainfall to a reasonable degree, but there are various limitations—most models underestimate extreme rainfall events and are unable to reproduce the regions of the landmass that are homogeneous with respect to rainfall.

**Keywords: CMIP6, Indian monsoon, spatial pattern, graphical model, simulation**

OPEN ACCESS

***Edited by:***
*Renata Goncalves Tedeschi, Vale Technological Institute, Brazil*

***Reviewed by:***
*Kanhu Charan Pattnayak, University of Leeds, United Kingdom*
*Feifei Luo, Institute of Atmospheric Physics, China*

***Correspondence:**
*Adway Mitra*
*adway@cai.iitkgp.ac.in*

***Specialty section:***
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

***Received:*** *17 January 2021*
***Accepted:*** *09 April 2021*
***Published:*** *07 May 2021*

***Citation:***
*Mitra A (2021) A Comparative Study on the Skill of CMIP6 Models to Preserve Daily Spatial Patterns of Monsoon Rainfall Over India. Front. Clim. 3:654763. doi: 10.3389/fclim.2021.654763*

## 1. INTRODUCTION

Every year, several countries in South Asia including India, Sri Lanka, Burma, Bangladesh, Pakistan receive heavy rainfall from the South Asian Monsoon system, roughly during the period June-September. It is caused by formation of a low-pressure region over North-western India, resulting in northward migration of the Intertropical Convergence Zone (ITCZ). Specifically in case of

India, the monsoon season accounts for about 80% of the annual rainfall in about 75% of the landmass, with the exception of some regions along the south-eastern coast and the hilly north-eastern region which receive substantial pre-monsoon (April–May) and post-monsoon (October–November) rainfall. Such rainfall is extremely important for sustenance of agriculture in India, which contributes to the lives and livelihood of over a billion people ( Gadgil and Gadgil, 2006). Indian monsoon is a highly complex phenomena, exhibiting significant spatial and temporal variations during its 4-month seasons, as discussed by Gadgil (2003) and Goswami and Chakravorty (2017). Indian monsoon is considered by many climate scientists to be linked to climatic phenomena in other parts of the world through teleconnections, such as El Nino-Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), and Madden-Julian Oscillations (MJO). However, it is well-known that in recent decades the monsoon circulation has significantly weakened, as pointed out by Ghosh et al. (2012) and Preethi et al. (2017), while at the same time extreme rainfall events have increased, according to Roxy et al. (2017). Naturally, there is significant concern among scientists and policy-makers about how Indian monsoon may be affected in future due to worldwide climate change. For this purpose, we need reliable simulation of future climate under different scenarios.

Over the last decade, various research groups across the world have developed global climate models such as General Circulation models (GCMs) under the aegis of Coupled Model Intercomparison Project (CMIP), with the aim of studying the impacts of various natural and anthropogenic forcings on past, present and future climate. Most of these models use physics-based differential equations about energy balance and coupling between land, ocean, and atmosphere. These models simulate global climate including many geophysical variables in the past and also the future under hypothetical scenarios related to greenhouse gas concentration in the environment (Representative Concentration Pathways) and socio-economic policies adopted by different countries (Shared Socioeconomic Pathways).

Simulations of the future by any model is hard to evaluate, since the ground truth is not known. In order to estimate the reliability of the future simulations by any model, it is necessary to evaluate its simulation of the historical period, for which we do have the ground truth. Usually, some important statistical measures are calculated from the model simulations, and compared to the corresponding measures calculated from the ground truth. The accuracy of a model is based on such comparisons. Various studies focus on various statistical measures for such evaluation. Simulations by the third phase of models (CMIP3) were not very accurate due to their course resolution and failure to take into account various environmental factors, but they were improved upon by the fifth phase of the models (CMIP5). Various studies such as Sperber et al. (2013) have compared the broad spatial and temporal characteristics of simulated monsoon rainfall in Southern Asia including India, and noticed a slight improvement in some aspects, though other aspects such as teleconnections are still not represented accurately. A similar study was done for simulation of daily maximum and minimum temperature and precipitation over China by models from both families in Sun et al. (2015). However, it was pointed out by various research groups (e.g., Saha et al., 2014; Shashikanth et al., 2014; Jayasankar et al., 2015; Pattnayak et al., 2017) that these models do not represent several characteristics of Indian Monsoon very accurately in their simulations of the historically observed period, and hence their future projections are less reliable. Singh et al. (2017) found that regionalized versions of these models, often called Regional Climate Model (RCM) could not help much. Some studies like Raju and Kumar (2014) have tried to combine the CMIP5 model simulations to improve the accuracy with respect to a few statistical measures, and identified a few models as suitable for India.

The sixth phase of models (CMIP6) which have been released recently, operate at much higher spatial resolutions and take into account more small-scale or localized processes. An excellent overview of these models is provided by Eyring et al. (2016). CMIP6 models such as Wu et al. (2019) have raised hopes of researchers. Some research such as Gusain et al. (2020), have already explored the improvements in the representation of Indian Monsoon and its different characteristics in the simulations of the historical period by some of these models. The monsoon characteristics studied by Gusain et al. (2020) from these model simulations include mean seasonal precipitation across Indian landmass, seasonal climatology of daily rainfall during the June-September period over the *Monsoon Zone* of Central India, extreme rainfall events across the landmass, and the duration and frequency of *active and break spells*—an important feature of Indian monsoon denoting intra-seasonal oscillations (studied in great details by many scientists such as Rajeevan et al., 2010; Nair et al., 2018). The study found that the CMIP6 models show a greater statistical consistency with observed data than models from CMIP5 or CMIP3 families, though the spatial variations are yet to be represented accurately. Similar studies have been made for other regions affected by monsoon systems, such as China (Xin et al., 2020). This study too focuses on the representation of various characteristics of rainfall over China, such as spatial correlation between simulated and observed data, mean seasonal precipitation across Chinese landmass, inter-annual trends of seasonal precipitation, and relation between rainfall and horizontal winds.

The aim of this work is to focus on *spatial distribution of daily monsoon rainfall* over India. Our aim is to see if the daily distribution of rainfall as simulated by these models have any similarity with the actual daily distribution of monsoon rainfall. However, since the model simulations are not synchronized with observations on daily basis, a direct day-by-day comparison between model simulations and observations is not possible. We wish to evaluate CMIP6 models by identifying frequent spatial patterns of daily rainfall in the simulations of monsoon by these models, and comparing these patterns with those identified from observed data. Spatial patterns have generally been considered as Empirical Orthogonal Functions (EOF) in the Earth Sciences community, including for spatial analysis of Indian Monsoon, such as Suhas et al. (2013). However, a different approach was considered in the recent works by Mitra et al. (2018), where a

model based on Machine Learning was used to create a binary representation of the precipitation data. This representation was used to create a few canonical spatial patterns, such that the spatial distribution (map) of rainfall on each day can be approximated using one of these patterns. Unlike EOF-based patterns, these patterns are not additive, and have both binary and real-valued representations. The binary representations are spatio-temporally coherent, and hence more comprehensive and suitable for studying different climatic variables, as done by Sharma et al. (2021). Additionally, the model is able to identify spatial clusters, i.e., compact regions on the landmass with similar intra-seasonal and inter-seasonal variation in rainfall. In this work we use the same approach to identify such canonical spatial patterns of daily rainfall and spatial clusters from monsoon rainfall data obtained from observations by India Meteorological Department (IMD) during the period 2000–2014. Next, we apply the same technique on the monsoon rainfall simulated by seven models from CMIP6 family, to identify spatial patterns and clusters from them. The patterns and clusters obtained from each of these models are compared to those obtained from the IMD observations. We define and evaluate several measures of compatibility between these patterns. On the basis of these measures, we classify the 7 CMIP6 models. It turns out that some of the models can capture the spatial patterns partially well, but not the others. None of the models are able to account for the heavy-to-extreme rainfall events. The regions of spatial homogeneity, as identified by the simulations from most of the models, are not very homogeneous with respect to the actual observations. Thus, we conclude that CMIP6 models are only somewhat accurate in reproducing daily spatial distribution of monsoon rainfall over India.

# 2. MATERIALS AND METHODS

## 2.1. Datasets

In this work, we consider precipitation data over the landmass of India during the monsoon months of June-September, for the period 2000–2014. The reason for considering this period is that it is recent and relatively less well-studied in literature. We obtain ground truth data from India Meteorological Department (IMD)[1]. We also collect the data related to simulation of Indian Monsoon rainfall by the following CMIP6 models: ACCESS-ESM1.5 developed by Australian Community Climate and Earth System Simulator (ACC) (see Ziehn et al., 2020 for details), Beijing Climate Center Climate System Model (BCC) (see Wu et al., 2019 for details), Canadian Earth System Model Version 5 (CAN) (see Swart et al., 2019 for details), Earth Consortium Model Version 3 (EC) (see Wyser et al., 2020 for details), Institute of Numerical Mathematics Climate Model Version 4.8 (INM) (see Volodin et al., 2018 for details), Max Plank Institute Earth System Model Version 1.2 (MPI) (see Mauritsen et al., 2019 for details), and Norwegian Climate Center Earth System Model Version 2.0 (NOR) (see Seland et al., 2020 for details).

The datasets specifically related to Indian Monsoon simulation by these models are available at https://zenodo.org/record/3873998#.X_g60dgzaUk. This dataset was created based on the study by Mishra et al. (2020). In these simulations, the precipitation data for Indian monsoon is available at a resolution of $0.25 \times 0.25^\circ$ (see Pai et al., 2014). For ease of computations, we coarsen them to $1 \times 1^\circ$ resolution using spatial averaging, following the same grid system as the widely-used dataset published by Rajeevan et al. (2006). For the ground-truth dataset also, we use the same grid structure. According to this grid system, the landmass of India consists of 357 grid-locations. For every location, we have daily rainfall data for the June-September season (122 days) of the period 2000–2014, i.e., we have totally $122 \times 15 = 1,830$ days. In case of the CMIP6 models mentioned above, the years are not synchronized to the actual years, so we must limit our analysis to statistics calculated across the years for a fair comparison.

## 2.2. Probabilistic Graphical Model

We fit a probabilistic graphical model developed by Mitra et al. (2018) on to each of these datasets. Consider $S$ locations and $T$ time-points, i.e., here $S = 357$ and $T = 1,830$. For every location $s$, we consider a set of neighboring locations $\Omega(s)$, which are within a distance of $1^\circ$ from $s$ along either latitudes or longitudes. We denote by $X(s,t)$, the precipitation at any spatial location $s$ on day $t$. For any day $t$, the vector $X(t) = \{X(s,t)\}_{s=1}^{S}$ is called the *spatial map* or spatial distribution of rainfall on that day. Similarly for any location $s$, the vector $X(s) = \{X(s,t)\}_{t=1}^{T}$ is called the *time-series* of rainfall at that location.

The model aims to find a binary representation $Z(s,t)$ of $X(s,t)$. $Z(s,t) = 1$ essentially indicates high value of rainfall at location $s$ on day $t$ (*wet day*), while $Z(s,t) = 2$ indicates a low value of the same (*dry day*). However, there is no hard threshold between high and low values. Assignment of $Z(s,t)$ is done based on local climatology of daily rainfall at $s$, and also on the influence of neighboring values $Z(s',t)$, where $s' \in \Omega(s)$ so that *spatio-temporal coherence* is maintained, i.e., neighboring locations are likely to have the same value of Z. Thus, for each day $t$ during the period under consideration, we have a *real-valued spatial map* $X(t)$ of rainfall as well as a *binary-valued spatial map* $Z(t)$ over the geographical domain. The model assigns each day $t$ to a cluster denoted by $U(t)$, such that days with "similar" spatial maps are assigned to the same cluster. A cluster is represented by a *binary spatial pattern* denoted by $\theta_d$ which is the *mode* of the binary spatial maps across the constituent days of that cluster, and also by a *real-valued spatial pattern* denoted by $\theta$ which is the *mean* of the real-valued spatial maps across the constituent days. Similarly, the model assigns a spatial cluster index $V(s)$ to each locations $s$, such that locations with "similar" time-series of $Z$-values over the entire period of $T$ days, are assigned to the same cluster. Also, neighboring locations are likely to be assigned to the same spatial clusters, so that spatial compactness of the clusters is maintained. Just like clusters of $U$, each cluster of $V$ is represented by a canonical time-series $\phi_d$ of binary values and $\phi$ of real values.

In this model, all of the variables $X$, $Z$, $U$, $V$, are considered as random variables, while $\theta$, $\theta_d$, $\phi$, and $\phi_d$ are considered as unknown parameters to be estimated. First of all, we construct a probabilistic graphical model using $(Z, U, V)$, which is shown

[1]https://imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html.

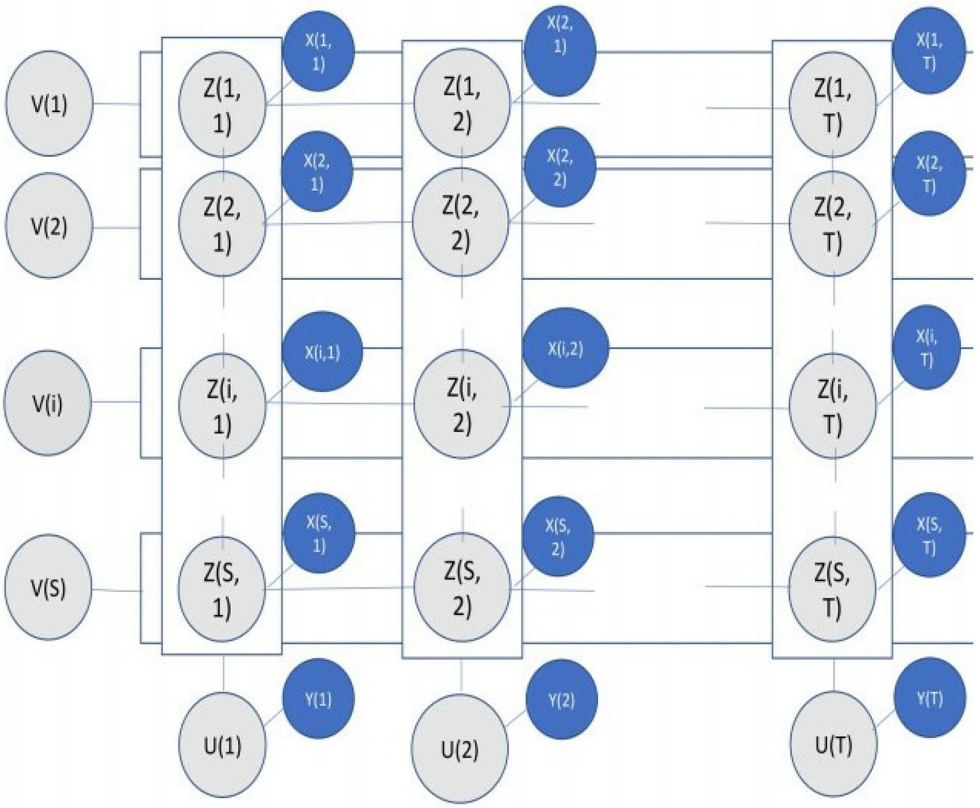



**FIGURE 1 |** The probabilistic graphical model. The rows indicate spatial locations $\{1, S\}$, and the columns indicate days $\{1, T\}$. The observation nodes for $X$ are marked in blue, while the latent variable nodes for $Z, U, V$ are marked in gray.

in **Figure 1**. Each node of the model represents a specific random variable, such as $X(s,t)$, $Z(s,t)$, $U(t)$, or $V(s)$. Any two $Z$-nodes are joined by *spatial edges* if they are spatially adjacent, for e.g., $Z(s,t)$ and $Z(s',t)$ where $s' \in \Omega(s)$. Again, two $Z$-nodes are joined by *temporal edges* if they are temporally adjacent, for e.g., $Z(s,t)$ and $Z(s,t+1)$. Again, $Z(s,t)$-node and $X(s,t)$-node are connected by *data edges*. Additionally, all $Z(s,t)$-nodes and $X(s,t)$-nodes for each day $t$ are connected to the node $U(t)$, though in the figure $U(t)$ is shown to be connected to $Y(t)$, which is a dummy node representing $Y(t) = \sum_{s=1}^{S} X(s,t)$. Similarly, all $Z(s,t)$-nodes of each location $s$ are connected to the corresponding $V(s)$-node. Each of the edges are provided with an *edge potential function* $\Psi$. The spatial edge potential functions are defined in such a way that it takes a high value when $Z(s,t) = Z(s',t)$, and low value otherwise. Temporal edge potentials are defined likewise. The data edge potentials between $Z(s,t)$ and $X(s,t)$ are defined as the PDF of a Gamma distribution on $X$, whose shape and scale parameters are specific to location $s$ and the value of $Z(s,t)$, i.e., we assume that $X(s,t) \sim Gamma(\alpha_{sk}, \beta_{sk})$ with $k = Z(s,t)$. This means that the rainfall at any location is modeled by a Gamma mixture distribution with two modes [as $Z(s,t)$ is binary], one for high values and one for low values. The daily binary spatial map $Z$ for any day $t$ is modeled as a Bernoulli-corrupted version of the corresponding spatial pattern $\theta_d[U(t)]$. Similarly, the binary time series $Z$ for any location $s$ is modeled as a Bernoulli-corrupted version of the canonical time-series $\phi_d[V(s)]$.

This graphical model forms a Markov Random Field (Kindermann, 1980). The joint distribution $P(X, Z, U, V)$ is the product of all the edge potentials functions. Among the random variables mentioned above, $X$ is observed. The computation of values of $(Z, U, V)$, is done with the aim of maximizing $P(X, Z, U, V)$. Clearly, a configuration with more spatial and temporal coherence, will have a higher probability, due to nature of the potential functions on the spatial and temporal edges. Since standard maximum-likelihood or Expectation-Maximization are not tractable here due to the complex coupling between the variables, we use the approach of Gibbs Sampling under the paradigm of Markov Chain Monte Carlo methods (Neal, 1993). In this approach, we iteratively sample each random variable from its conditional distribution, holding all the other variables constant at their current values. The process is repeated hundreds of times, samples collected at regular intervals, and the modal values of these samples are used as Maximum A-posteriori (MAP) estimates of the random variables.

Interpretation of the values of $Z$, $\theta$, $\theta_d$, and $V$ are important to understanding the results of this model. First of all, $Z$ provides

**TABLE 1 |** The spatial statistics of $X$ and $Z$, as computed from CMIP6 model simulations and compared against the IMD observations.

| | IMD | ACC | BCC | CAN | EC3 | INM | MPI | NOR |
|---|---|---|---|---|---|---|---|---|
| $nZ_1 - \ell_2$ | 0 | 2.3 | **1.3** | 2.5 | 1.8 | 1.9 | 2.0 | 1.7 |
| $nZ_1 - cr$ | 1 | 0.77 | 0.91 | 0.78 | **0.92** | 0.89 | 0.87 | 0.9 |
| $X_1 - \ell_2$ | 0 | 103 | 68 | 135 | 60 | 90 | 74 | **59** |
| $X_1 - cr$ | 1 | 0.62 | 0.59 | 0.58 | **0.79** | 0.62 | 0.66 | 0.73 |
| $X_2 - \ell_2$ | 0 | 18 | **9** | 22 | 12 | 10 | **9** | 10 |
| $X_2 - cr$ | 1 | 0.55 | 0.72 | 0.58 | 0.74 | 0.78 | **0.80** | 0.77 |
| spCh | 0.85 | 0.93 | 0.91 | 0.94 | 0.92 | 0.91 | 0.9 | 0.92 |
| spCr | 0.59 | 0.89 | 0.88 | 0.93 | 0.86 | 0.89 | 0.86 | 0.9 |

*The measures are explained in section 3.1. For each measure, the CMIP6 model whose results are closest to the IMD data are highlighted.*

a binary representation of the observed data $X$. If we look at the spatial map of $X(t)$ on a particular day $t$, then its corresponding binary representation $Z(t)$ is a binary map where locations having *high* rainfall have $Z(t) = 1$, while those locations having *low* rainfall have $Z(t) = 2$. These binary maps are more spatially coherent than the real-valued ones, where high-rainfall and low-rainfall regions are more clearly demarcated. Coming to the spatial patterns $\theta, \theta_d$, each day's spatial map $X(t)$ can be approximated with a real-valued $\theta$-pattern while each day's binary spatial map $Z(t)$ can be approximated with a binary $\theta_d$-pattern.

The number of canonical patterns is not fixed, but estimated by the model based on the data. There are user-tuneable hyperparameters (mentioned in Mitra et al., 2018) which indicate how closely a canonical pattern must approximate the daily spatial maps/patterns, which have the effect of increasing or decreasing the number of canonical patterns. But generally about 10 patterns, each of which account for at least 60 of the 1,830 days from at least 8 of the 15 years, can cover 70–90% of the days. We call such patterns as *prominent patterns*, and these patterns contain the information regarding the *usual* daily spatial distribution of rainfall. The remaining days which are assigned to non-prominent or rare patterns are days with unusually high rainfall, spread over large parts of the country.

## 3. RESULTS

Now, we come to the comparison of the different CMIP6 models with the observed data, as obtained from India Meteorological department. For this purpose, we fit the probabilistic graphical model discussed above to the daily rainfall observations $X$ from all these datasets (observations and CMIP6 model simulations). Parameters and hyperparameters used for the model (as listed by Mitra et al., 2018) are the same for each of the datasets, for meaningful comparison. Let us denote by $X^{IMD}$, $Z^{IMD}$, $\theta^{IMD}$, and $\theta_d^{IMD}$ the IMD observations and corresponding binary representation and spatial patterns. Similarly, we denote by $X^{MODEL}$, $Z^{MODEL}$, $\theta^{MODEL}$, and $\theta_d^{MODEL}$ the daily rainfall values and corresponding binary representation and spatial patterns from any CMIP6 model (for specific models, we will use $X^{BCC}$, $\theta^{NOR}$ etc).

### 3.1. Quantitative Analysis of the Binary Representations

We begin our quantitative comparison of the CMIP6 model simulations and the actual observations from IMD in terms of interpreting the binary representations, i.e., $Z^{IMD}$ and $Z^{MODEL}$. For each of the 357 locations, we compute the mean rainfall values across all *wet days* for which $Z = 1$ and also for all *dry days* for which $Z = 2$ separately. We also calculate the fraction of wet days ($Z = 1$) at each location over the study period. These indicate how wet are the *wet and dry days* in different locations, in the actual IMD dataset as well as in the simulations by CMIP6 models.

We denote these quantities by $nZ_1(s) = \frac{1}{T}\sum_{t=1}^{T} I(Z(s,t) = 1)$, $X_1(s) = \frac{\sum_{t=1}^{T} X(s,t) I(Z(s,t)=1)}{\sum_{t=1}^{T} I(Z(s,t)=1)}$, $X_2(s) = \frac{\sum_{t=1}^{T} X(s,t) I(Z(s,t)=2)}{\sum_{t=1}^{T} I(Z(s,t)=2)}$ where $I$ is the indicator function. Since these quantities are calculated at all $S$ locations, we compare them between the model simulation datasets and the actual IMD observations using $\ell_2$ norm (denoted by $nZ_1 - \ell_2, X_1 - \ell_2, X_2 - \ell_2$) and correlation coefficient (denoted by $nZ_1 - cr, X_1 - cr, X_2 - cr$). The results are shown in **Table 1**.

Another issue which we consider is *spatial correlation*—whether the rainfall volume at adjacent locations are correlated or not. For every pair of locations $s, s'$ such that $s' \in \Omega(s)$, we compute the spatial correlation coefficient between the time-series $X(s)$ and $X(s')$, denoted by $spCr(s, s')$. Similarly, we compute the spatial coherence between the binary time-series $Z(s)$ and $Z(s')$, i.e., $spCh(s, s') = \frac{1}{T}\sum_{t=1}^{T} I[Z(s,t) = Z(s',t)]$. The mean values of these quantities are then computed over all pairs of $(s, s')$. These are repeated for both the IMD dataset as well the CMIP6 simulation datasets, and shown in **Table 1**.

It turns out that for most of the measures related to local statistics, the models BCC-CSM, EC-3, MPI-ESM-1,2, and NOR-ESM-2 perform comparable to each other, while ACC-ESM-1.5, CAN-ESM-5, and INM-CM-4.8 are found to lag behind. The spatial coherence and spatial correlation are *significantly overestimated* in all the models, which may indicate that only processes at larger spatial scales are simulated by these models.

### 3.2. Visual Analysis of Prominent Spatial Patterns

In this paper, we are specifically interested in the spatial patterns of daily rainfall, as obtained from the different CMIP6 models

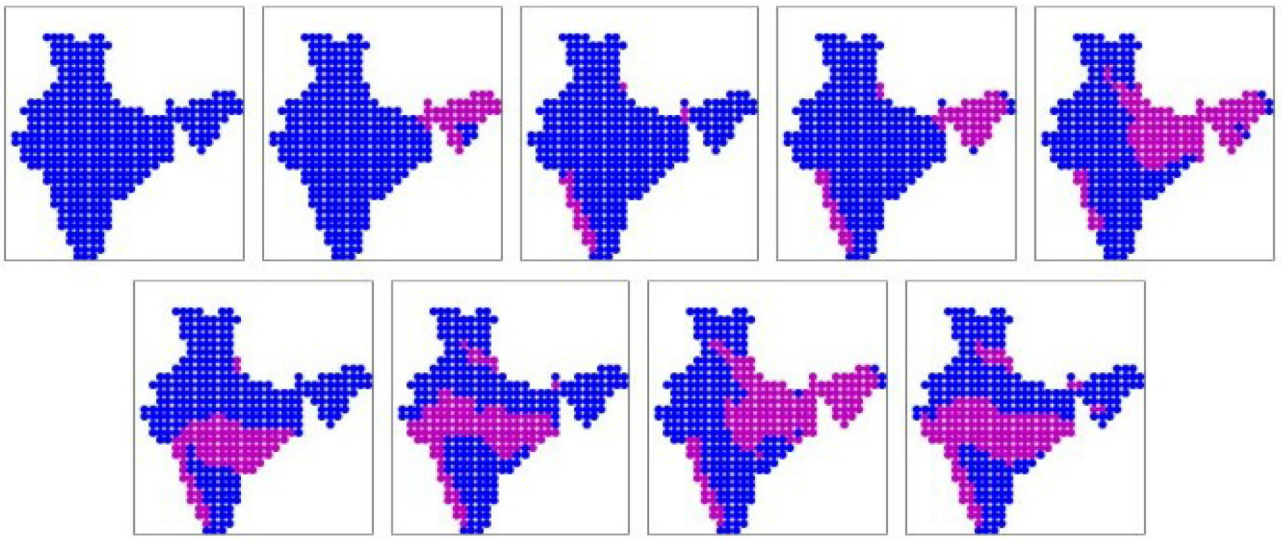


**FIGURE 2 |** The nine prominent binary spatial patterns found from the observations from India Meteorological Department (IMD). Pink indicates *wet* locations and blue indicates *dry* locations. Patterns are numbered from left to right across each row, top to bottom.

**TABLE 2 |** Mean daily rainfall for prominent spatial patterns identified from different CMIP6 models in mm/day/location, sorted in ascending order.

| Model | Pat1 | Pat2 | Pat3 | Pat4 | Pat5 | Pat 6 | Pat7 | Pat8 | Pat9 | Pat10 | Pat11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| IMD | 2.4 | 4.4 | 5.5 | 7.7 | 7.9 | 7.9 | 8.5 | 11.5 | 11.6 | – | – |
| ACC | 1.2 | 2.9 | 4.8 | 5.3 | 7.2 | 7.8 | 11.0 | – | – | – | – |
| BCC | 1.5 | 3.4 | 4.7 | 6.6 | 7.6 | 7.6 | 9.5 | 9.9 | 9.9 | 13.1 | – |
| CAN | 1.2 | 2.2 | 2.7 | 3.2 | 4.0 | 4.7 | 6.1 | 7.4 | 9.8 | – | – |
| EC3 | 1.7 | 3.8 | 6.0 | 8.0 | 8.8 | 9.0 | 11.6 | 11.6 | 14.8 | – | – |
| INM | 2.1 | 3.2 | 6.0 | 7.5 | 7.7 | 8.1 | 8.4 | 10.0 | 10.9 | 11.2 | 12.2 |
| MPI | 1.6 | 3.8 | 5.8 | 5.9 | 6.3 | 8.2 | 8.4 | 10.7 | 12.0 | – | – |
| NOR | 1.7 | 2.8 | 5.2 | 5.6 | 7.2 | 8.6 | 9.2 | 10.3 | 10.6 | 13.0 | – |

and from the observations. For each dataset, we identify the spatial patterns $\theta_d, \theta$ as mentioned in section 2. For each dataset, we focus on the set of *prominent* spatial patterns, which are a subset of the spatial patterns identified by the probabilistic graphical model. As already mentioned, a prominent spatial pattern appears on at least 60 of the 1,830 days, from at least 8 of the 15 years during the study period considered. It turns out that for all datasets, there are 7–11 prominent spatial patterns.

### 3.2.1. IMD Dataset

The prominent patterns (binary) obtained from the IMD dataset are shown in **Figure 2**. There are nine prominent spatial patterns, which cover 94% of the 1,830 days in the study period. The patterns are sorted in ascending order of mean all-India rainfall, as indicated in **Table 2**. The first three patterns are associated with low rainfall, either scattered or limited to the North-eastern region (pattern 2) or the western coast (pattern 3). In pattern 4 too the rainfall is mostly limited to the western coast and North-east, though it is heavier magnitude. In patterns 5 and 8, the wet areas are mostly in the Gangetic plain and foot-hills of the Himalayas, while in patterns 6, 7, and 9 the rainfall is concentrated in the monsoon zone of Central India. These patterns seem to indicate that rainfall does not happen simultaneously in Gangetic plains and Central India. Although the patterns 5, 8 as well as 6, 7, 9 look nearly identical in the binary representation, they differ in the volumes of rainfall associated with them. These rainfall volumes are indicated in the real-valued versions of these prominent patterns are shown in the **Supplementary Figure 1**.

### 3.2.2. ACCESS-ESM-1.5 Dataset

The seven prominent patterns (binary) obtained from the ACCESS simulation dataset are shown in **Figure 3**. The patterns are sorted according to the mean daily rainfall rates (all-India), as given in **Table 2**. These seven patterns cover 79% of the days. Here we find 2 patterns where most of the landmass is dry, just like the IMD dataset. In patterns 3, 4, 5, 6, 7 we find the rainfall concentrated along the western coast and the foothills of the Himalayas, North of the Gangetic planes. None of the patterns show much rainfall along major parts of the Gangetic

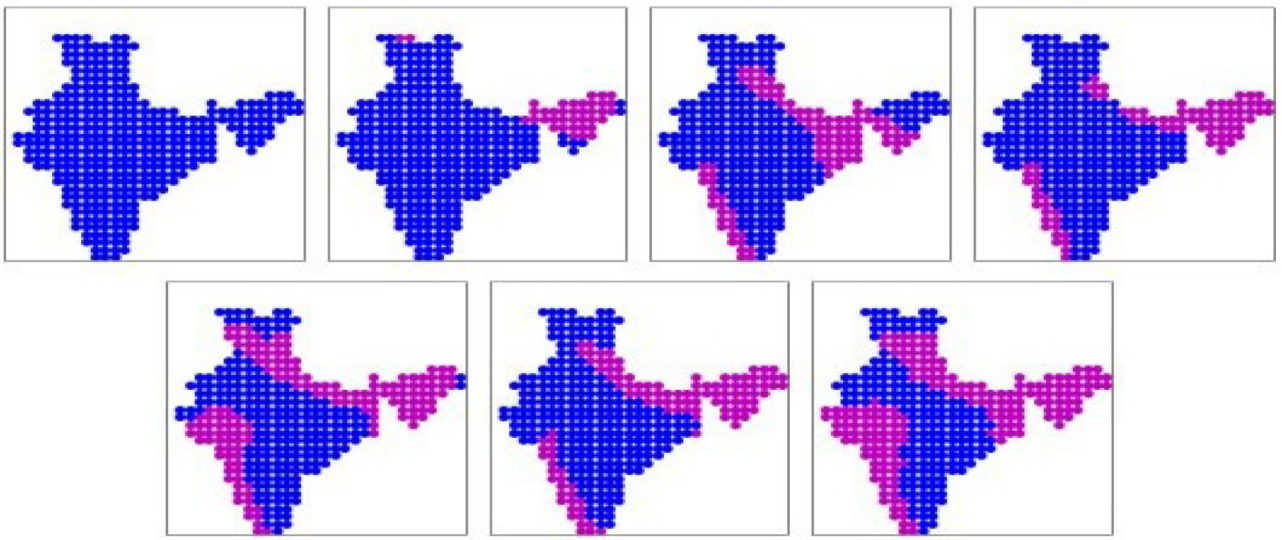


**FIGURE 3 |** The seven prominent binary spatial patterns found from ACCESS-ESM-1.5 model. Pink indicates *wet* locations and blue indicates *dry* locations. Patterns are numbered from left to right across each row, top to bottom.

plains, Central India and the Eastern coast. Some parts of the central India are covered in patterns 5 and 7, but simultaneously with Himalayan foothills. So many of the patterns obtained from simulations of this model do not agree well with IMD dataset. The real-valued versions of these prominent patterns are shown in **Supplementary Figure 2**.

### 3.2.3. BCC-CSM Dataset

The 10 prominent patterns (binary) obtained from the BCC-CSM simulation dataset are shown in **Figure 4**. The patterns are sorted in ascending order of mean all-India rainfall rates, as given in **Table 2**. These 10 patterns cover 85% of the days. Here once again we find the first four patterns corresponding to low all-India rainfall, which are concentrated in the western coast and North-East. Pattern 5 covers the entire peninsular region including the south-eastern parts, which are known to remain dry during this period. This pattern is in disagreement with the patterns from the IMD dataset. Similarly pattern 8, which shows rainfall only in the eastern side (roughly the states of Bihar, Odisha, Bengal, and the North-east), is not found in the IMD dataset. Patterns 6, 9, 10 show rainfall in Central India, and 9, 10 include the eastern coast as well. These are broadly in agreement with the patterns 6, 7, 9 of the IMD dataset, though located a bit Northward. Pattern 7 shows rainfall in the Gangetic plain, much like patterns 5, 8 of the IMD dataset. The real-valued versions of these prominent patterns are shown in **Supplementary Figure 3**.

### 3.2.4. CAN-ESM-5 Dataset

The nine prominent patterns (binary) obtained from the CAN-ESM-5 simulation dataset are shown in **Figure 5**. The patterns are sorted in ascending order of mean all-India rainfall rates, as given in **Table 2**. These nine patterns cover 85% of the days. The patterns 1, 2, 3 resemble the corresponding dry patterns from the IMD dataset. But patterns 4, 5, 7, 8 show rainfall concentrated only in the eastern and north-eastern region, while in case of pattern 9 most of the eastern coast and the peninsular India are simultaneously wet. These patterns seem to be in disagreement with the IMD patterns. There is no pattern which covers Central India (except 9, though only partially) and western parts of the Gangetic plain including the foothills of Himalayas. The real-valued versions of these prominent patterns are shown in **Supplementary Figure 4**.

### 3.2.5. EC-3 Dataset

The nine prominent patterns (binary) obtained from the EC-3 simulation dataset are shown in **Figure 6**. The patterns are sorted in ascending order of mean all-India rainfall rates, as given in **Table 2**. These 11 patterns cover 75% of the days. Once again, patterns 1, 2, 3 show rainfall limited to the north-eastern region and western coast, like the IMD dataset. Patterns 4, 5, and 7 show rainfall over the Gangetic plain and foothills of the Himalayas, like patterns 5, 8 from the IMD dataset. But patterns 6, 8, 9 show rainfall occurring simultaneously over Central India and large parts of the Gangetic plain, which is in disagreement with the patterns from the IMD dataset. The real-valued versions of these prominent patterns are shown in **Supplementary Figure 5**.

### 3.2.6. INM-CM-4.8 Dataset

The 11 prominent patterns (binary) obtained from the INM-CM-4.8 simulation dataset are shown in **Figure 7**. The patterns are sorted in ascending order of mean all-India rainfall rates, as given in **Table 2**. These 11 patterns cover 82% of the days. Here we have dry patterns 1, 2, 3 like all other datasets. Pattern 5 shows rainfall over the Gangetic plain, while patterns 6, 11 shows rainfall over Central India, like the patterns from the IMD dataset. However, patterns 7 shows rainfall along the entire eastern coast, which is not observed in the IMD dataset. In patterns 4, 5, 8, 9, 10 we see rainfall simultaneously in Central India and the Gangetic

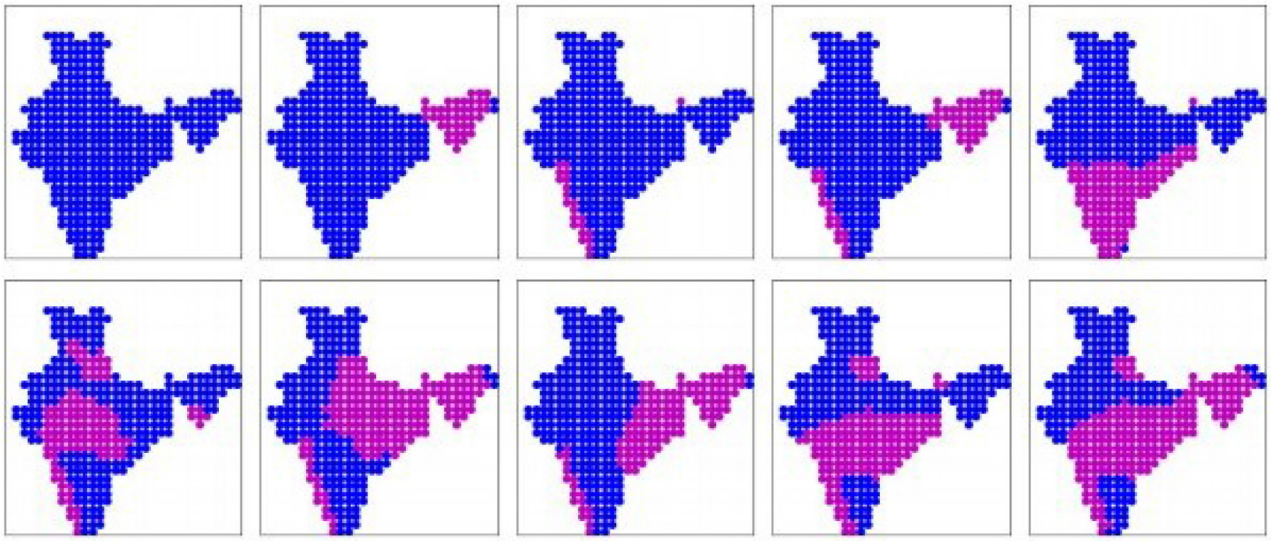


**FIGURE 4 |** The 10 prominent binary spatial patterns found from BCC-CSM model. Pink indicates *wet* locations and blue indicates *dry* locations. Patterns are numbered from left to right across each row, top to bottom.

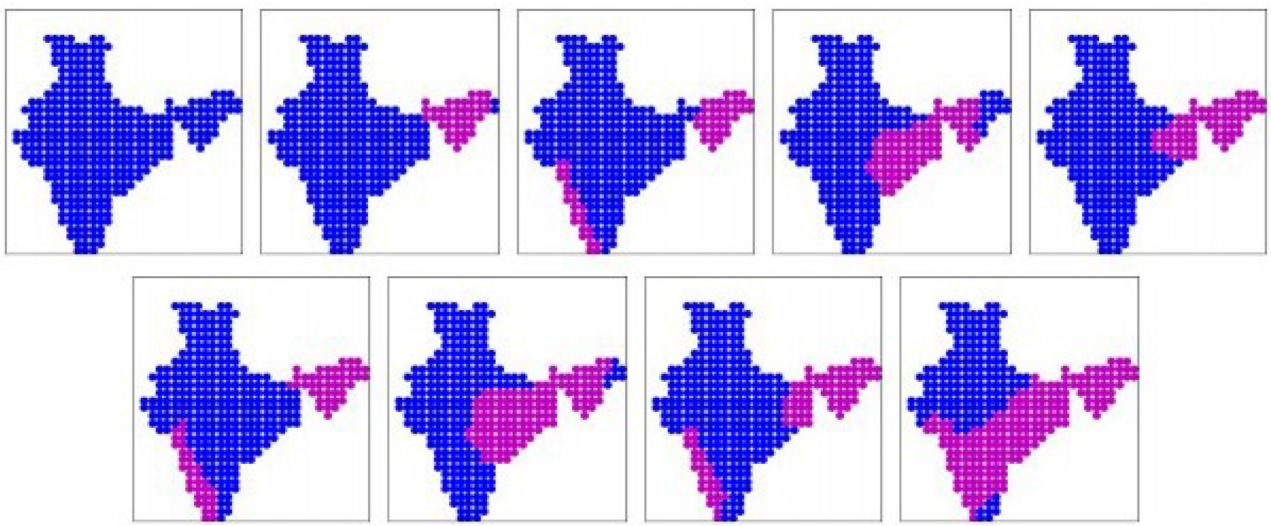


**FIGURE 5 |** The nine prominent binary spatial patterns found from CAN-ESM-5 model. Pink indicates *wet* locations and blue indicates *dry* locations. Patterns are numbered from left to right across each row, top to bottom.

plain, which too is in disagreement with the IMD patterns. The real-valued versions of these prominent patterns are shown in **Supplementary Figure 6**.

### 3.2.7. MPI-ESM-1.2 Dataset

The nine prominent patterns (binary) obtained from the MPI-ESM-1.2 simulation dataset are shown in **Figure 8**. The patterns are sorted in ascending order of mean all-India rainfall rates, as given in **Table 2**. These nine patterns cover 75% of the days. Just like all other datasets, we have 3 dry patterns, but unlike other models, none of them show rainfall along the Western coast only. Patterns 6 and 9 show rainfall mostly over the Gangetic plain, like patterns 5, 8 from the IMD dataset. Pattern 8 shows rainfall over central India. But pattern 4 where the rainfall is over peninsular India only, as well as patterns 5,7 where the rainfall occurs from eastern coast (around Odisha state) till western parts of the Gangetic plains excluding the eastern parts of the Gangetic plain, are in disagreement with the IMD dataset. The real-valued versions of these prominent patterns are shown in **Supplementary Figure 7**.

### 3.2.8. NOR-ESM-2 Dataset

The 10 prominent patterns (binary) obtained from the NOR-ESM-2 simulation dataset are shown in **Figure 9**. The patterns

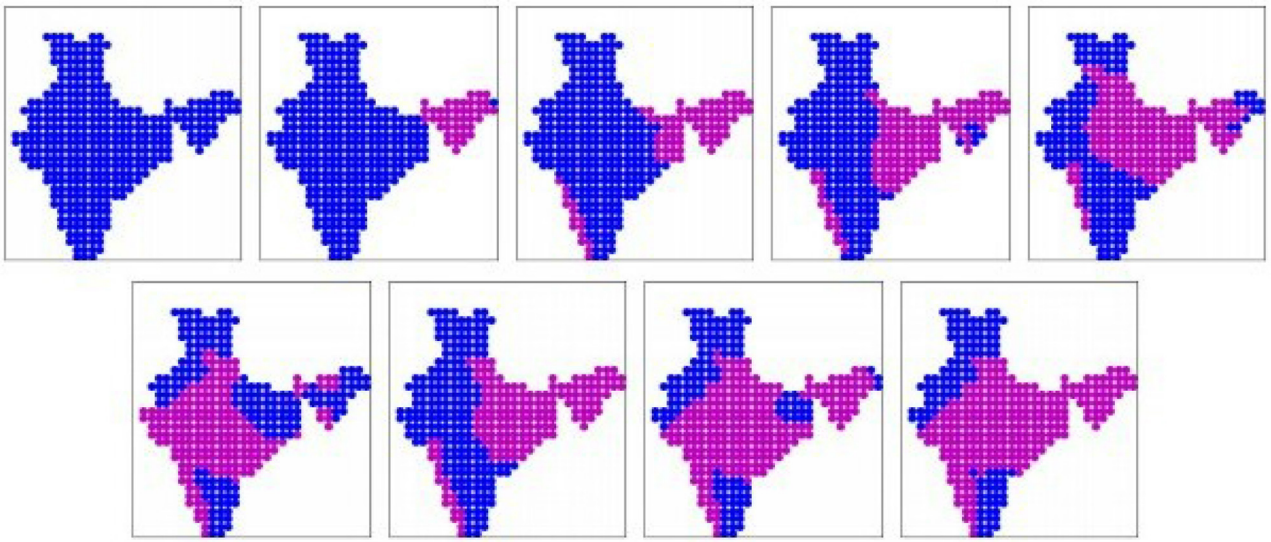


**FIGURE 6 |** The nine prominent binary spatial patterns found from EC-3 model. Pink indicates *wet* locations and blue indicates *dry* locations. Patterns are numbered from left to right across each row, top to bottom.

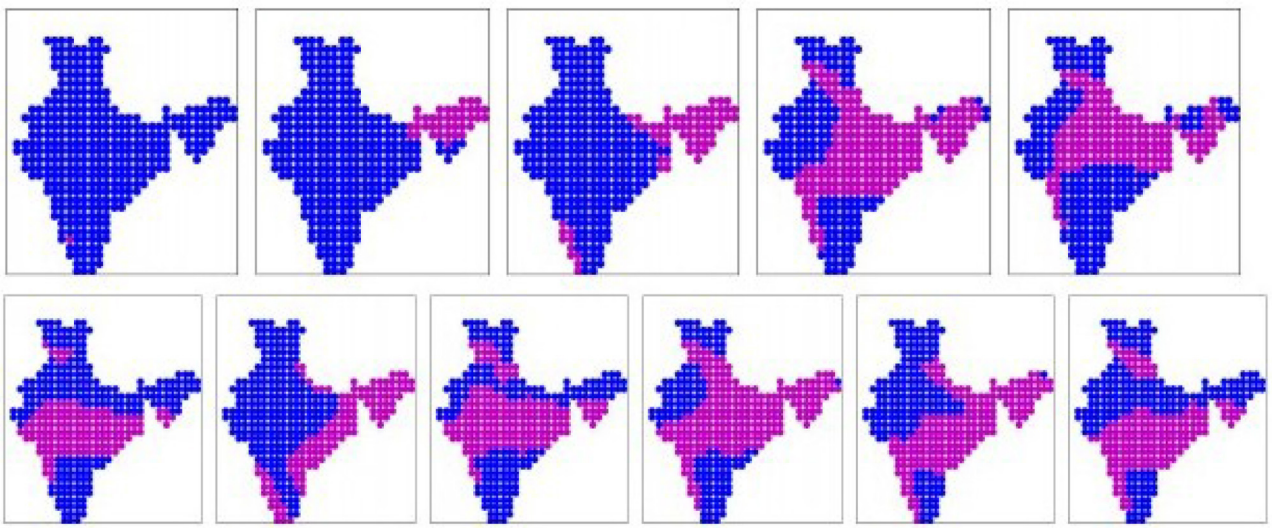


**FIGURE 7 |** The 11 prominent binary spatial patterns found from INM-CM-4.8 model. Pink indicates *wet* locations and blue indicates *dry* locations. Patterns are numbered from left to right across each row, top to bottom.

are sorted in ascending order of mean all-India rainfall rates, as given in **Table 2**. These 10 patterns cover 76% of the days. Here too we have three dry patterns like the IMD dataset. Patterns 5,8,9 show rainfall over the Gangetic plains including Himalayan foothills. However, pattern 4 which show rainfall over the Eastern region only, and patterns 6 and 10 where rainfall covers Central India and Gangetic plain simultaneously, are in disagreement with the patterns from the IMD dataset. The real-valued versions of these prominent patterns are shown in **Supplementary Figure 8**.

## 3.3. Quantitative Analysis of Prominent Spatial Patterns

We now carry out a quantitative analysis of the spatial patterns obtained from the different CMIP6 models and the ground truth data. The first analysis is to see how well the spatial patterns from each CMIP6 model can fit the ground truth data. For each day, we choose among the prominent spatial patterns from a particular CMIP6 model, that pattern which is the closest approximation of the spatial maps $X_{IMD}(t)$ and $Z_{IMD}(t)$. For this we calculate $\tau^{MODEL}(t) = \min ||X^{IMD}(t) - \theta^{MODEL}||$ and $\tau_d^{MODEL}(t) =$

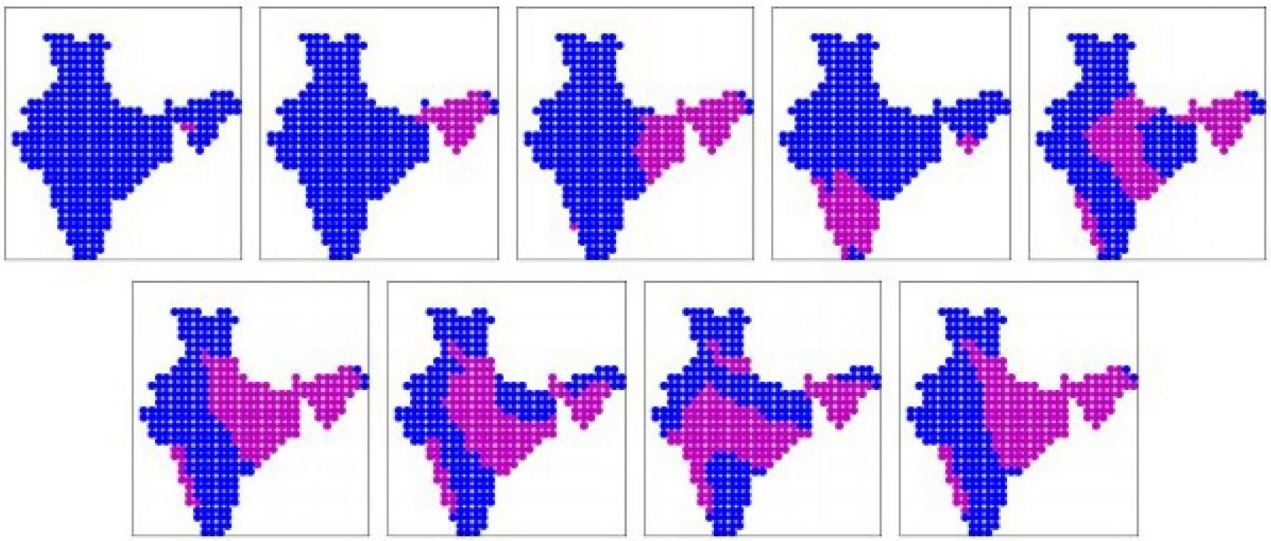


**FIGURE 8 |** The nine prominent binary spatial patterns found from MPI-ESM-1.2 model. Pink indicates *wet* locations and blue indicates *dry* locations. Patterns are numbered from left to right across each row, top to bottom.

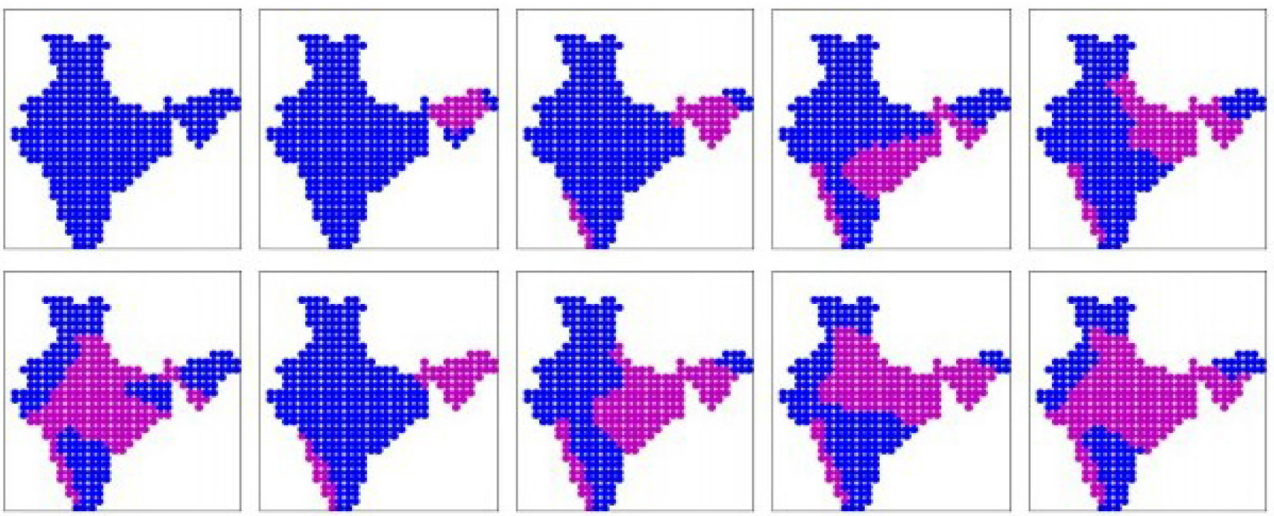


**FIGURE 9 |** The 10 prominent binary spatial patterns found from NOR-ESM2 model. Pink indicates *wet* locations and blue indicates *dry* locations. Patterns are numbered from left to right across each row, top to bottom.

$\min ||Z^{IMD}(t) - \theta_d^{MODEL}||$, where $||$ denotes a suitable distance measure. We use $\ell_2$ norm for $\tau$ and Hamming distance for $\tau_d$. We take the mean value of $\tau^{MODEL}$ and $\tau_d^{MODEL}$ over all the days, and denote these by *PatternScore* and *dPatternScore*. This essentially indicates, how well the prominent spatial patterns identified from the models can describe actual spatial maps of daily rainfall.

Analogously, for each day's rainfall map $X^{MODEL}$ as simulated by the models, we try to approximate them with the prominent spatial patterns identified from the IMD dataset. For each simulated day, we calculate $\kappa^{MODEL}(t) = \min ||X^{MODEL}(t) - \theta^{IMD}||$ and $\kappa_d^{MODEL}(t) = \min ||Z^{MODEL}(t) - \theta_d^{IMD}||$, where $||$ denotes a suitable distance measure. We use $\ell_2$ norm for $X$ and Hamming distance for $Z$. We take the mean value of $\kappa$ and $\kappa_d$ over all the simulated days, and denote it by *MapScore* and *dMapScore*. This essentially indicates how well the daily spatial maps simulated by the models can resemble the prominent spatial patterns identified from the actual data.

The results may be seen in **Table 3**. We find that for the IMD dataset, *dPatternScore* is 0.81, which means that on any day during the study period, the binarized rainfall agrees with value predicted by the corresponding spatial pattern (binary) at about 81% of the locations on average. For other models, this

**TABLE 3 |** Quantitative comparison of the spatial patterns and daily rainfall maps obtained from the IMD dataset and those obtained from different CMIP6 model simulations.

| | IMD | ACC | BCC | CAN | EC3 | INM | MPI | NOR |
|---|---|---|---|---|---|---|---|---|
| PatternScore | 251 | 266 | **254** | 265 | 257 | 257 | 257 | 257 |
| dPatternScore | 0.81 | 0.75 | 0.78 | 0.77 | **0.79** | **0.79** | 0.78 | **0.79** |
| MapScore | 251 | 173 | 214 | **144** | 209 | 183 | 200 | 224 |
| dMapScore | 0.81 | 0.83 | 0.81 | **0.84** | 0.81 | 0.80 | 0.79 | 0.80 |

*The model giving best result with respect to the corresponding measure is shown in bold.*

**TABLE 4 |** The number of extreme rainfall events in different datasets, i.e., the number of (*location*, *day*) pairs at which the volume of rainfall exceeds thresholds as mentioned in the leftmost column.

| | IMD | ACC | BCC | CAN | EC3 | INM | MPI | NOR |
|---|---|---|---|---|---|---|---|---|
| >200 mm | 278 | 89 | 93 | 94 | 126 | 45 | 62 | 194 |
| >150 mm | 793 | 358 | 497 | 304 | 514 | 245 | 319 | 634 |
| >100 mm | 3,133 | 1,578 | 2,286 | 1,174 | 2,223 | 1,218 | 1,626 | 2,502 |
| >50 mm | 16,941 | 11,190 | 13,925 | 6,502 | 14,954 | 10,090 | 12,749 | 15,519 |

score is somewhat lesser. The same trend holds when we consider *PatternScore*, where the actual rainfall values at each location are compared to that predicted by the corresponding spatial pattern. This indicates that the CMIP6 models have not been very effective in recognizing that there exist spatial patterns of daily rainfall. Not only do the patterns extracted from the simulation datasets not resemble those obtained from the IMD dataset, in fact the patterns are not very pronounced in their own rainfall maps.

When we consider the daily spatial maps of rainfall as simulated by the models, we see somewhat unexpected results. We find that ACCESS-ESM-1.5 and CAN-ESM-5 models, which have generally been less accurate, return the least values of *MapScore*, indicating that the daily rainfall maps simulated by them resemble the actual spatial patterns (obtained from the real data) better, compared to other models. In fact, the MapScore is worst from the IMD dataset itself. Yet, on deeper investigation, we realize that this is due to the presence of many more *extreme rainfall events* in the real dataset than those simulated by the models, as shown in **Table 4**. We find that CAN-ESM-5 model has the least number of rainfall events above 50 mm, while for NOR-ESM-2 this number is the highest, and closest to the actual number. The real-valued spatial patterns ($\theta$) do not contain high values of rainfall, due to which the $\ell_2$ norms used for calculation of *MapScore* is high.

## 3.4. Analysis of Spatial Clusters

Now, we turn our attention to the *spatial clusters* obtained from the $V$-variable of the probabilistic graphical model. All locations assigned the same value of $V$ constitute a spatial cluster, implying that their rainfall time-series are nearly identical. The number of spatial clusters is not fixed by the user, but determined by the model from the data. The maps showing the spatial clusters may be seen in **Figure 10**, where each cluster is indicated with a different color. It may be observed that all the spatial clusters obtained are spatially contiguous. The number of clusters, as indicated in **Table 5**, varies in the range 32–55. We find that in all cases, certain geographically special regions such as the Thar desert and the Rann of Kutch in the west, come as single clusters. In case of the clustering obtained from IMD dataset, the rain shadow region of the Deccan plateau in the south comes as a single cluster, though this feature does not show up in other models.

Now, we investigate the quality of these spatial clusters. A cluster is useful only if it is homogeneous. For any cluster $k$, we calculate the mean time-series $X_k^{MODEL}$ of rainfall. We then calculate the correlation coefficients between $X_k^{MODEL}$ and the rainfall time-series $X^{MODEL}(s)$ for each location $s$ within that cluster, i.e., $V^{MODEL}(s) = k$, which indicates how well the rainfall time-series at each location is correlated to that of its spatial cluster. We take the mean of these correlation coefficients for all locations and refer to it as *model-smoothness*. This process is repeated for all the datasets, including IMD and CMIP6 simulations. Next, we examine if the spatial clusters obtained from the CMIP6 model simulations make any physical sense, i.e., are those regions homogeneous with respect to actual rainfall data? For this purpose, we repeat the same exercise as above using $X^{IMD}$ as rainfall time-series, but $V^{MODEL}$ to define the clusters, for each model separately. We call these statistics as *data-smoothness*.

These statistics are available in **Table 5**. We find that some models like ACCESS-ESM-1.5 and CAN-ESM-5 produce significantly less number of spatial clusters than the IMD dataset. The values of *model-smoothness* indicate that clusters from all datasets are relatively homogeneous. However, there may not be a strong physical basis of the spatial clusters obtained from the CMIP6 models, as the *data-smoothness* values from them are relatively low compared to that from the IMD dataset. The numbers are particularly poor for ACCESS-ESM-1.5 and CAN-ESM-5 models, and relatively better for MPI-ESM-1.2.

# 4. DISCUSSIONS

Finally, we come to a discussion of the results. First of all, a major observation is that spatial correlation of monsoon rainfall is heavily overestimated in all the CMIP6 models, indicating

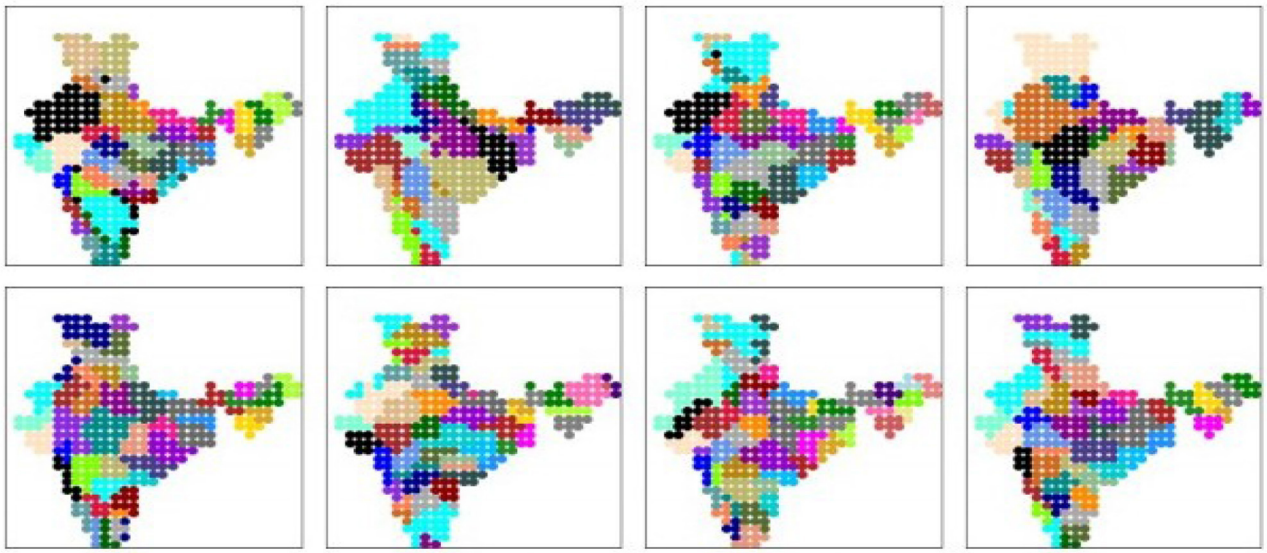


**FIGURE 10 |** The spatial clusters obtained from IMD dataset and all CMIP6 model simulation datasets. Each color denotes one cluster. **(Top)** (left to right): IMD dataset, ACCESS-ESM-1.5, BCC-CSM, CAN-ESM-5; **(Bottom)** (left to right): EC-3, INM-CM-4.8, MPI-ESM-1.2, NOR-ESM-2.

**TABLE 5 |** Statistics of the spatial clusters computed from the IMD dataset and different CMIP6 model simulations, as mentioned in section 3.4.

| | IMD | ACC | BCC | CAN | EC3 | INM | MPI | NOR |
|---|---|---|---|---|---|---|---|---|
| #sp.clus | 48 | 32 | 50 | 34 | 47 | 51 | 55 | 46 |
| Model-smoothness | 0.92 | 0.89 | 0.9 | 0.92 | 0.87 | 0.91 | 0.9 | 0.9 |
| Data-smoothness | 0.92 | 0.69 | 0.75 | 0.7 | 0.73 | 0.74 | 0.76 | 0.72 |

that rainfall is spatially smoother in the model simulations than in the actual data. This may indicate that the models simulate monsoon rainfall through large-scale processes which cover many grids, leading to such spatial correlation. They may be missing out on localized convective events, or missing out on spatial heterogeneity of large-scale processes during monsoon. It turns out that the probabilistic graphical model is able to identify a small number (7–11) or prominent spatial patterns from each of the datasets including the IMD observations and the CMIP6 simulations. However, there are two major differences: (i) While the rainfall maps on a large fraction of the days conform to any of these prominent spatial patterns in case of the IMD dataset, this fraction is less in case of all the CMIP6 model simulations; (ii) The spatial patterns identified from the model simulations do not match well with those from the IMD dataset, and hence these model-based patterns cannot fit the spatial maps of actual daily rainfall. The second point is made based on both the visual inspection of section 3.2 and the quantitative analysis of section 3.3 using the *PatternScore* and *dPatternScore* measures. Additionally, we also see that most of the models seriously underestimate the number of extreme rainfall events. Coming to spatial clusters, we can find a number of reasonably homogeneous spatial clusters from the IMD datasets as well as from the CMIP6 model simulations, but most of the clusters formed by models do not seem to have a strong physical basis, as they are not very homogeneous with respect to actual rainfall from IMD observations, as indicated by the *data-smoothness* measure.

Among the different CMIP6 models we considered, we find that there is a variation in performance. The ACCESS-ESM-1.5 and CAN-ESM-5 are found to be unsuitable on almost all the measures we considered, including the spatial statistics (**Table 1**), number of extreme events, and suitability of their spatial patterns and spatial clusters to actual rainfall data. INM-CM-4.8 is found to perform poorly on some measures, especially the local statistics and number of extreme events, but decently on other measures. Several of the spatial patterns identified from this model are in disagreement with the spatial patterns from IMD dataset, but this is somehow not captured by the *dPatternScore* measure where INM-CM-4.8 does well. But the other models, namely BCC-CSM, EC-3, MPI-ESM-1.2, and NOR-ESM-2 are found to be quite robust on most of the measures that we consider.

More than hindcasting of historical data, the biggest utility of these GCMs is in simulation of future climate. It is well-known that no individual model is reliable enough for simulations, which is why multi-model ensembles are considered for most studies regarding simulation of future scenarios. However, it is important to assign weights to the ensemble members based on their performance on the historical period, since that can be evaluated against ground truth. We hope this work will enable such weighing of climate models, especially for studies about spatial properties of Indian monsoon under future scenarios.

## DATA AVAILABILITY STATEMENT

The datasets used in this study are all available in public domain. The observational dataset provided by India Meteorological Department are available at: https://imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html. The datasets simulated by different CMIP6 models are available at: https://zenodo.org/record/3873998#.X_g60dgzaUk.

## AUTHOR CONTRIBUTIONS

The problem formulation, experimentation, and writing of the paper were all done by AM.

## FUNDING

This research was partially funded by Grant Number IIT/SRIC/AI/STF/2020-2021/067, provided by Sponsored Research and Industrial Consultancy (SRIC), Indian Institute of Technology Kharagpur.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fclim.2021.654763/full#supplementary-material

## REFERENCES

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., et al. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* 9, 1937–1958. doi: 10.5194/gmd-9-1937-2016

Gadgil, S. (2003). The Indian monsoon and its variability. *Annu. Rev. Earth Planet. Sci.* 31, 429–467. doi: 10.1146/annurev.earth.31.100901.141251

Gadgil, S., and Gadgil, S. (2006). The Indian monsoon, GDP and agriculture. *Econ. Polit. Week*. 4887–4895. Available online at: http://www.jstor.org/stable/4418949

Ghosh, S., Das, D., Kao, S.-C., and Ganguly, A. R. (2012). Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. *Nat. Clim. Change* 2, 86–91. doi: 10.1038/nclimate1327

Goswami, B., and Chakravorty, S. (2017). "Dynamics of the Indian summer monsoon climate," in *Oxford Research Encyclopedia of Climate Science*. doi: 10.1093/acrefore/9780190228620.013.613

Gusain, A., Ghosh, S., and Karmakar, S. (2020). Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmos. Res.* 232:104680. doi: 10.1016/j.atmosres.2019.104680

Jayasankar, C., Surendran, S., and Rajendran, K. (2015). Robust signals of future projections of Indian summer monsoon rainfall by IPCC AR5 climate models: role of seasonal cycle and interannual variability. *Geophys. Res. Lett*. 42, 3513–3520. doi: 10.1002/2015GL063659

Kindermann, R. (1980). *Markov Random Fields and Their Applications*. Providence, RI: American Mathematical Society. doi: 10.1090/conm/001

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M earth system model version 1.2 (MPI-ESM1. 2) and its response to increasing CO2. *J. Adv. Model. Earth Syst*. 11, 998–1038. doi: 10.1029/2018MS001400

Mishra, V., Bhatia, U., and Tiwari, A. D. (2020). Bias-corrected climate projections for South Asia from coupled model intercomparison project-6. *Sci. Data* 7:338. doi: 10.1038/s41597-020-00681-1

Mitra, A., Apte, A., Govindarajan, R., Vasan, V., and Vadlamani, S. (2018). A discrete view of the Indian monsoon to identify spatial patterns of rainfall. *Dyn. Stat. Clim. Syst*. 3:dzy009. doi: 10.1093/climsys/dzy009

Nair, P., Chakraborty, A., Varikoden, H., Francis, P., and Kuttippurath, J. (2018). The local and global climate forcings induced inhomogeneity of Indian rainfall. *Sci. Rep*. 8, 1–12. doi: 10.1038/s41598-018-24021-x

Neal, R. M. (1993). *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Toronto, ON: Department of Computer Science; University of Toronto.

Pai, D., Sridhar, L., Rajeevan, M., Sreejith, O., Satbhai, N., and Mukhopadhyay, B. (2014). Development of a new high spatial resolution (0.25 × 0.25) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam* 65, 1–18. Available online at: https://metnet.imd.gov.in/mausamdocs/16511_F.pdf

Pattnayak, K., Kar, S., Dalal, M., and Pattnayak, R. (2017). Projections of annual rainfall and surface temperature from CMIP5 models over the BIMSTEC countries. *Glob. Planet. Change* 152, 152–166. doi: 10.1016/j.gloplacha.2017.03.005

Preethi, B., Mujumdar, M., Kripalani, R., Prabhu, A., and Krishnan, R. (2017). Recent trends and tele-connections among south and East Asian summer monsoons in a warming environment. *Clim. Dyn*. 48, 2489–2505. doi: 10.1007/s00382-016-3218-0

Rajeevan, M., Bhate, J., Kale, J., and Lal, B. (2006). High resolution daily gridded rainfall data for the Indian region: analysis of break and active. *Curr. Sci*. 91, 296–306. Available online at: https://www.jstor.org/stable/24094135

Rajeevan, M., Gadgil, S., and Bhate, J. (2010). Active and break spells of the Indian summer monsoon. *J. Earth Syst. Sci*. 119, 229–247. doi: 10.1007/s12040-010-0019-4

Raju, K. S., and Kumar, D. N. (2014). Ranking of global climate models for India using multicriterion analysis. *Clim. Res.* 60, 103–117. doi: 10.3354/cr01222

Roxy, M. K., Ghosh, S., Pathak, A., Athulya, R., Mujumdar, M., Murtugudde, R., et al. (2017). A threefold rise in widespread extreme rain events over central India. *Nat. Commun*. 8, 1–11. doi: 10.1038/s41467-017-00744-9

Saha, A., Ghosh, S., Sahana, A., and Rao, E. (2014). Failure of CMIP5 climate models in simulating post-1950 decreasing trend of Indian monsoon. *Geophys. Res. Lett*. 41, 7323–7330. doi: 10.1002/2014GL061573

Seland, Ø., Bensten, M., Graff, L., Olivié, D., Toniazzo, T., Gjermundsen, A., et al. (2020). The Norwegian earth system model, NORESM2-evaluation of the CMIP6 deck and historical simulations. *Geosci. Model Dev. Discuss*. 13, 6165–6200. doi: 10.5194/gmd-13-6165-2020

Sharma, A., Mitra, A., Vasan, V., and Govindarajan, R. (2021). Spatio-temporal relationships between rainfall and convective clouds during Indian monsoon through a discrete lens. *Int. J. Climatol*. 41, 1351–1368. doi: 10.1002/joc.6812

Shashikanth, K., Salvi, K., Ghosh, S., and Rajendran, K. (2014). Do CMIP5 simulations of Indian summer monsoon rainfall differ from those of CMIP3? *Atmosph. Sci. Lett*. 15, 79–85. doi: 10.1002/asl2.466

Singh, S., Ghosh, S., Sahana, A., Vittal, H., and Karmakar, S. (2017). Do dynamic regional models add value to the global model projections of indian monsoon? *Clim. Dyn*. 48, 1375–1397. doi: 10.1007/s00382-016-3147-y

Sperber, K., Annamalai, H., Kang, I.-S., Kitoh, A., Moise, A., Turner, A., et al. (2013). The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Clim. Dyn.* 41, 2711–2744. doi: 10.1007/s00382-012-1607-6

Suhas, E., Neena, J., and Goswami, B. (2013). An Indian monsoon intraseasonal oscillations (miso) index for real time monitoring and forecast verification. *Clim. Dyn*. 40, 2605–2616. doi: 10.1007/s00382-012-1462-5

Sun, Q., Miao, C., and Duan, Q. (2015). Comparative analysis of CMIP3 and CMIP5 global climate models for simulating the daily mean, maximum, and minimum temperatures and daily precipitation over China. *J. Geophys. Res.* 120, 4806–4824. doi: 10.1002/2014JD022994

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian earth system model version 5 (CANESM5. 0.3). *Geosci. Model Dev*. 12, 4823–4873. doi: 10.5194/gmd-12-4823-2019

Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., et al. (2018). Simulation of the modern climate using the INM-CM48 climate model. *Russ. J. Numer. Anal. Math. Model.* 33, 367–374. doi: 10.1515/rnam-2018-0032

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., et al. (2019). The Beijing climate center climate system model (BCC-CSM): The main progress from CMIP5 to CMIP6. *Geosci. Model Dev*. 12, 1573–1600. doi: 10.5194/gmd-12-1573-2019

Wyser, K., Noije, T. V., Yang, S., Hardenberg, J. V., O'Donnell, D., and Döscher, R. (2020). On the increased climate sensitivity in the ec-earth model from CMIP5 to CMIP6. *Geosci. Model Dev*. 13, 3465–3474. doi: 10.5194/gmd-13-3465-2020

Xin, X., Wu, T., Zhang, J., Yao, J., and Fang, Y. (2020). Comparison of CMIP6 and CMIP5 simulations of precipitation in China and the East Asian summer monsoon. *Int. J. Climatol*. 40, 6423–6440. doi: 10.1002/joc.6590

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., et al. (2020). The Australian earth system model: access-ESM1. 5. *J. South. Hemis. Earth Syst. Sci.* 70, 193–214. doi: 10.1071/ES19035

**Conflict of Interest:** The author declares 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 Climate

ORIGINAL RESEARCH
published: 17 May 2021
doi: 10.3389/fclim.2021.675840

# Information Entropy as Quantifier of Potential Predictability in the Tropical Indo-Pacific Basin

*Olawale J. Ikuyajolu*[1,2], *Fabrizio Falasca*[1] *and Annalisa Bracco*[1,2]*

[1] School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, United States, [2] Program in Ocean Science and Engineering, Georgia Institute of Technology, Atlanta, GA, United States

Global warming is posed to modify the modes of variability that control much of the climate predictability at seasonal to interannual scales. The quantification of changes in climate predictability over any given amount of time, however, remains challenging. Here we build upon recent advances in non-linear dynamical systems theory and introduce the climate community to an information entropy quantifier based on recurrence. The entropy, or complexity of a system is associated with microstates that recur over time in the time-series that define the system, and therefore to its predictability potential. A computationally fast method to evaluate the entropy is applied to the investigation of the information entropy of sea surface temperature in the tropical Pacific and Indian Oceans, focusing on boreal fall. In this season the predictability of the basins is controlled by two regularly varying non-linear oscillations, the El Niño-Southern Oscillation and the Indian Ocean Dipole. We compute and compare the entropy in simulations from the CMIP5 catalog from the historical period and RCP8.5 scenario, and in reanalysis datasets. Discrepancies are found between the models and the reanalysis, and no robust changes in predictability can be identified in future projections. The Indian Ocean and the equatorial Pacific emerge as troublesome areas where the modeled entropy differs the most from that of the reanalysis in many models. A brief investigation of the source of the bias points to a poor representation of the ocean mean state and basins' connectivity at the Indonesian Throughflow.

**Keywords: tropical climate, predictability, entropy, ENSO, IOD**

OPEN ACCESS

***Edited by:***
*Sarah Kang,*
*Ulsan National Institute of Science and Technology, South Korea*

***Reviewed by:***
*Tomomichi Ogata,*
*Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan*
*Kaiming Hu,*
*Institute of Atmospheric Physics (CAS), China*

***Correspondence:**
*Annalisa Bracco*
*abracco@gatech.edu*

***Specialty section:***
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

***Received:*** *04 March 2021*
***Accepted:*** *13 April 2021*
***Published:*** *17 May 2021*

***Citation:***
*Ikuyajolu OJ, Falasca F and Bracco A (2021) Information Entropy as Quantifier of Potential Predictability in the Tropical Indo-Pacific Basin. Front. Clim. 3:675840. doi: 10.3389/fclim.2021.675840*

## INTRODUCTION

In the past decade, our theoretical understanding of the physics of the climate system has advanced in fundamental ways. These advancements proceeded in parallel with model improvements and computing capabilities. Understanding and especially predicting climate change at regional or local scales—the scales that are relevant to society—remains, however, challenging: regional climate change is influenced by the large-scale climate and, at the same time, feeds back to the global scale but these interactions are poorly represented in models.

In this work, we focus on the tropical Pacific and Indian Oceans, where the El Niño Southern Oscillation (ENSO; Bjerknes, 1966, 1969) and the Indian Ocean Dipole (IOD; Saji et al., 1999; Webster et al., 1999) control the largest portions of the variance at interannual scale (**Figure 1**). They impact key variables of societal relevance, from surface temperature to precipitation and

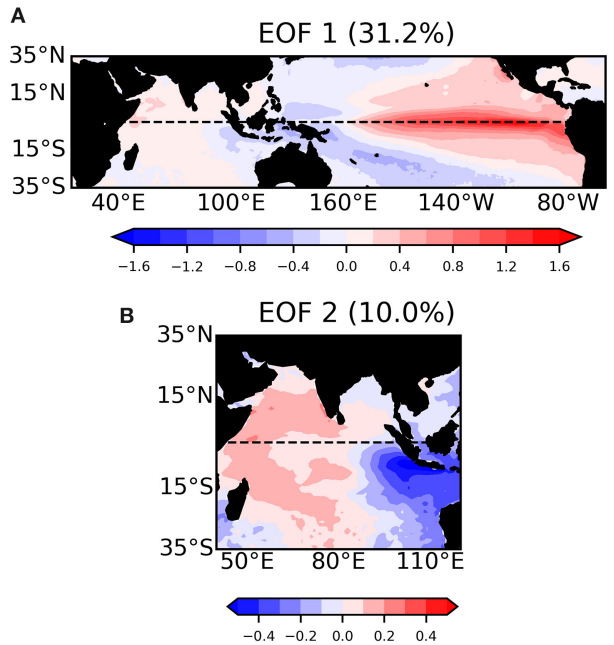



**FIGURE 1 | (A)** First EOF mode of the tropical Indo-Pacific SST monthly anomalies, showing the observed ENSO pattern. **(B)** Second EOF mode of tropical Indian Ocean SST monthly anomalies highlighting the IOD patterns. The monthly data cover the period 1980–2018 and are obtained from SODA 3.4.2 reanalysis. Percentages of variance explained by the EOF patterns are included. The black dash line depicts the equator.

the frequency of extreme events such as tropical cyclones, typhoons and droughts. These climate modes are of the uttermost importance not only for water and food security, but also for global health, as they modulate, for example, malaria occurrences in India (Dhiman and Sarkar, 2017; Anyamba et al., 2019), Indonesia (Kovats, 2000) and Africa (Hashizume et al., 2012; Kreppel et al., 2019). Assessing their potential predictability and quantifying how this predictability is simulated in climate models and projected to change in the future, is a priority.

ENSO and IOD play a crucial role in global climate variability. In light of this importance, their interactions have been the subject of numerous studies. A recent review of ENSO teleconnections can be found in Yeh et al. (2018). The IOD develops in late summer with a November maximum. Since its discovery, observational and modeling studies have focused on its linkages with the Asian Summer Monsoon (Behera et al., 1999; Ashok et al., 2001; Guan et al., 2003; Saji and Yamagata, 2003), its modulation of the Indian Summer Monsoon and ENSO (Ashok et al., 2001); and its teleconnections over North America, Australia, South-Africa (Li and Mu, 2001) and East Africa (Black et al., 2003). The regional and global influences of these modes cannot be overemphasized, yet a realistic simulation of the characteristics and teleconnections of these modes remains challenging in state-of-the-art climate models (Weller and Cai, 2013), precluding the reliability of future projections.

Here, we introduce the climate science community to a quantifier of complexity, or entropy, recently proposed in the literature that allows for distinguishing regular, chaotic, and random behaviors in time-series. It provides a quick framework to investigate potential predictability and verify how well climate models represent it. We test it using an observationally-based reanalysis product and model data-sets, we verify its robustness and finally use it to exemplify the challenges implicit in using global climate models—here the integrations from the Coupled Model Intercomparison Project phase 5, CMIP5—(Taylor et al., 2012) to investigate present and future predictability. We focus on the Indo-Pacific Ocean in boreal fall, when the IOD variance is maximized and many countries are impacted by its and ENSO modulation of rainfall and temperatures over the surrounding land masses.

The rest of this paper is organized as follows: section Data provides information about the observational data and models used; section The Entropy Quantifier: Recurrence Plots and Information Entropy focuses the information entropy quantifier; sections Entropy in Reanalysis Data, CMIP5 Historical Runs and CMIP5 Projections discusses the entropy of historical and future projection focusing on the boreal fall season, and section Entropy and the ENSO-IOD Relation analyzes possible sources of model bias in the historical period. A summary of the findings concludes the work.

## DATA

We analyze monthly mean outputs from 15 CMIP5 models and consider both historical runs and future projections following the representative concentration pathway (RCP) 8.5 scenario. For each model, we use at least 1 member for the historical analysis and its evolution in the RCP case. For 10 models we verified that internal model variability does not impact the outcome of our analysis by exploring 3 (or 2 when three runs were not readily available) members in the historical period. The output variable we focus on is sea surface temperature (SST), and further consider sea level pressure (SLP) and thermocline depth (Z20) in investigating model divergence.

In terms of observational datasets, we use the monthly mean subsurface temperature from the Simple Ocean Data Assimilation (SODA) version 3.4.2, which was forced by the European Center for Medium-range Weather Forecasts (ECMWF) re-analysis ERA-Interim (Carton et al., 2018). SODA has a spatial resolution of 0.25° x 0.25° with 50 vertical levels, and spans the 1980- 2017 period. For cross validation, we also considered fields from the Ocean Re-Analysis System 4 (ORAS4) (Zuo et al., 2019). Results obtained using SODA and ORAS4 are nearly identical and consistent. For brevity, we focus our discussion on SODA. Sea level pressure is from the monthly mean ERA-Interim (Dee et al., 2011). All data have been de-trended and re-gridded into a uniform 1° longitude x 1° latitude spatial resolution before analysis. For the observational datasets, we look at the period 1980–2018 for temporal consistency among them, while in the CMIP5 models we consider the last 39 years of the historical period (1967–2005)—same period length as in

**TABLE 1 |** Models and ensemble members analyzed in this work.

| Models acronym | Model | Institute, country | Ensemble members | |
|---|---|---|---|---|
| | | | Hist | RCP 8.5 |
| ACCESS 1.0 | Australian community climate and earth-system simulator, version 1.0 | Commonwealth Scientific and Industrial Research Organisation (CSIRO)–Bureau of meteorology | 2 | 1 |
| ACCESS 1.3 | Australian community climate and earth-system simulator, version 1.3 | BOM, Australia | 1 | 1 |
| CanESM2 | Second generation Canadian earth system model | Canadian Centre for Climate Modeling and Analysis (CCCma), Canada | 3 | 1 |
| CCSM4 | Community climate system model, version 4 | NCAR, United States | 3 | 1 |
| CMCC-CESM | Centro euro-mediterraneo cambiamenti climatici climate model | Centro Euro-Mediterraneo per I Cambiamenti Climatici (CMCC), Italy | 1 | 1 |
| CNRM-CM5 | Centre national de recherches meteorologiques coupled global climate model, version 5 | Centre National de Recherches Meteorologiques (CNRM)–Centre Europeen de recherche et de | 3 | 1 |
| GFDL-CM3 | Geophysical fluid dynamics laboratory climate model, version 3 | National Oceanic and Atmospheric Administration (NOAA)/geophysical fluid | 3 | 1 |
| GISS E2-H | Goddard institute for space studies model E2, coupled with the Hybrid Coordinate Ocean Model (HYCOM) | National Aeronautics and Space Administration (NASA) Goddard institute for space | 3 | 1 |
| GISS E2-R | Goddard institute for space studies model E2, coupled with the Russell ocean model | NASA GISS, United States | 3 | 1 |
| HadGEM2-ES | Hadley centre global environment model, version 2–earth system | UKMO Hadley Centre, United Kingdom | 3 | 1 |
| INM-CM4 | Institute of numerical mathematics coupled model, version 4.0 | Institute of Numerical Mathematics (INM), Russia | 1 | 1 |
| IPSL-CM5A-LR | L'Institut pierre-simon laplace coupled model, version 5A, coupled with NEMO, low resolution | L'Institut Pierre-Simon Laplace (IPSL), France | 3 | 1 |
| IPSL-CM5A-MR | L'Institut pierre-simon laplace coupled model, version 5A, coupled with NEMO, mid resolution | IPSL, France | 1 | 1 |
| MPI-ESM-LR | Max planck institute earth system model, low resolution | Max Planck Institute for Meteorology (MPI-M), Germany | 1 | 1 |
| MRI-CGM3 | Meteorological research institute coupled atmosphere–ocean general circulation model, version 3 | Meteorological Research Institute (MRI), Japan | 3 | 1 |

the reanalyses—and the last 30 years of the XXI century (2071–2100) as representative of future climate. **Table 1** summarized the model runs used.

In the following, the IOD strength and variability is quantified by the IOD Index, which is the difference in the SST monthly anomalies (SSTAs) averaged over the western tropical Indian Ocean (WTIO) (50°–70°E, 10°S−10°N) and in the SETIO region (90°–100°E, 0°S−10°N) (Saji et al., 1999). For ENSO we adopt the Niño-3.4 index, defined as the area-averaged SSTAs over 170°W−120°W, 5°S−5°N. To examine the strength of oceanic linkages between the Pacific and Indian oceans and the relationship between ENSO and IOD, we perform a correlation analysis using the depth anomalies in the 20°C isotherm (Z20), a proxy for the tropical thermocline depth, in the ITF region (area average over 120°E−131.5° E, 7.5°-8.5°S) (Bracco et al., 2005) or the sea level difference between the western Pacific (WP; 0–10°N; 125°E−145°E) and the Eastern Indian Ocean (EIO; 10°S−20°S; 110°E−130°E) (Mayer et al., 2018).

## THE ENTROPY QUANTIFIER: RECURRENCE PLOTS AND INFORMATION ENTROPY

Given a dynamical system, several measures of complexity have been proposed to distinguish regular (e.g., periodic), chaotic and random behaviors in the time-series that describes it. Among those complexity quantifiers, the most common are Lyapunov

exponents, fractal dimensions, scaling exponents, divergence rates and entropies (Manneville, 1990; Shi, 2007). Most of these quantifiers work very well in the case of low dimensional dynamical systems but their application is complicated for *noisy* time series coming from high-dimensional real-world systems. Bandt and Pompe (2002) proposed the *permutation entropy* quantifier to address this issue, and along with minor modifications this tool has been used in climate science for the analysis of a proxy record of ENSO spanning the Holocene (Saco et al., 2010) and for ice core records by Garland et al. (2019). The permutation entropy depends on two parameters and at least for low dimensional systems (for example a Logistic Map) has a good, but not optimal, correlation with the Lyapunov exponents of the system under investigation.

Here we adopt instead an entropy quantifier based on the distribution of microstates in a recurrence plot (Prado et al., 2020). All complexity quantifiers are based on fundamental phase space properties of ergodic dynamical systems, such as trajectory recurrence (Poincaré, 1890; Cvitanović et al., 2016) and this recurrence can be visualized, for a given time-series, using the recurrence plot (RP) method introduced by Eckmann et al. (1987). Given a trajectory $x_i$ in a dynamical system in its $d$-dimensional phase space at time $i$, its RP is given by an $N \times N$ matrix of 1 and 0 such as:

$$RP_{i,j}(\varepsilon) = \Theta\left(\varepsilon - \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|\right),\ \boldsymbol{x}_i \in \mathbb{R}^d,\ i,j = 1,\ldots,N \quad (1)$$

where $\varepsilon$ is the threshold distance and defines the neighborhood of a state $\boldsymbol{x}_i$, $\Theta$ is the Heaviside function, $\|\cdot\|$ is a norm (Euclidean distance in our case) and $N$ is the number of states considered.

The analysis of structures (such as diagonal, vertical or horizontal lines) in a RP is known as recurrence quantification analysis (RQA) and has found numerous applications in biology, neuroscience, physics, geosciences and economics, among other disciplines (Webber and Marwan, 2015). In essence, given a time series it is possible to reconstruct a state space representation using the embedding theorem (Takens, 1981), and then compute the recurrence plot (which is a matrix with "1" if two states are recurrent, "0" if they are not recurrent). Most methods to estimate complexity of time series are directly related to RQA (Marwan et al., 2007), recurrence network analysis (Donner et al., 2011) and information theory (Bandt and Pompe, 2002; Balasis et al., 2013). Some of these quantifiers have been applied to climate science to analyze regime shifts and tipping points in single time-series (e.g., Donges et al., 2011). All these methods require "phase space reconstruction," as mentioned, or, in other words, they require embedding the time series in a higher dimensional space. Phase space reconstruction is computationally time consuming and sensitive to both the choice of parameters and to the system's noise and dimensionality (Gilpin, 2020), and is therefore not advisable for real-world spatiotemporal fields as it may lead to spurious results (Marwan, 2011; Riedl, 2013).

To remedy these issues, Corso et al. (2018) proposed a *recurrence entropy* quantifier that does not require phase space reconstruction and can be safely applied to fields of any dimension and complexity. The idea behind is that the entropy of a time series can be computed by the probability of occurrence of microstates in its RP. A microstate of size $M$ is defined as an $M \times M$ matrix inside the RP. The total number of configurations of 1 and 0 in a microstate of size $M$ is $M^* = 2^{M^2}$ and is possible to define a probability of occurrence $P_k$ of a microstate $k$ as $P_k = \frac{n_k}{M^*}$ , with $n_k$ being the number of occurrences of the $k$-th microstate in the RP. The information entropy of the time series is then given by:

$$S\left(M^*\right) = -\sum_{k=1}^{M^*} P_k \ln P_k. \quad (2)$$

It is possible to normalize the recurrence entropy by its maximum possible value $M^2 \ln 2$ which corresponds to the case in which all microstates appear with the same probability $P_k = \frac{1}{M^*}$; in this case $S = 0$ (1) implies perfect predictability (unpredictability) of the system dynamics.

The number of admissible microstates $M^*$ grows exponentially as a function of $M$ but Corso et al. (2018) showed that only few of them populate the RP. It is therefore reasonable to consider only a random subsample $\overline{M}$ of the possible microstates in the RP and expect a rapid convergence to $S(M^*)$ (see **Supplementary Material**). In this case, the entropy is normalized by $\ln \overline{M}$.

In layman's terms, the information entropy based on the recurrence of microstates can be computed in two steps by calculating firstly the recurrence plot and then the Shannon Entropy based on the distribution of microstates. If the Euclidean distance between states x(t = a) and x(t = b) is less than epsilon then they are considered neighbors and described by "1" in the recurrence plot matrix, vice versa they are described by "0."

The recurrence entropy, in whichever way it is calculated, depends on the choice of the distance threshold $\varepsilon$. Several heuristics have been proposed for selecting $\varepsilon$ : ∼5% of the maximal phase space diameter (Mindlin and Gilmore, 1992); no more than 10% of the mean (or maximum) phase space diameter (Koebbe and Mayer-Kress, 1992; Zbilut and Webber, 1992); a value ensuring a recurrence point density of ∼1% (Zbilut et al., 2002). More recently, Prado et al. (2020) relied on the "maximum entropy principle" to free the analysis from having to select $\varepsilon$. Given a RP and microstates of size $M$, they examined the dependence of the recurrence entropy $S$ on $\varepsilon$ and found that, typically, the curve $S(\varepsilon)$ has a well-defined maximum, $S_{Max}$, that does not vary significantly for a range of $\varepsilon$ values. In simple deterministic and stochastic systems, the maximum entropy $S_{Max}$ calculated by Prado and colleagues is highly correlated with the Lyapunov exponent, further improving the complexity quantifier proposed by Corso et al. (2018). $S_{Max}$ is therefore our preferred choice. Technical details for the heuristic used to determine it and an example of the convergence when considering only a subsample $\overline{M}$ of all possible microstates in the RP can be found in the **Supplementary Material**.

The information entropy so computed measures the complexity of the time series examined, and the two terms, entropy and complexity will be used interchangeably in the reminder of the paper. The higher the entropy, the more

complex and less predictable is the system, and vice versa. Examples from simple systems such as a logistic map (May, 1976) can be found in both Corso et al. (2018) and Prado et al. (2020). In the **Supplementary Material** we quantify the information entropy of the Lorenz system (Lorenz, 1963), while a first application to noisy, high-dimensional data can be found in Falasca et al. (2020) in which the authors used the entropy quantifier to identify abrupt regime shifts in paleoclimate simulations.

The information entropy when applied to a climate field quantifies the degree of complexity of a given region (or grid point) and depends only on the size of a microstate $M$. In the next section we will explore its sensitivity from $M = 2$ to $M = 8$. In summary, for a given climate field $\boldsymbol{X}(t)$, in this work we compute the information entropy and its variability for all the time series in the spatial grid to derive the spatiotemporal entropy field $\boldsymbol{S_X}(t)$ without having to rely on embedding.

## ENTROPY IN REANALYSIS DATA, CMIP5 HISTORICAL RUNS AND CMIP5 PROJECTIONS

The entropy fields for the SST in the SODA/ERA-Interim reanalysis and two representative CMIP5 models, CCSM4 and HadGEM2-ES are shown in **Figure 2** using monthly average fields and all months, and for different values of $M$. With this first analysis we explore the robustness of metric to the size of the microstates. A microstate of size M = 3, for example, quantifies the recurrence between $x(t_1)$, $x(t_1+1$ *month)*, $x(t_1+2$ *months)* and another random sequence $x(t_2)$, $x(t_2+1$ *month)*, $x(t_2+2$ *months)*, while if M = 2 the recurrence will be evaluated between $x(t_1)$, $x(t_1+1$ *month)* and another, shorter, random sequence $x(t_2)$, $x(t_2+1$ *month)*, and so on.

The patterns in the entropy plots do not vary significantly with $M$, but the entropy values do, with generally lower predictability (higher entropy values) and smaller chances of recurrence for increasing $M$, as to be expected given that we are evaluating the recurrence of a longer (in time) microstate, and very high, unstructured, predictability if $M$ is too small. This dependence can be exploited for studies focusing on different time-scales and output frequency, for example by using daily data for evaluating subseasonal predictability.

When all months are considered, in SODA/ERA-Interim the highest predictability is found in the central tropical Pacific in the ENSO impacted area, as expected. Upwelling regions, both along South America in the cold tongue and in the Arabian Sea, are characterized by higher complexity near the coasts that in offshore waters, and the IOD and SIOD action regions appear slightly more predictable than the remaining IO.

CCSM4 overestimates predictability (underestimates entropy) nearly everywhere. In terms of patterns, it is close to the observed ones, but it does not capture the lower than surrounding predictability of the cold tongue in the Pacific and in the western Arabian Sea upwelling, and underestimates slightly that of the ocean region to the west of the tip of India, that participate in the IOD dynamics. HadGEM2-ES, on the other hand, does not capture the ENSO predictability potential around the Equator (10°N−10°S), and overestimates predictability in the upwelling systems in the Pacific and to the west of Australia.

The usefulness of the entropy is also in highlighting changes in predictability among months, seasons and subseasonal time scales if daily or higher frequency data are available. Given the monthly averages used in this study, we show next the entropy in boreal spring (March to May, MAM) (**Figure 3**), when complexity is expected to be higher due to the spring predictability barrier to ENSO, and in the extended fall season, August to November (ASON) (**Figure 4**), when the impacts on the rainfall variability in the continents surrounding the Indian Ocean are greatest and the variance explained by the IOD is largest.

In both figures the entropy is shown in SODA/ERA-Interim and all models using $M = 3$, chosen in light of the seasonal scope of our analysis (3 or 4 instead of 12 months in each year). In all CMIP5 models the complexity is higher in boreal spring in the Pacific, indicating that they generally capture the spring barrier to ENSO predictability (e.g., Torrence and Webster, 1998; Duan and Wei, 2013). Little agreement is found among models on the spring predictability of IO.

In boreal fall, models with a realistically shaped—but underestimated—entropy in the ENSO region, especially in the southern hemisphere (CCSM4, CanESM2, MPI) tend to overestimate the predictability in the IO, west of Sumatra for CCSM4 and west of Australia in CanESM2 and MPI. This bias was not as evident in spring. Other models share the bias in the IO but also have high (generally too high) predictability in the tropical Pacific and lower than observed at the equator (ACCESS1.0 and 1.3, CMCC-CESM). CNRM-CM5 reproduces best the entropy patterns in both basins. GFDL-CM3 underestimates the complexity of the equatorial Pacific, while the IPSL model, in both its version, is characterized by an ENSO pattern protruding too far west into the warm pool area and very high predictability north and south of the equator in the western Pacific. The remaining models tend to underestimate the SODA/ERA predictability in the Indo-Pacific. In the IO, many models underestimate the entropy in the 15°S−35°S band and overestimate it along the equatorial ocean, missing the IOD-related predictability and displaying little intra-model agreement on the overall patterns.

Since the IOD discovery, Saji et al. (1999) reported the existence of a correlation between the IOD and ENSO indices. This correlation reaches 0.62 in the ASON season over the period considered for the reanalysis, consistent with the SODA/ERA-Interim entropy pattern, which suggests some predictability potential in the eastern, equatorial and part of the western IO. In the Arabian Sea, on the other hand, the complexity tends to be larger, likely due to the energetic mesoscale field characterizing this upwelling system in fall.

Since correlation does not imply causation, many studies have questioned whether the IOD can occur independently of ENSO and by which mechanism (e.g., Allan et al., 2001; Ashok et al., 2003; Gualdi et al., 2003; Saji and Yamagata, 2003). The general consensus is that the IOD is an ocean-atmospheric coupled mode of climate variability intrinsic to the IO that can be excited

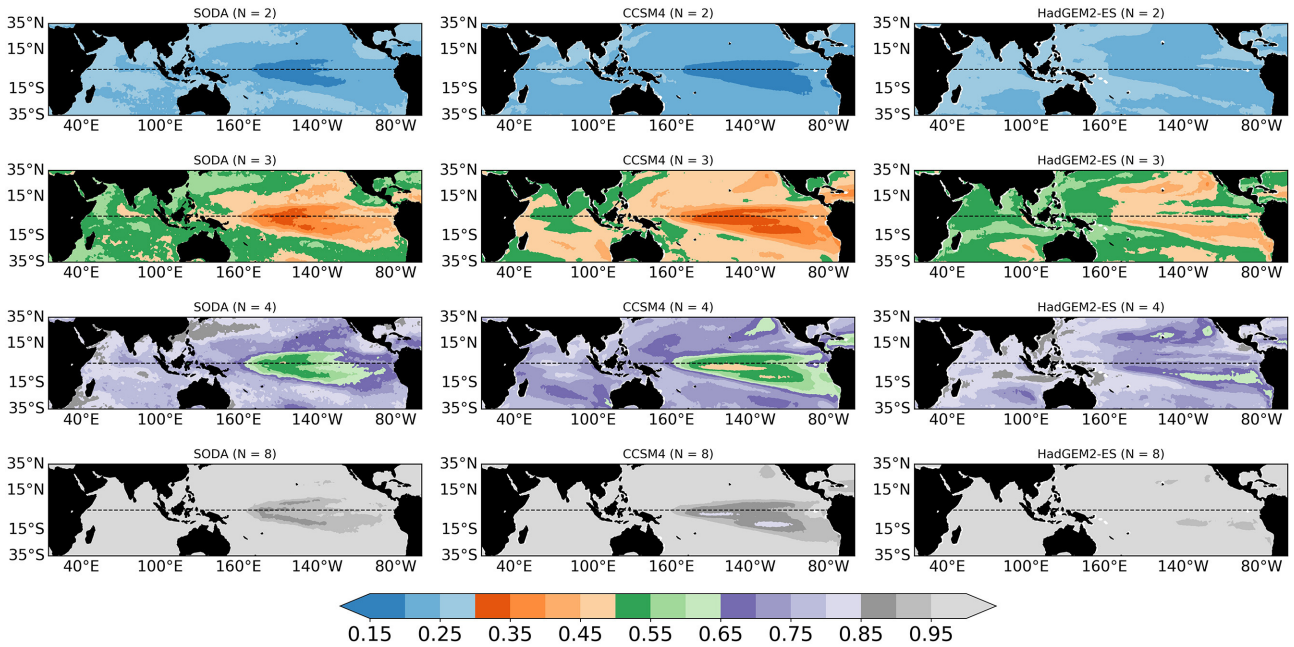



**FIGURE 2 |** SST entropy fields with microstates, $M = 2, 3, 4$, and 8 for SODA reanalysis (left column), CCSM4 (middle column) and HadGEM2-ES (right column) CMIP5 historical runs. The data span over the period 1967–2005 for models and 1980–2018 for SODA reanalysis. All months are considered.

by ENSO (see also Webster et al., 1999; Cai et al., 2009). As a result, its predictability does depend on ENSO through an atmospheric teleconnection and possibly through an oceanic bridge. The atmospheric teleconnection is indisputable and is achieved through changes in the Walker circulation over the western Pacific Ocean during the development of ENSO events (e.g., Lau and Nath, 2003; Annamalai et al., 2005; Lau et al., 2005; Behera et al., 2006; Kug et al., 2006; Izumo et al., 2010; Kajtar et al., 2015). Some studies suggest that the ENSO-IOD interaction may also be modulated by an oceanic bridge through changes in the Indonesian Throughflow (ITF) transport between the two basins (Bracco et al., 2005; Yuan et al., 2011; Zhou et al., 2015). The ENSO modulations of the thermocline depth in the Pacific Ocean, and particularly in the Warm Pool region, propagate through the ITF to the northwestern Australia coast and then into the IO (Cai et al., 2005; Behera et al., 2006). The ITF signal is then transported to the southeastern tropical IO (SETIO) by coastal Kelvin waves that develop off south Java (Sprintall et al., 1999). Modeling studies found indeed that closing the ITF annihilates the ENSO-IOD relationship (Wajsowicz and Schneider, 2001; Bracco et al., 2005; Song et al., 2007; Santoso et al., 2011; Kajtar et al., 2015) and van Sebille et al. (2014) confirmed that the ITF transport correlates with ENSO using an eddy-permitting ocean model. In section Entropy and the ENSO-IOD Relation we will briefly investigate these atmospheric and oceanic connections in light of the large discrepancies found among models and between models and reanalysis in the entropy field.

We focus next on future projections. The multi-model mean consensus on the evolution of ENSO under global warming in CMIP5 is toward a strengthening of its amplitude and permanent El-Niño-like conditions (IPCC, 2014; Cai et al., 2015). At the same time the IO warms up but not uniformly. The predicted warming differs significantly among models, but in most the western Indian Ocean warms more than the east side (Di Nezio et al., 2020). In all models but the two versions of GISS-E2, CNRM-CM5 and INMCM4, the warming pattern over the IO resembles that of a positive IOD, which is associated with above normal precipitations over East Africa in fall and greater chances of fires in Australia and Indonesia (not shown). Despite agreement among many models in the warming patterns, no robust behavior is found in the model representation of tropical SST predictability in the RCP8.5 global warming scenarios (**Figure 5**). In the Pacific, ACCESS1.0, CanESM, IPSL-CM5A-LR are characterized by a decrease in entropy, the opposite is verified in ACCESS1.3, CMCC-CESM, GFDL-CM3, both versions of the GISS model, HadGEM2-ES, IPSL-CM5A-MR, MPI-ESM-LR and INMCM4. The remaining models are almost invariant. In the Indian Ocean the behavior may differ from the Pacific. The entropy decreases in the ACCESS and IPSL runs, all four of them, in CanESM, MRI and in the GFDL model south of the Equator, increases in the CNRM-CM5 run, which had the closest representation to the reanalysis in the historical period, and in the GISS-E2-R model, and is nearly unchanged in the rest of the cases.

The historical and future entropies suggest that many CMIP5 models misrepresent both ENSO and the IOD in boreal fall, and their relationship, and generally underestimate the complexity of the Indian Ocean SST variability, which is neither a mere response to ENSO or independent of it.

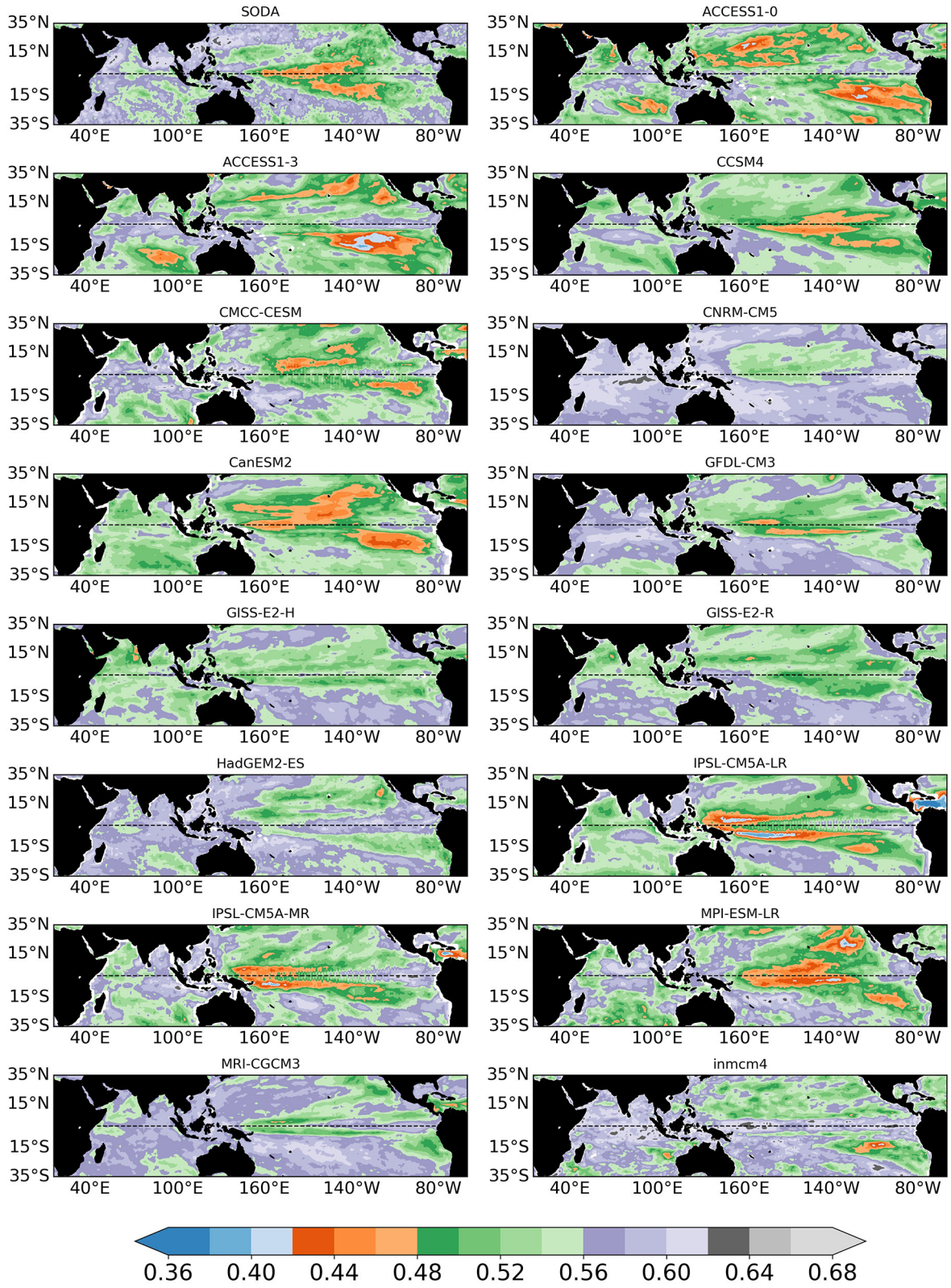



**FIGURE 3 |** Boreal spring (March–April–May) SST entropy fields with microstates, $M = 3$ during 1967–2005 for models and during 1980–2018 for SODA reanalysis.

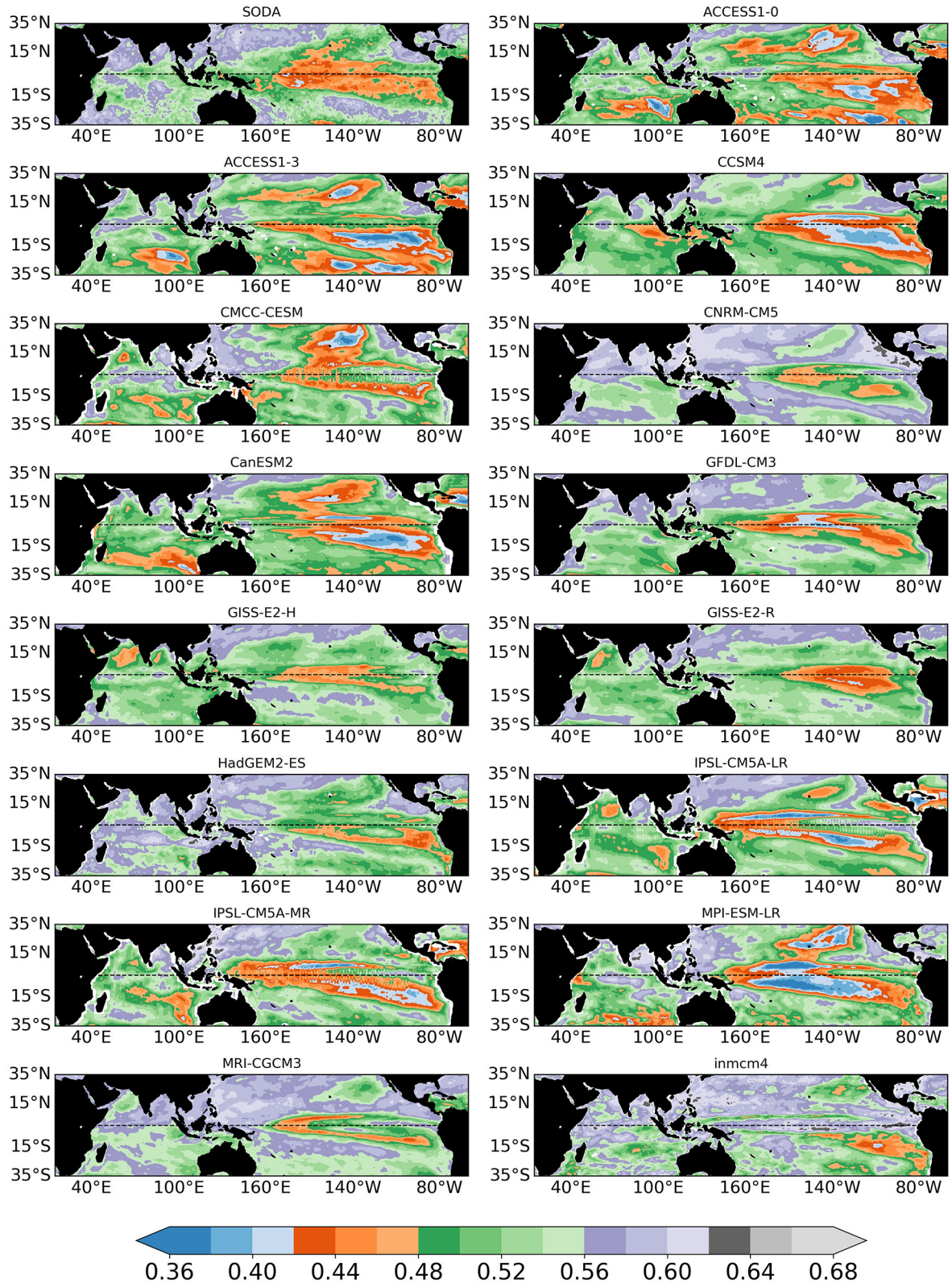



**FIGURE 4 |** Extended boreal fall (August–September–October–November, ASON) SST entropy fields with microstates $M = 3$ during 1967–2005 for CMIP5 models and during 1980–2018 for SODA reanalysis.

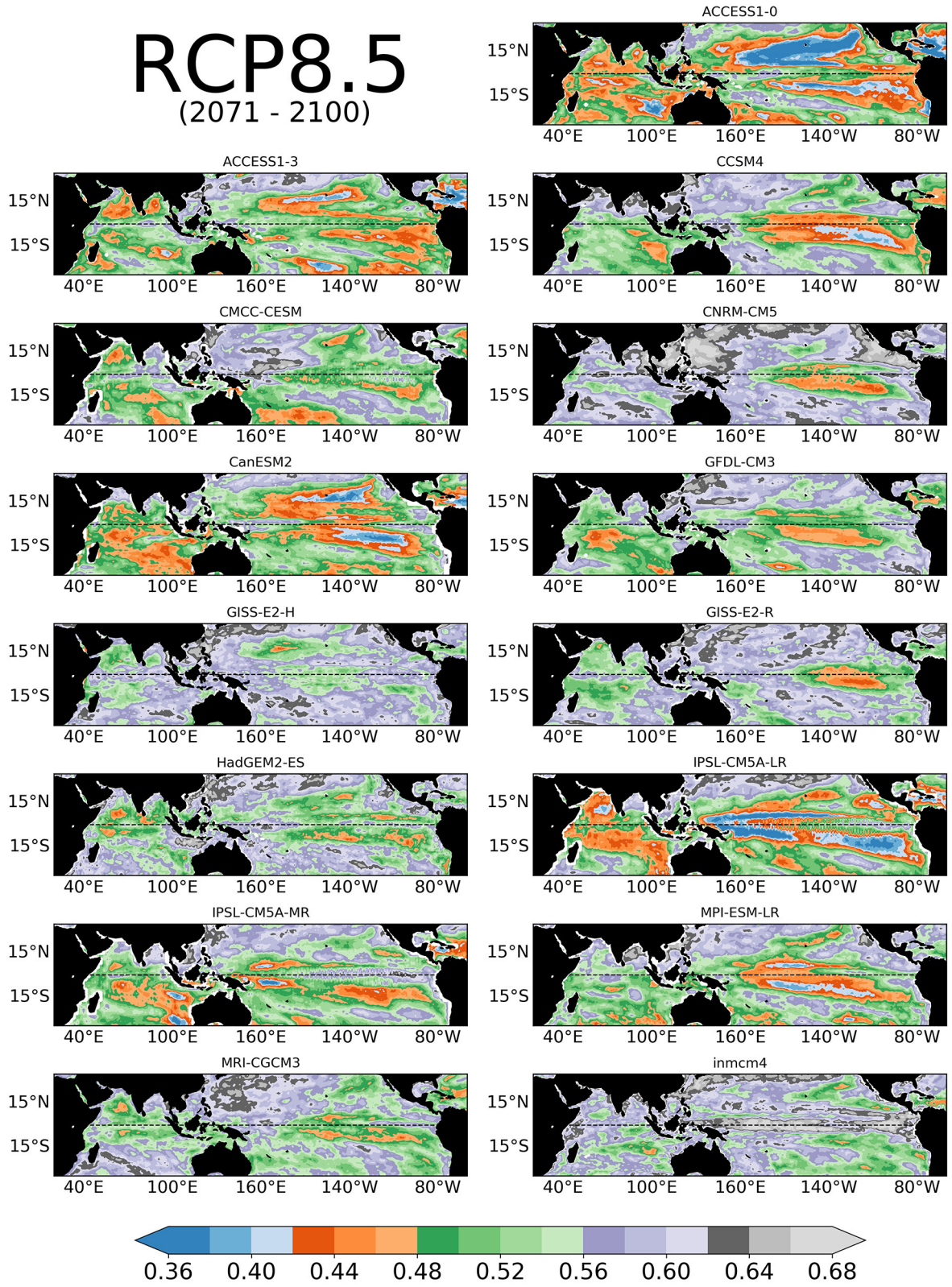



**FIGURE 5 |** Extended boreal fall (ASON) SST entropy fields with microstates, $M = 3$ for CMIP5 RCP8.5 runs used in this study. The model data span the period 2071–2100.

## ENTROPY AND THE ENSO-IOD RELATION

In this section, we first discuss how the entropy measure relates to more traditional analysis methods and then we delve into the model biases in the ENSO-IOD representation. We focus on the historical period, for which reanalysis products are available for comparison.

The information entropy is a non-linear time series analysis tool that allow for extracting and accounting for non-linear information that cannot be resolved using traditional linear methods, such as Empirical Orthogonal Function (EOFs), or power spectra. At the same time, the entropy bias can be anticipated in part from a traditional EOF analysis of SST anomalies (**Figure 6**). Indeed, the entropy identifies major problems in the representation of climate variability patterns, as done by the EOFs, if these biases influence recurrence. Additionally, the information entropy introduces information on the complexity of the climate modes in different regions. Too large complexity in the equatorial Pacific, for example, is indicative of a structural problem common to several models which is not (or not only) linked to the modeled SST ENSO patterns *per se*, but to the representation of equatorial dynamics both in the atmosphere and in the ocean. The Indo-Pacific EOF patterns are generally more in agreement with reanalysis data than the entropy field. ENSO is often modeled with the observed strength and maximum load at the equator even when it extends too far into the warm pool region (e.g., ACCESS1.3, CCSM4, GFDL, GISS-E2-H, INMCM4, both IPSL versions and MPI) and has too much strength in the central and western portion of the Pacific basin compared to the eastern one (Bellenger et al., 2014; Chen et al., 2017). It is however too strong in CMCC-CESM and CanESM, and too weak in HadGEM2, MRI and INMCM4. In the reanalysis, the SST anomaly patterns in the IO are characterized by a modest negative signal into the SETIO region that extends into the IO from the warm pool region, and by a positive signal elsewhere with slightly larger intensity in the Arabian Sea. The SETIO pattern is often underestimated (this is the case for CanESM and HadGEM-ES, both versions of the GISS and IPSL models), and is stronger than observed but equatorially confined in GFDL. The positive signal, on the other hand, is overestimated in CMCC, GFLD, IPSL and MPI. The EOF patterns, again, do not correspond to the entropy ones, especially south of 15°S, where the SST predictability is often greatly overestimated by the models.

The complexity of ENSO is also—but not exclusively—linked to its power spectrum and therefore its periodic recurrence in time. **Figure 7** compares the spectra among models and the reanalysis based on the (monthly) Nino3.4 index. The bias in the modeled spectra is reflected in that of the entropy field along the equatorial Pacific, with the caveat that several models overestimate the complexity of ENSO at the equator but display highly predictable (and therefore too regular) dynamics away from it, especially in the southern hemisphere (see e.g., the ACCESS models). Noticeably, models with too much power that spreads across multiple time scales maintain relatively low predictability, being the last related to the repetitiveness of the microstates (see e.g., CMCC-CESM). Given the interval considered, low entropy characterizes also models with ENSO power on time scales that are too long (e.g., HadGEM2-ES, for which ENSO peaks at 8–10 years).

In **Figure 8** we show the variance of the Niño-3.4 and IOD indices across the four seasons. In many models, the variances of the two indices are comparable in magnitude, while being characterized by an approximately 3:1 ratio in the observational dataset. Additionally, the Niño-3.4 -IOD cross-correlation (**Supplementary Figure 3**), which is linked to the directionality of the connections between the two modes and therefore the predictability potential across the basins, indicates that the Niño-3.4 variability lags, or co-occurs with, the IOD one in most models, while it leads by 2 months in the reanalysis. CNRM-CM5 emerges again as the model with the closest representation to the reanalysis. Cross-correlations between the Nino3.4 index and the SETIO and WTIO indices separately point to the eastern region of the IO as source of the bias (not shown).

To better highlight what limits the reliability of model climate predictions in the IO, we next examine briefly atmospheric and oceanic connections induced by developing ENSO events affect the IO and vice versa. We use standard correlation and regression analysis and our investigation does not aim at being exhaustive, but simply identifies possible biases that are shared among many models and appear to influence how predictability is simulated.

### Atmospheric Teleconnection

The sea level pressure anomalies (SLPAs) over the SETIO region can be used as a proxy for the strength of atmospheric changes to the Walker circulation induced by ENSO over this area. A simple cross-correlation analysis shows that the modeled SLPa-ENSO relationship is captured by all models (**Figure 9**). The same is verified for the WTIO region (not shown). Interestingly CNRM-CM5 underestimates the signal, which is better captured by the GISS runs, CanESM or GFDL.

From this simple analysis, we can infer that the bias in representing the IOD-ENSO interaction is not caused, at a first order, by a deficient representation of the atmospheric teleconnection from the Pacific into the IO. The IOD can independently modify the Walker circulation through air-sea interactions in SETIO (Izumo et al., 2010) but the ENSO influence on the Walker circulation is captured relatively well in all models. We will show later that the same cannot be said if we zoom onto the ITF region.

### Oceanic Bridge

ENSO influences the IO ocean circulation by modulating the transport of warm and fresh water from the Pacific Warm Pool region into the IO via the ITF. Positive transport anomalies occur during La Niña events and vice versa in El Niño years (Meyers, 1996). England and Huang (2005) found the ITF transport, defined as the depth integrated transport over the whole water column at 8°S, 120°–131.5°E, to be anticorrelated with the Niño-3 index (SSTa averaged over 150°W−90°W, 5°S−5°N), with a maximum of −0.32 when the ITF lags ENSO by 9 months. Bracco et al. (2005) focused instead on the variability of the 20°C isotherm (Z20) that may more directly impact the upwelling in the SETIO region and found that that the fall correlation

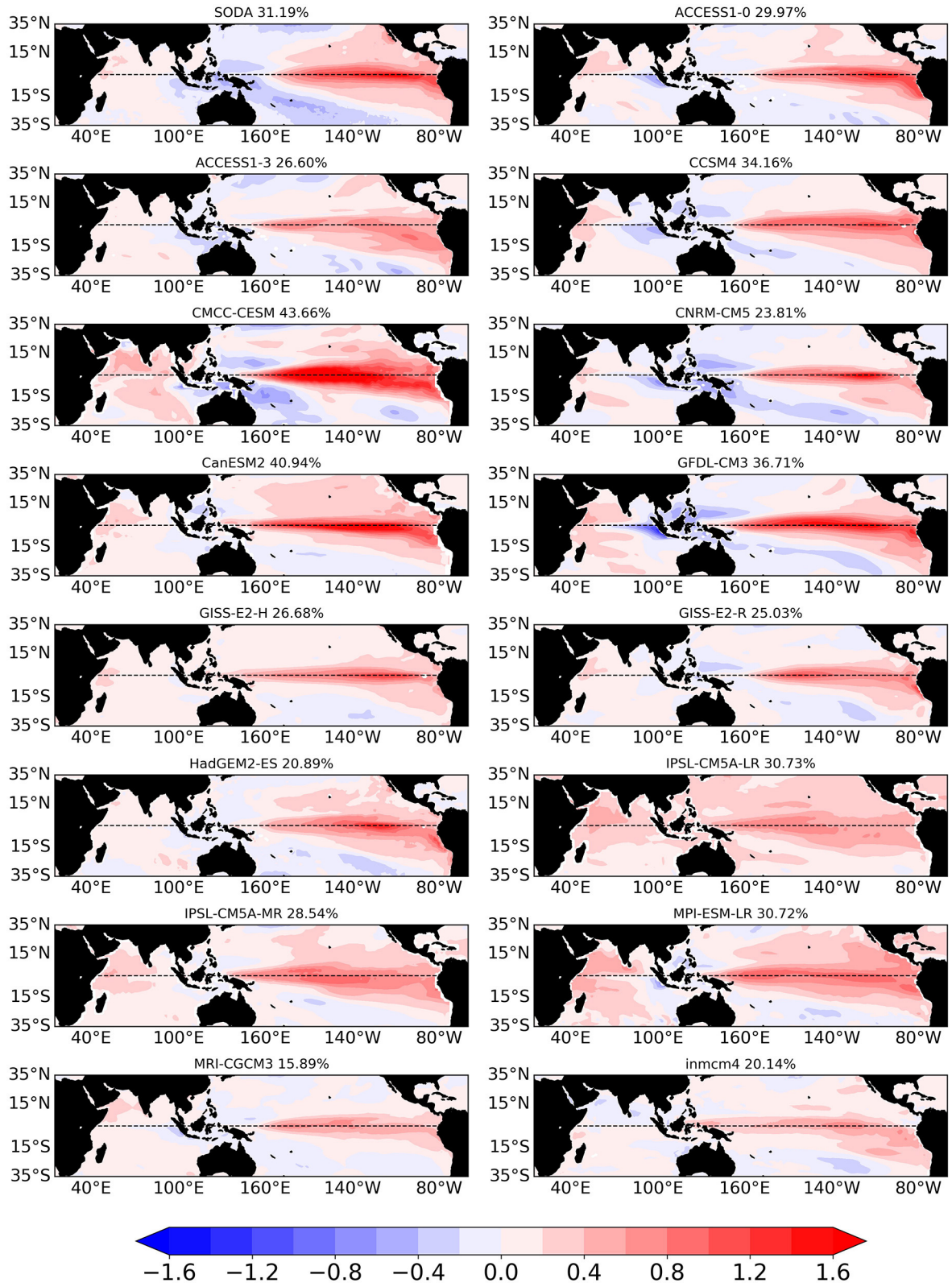



**FIGURE 6 |** The first EOF patterns of tropical Indo-Pacific sea surface temperature anomalies during the extended boreal fall season (ASON) in 1967–2005 for models and 1980–2018 for SODA reanalysis. Percentages of variance explained by the EOF patterns are included.

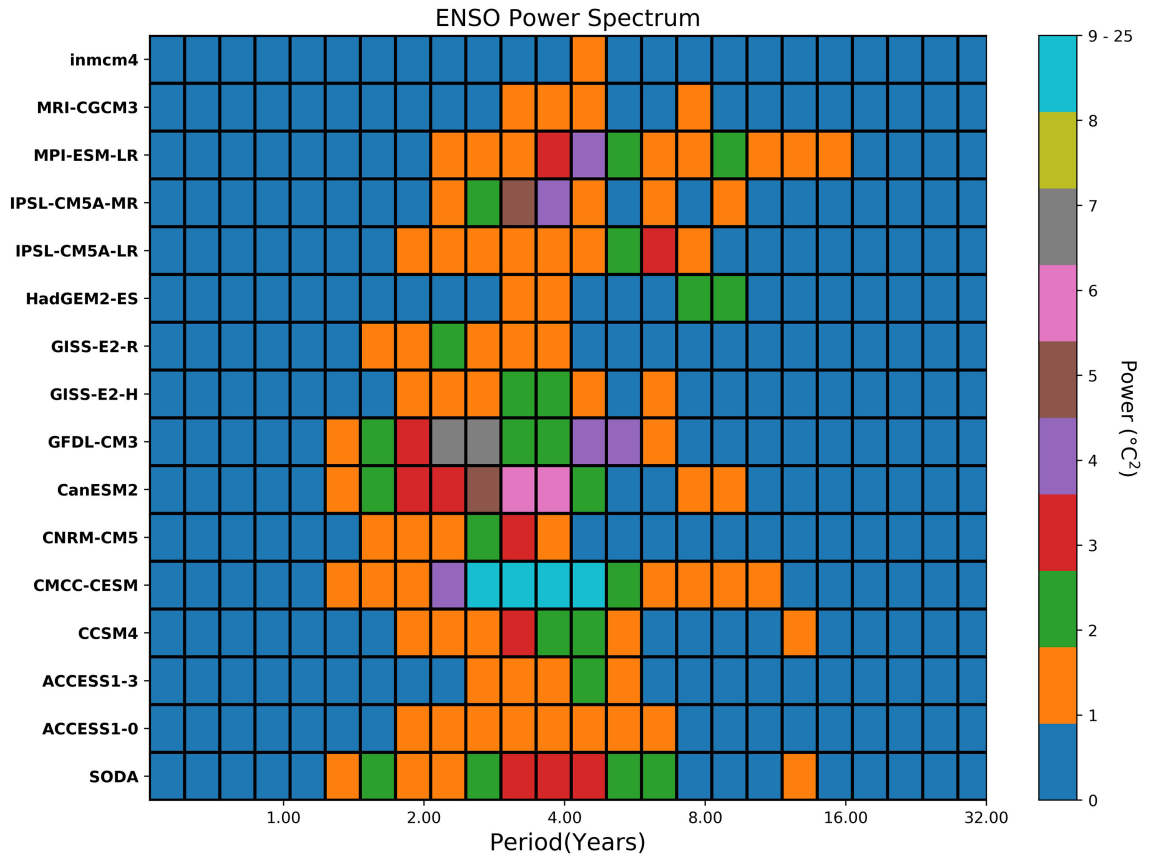



**FIGURE 7 |** Power spectrum of Niño3.4 index during 1967–2005 for models and during 1980–2018 for SODA reanalysis. The period (in years) is on the x-axis and the power is shaded.

between the Z20 anomalies and Niño-3.4 reached −0.62 over the period 1958–2002.

In SODA, the maximum (minimum) transport occurs in boreal spring (fall), and the thermocline depth oscillates between 123 and 135 m. In the models, the annual cycle varies greatly in both amplitude and phase (**Supplementary Figure 4**). Several models have a thermocline that is 30 m or more too deep. Many, including INMCM4, GISS-E2- R, GISS-E2-H, MPI and CANESM, display an out of phase seasonal cycle (minimum in spring and maximum in fall) or shaped with two peaks. CMCC-CESM stands out for having a thermocline that is too shallow (about 20 m shallower than in the reanalysis in fall). CCSM, GFDL and IPSL-MR provide the closest representation to that in the reanalysis, with CCSM being the most realistic.

The cross-correlations between the Z20 anomalies averaged over the ITF region and Nino3.4 indices are shown in **Figure 10**. In the reanalysis, the maximum anticorrelation (−0.71) is found for the Niño-3.4 index lagging the ITF thermocline by 1 month. Statistically significant anticorrelations are found in all models but INMCM. However, the modeled anticorrelations are consistently maximized at about 3–5 months lags. The relationship between the Niño-3.4 index and the ITF is misrepresented by the models also when using the sea level difference between the western Pacific (WP; 0–10°N; 125°E−145°E) and the Eastern Indian Ocean (EIO; 10°S−20°S; 110°E−130°E) as ITF proxy (Mayer et al., 2018). Correlation coefficients are smaller than for Z20 but significant in the reanalyses, while the linkage is missed entirely in most models (**Supplementary Figure 5**). This is indicative of problems in the representation of both oceanic and atmospheric processes at the equator.

Finally, the regression maps of the SST over the tropics and of subsurface temperatures averages between 5°N and 5°S onto the IOD index are shown in **Figures 11**, **12**. At the surface the link between the IOD and the ENSO anomalies are captured by most models but INMCM4 and underestimated by ACCESS1.3, HadGEM2-ES, followed by CCSM4, MPI and MRI and CNRM-CM5. At the ocean surbsurface, on the other hand, the regression patterns are generally strong underneath the warm pool, where they have the same sign of the SETIO region, but underestimated in the central and eastern Pacific. At the subsurface CNRM-CM5 emerges as closest to the reanalysis in the pattern representation. Modeled IODs somewhat mimic the characteristics of the *unseasonal* IOD events identified in the observations by Du et al. (2013), but for their seasonality. The *unseasonal* IOD are relevant in late spring and summer

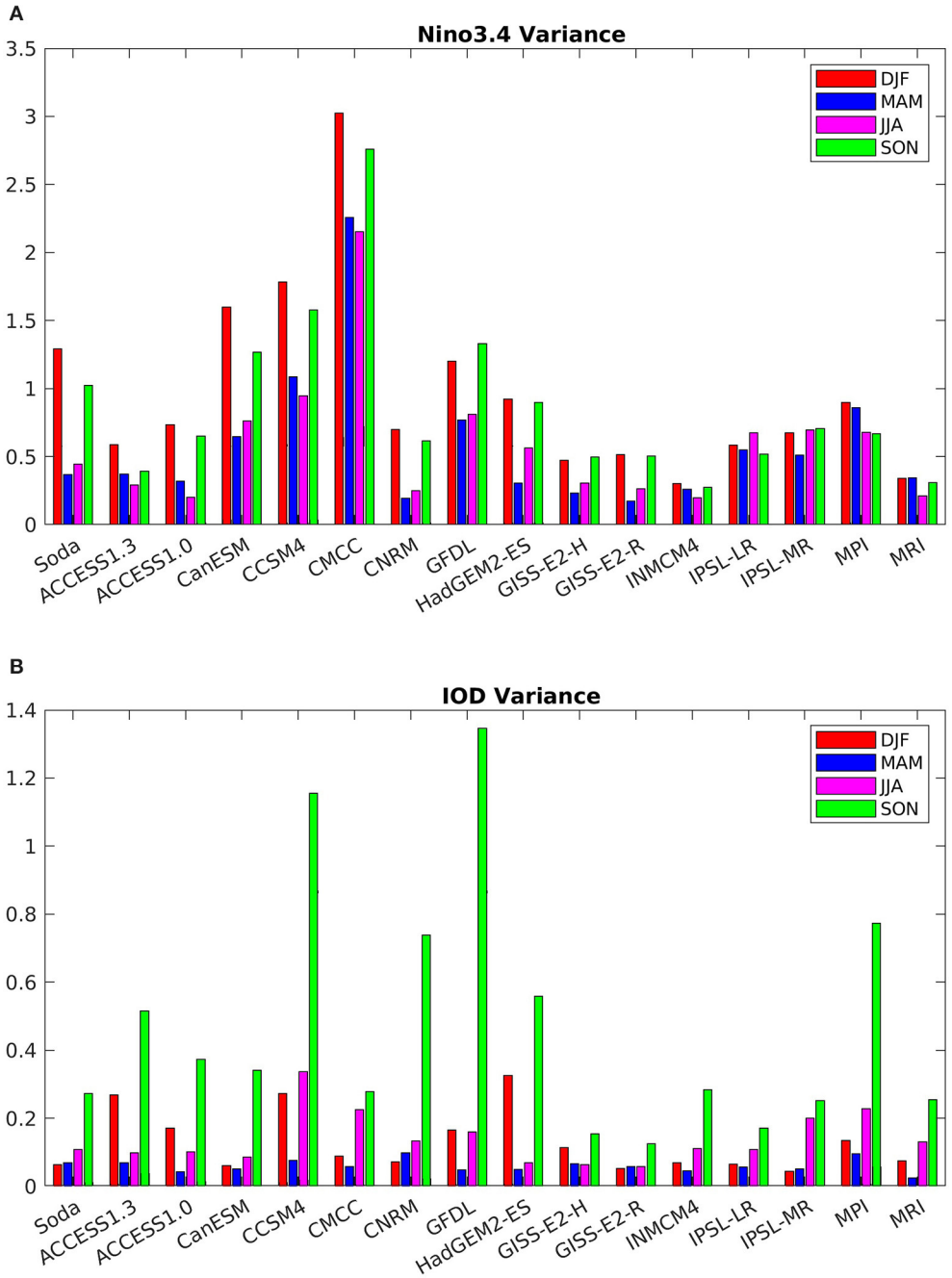



**FIGURE 8 |** Seasonal variance of Niño3.4 **(A)** and IOD **(B)** indices during 1967–2005 for models and during 1980–2018 for SODA reanalysis.

since the 1970's, and resemble the CMPI5 ones by being mostly independent of ENSO and forced by (too strong for fall in the CMIP5 case) equatorial winds. This figure also points to the differences in the representation of the ITF bathymetry among the models.

The representation of the relative seasonal evolution of the two modes identified in the cross-correlations and the greater independence of the modeled IOD from the Pacific evolution emerge as problematic in all models. Our brief analysis extends the results by Cai and Cowan (2009) that highlighted other systematic biases in the IO and in particular (a) the western IO thermocline being deeper than the eastern IO while the opposite is verified in the observations so that the wave dynamics that transport the IOD temperature anomalies are not reproduced correctly; (b) slightly warmer SST in the western IO than eastern IO, while a much warmer eastern IO is observed; (c) stronger than observed climatological easterly winds over the equatorial Pacific.

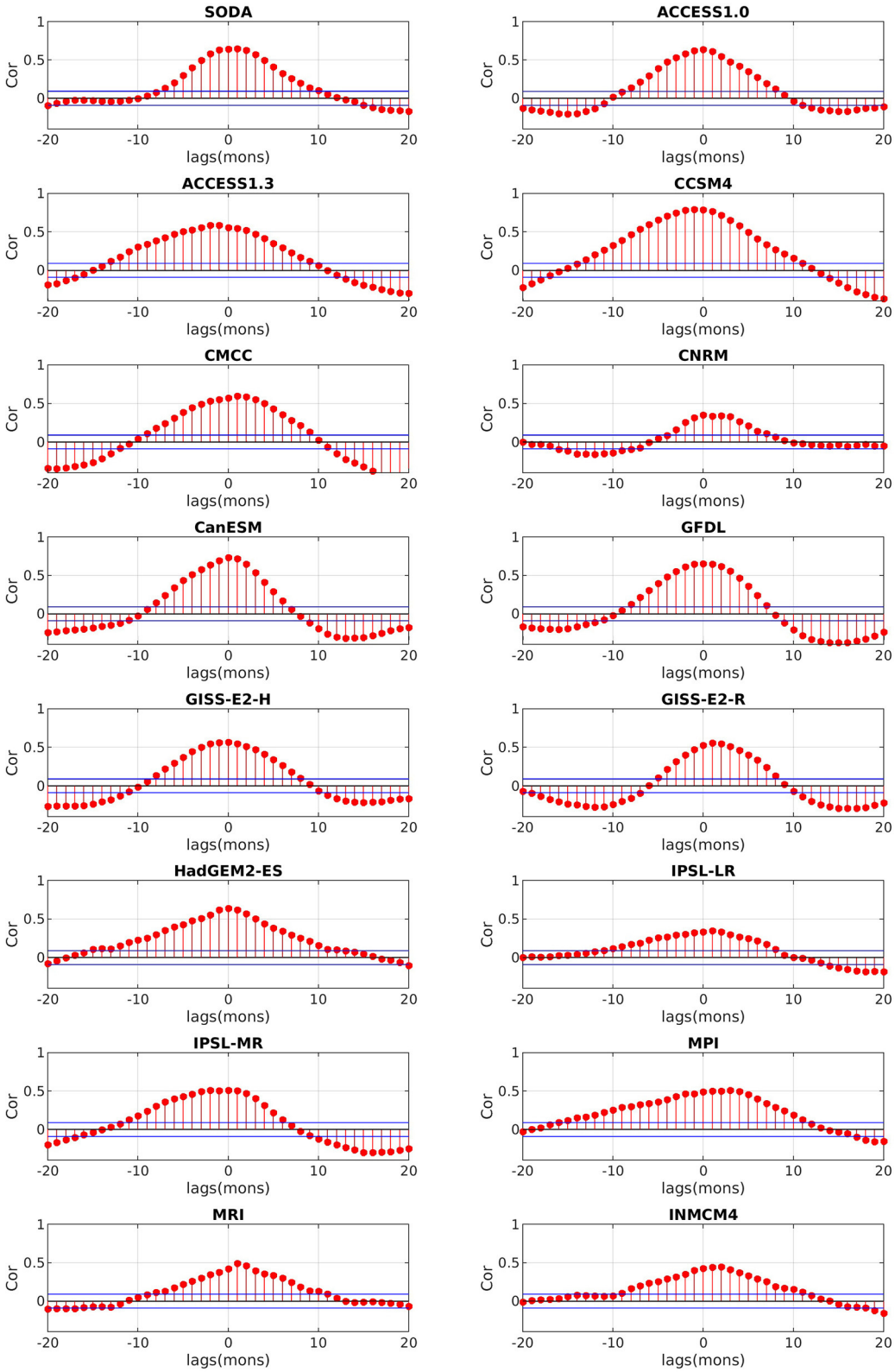



**FIGURE 9 |** Cross correlation between the sea level pressure anomalies (SLPAs) over the SETIO region and the Niño3.4 Index during 1967–2005 for models and during 1980–2018 for SODA/ERA-interim reanalysis. Positive (negative) lags indicate SLPa lags (leads) the Niño3.4 Index.

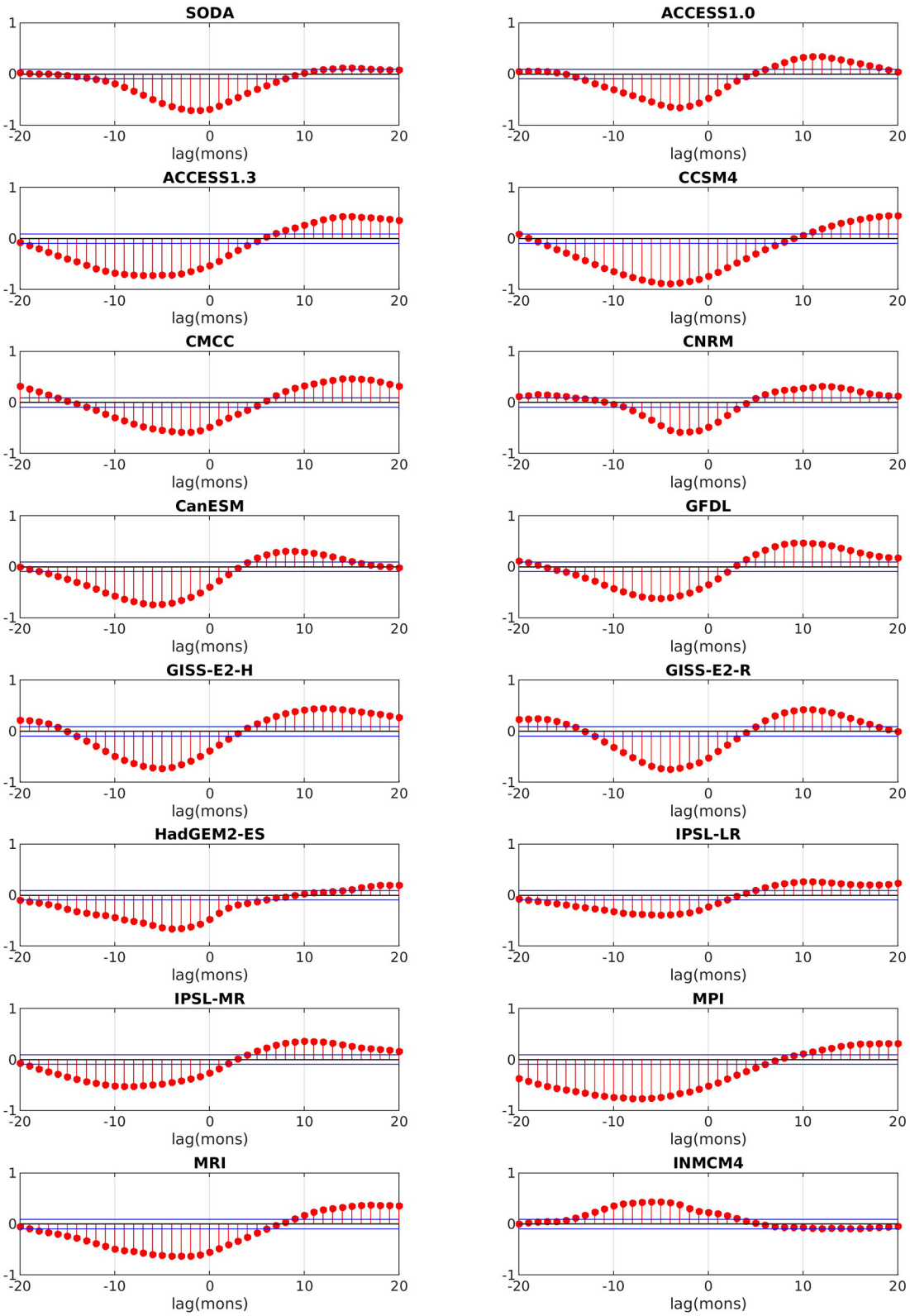



**FIGURE 10 |** Cross correlation between the Indonesian Throughflow (ITF) thermocline depth anomaly and the Niño3.4 index during 1967–2005 for models and during 1980–2018 for SODA reanalysis. Positive (negative) lags indicate ITF lags (leads) the Niño3.4 index.

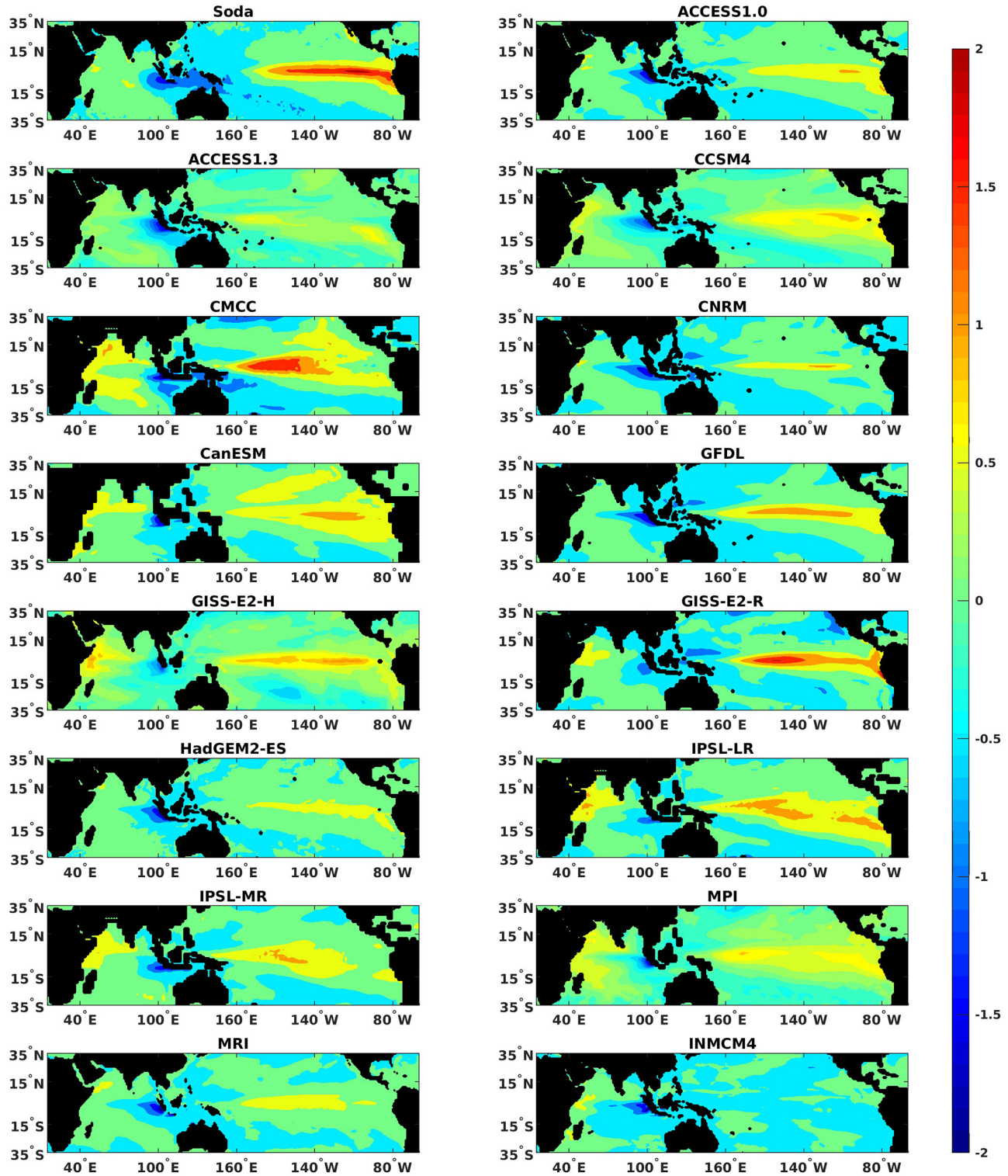



**FIGURE 11 |** Regression maps of tropical Indo-Pacific sea surface temperatures on the IOD during the extended boreal fall season (ASON) of 1967–2005 for models and 1980–2018 for SODA reanalysis.

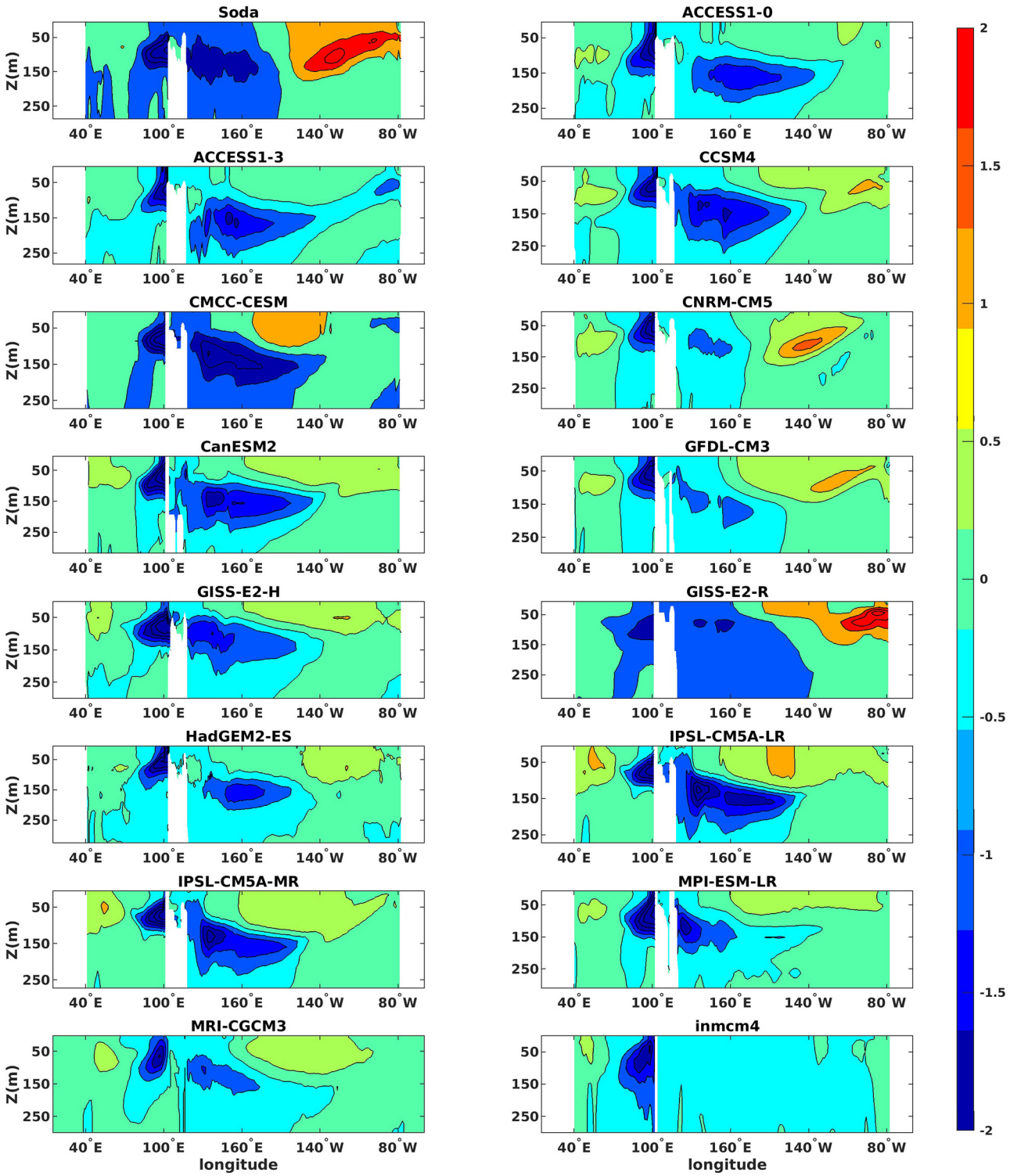



**FIGURE 12 |** Regression maps of the tropical Indo-Pacific sea subsurface temperatures averaged between 5°N and 5°S onto the IOD during the extended boreal fall season (ASON) of 1967–2005 for models and 1980–2018 for SODA reanalysis.

## SUMMARY

Weather prediction is reliable to <10 days due to the atmosphere chaotic behavior, the sensitivity on the initial conditions and model limitations. Climate prediction, on the other hand, has the potential to extend to longer time scales due to the modulations exerted by its slowly varying component, the ocean. Climate modes that manifest as non-linear oscillations, such as the El Niño-Southern Oscillation, the Pacific Decadal Oscillation, the Atlantic Multi-Decadal Oscillation, the Indian Ocean Dipole, etc. further contribute to the extended range of potential predictability of the climate system.

Global warming is posed to modify these modes of natural variability that control much of the predictability at seasonal to interannual scales. Examples include the intensification of the IOD (Di Nezio et al., 2020), and a higher frequency of Central Pacific El Niños (Freund et al., 2019) observed in the last decade and projected to continue in the future, and shifts in its global-reaching teleconnections following the expansion of the Hadley cells (Kang and Lu, 2012). In light of the societal impacts that these modes exert by modulating surface temperatures, storm frequency, droughts and floods, their current and future predictability potential is of the uttermost importance. It remains, however, elusive due to the complexity of the climate system, the biases in climate models, the challenges implicit in quantifying it.

In this work, we introduced a new method, building upon tools developed within the non-linear dynamical systems community, to quantify predictability in terms of information entropy. The information entropy of a climate field quantifies the degree of complexity of a given region or grid point in terms of recurrence (how many times a phase space trajectory visits roughly the same area in the phase space over a given period) and evaluate how it varies over time. Given that the evolution of a climate field such as SST results from interactions of many fields, from atmospheric winds and heat fluxes to ocean currents, its space phase evolution accounts for all these interactions, linear and non-linear.

Most studies employing a comparable non-linear data analysis technique in climate science have focused on a limited number of 1-dimensional time series (i.e., paleoclimate proxies—one or few at the time—or their modeled equivalent) (Donges et al., 2011, 2015). The possibility to use it on large, multi-dimensional fields has been opened by Corso et al. (2018) by introducing a new entropy quantifier applicable to d-dimensional fields using the recurrent plot (RP) technique (Eckmann et al., 1987) without the need for embedding or phase space reconstruction. We applied it to the SST field in the tropical Pacific and Indian Ocean, comparing CMIP5 models and reanalysis data, over 39 years in the historical period and in the last 30 years of the XXI century RCP8.5 projections. We focused on the boreal fall season, when the ENSO signature is strong and the IOD has maximum variance in the Indian Ocean. Understanding the ENSO—IOD relation and the influence of the IOD on all countries facing the Indian Ocean is essential to improve prediction of extreme rainfall and droughts over Australia, India, Indonesia and East Africa. In these highly populated regions downscaled regional experiments (Hashizume et al., 2012) are limited by the representation of the large scale atmospheric and oceanic circulation.

Our analysis highlights the limitations that coupled climate models still face in addressing climate predictability and its potential changes. No robust signal is found in the models, in both basins, as to how predictability will evolve in the future when considering the projections for the last 30 years of the XXI century. In the historical period, very few models capture both pattern and intensity of the entropy signal of the reanalysis, and CNRM-CM5 outperforms all others in this regard. Interestingly, in an intercomparison performed using complex network analysis (Fountalis et al., 2015), CNRM-CM5 outperformed all other CMIP5 models also in the representation of rainfall patterns and their connectivities, but performed as well as several others in representing the (global) SST network. This again points to the power of the entropy quantifier as a measure of topological similarity linked to predictive power. The entropy quantifier, in summary, is a computationally efficient performance metric for coupled climate models, capable of capturing dynamical characteristics, both linear and non-linear, that goes beyond the climatology and local variability of the field under investigation, and delves into its topological properties.

We also showed that a biased representation of coupled equatorial dynamics and of the atmospheric and subsurface oceanic bridge between the Pacific and Indian Oceans via the ITF contributes to the poor representation of the Indo-Pacific entropy in fall. Incidentally, we verified that the intra-model variability is relatively small in the historical runs in the statistics presented, from the entropy to the cross-correlations, and each model behavior is consistent across ensemble members. Overall, many models struggle at the equator, in both basins, and display unrealistic regular dynamics in the IO, especially south of the equator and/or in the Arabian Sea.

Our results exemplify how information entropy may contribute a new powerful tool to investigate the potential predictability of the climate system. More work is needed to explore its relevance on time scales other than interannual, using higher or lower frequency time series as input to explore, for example, subseasonal phenomena o changes across millennia, and the entropy usefulness in analyzing fields other than SSTs. In the near future, we aim at applying it across time scales and with high frequency (at least daily) data to quantify how and where climate predictability emerges from the weather noise.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The SODA reanalysis data can be obtained from: https://www2.atmos.umd.edu/~ocean/index_files/soda3.4.2_mn_download_b.htm. ERA-Interim can be obtained from https://apps.ecmwf.int/datasets/data/interim-full-mnth/levtype=sfc/. Monthly mean output from Coupled Model Intercomparison Project Phase 5 (CMIP5) historical and RCP8.5 experiments are publicly available on https://esgf-node.llnl.gov/search/cmip5/. The entropy code can be found at https://github.com/FabriFalasca/NonLinear_TimeSeries_Analysis.

## AUTHOR CONTRIBUTIONS

OI performed all analyses. FF developed and optimized the software for the information entropy quantifier. AB coordinated the analyses and wrote the manuscript. All authors contributed to the interpretation of the results and approved the submitted version.

## FUNDING

AB and FF were supported by the Department of Energy, Regional and Global Model Analysis (RGMA) Program, Grant No.: 0000253789.

## ACKNOWLEDGMENTS

This research resulted from a term project during a course in the School of Earth and Atmospheric Sciences at Georgia Tech. We thank the School for contributing toward the publication costs of this work. We also thank the two reviewers who contributed to improving the original submission.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fclim.2021.675840/full#supplementary-material

## REFERENCES

Allan, R., Chambers, D. P., Drosdowsky, W. L., Hendon, H. A. H., Latif, M. Z., Nicholls, N. S., et al. (2001). Is there an indian ocean dipole and is it independent of the el niño-southern oscillation. *Clivar. Exch.* 6, 18–22. doi: 10.5194/bg-10-6677-2013

Annamalai, H., Xie, S. P., McCreary, J. P., and Murtugudde, R. (2005). Impact of indian ocean sea surface temperature on developing el niño. *J. Clim.* 18, 302–319. doi: 10.1175/JCLI-3268.1

Anyamba, A., Chretien, J.-P., Britch, S. C., Soebiyanto, R. P., Small, J. L., Jepsen, R., et al. (2019). Global disease outbreaks associated with the 2015–2016 El Niño event. *Sci. Rep.* 9:1930. doi: 10.1038/s41598-018-38034-z

Ashok, K., Guan, Z., and Yamagata, T. (2001). Impact of the indian ocean dipole on the relationship between the indian monsoon rainfall and enso. *Geophys. Res. Lett.* 28, 4499–4502. doi: 10.1029/2001GL013294

Ashok, K., Guan, Z., and Yamagata, T. (2003). A look at the relationship between the enso and the indian ocean dipole. *J. Meteorol. Soc. Japan* 81, 41–56. doi: 10.2151/jmsj.81.41

Balasis, G., Donner, R. V., Potirakis, S. M., Runge, J., Papadimitriou, C., Daglis, I. A., et al. (2013). Statistical mechanics and information-theoretic perspectives on complexity in the earth system. *Entropy* 15, 4844–4888. doi: 10.3390/e15114844

Bandt, C., and Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* 88:174102. doi: 10.1103/PhysRevLett.88.174102

Behera, S. K., Krishnan, R., and Yamagata, T. (1999). Unusual ocean-atmosphere conditions in the tropical indian ocean during 1994. *Geophys. Res. Lett.* 26, 3001–3004. doi: 10.1029/1999GL010434

Behera, S. K., Luo, J. J., Masson, S., Rao, S. A., Sakuma, H., and Yamagata, T. (2006). A CGCM study on the interaction between IOD and ENSO. *J. Clim.* 19, 1688–1705. doi: 10.1175/JCLI3797.1

Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M., and Vialard, J. (2014). ENSO representation in climate models: from CMIP3 to CMIP5. *Clim. Dyn.* 42, 1999–2018. doi: 10.1007/s00382-013-1783-z

Bjerknes, J. (1966). A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus* 18, 820–829. doi: 10.3402/tellusa.v18i4.9712

Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Rev.* 97, 163–172. doi: 10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2

Black, E., Slingo, J., and Sperber, K. R. (2003). An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean sst. *Mon. Wea. Rev.* 131, 74–94. doi: 10.1175/1520-0493(2003)131<0074:AOSOTR>2.0.CO;2

Bracco, A., Kucharski, F., Molteni, F., Hazeleger, W., and Severijns, C. (2005). Internal and forced modes of variability in the Indian Ocean. *Geophys. Res. Lett.* 32:L12707. doi: 10.1029/2005GL023154

Cai, W., and Cowan, T. (2009). Why is the amplitude of the indian ocean dipole overly large in cmip3 and CMIP5 climate models? *Geophys. Res. Lett.* 40, 1200–1205. doi: 10.1002/grl.50208

Cai, W., Hendon, H. H., and Meyers, G. (2005). Indian ocean dipolelike variability in the csiro mark 3 coupled climate model. *J. Clim.* 18, 1449–1468. doi: 10.1175/JCLI3332.1

Cai, W., Pan, A., Roemmich, D., Cowan, T., and Guo, X. (2009). Argo proles a rare occurrence of three consecutive positive indian ocean dipole events, 2006–2008. *Geophys. Res. Lett.* 36. doi: 10.1029/2008GL037038

Cai, W., Santoso, A., Wang, G., Yeh, S.-W., An, S.-I., Cobb, K. M., et al. (2015). ENSO and greenhouse warming. *Nat. Clim. Change* 5, 849–859. doi: 10.1038/nclimate2743

Carton, J. A., Chepurin, G. A., and Chen, L. (2018). Soda3: a new ocean climate reanalysis. *J. Clim.* 31, 6967–6983. doi: 10.1175/JCLI-D-18-0149.1

Chen, C., Cane, M., Wittenberg, A., and Chen, D. (2017). ENSO in the CMIP5 simulations: life cycles, diversity, and responses to climate change. *J. Clim.* 30, 775–801. doi: 10.1175/JCLI-D-15-0901.1

Corso, G., Prado, T., Lima, G., Kurths, J., and Lopes, S. (2018). Quantifying entropy using recurrence matrix microstates. *Chaos* 28:083108. doi: 10.1063/1.5042026

Cvitanović, P., Artuso, R., Mainieri, R., Tanner, G., and Vattay, G. (2016). *Chaos: Classical and Quantum,* ChaosBook.org. Copenhagen: Niels Bohr Institute.

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

Dhiman, R. C., and Sarkar, S. (2017). El Niño southern oscillation as an early warning tool for malaria outbreaks in India. *Malaria J.* 16:122. doi: 10.1186/s12936-017-1779-y

Di Nezio, P., Puy, M., Thirumalai, K., Jin, F. F., and Tierney, J. (2020). Emergence of an equatorial mode of climate variability in the Indian Ocean. *Sci. Adv.* 6:7684. doi: 10.1126/sciadv.aay7684

Donges, J. F., Donner, R. V., Marwan, N., Breitenbach, S. F. M., Rehfeld, K., and Kurths, J. (2015). Non-linear regime shifts in Holocene Asian monsoon variability: potential impacts on cultural change and migratory patterns. *Clim. Past* 11, 709–741. doi: 10.5194/cp-11-709-2015

Donges, J. F., Donner, R. V., Trauth, M. H., Marwan, N., Schellnhuber, H. J., and Kurths, J. (2011). Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20422–20427. doi: 10.1073/pnas.1117052108

Donner, R., Small, M., Donges, J., Marwan, N., Zou, Y., Xiang, R., et al. (2011). Recurrence-based time series analysis by means of complex network methods. *Int. J. Bifurcat. Chaos* 21, 1019–1046. doi: 10.1142/S0218127411029021

Du, Y., Cai, W., and Wu, Y. (2013). A new type of the Indian Ocean dipole since the Mid-1970s. *J. Clim.* 26, 959–972. doi: 10.1175/JCLI-D-12-00047.1

Duan, W., and Wei, C. (2013). The 'spring predictability barrier' for ENSO predictions and its possible mechanism: results from a fully coupled model. *Int. J. Climatol.* 33, 1280–1292. doi: 10.1002/joc.3513

Eckmann, J.-P., Kamphorst, S. O., and Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhys. Lett.* 5, 973–977. doi: 10.1209/0295-5075/4/9/004

England, M. H., and Huang, F. (2005). On the interannual variability of the Indonesian throughflow and its linkage with ENSO. *J. Clim.* 18, 1435–1444. doi: 10.1175/JCLI3322.1

Falasca, F., Crétat, J., Braconnot, P., and Bracco, A. (2020). Spatiotemporal complexity and time-dependent networks in sea surface temperature from mid- to late Holocene. *Eur. Phys. J. Plus* 135:392. doi: 10.1140/epjp/s13360-020-00403-x

Fountalis, I., Bracco, A., and Dovrolis, C. (2015). ENSO in CMIP5 simulations: network connectivity from the recent past to the twenty-third century. *Clim. Dyn.* 45, 511–538. doi: 10.1007/s00382-014-2412-1

Freund, M. B., Henley, B. J., Karoly, D. J., McGregor, H. V., Abram, N. J., and Dommenget, D. (2019). Higher frequency of central Pacific El Niño events in recent decades relative to past centuries. *Nat. Geosci.* 12, 450–455. doi: 10.1038/s41561-019-0353-3

Garland, J., Jones, T. R., Neuder, M., White, J. W. C., and Bradley, E. (2019). An information-theoretic approach to extracting climate signals from deep polar ice cores. *Chaos* 29:101105. doi: 10.1063/1.5127211

Gilpin, W. (2020). *Deep Reconstruction of Strange Attractors From Time Series*. Cambridge, MA: Harvard University.

Gualdi, S., Guilyardi, E., Navarra, A., Masina, S., and Delecluse, P. (2003). The interannual variability in the tropical Indian Ocean as simulated by a CGCM. *Clim. Dyn.* 20, 567–582. doi: 10.1007/s00382-002-0295-z

Guan, Z., Ashok, K., and Yamagata, T. (2003). Summertime response of the tropical atmosphere to the Indian Ocean Dipole sea surface temperature anomalies. *J. Meteorol. Society of Japan. Ser. II* 81, 533–561. doi: 10.2151/jmsj.81.533

Hashizume, M., Chaves, L. F., and Minakawa, N. (2012). Indian Ocean Dipole drives malaria resurgence in east african highlands. *Sci. Rep.* 2:269. doi: 10.1038/srep00269

IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Technical report, IPCC, Geneva, Switzerland.

Izumo, T., Vialard, J., Lengaigne, M., de Boyer Montegut, C., Behera, S. K., Luo, J.-J., et al. (2010). Influence of the state of the Indian Ocean Dipole on the following year's El Niño. *Nat. Geosci.* 3:168. doi: 10.1038/ngeo760

Kajtar, J. B., Santoso, A., and England, M. H. W. (2015). Indo-pacific climate interactions in the absence of an indonesian throughflow. *J. Clim.* 28, 5017–5029. doi: 10.1175/JCLI-D-14-00114.1

Kang, S. M., and Lu, J. (2012). Expansion of the Hadley cell under global warming: winter versus summer. *J. Clim.* 25, 8387–8393. doi: 10.1175/JCLI-D-12-00323.1

Koebbe, M., and Mayer-Kress, G. (1992). "Use of recurrence plots in the analysis of time-series data," in *Nonlinear Modeling and Forecasting*, eds M. Casdagli and S. Eubank (New York, NY: Addison Wesley).

Kovats, R. S. (2000). El Niño and human health. *Bull World Health Organ.*, 78, 1127–1135. Available online at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2560836/pdf/11019461.pdf

Kreppel, K., Caminade, C., Govella, N., Morse, A. P., Ferguson, H. M., and Baylis, M. (2019). Impact of ENSO 2016–17 on regional climate and malaria vector dynamics in tanzania. *Environ. Res. Lett.* 14:075009.075014 doi: 10.1088/1748-9326/ab26c7

Kug, J., Li, T., An, S.-I., Kang, I., Luo, J., Masson, S., et al. (2006). Role of the ensoindian ocean coupling on enso variability in a coupled gcm. *Geophys. Res. Lett.* 33:24916. doi: 10.1029/2005GL024916

Lau, N.-C., Leetmaa, A., Nath, M. J., and Wang, H.-L. (2005). Influences of ENSO-induced Indo–western Pacific SST anomalies on extratropical atmospheric variability during the boreal summer. *J. Clim.* 18, 2922–2942. doi: 10.1175/JCLI3445.1

Lau, N.-C., and Nath, M. J. (2003). Atmosphere–ocean variations in the Indo-Oacific sector during ENSO episodes. *J. Clim.* 16, 3–20. doi: 10.1175/1520-0442(2003)016<0003:AOVITI>2.0.CO;2

Li, C., and Mu, C. (2001). The influence of the Indian Ocean Dipole on atmospheric circulation and climate. *Adv. Atmosphere. Sci.* 18, 831–843. doi: 10.1007/BF03403506

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *J. Atmosphere. Sci.* 20, 130–141. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Manneville, P. (1990). "Characterization of temporal chaos," in *Dissipative Structures and Weak Turbulence*, ed P. Manneville (Boston, MA: Academic Press). doi: 10.1016/B978-0-08-092445-8.50011-0

Marwan, N. (2011). How to avoid potential pitfalls in recurrence plot based data analysis. *Int. J. Bifurcat. Chaos* 21, 1003–1017. doi: 10.1142/S0218127411029008

Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Phys. Rep.* 438, 237–329. doi: 10.1016/j.physrep.2006.11.001

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature* 261, 459–467. doi: 10.1038/261459a0

Mayer, M., Alonso Balmaseda, M., and Haimberger, L. (2018). Unprecedented 2015/2016 Indopacific heat transfer speeds up tropical pacific heat recharge. *Geophys. Res. Lett.* 45, 3274–3284. doi: 10.1002/2018GL077106

Meyers, G. (1996). Variation of Indonesian throughflow and the El Niño-Southern oscillation. *J. Geophys. Res. Oceans* 101, 12255–12263. doi: 10.1029/95JC03729

Mindlin, G. M., and Gilmore, R. (1992). Topological analysis and synthesis of chaotic time series. *Phys. D Nonlinear Phenomena* 58, 229–242. doi: 10.1016/0167-2789(92)90111-Y

Poincaré, P. (1890). Sur le probleme des trois corps et les équations de la dynamique. *Acta Mathe.* 13, 1–271. doi: 10.1007/BF02392514

Prado, T., Corso, G., Santos Lima, G., Budzinski, R., Boaretto, B., Ferrari, F., et al. (2020). Maximum entropy principle in recurrence plot analysis on stochastic and chaotic systems. *Chaos* 30:043123. doi: 10.1063/1.5125921

Riedl, N. (2013). Practical considerations of permutation entropy. *Eur. Phys. J. Spec. Topics* 222, 249–262. doi: 10.1140/epjst/e2013-01862-7

Saco, P. M., Carpi, L. C., Figliola, A., Serrano, E., and Rosso, O. A. (2010). Entropy analysis of the dynamics of El Niño/Southern oscillation during the holocene. *Phys. A: Statist. Mech. Appl.* 389, 5022–5027. doi: 10.1016/j.physa.2010.07.006

Saji, N. H., Goswami, B. N., and Vinayachandran, P. N., and Yamagata, T. (1999). A dipole mode in the tropical indian ocean. *Nature* 401, 360–363. doi: 10.1038/43854

Saji, N. H., and Yamagata, T. (2003). Possible impacts of Indian Ocean Dipole mode events on global climate. *Clim. Res.* 25, 151–169. doi: 10.3354/cr025151

Santoso, A., W., England, M. H., and Phipps, S. J. (2011). The role of the Indonesian throughflow on ENSO dynamics in a coupled climate model. *J. Clim.* 24, 585–601. doi: 10.1175/2010JCLI3745.1

Shi, W. (2007). "Lyapunov exponent analysis to chaotic phenomena of marine power system," in *Fault Detection, Supervision and Safety of Technical Processes 2006*, ed H. Y. Zhang (Oxford: Elsevier Science Ltd.). doi: 10.1016/B978-008044485-7/50251-7

Song, Q., Vecchi, G. A., and Rosati, A. J. (2007). The role of the Indonesian throughflow in the Indo–Pacific climate variability in the GFDL coupled climate model. *J. Clim.* 20, 2434–2451. doi: 10.1175/JCLI4133.1

Sprintall, J., Chong, J., Syamsudin, F., Morawitz, W., Hautala, S., Bray, N., et al. (1999). Dynamics of the south Java current in the Indo-Australian basin. *Geophys. Res. Lett.* 26, 2493–2496. doi: 10.1029/1999GL002320

Takens, F. (1981). "Detecting strange attractors in fluid turbulence," in *Dynamical System and Turbulence, Lecture Notes in Mathematics,* eds D. A. Rand and L. S. Young (Berlin: Springer Verlag). doi: 10.1007/BFb0091924

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* 93, 485–498. doi: 10.1175/BAMS-D-11-00094.1

Torrence, C., and Webster, P. J. (1998). The annual cycle of persistence in the El Niño-Southern oscillation. *Q. J. R. Meteorol. Soc.* 124, 1985–2004. doi: 10.1256/smsqj.55009

van Sebille, E., Sprintall, J., Schwarzkopf, F. U., Sen Gupta, A., Santoso, A., England, M. H., et al. (2014). Pacific-to-indian ocean connectivity: Tasman leakage, Indonesian Throughflow, and the role of ENSO. *J. Geophys. Res. Oceans* 119, 1365–1382. doi: 10.1002/2013JC009525

Wajsowicz, R. C., and Schneider, E. K. (2001). The Indonesian Throughflow's effect on global climate determined from the cola coupled climate system. *J. Clim.* 14, 3029–3042. doi: 10.1175/1520-0442(2001)014<3029:TITSEO>2.0.CO;2

Webber, C., and Marwan, N. (eds.). (2015). *Recurrence Quantification Analysis. Understanding Complex Systems*. Cham: Springer. doi: 10.1007/978-3-319-07155-8

Webster, P. J., Moore, A. M., Loschnigg, J. P., and Leben, R. R. (1999). Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997-98. *Nature* 401, 356–360. doi: 10.1038/43848

Weller, E., and Cai, W. (2013). Realism of the Indian Ocean Dipole in CMIP5 models: the implications for climate projections. *J. Clim.* 26, 6649–6659. doi: 10.1175/JCLI-D-12-00807.1

Yeh, S.-W., Cai, W., Min, S.-K., McPhaden, M. J., Dommenget, D., Dewitte, B., et al. (2018). ENSO atmospheric teleconnections and their response to greenhouse gas forcing. *Rev. Geophys.* 56, 185–206. doi: 10.1002/2017RG000568

Yuan, D., Wang, J., Xu, T., Xu, P., Hui, Z., Zhao, X., et al. (2011). Forcing of the Indian Ocean Dipole on the interannual variations of the tropical Pacific Ocean: roles of the Indonesian Throughflow. *J. Clim.* 24, 3593–3608. doi: 10.1175/2011JCLI3649.1

Zbilut, J. P., and Webber, C. L. (1992). Embeddings and delays as derived quantification of recurrence plots. *Phys. Lett. A* 171, 199–203. doi: 10.1016/0375-9601(92)90426-M

Zbilut, J. P., Zaldivar-Comenges, J. M., and Strozzi, F. (2002). Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data. *Phys. Lett. A* 297, 173–181. doi: 10.1016/S0375-9601(02)00436-X

Zhou, Q., Duan, W., Mu, M., and Feng, R. (2015). Influence of positive and negative indian ocean dipoles on ENSO via the Indonesian Throughflow: results from sensitivity experiments. *Adv. Atmosphere. Sci.* 32, 783–793. doi: 10.1007/s00376-014-4141-0

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M. (2019). The ecmwf operational ensemble reanalysis analysis system for ocean and sea ice: a description of the system and assessment. *Ocean Sci.* 15, 779–808. doi: 10.5194/os-15-779-2019

**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 Climate

ORIGINAL RESEARCH
published: 21 May 2021
doi: 10.3389/fclim.2021.655919

# Multi-Model Multi-Physics Ensemble: A Futuristic Way to Extended Range Prediction System

*Atul K. Sahai[1]*, Manpreet Kaur[1,2], Susmitha Joseph[1], Avijit Dey[1], R. Phani[1], Raju Mandal[1] and Rajib Chattopadhyay[1,3]*

[1] Indian Institute of Tropical Meteorology, Ministry of Earth Sciences, Pune, India, [2] Department of Atmospheric and Space Sciences, Savitribai Phule Pune University, Pune, India, [3] India Meteorological Department, Ministry of Earth Sciences, Pune, India

In an endeavor to design better forecasting tools for real-time prediction, the present work highlights the strength of the multi-model multi-physics ensemble over its operational predecessor version. The exiting operational extended range prediction system (ERPv1) combines the coupled, and its bias-corrected sea-surface temperature forced atmospheric model running at two resolutions with perturbed initial condition ensemble. This system had accomplished important goals on the sub-seasonal scale skillful forecast; however, the skill of the system is limited only up to 2 weeks. The next version of this ERP system is seamless in resolution and based on a multi-physics multi-model ensemble (MPMME). Similar to the earlier version, this system includes coupled climate forecast system version 2 (CFSv2) and atmospheric global forecast system forced with real-time bias-corrected sea-surface temperature from CFSv2. In the newer version, model integrations are performed six times in a month for real-time prediction, selecting the combination of convective and microphysics parameterization schemes. Additionally, more than 15 years hindcast are also generated for these initial conditions. The preliminary results from this system demonstrate appreciable improvements over its predecessor in predicting the large-scale low variability signal and weekly mean rainfall up to 3 weeks lead. The subdivision-wise skill analysis shows that MPMME performs better, especially in the northwest and central parts of India.

**Keywords: multi-physics, multi-model, extended range prediction, monsoon, ensemble prediction**

OPEN ACCESS

***Edited by:***
*Marcelo Barreiro, Universidad de la República, Uruguay*

***Reviewed by:***
*Laura Ferranti, European Center for Medium-Range Weather Forecasts, United Kingdom*
*Carmen Alvarez-Castro, Universidad Pablo de Olavide, Spain*

****Correspondence:***
*Atul K. Sahai*
*sahai@tropmet.res.in*

***Specialty section:***
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

***Received:*** *19 January 2021*
***Accepted:*** *30 March 2021*
***Published:*** *21 May 2021*

***Citation:***
*Sahai AK, Kaur M, Joseph S, Dey A, Phani R, Mandal R and Chattopadhyay R (2021) Multi-Model Multi-Physics Ensemble: A Futuristic Way to Extended Range Prediction System. Front. Clim. 3:655919. doi: 10.3389/fclim.2021.655919*

## 1. INTRODUCTION

The Indian summer monsoon is an economically prodigious phenomenon accountable for the gross domestic product (GDP) of the world's second-largest populated country (Gadgil and Gadgil, 2006). A voluminous scientific literature unveils the manifold aspects and theories concatenating the events of this significant annual occurrence (Raghavan, 1973; Rao, 1976; Sikka and Gadgil, 1980; Parthasarathy et al., 1992; Webster and Yang, 1992; Goswami et al., 1999; Wang and Fan, 1999; Jiang et al., 2004; Joseph and Sijikumar, 2004; Goswami, 2005; Annamalai, 2010; Rajeevan et al., 2010). Apart from being a decisive economic factor, the monsoon has perpetuated the research in recent decades to undertake the emanating climate changes and accompanied extreme weather conditions (Goswami et al., 2006, 2019; Ajayamohan and Rao, 2008; Guhathakurta and Rajeevan, 2008; Rajeevan et al., 2008; Joseph et al., 2015; Parker et al., 2016; Sooraj et al., 2016; Houze et al., 2017; Roxy et al., 2017). In the above view, the prediction of monsoon is not only exigent but is

highly inevitable. The significant rainfall contribution from intra-seasonal scale variability in the monsoon highlights the stature of sub-seasonal to seasonal (S2S) scale prediction (Abhilash et al., 2013, 2014b,c; Vitart and Robertson, 2018; Robertson et al., 2019).

The deterministic prediction on the S2S scale has limitations, and therefore probabilistic methods or ensemble prediction systems are considered (Molteni et al., 1996; Buizza et al., 2007, 2008; Vitart and Molteni, 2009; Rashid et al., 2011). The prediction from an ensemble of perturbed initial conditions (ICs) is one of the popular techniques. An idea initiated from the extra-tropical cyclogenesis problem (Bjerknes and Solberg, 1922) and an instigating theory of baroclinic instability (Charney, 1947; Eady, 1949), the perturbations in the atmospheric flows became a central solution to the initial value problem of numerical weather prediction (O'Malley, 1988). Later, it was polished into a well-versed technique to generate the ensemble of ICs to enhance prediction skill across various weather scales (Toth and Kalnay, 1993, 1997; Buizza and Palmer, 1995, 1998). The atmospheric lagged average is another traditional ensemble generation method (Hoffman and Kalnay, 1983; Kalnay and Dalcher, 1987; Chen et al., 2013) where the forecast from different initialization for the same target period is amalgamated into ensemble mean. These two techniques are famously known to address the uncertainties sourced from ICs.

Some of the recent literature incline toward grand ensemble based on multiple models (Krishnamurti et al., 2000; Sahai et al., 2013; Abhilash et al., 2015; Kalnay, 2019). The advantages in one or more aspects of one model formulation over the other could provide better assistant in the multi-model approach. The concept of inter-model diversity arises from the need to address another class of errors recognized as model-errors. Although there is a varying perspective on the nature and origin of these errors, they are largely attributed to the representation of physical processes in the model. The approximations considered while formulating parameterization schemes and misrepresentation of significant sub-grid scale phenomena in the model could cause biases in the predicted fields. Further, it is proposed that the multi-physics ensemble scheme can be an alternative to account for these model-errors (Richardson, 1997; Harrison et al., 1999; Orrell et al., 2001). The intra-model diversification introduced by using more than one physical parameterization showed significant improvement over single physics predictions (Stensrud and Fritsch, 1994; Berner et al., 2011; Tapiador et al., 2012; Greybush et al., 2017; Xu et al., 2020).

The above-mentioned ensemble prediction techniques have advantages as well as limitations when it comes to real-time prediction. For example, the perturbed initial conditions based ensemble could palliate the growth of initial errors, but such ensemble tends to be under dispersive, leading to presumptuous probabilistic prediction and underestimated larger weather anomalies (Stensrud et al., 2000). Similarly, lagged ensembles with improper weights from older initializations can debase the mean forecast (Abhilash et al., 2014b). Further, the model-error ensemble techniques require physical consistency among the members in terms of errors but are known to increase the ensemble spread (Green et al., 2017). Therefore, careful examination of these techniques is required to achieve the desired improvement.

The efficacy of any prediction tool is determined by its validity and reliability measured as the forecast skill (Murphy, 1991; Casati et al., 2008). Many skill assessment and verification methods are available to evaluate and compare various prediction strategies (Ghelli and Ebert, 2008; Jolliffe and Stephenson, 2011; Ebert et al., 2013). These methods increase the confidence in any prediction approach and motivate to understand and improve the limitations in the hypothesis formulation.

The skill analysis is vital, especially for the complex monsoon systems giving a significant annual rain share. In the present study, we evaluate the skill of a multi-model multi-physics ensemble prediction strategy for the Indian summer monsoon. This strategy is a part of developing a new extended range ensemble forecasting framework and here we will compare it to its current operational version. The functional version is only a multi-model ensemble prediction system (Sahai et al., 2013, 2016; Abhilash et al., 2014c, 2015) developed under the "National Monsoon Mission(NMM)" project (Rao et al., 2020) and has received acclaim on its successful implementation in 2016. This operational version is being used for extended range prediction (ERP) at the India Meteorological Department (IMD) and provides outlooks for rainfall, heatwaves, cyclones, and other meteorological parameters for various sectoral applications (Pattanaik et al., 2019). The next ERP version under development uses a multi-physics approach along with the multi-model framework. The results presented here are from the preliminary runs of this new version generated from unperturbed ICs. The comparison with the older version highlights its usefulness and drawbacks. This documentation will be handy for further improvements and modifications in the new framework.

The next section elaborates more on both the prediction systems as well as methodologies and datasets utilized in the study. The skill of ERP systems is discussed in the subsequent section, followed by conclusions.

## 2. DATA AND METHODOLOGY

The operational ERP system at IMD is a multi-model ensemble framework (Pattanaik et al., 2019). It comprises of two horizontal resolution variants (with 382 and 126 truncations) of two models; climate forecast system version 2 (CFSv2) and atmospheric global forecast system (GFS) from National Centers for Environmental Prediction (NCEP) (Saha et al., 2014). Further, this total of four variants run with a four-member ensemble of perturbed atmospheric ICs. These atmospheric ICs are obtained from National Center for Medium Range Weather Forecasting and oceanic ICs from Indian National Center for Ocean Information Services for CFSv2. Additionally, the real-time sea-surface temperature (SST) from CFSv2 after bias-correction is used as forcing to GFS (detailed technique can be seen in Abhilash et al., 2014a, Mandal et al., 2019, and Kaur et al., 2020). This ERP system was developed and thoroughly tested for skill at Indian Institute of Tropical Meteorology (IITM) under NMM. The operational forecasts are generated every week with Wednesday

ICs for the next 32 days, also on the fly hindcast for 2003 to 2015 is produced for each IC. This system is henceforth addressed as ERPv1 in the paper.

The successor version of the above-mentioned prediction system is in the final development stage. This new ERP system also has two model variants CFSv2 and GFS, but the two resolution variants are now replaced with one seamless mode where the horizontal resolution of T574 transitions into the coarser T382 resolution after 15 days. Additionally, a multi-physics strategy is adopted for generating ensemble. We have used three convective parameterization permutations with two micro-physics parameterizations. These convection schemes include Simplified-Arakawa Shubert (SAS) (Pan and Wu, 1995), revised deep-convection SAS (NSAS) (Han and Pan, 2011), and revised SAS with modified shallow-convection (NSAS_SC) (Han and Pan, 2011). Zhao and Carr (ZC) (Zhao and Carr, 1997) and Ferrier (FER) (Ferrier et al., 2002) are the two micro-physics schemes incorporated in the new formulation. The resultant six physics combinations are SASZC, SASFER, NSASZC, NSASFER, NSASZC_SC, and NSASFER_SC. CFSv2 runs with all six combinations, whereas GFS has only four and does not include SASZC and SASFER. Similar to ERPv1, GFS is forced with bias-corrected CFSv2 real-time SST. The NCEP climate forecast system reanalysis ICs are utilized for both CFSv2 and GFS. The new multi-physics multi-model prediction contains 36 days forecast initialized on 1st, 6th, 11th, 16th, 21st, and 26th of each month for hindcast period 2001–2015. We are going to label this physics-based multi-model ensemble as MPMME hereafter.

The anomaly correlation coefficient (ACC), Pearson correlation, Heidke skill score (HSS) (Barnston, 1992), root mean square error (RMSE), root mean square skill score (RMSS), and Brier skill score (BSS) (Brier, 1950) are the verification matrices used to analogize MPMME skill with ERPv1. The verification is done for the weekly mean rainfall forecast at 4-week leads. The week 1 lead corresponds to the initial 7 days forecast, subsequent 8–14 days constitute week 2; similarly, 15–21 and 22–28 days forecast defines week 3 and 4, respectively, using common hindcast 2003–2015 from both the versions of ERP system. The sample size considered for ERPv1 is 22 weeks × 13 years = 286 forecasts for each lead. Similarly, MPMME has a sample of 24 weeks × 13 years = 312 forecasts. The skill scores are computed against observed daily rainfall from Tropical Rain Measure Mission (TRMM) merged rainfall provided by IMD (Mitra et al., 2009; Pai et al., 2014). The Monsoon Intraseasonal Oscillation (MISO) indices are computed following Sahai et al. (2013) and Suhas et al. (2013). It is specified that the MPMME includes only control runs (i.e., six members from CFSv2 and four members from GFS). Therefore, We have selected only 10 members from ERPv1 (three from each variant of CFS model and two from each GFS variant) for a fair comparison with 10-member MPMME.

## 3. RESULTS

The hindcast from both ERPv1 and MPMME is analyzed for skill in predicting Indian summer monsoon weekly mean rainfall (ISMR) from June to September.

**TABLE 1 |** Anomaly correlation coefficient (ACC) of predicted weekly mean rainfall over the monsoon zone.

| | Lead | ERPv1 | MPMME |
|---|---|---|---|
| ACC | W1 | 0.775 | 0.856 |
| | W2 | 0.627 | 0.706 |
| | W3 | 0.378 | 0.480 |
| | W4 | 0.248 | 0.173 |

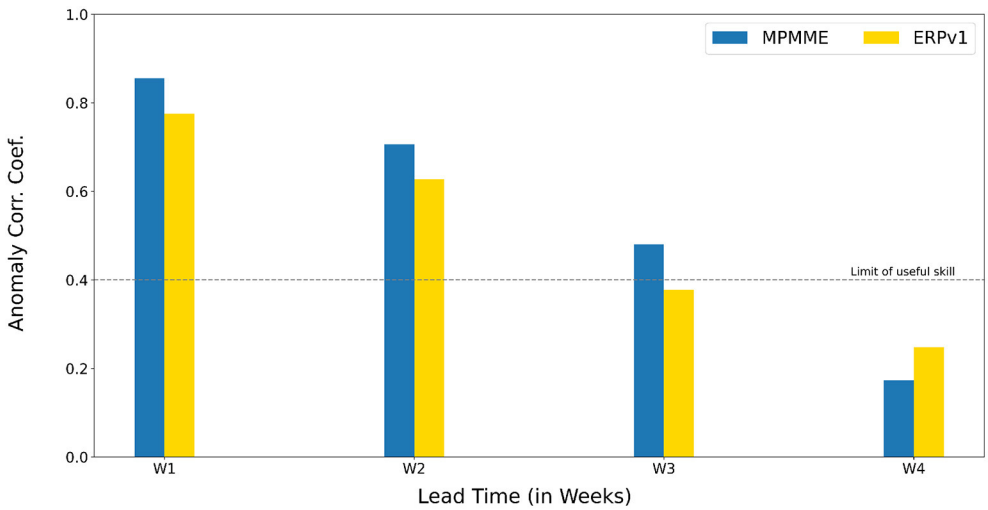



**FIGURE 1 |** Anomaly correlation coefficient for weekly mean rainfall (mm/day) over Monsoon Zone of India during summer monsoon.

### 3.1. Monsoon Zone Weekly Rainfall Prediction Skill

ISMR exhibit large spatiotemporal heterogeneity during the season with positive (negative) rainfall anomalies over central India representing active (break) monsoon. A box defined in Rajeevan et al. (2010) over central India is the widely used prototypical monsoon region known as the core monsoon zone. **Figure 1** and **Table 1** illustrate the skill of predicted weekly mean rainfall averaged over the monsoon zone at 4-week leads. The ERPv1 has 0.78 and 0.63 ACC in week 1 and 2 lead, respectively. It is improved by almost a factor of 0.1 in MPMME for both the weeks. Although the skill is dropped in the third, it is still above 0.4, the practical skill limit. In the 4th week, the skill further declined. The difference between the deterministic prediction skill of both systems over the monsoon zone is statistically significant at 99.9, 95%, and above 90% confidence level for week 1, 2, and 3, respectively, and the difference in skill is not significant in the 4th week.

Apart from spatial non-uniformity, the monsoonal rainfall has well-documented temporal variability that arises from intra-seasonal fluctuations. These fluctuations are recognized as spells of increased and minimum to no rain conditions over the monsoon zone. The transitions between these two spells are challenging but crucial, and models would have difficulties predicting such transitions, limiting the predictability of monthly rainfall. **Figure 2** compares the monthly skill of weekly averaged rainfall over the monsoon zone for both systems. June and September have higher skill than July and August in both systems, which could be attributed to model inefficiency to predict frequent synoptic-scale systems in later months. However, the coefficient values are >0.6 for both systems in the first 2 weeks, which are reduced in following leads.

Regarding improvement, the month of June (**Figure 2A**) record the highest increase in the skill where at all 4 leads, MPMME shows 10, 18, 32, and 12% improvement over ERPv1, which is significant at 95% confidence level. The significant phenomena during June, such as monsoon onset and cyclonic system genesis impact the subsequent progress of the monsoon. These events are important to be predicted especially for dam management for releasing and storage planning of water, for agro-met services to begin sowing, and for disaster mitigation due to extreme rainfall activities. Hence, improvement in prediction skill of June will be highly beneficial for real-time ERP of monsoon onset and extreme rainfall conditions. Further, July, August, and September witness an increase in ACC up to week 3 lead (except week 2 lead during August) for MPMME. Relatively less skill is seen for the 4th week for these months than ERPv1, but the difference is insignificant as ERPv1 skill is also <0.4.

HSS gives fractional betterment of the forecast over a reference forecast, which is climatology in our case. HSS for deterministic forecast verification of various thresholds for weekly mean rainfall over monsoon zone is plotted in **Figure 3**. The skill decreases for higher rainfall thresholds at all weekly leads, indicating both versions' limitation in predicting heavy rainfall. However, the MPMME could perform better than reference forecast minimum up to 3 weeks leads for the given thresholds. The figure affirms the improvement in MPMME performance over ERPv1.

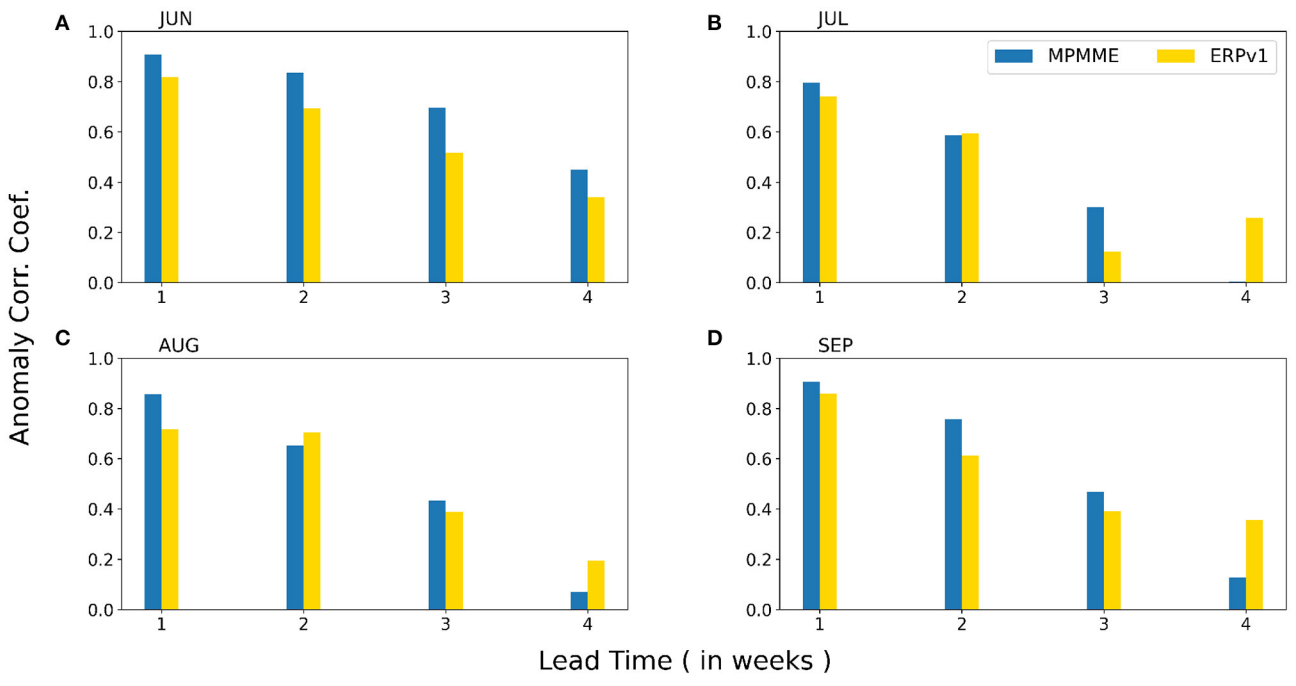



**FIGURE 2 |** Anomaly correlation coefficient for weekly mean rainfall over monsoon zone of India for the month of **(A)** June, **(B)** July, **(C)** August, and **(D)** September at four week leads.

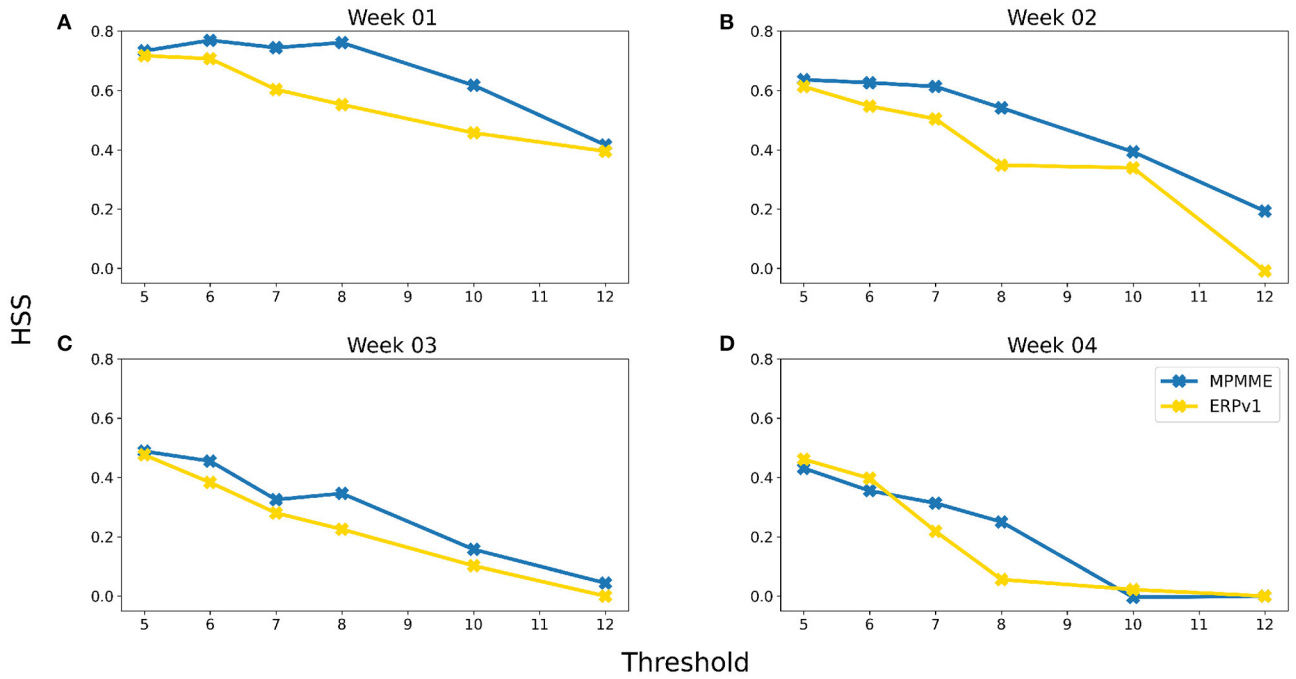



**FIGURE 3 |** Heidke Skill score at different thresholds of weekly mean rainfall over Monsoon Zone of India for **(A)** week 1, **(B)** week 2, **(C)** week 3, and **(D)** week 4 lead.

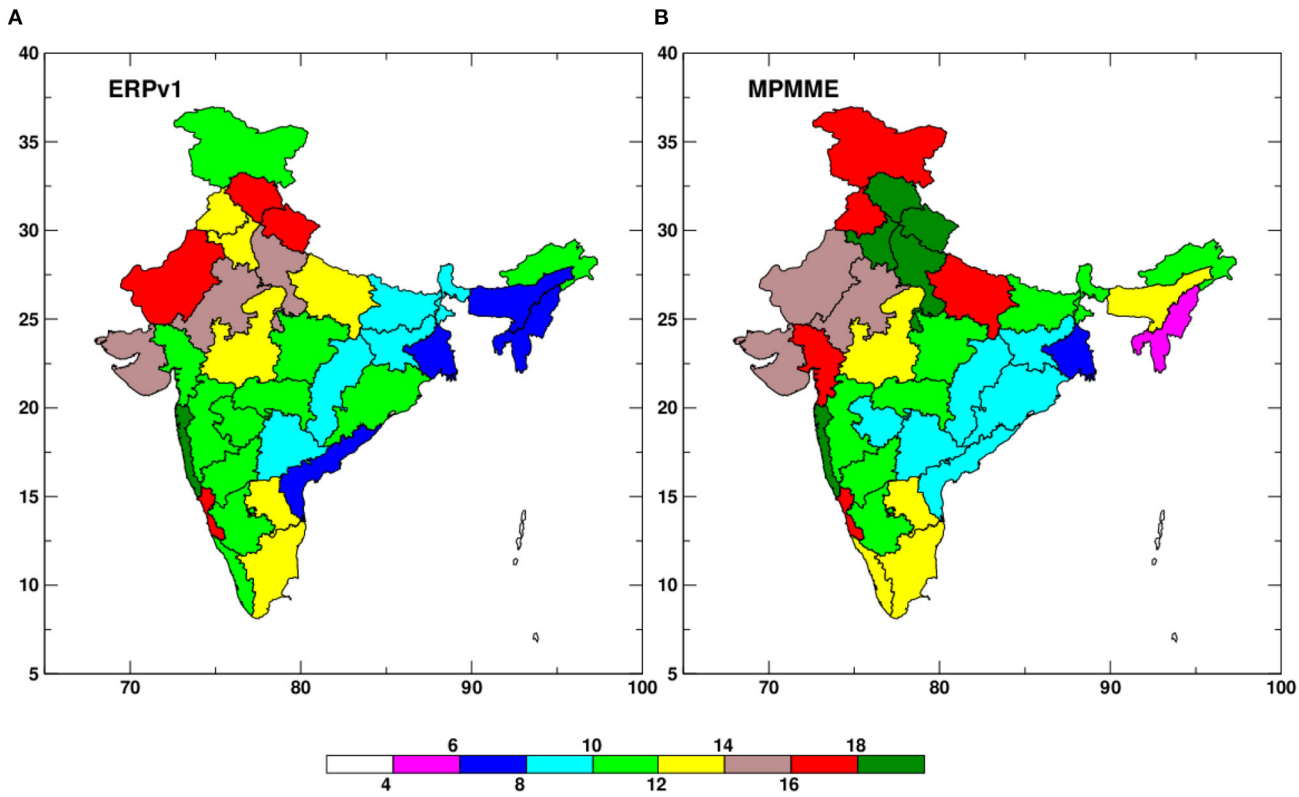



**FIGURE 4 |** Subdivision-wise predictability limit for **(A)** extended range prediction system (ERPv1) and **(B)** multi-physics multi-model ensemble (MPMME) during monsoon.

## 3.2. Skill Assessment Over Meteorological Subdivisions

The ERP broadens the application spectrum of the meteorological forecast. These different sector-specific applications stipulate meteorological information at the finer spatial scale. However, the generation of stakeholder requisite forecast products requires skill assessment at the smaller spatial scale because the area-averaged precipitation skill will not be sufficient. Therefore, we will look into the skill for the meteorological subdivisions of India. There are 36 such subdivisions across the country; for further details, please refer Joseph et al. (2019).

The standard signal to noise ratio (SNR) is considered here to find out the limit of rainfall predictability (lead at which SNR becomes one) for these subdivisions. **Figure 4** shows the spatial map of predictability with color indicating the number of predictable days. In both ERPv1 (**Figure 4A**) and MPMME (**Figure 4B**), maximum subdivisions show predictability of 10–14 days. A very few numbers of subdivisions, i.e., 5, have predictability >16 days. The number of such subdivisions with predictability higher than 16 days is almost doubled in MPMME. Simultaneously, the number is reduced by two for subdivisions with <8 predictable days in MPMME compared to ERPv1. In total, more than 12 subdivision show improvement in predictability by 2–4 days in MPMME, these subdivisions fall into north and northwest India. A similar increment is also seen for a few subdivisions in southern peninsular and northeast India. In contrast, for many subdivisions in central India, the predictability remains unchanged in MPMME, except a very few subdivision (i.e., 4) where predictability dropped by 1–2 days.

The week-wise anomaly correlations for subdivisions are shown in **Figure 5**, where ACC>0.2 is statistically significant at a level of 99.9%. Both ERPv1 and MPMME have good skill in the week 1 and 2 forecast, with MPMME outperforming ERPv1 for maximum subdivisions. The lead-in prediction skill is maintained in week 3 by MPMME, where many subdivisions have ACC >0.2 and 0.3 in contrast to ERPv1. Week 4 is less skillful than the first 3 weeks in both ERPv1 and MPMME, where most of the subdivisions shows ACC smaller than 0.2. **Figure 6** illustrates the RMSS values from ERPv1 and MPMME at 4 leads for meteorological subdivisions; the shaded values (i.e., >0) indicates reasonable prediction skill. Furthermore, similar to ACC, RMSS is also better in MPMME than ERPv1 for up to 3 weeks. Subdivisions in northeast India are less skillful in both ERPv1 and MPMME; the previous authors have linked lower predictability to more rainfall contribution from less predictable synoptic systems over the northeast regions (Abhilash et al., 2018; Joseph et al., 2019).

Overall, MPMME show reasonable improvement in deterministic skill over northwest and central India compared to ERPv1. The studies have reported the more frequent occurrence of extreme rainfall over these regions (Singh et al., 2011; Woo et al., 2019; Joseph et al., under revision; Rai et al., 2020). Therefore, improved predictability and prediction skill for these regions in MPMME can effectively improve the extreme event prediction (will be addressed in a separate study).

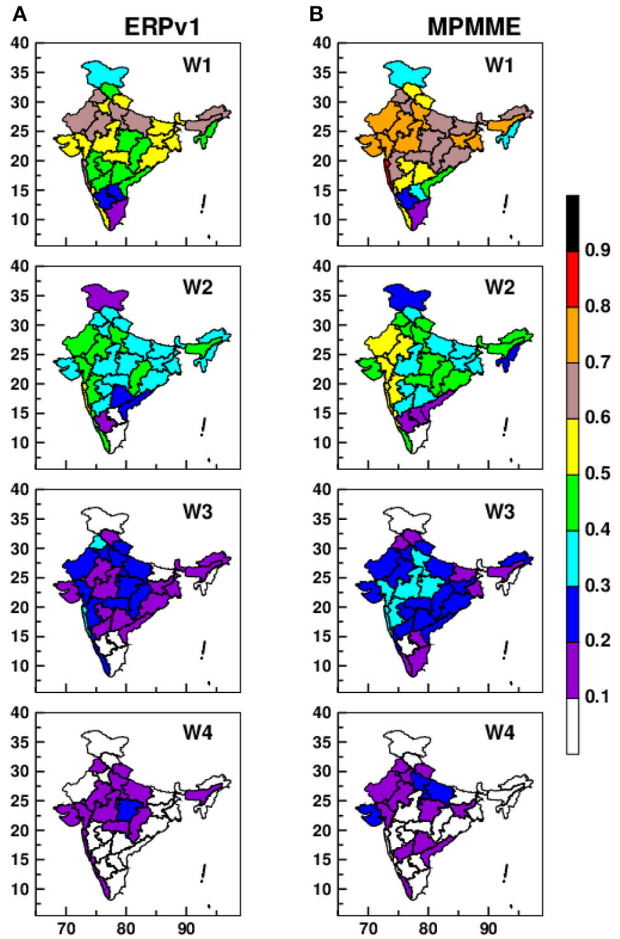



**FIGURE 5 |** Sub-divisional skill in terms of anomaly correlation coefficient for **(A)** extended range prediction system (ERPv1) and **(B)** multi-physics multi-model ensemble (MPMME) during monsoon.

## 3.3. Prediction Skill for Monsoon Intraseasonal Oscillation

MISO is one of the most dominant mode of low-frequency intraseasonal variability, known to provide predictability in the extended range during the Indian summer monsoon. The enhanced skill witnessed in earlier sections could be explained by analyzing the model's ability to capture this large-scale signal.

The MISO prediction skill is computed in terms of bivariate anomaly correlation coefficient (BVCC) and root mean square error as mentioned in Rashid et al. (2011) of predicted MISO Indices from all ICs with the observed. The leading pair of model predicted MISO Indices from all ICs with the observed counterpart are utilized for ACC and RMSE computation. The BVCC is plotted in **Figure 7** along with the RMSE for MISO indices as a function of lead days. We consider BVCC >0.5 and RMSE lower than 1.4 as a threshold for skillful MISO prediction. The horizontal line intersects the BVCC axis at 0.5 and RMSE axis at 1.41 to track the significant skill and error limit. The black line represents the combined skill for

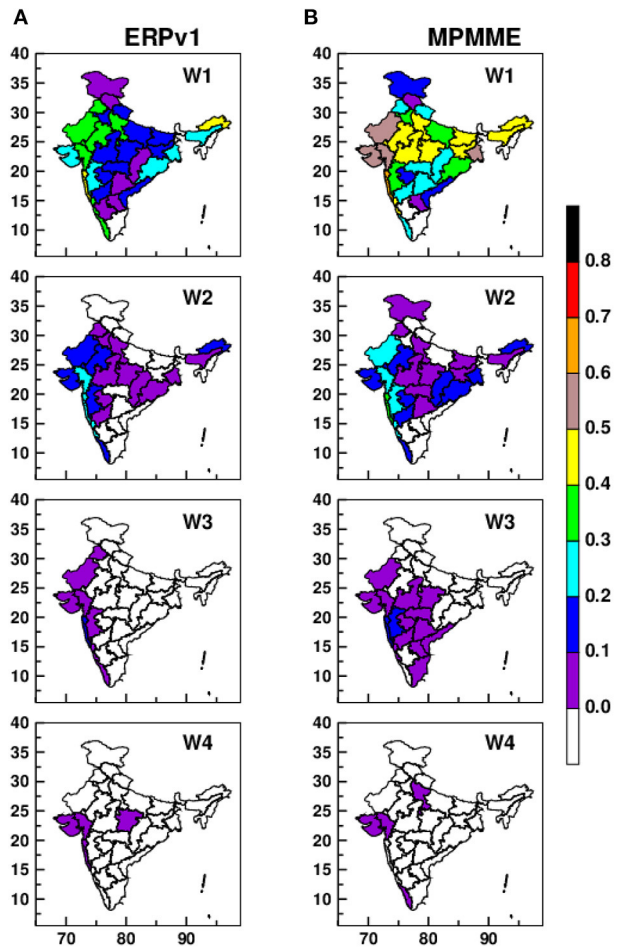



**FIGURE 6 |** Sub-divisional skill in terms of root mean square score for **(A)** extended range prediction system (ERPv1) and **(B)** multi-physics multi-model ensemble (MPMME) during monsoon.

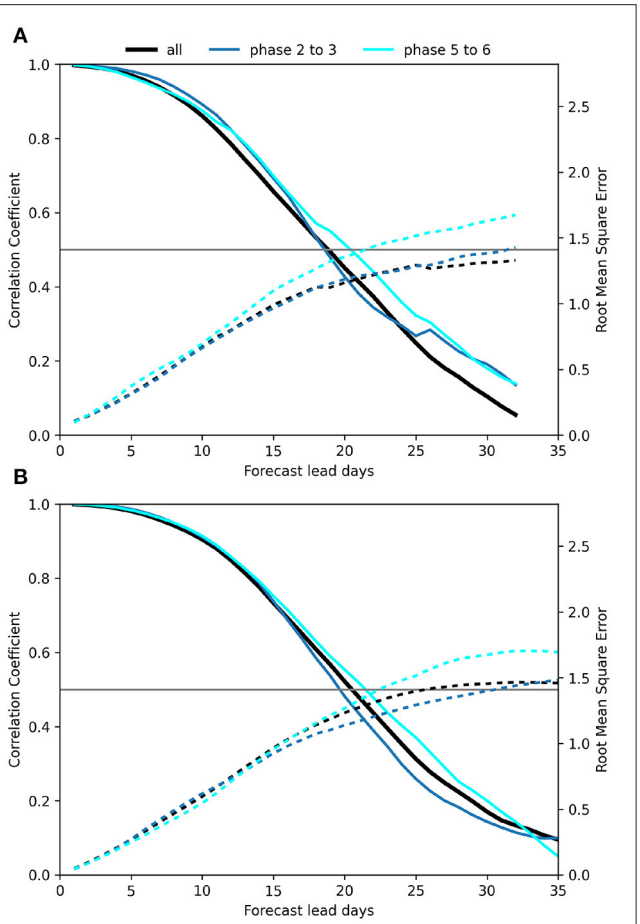



**FIGURE 7 |** Monsoon Intraseasonal Oscillation (MISO) prediction skill for **(A)** extended range prediction system (ERPv1) and **(B)** multi-physics multi-model ensemble (MPMME), where solid lines are for bivariate anomaly correlation coefficient (BVCC) and dotted line represent root mean square error (RMSE).

all MISO phases, whereas blue and cyan show the evolution of CC and RMSE for the transition to the active and break phase, respectively. The figure clearly shows a gain in skill for MPMME over ERPv1. The ERPv1 reaches the prediction limit in around 19 days, whereas the MPMME has this limit beyond 21 days.

The study from Goswami and Xavier (2003) reveals that the potential predictability for break (less to no rain) conditions during monsoon is high compared to the active. They also suggested that higher predictability of transition to break phase is due to governance of error growth in this phase by low frequency (30–60 days) signal. Abhilash et al. (2014b) also showed that ERP (from CFSv2-based 11-member ensemble) of breaks are more skillful. Similar inference can be made from **Figure 7** for phase-dependent prediction skill of both systems; ERPv1 and MPMME also show slightly better predictability for break transition (Cyan line in **Figures 7A,B**) than active (blue). MPMME have improved skill for both (active and break) phases in comparison to ERPv1.

The higher predictability in northwest regions of India is associated with low-frequency monsoon oscillations (Joseph et al., 2019), which is evident in **Figure 5**. Therefore, the 2–4 days increase in predictability in MPMME over these regions can be attributed to about 2 days gain in the skill of this low variability signal skill, i.e., MISO.

## 3.4. Probabilistic Forecast Skill

In the previous sections, we have evaluated the deterministic prediction skill from ERPv1 and MPMME; in this section, we look into some probabilistic verification. The BSS is calculated for categorical rainfall probabilistic prediction (**Figure 8**). Based on the tercile method, three categories are defined as above normal (when the rainfall amount is more than upper tercile value), near normal (when it is between upper tercile and lower tercile values), and below normal (when it is below the lower tercile value). The probabilities for either category are 100% for the observation.

The BSS compares the brier score of the forecast with the reference forecast (climatology), assuming a 33% equal occurrence probability for each category. A BSS value >0

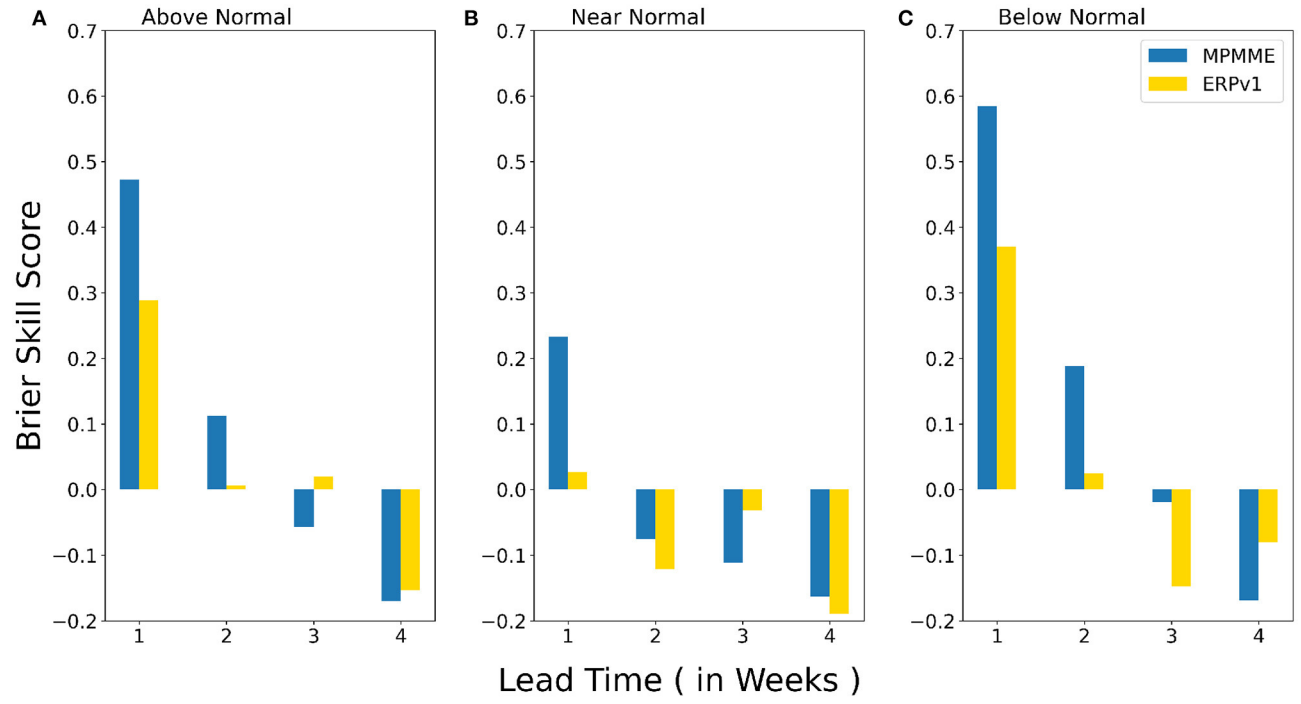



**FIGURE 8 |** Brier skill score from extended range prediction system (ERPv1) and multi-physics multi-model ensemble (MPMME) for **(A)** above normal, **(B)** near normal, and **(C)** below normal categorical rainfall forecast over monsoon zone.

indicates an improvement over climatology. Both ERPv1 and MPMME have better skill in predicting the above normal (**Figure 8A**) and below normal (**Figure 8C**) categories up to 2 weeks. The near-normal (**Figure 8B**) rainfall predictions are comparatively less skillful in both systems. It is interesting to note from the figure that the considerable improvement is there for MPMME over ERPv1 in almost all categories in the first 2-week leads. The skill of both system reduces at longer leads.

The analysis presented in this section favors MPMME until almost 21-day lead. However, the current results are only from multi-physics ensemble, i.e., addressing model-errors to some extent. Nevertheless, the model also suffers from initial condition errors at longer leads. A few earlier studies have also concluded that physics ensemble along with perturbed initial conditions ensemble could provide better skill by addressing two major bias components (Stensrud and Fritsch, 1994; Stensrud et al., 2000). Therefore, the results of this study can be ameliorated further with the careful selection of physics and initial condition ensemble.

## 4. CONCLUSIONS

The present work highlights the improvements of a physics-based multi-model extended range ensemble prediction system over its predecessor operational version in predicting ISMR. This new MPMME framework distinguishes itself from the ERPv1 in its single seamless horizontal resolution, and most importantly, for considering multiple realizations of atmospheric dynamics achieved by permutations of convection and microphysics parameterizations.

The skill of MPMME and ERPv1 for ISMR is compared using different verification scores for spatiotemporal forecast evolution at weekly leads. ACC for hindcast of 2003–2015 from MPMME signifies an improvement for monsoon zone rainfall over ERPv1 up to 3-week lead. The MPMME prediction skill for 4 months from June to September also witnesses increment up to 3-week lead. The month of June has the highest skill for MPMME at all 4-week leads, which will come in handy for predicting monsoon onset and extreme rain-producing systems in the onset phase. The HSS for rainfall over monsoon zone elucidates the enhanced skill at all thresholds of the weekly mean rainfall for MPMME over ERPv1.

The MPMME extends the predictability limit by 2–4 days compared to ERPv1 as indicated by sub-division map. Similarly, the conclusion drawn from subdivision-wise ACC and RMSS favors MPMME in most regions through 7, 14, and 21-day leads. The subdivisions in the northwest and central parts of the country exhibit a maximum increase in the skill. All the phases of large-scale MISOs have a better prediction from MPMME than ERPv1, reflecting in the overall gain in predictability at the sub-division level. The tercile-based categorical rainfall prediction is verified for the probabilistic skill of both systems. BSS for these categorical rainfall occurrences exhibits the superiority of physics-based MPMME over ERPv1.

Although the results presented here are from the preliminary development stage, the different verifications used in the study

support the MPMME over its operational version ERPv1 up to 21 days lead. Further assessment of signal and noise added per physics combination could assist in considering the weighted average of these ensembles to generate the forecast products. Since we have only considered the unperturbed control initial condition in MPMME, adding a few perturbed IC members could further help improve the prediction, especially at longer leads, by controlling the growth of initial condition uncertainty (Stensrud and Fritsch, 1994; Stensrud et al., 2000). The study proffers the utility of physics-based ensemble and finds its scope in further exploration. It is anticipated that the enhanced temporal skill for June and spatial skill for northwest and central regions of India could probably improve the extreme event prediction.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

AS, SJ, RP, and RC have conceptualized the MPMME strategy. RP, RM, AD, and MK were involved in the model runs, data-handling, and processing. MK has done the analysis, plotting, and structured the manuscript draft. All authors have contributed in writing and editing the final manuscript.

## ACKNOWLEDGMENTS

Research at IITM was fully supported by the Ministry of Earth Sciences and authors sincerely acknowledge it. The analysis and model integrations are performed on Aditya and Pratyush HPCS. We were thankful to Dr. D. R. Pattanaik for ERPv1 runs. The present work was part of Ph.D. thesis of MK. We express our gratitude to two reviewers for their insightful review, which contributed significantly to our manuscript's quality.

## REFERENCES

Abhilash, S., Mandal, R., Dey, A., Phani, R., Joseph, S., Chattopadhyay, R., et al. (2018). Role of enhanced synoptic activity and its interaction with intra-seasonal oscillations on the lower extended range prediction skill during 2015 monsoon season. *Clim. Dyn*. 51, 3435–3446. doi: 10.1007/s00382-018-4089-3

Abhilash, S., Sahai, A. K., Borah, N., Chattopadhyay, R., Joseph, S., Sharmila, S., et al. (2014a). Does bias correction in the forecasted SST improve the extended range prediction skill of active-break spells of Indian summer monsoon rainfall? *Atmos. Sci. Lett*. 15, 114–119. doi: 10.1002/asl2.477

Abhilash, S., Sahai, A. K., Borah, N., Chattopadhyay, R., Joseph, S., Sharmila, S., et al. (2014b). Prediction and monitoring of monsoon intraseasonal oscillations over Indian monsoon region in an ensemble prediction system using CFSv2. *Clim. Dyn*. 42, 2801–2815. doi: 10.1007/s00382-013-2045-9

Abhilash, S., Sahai, A. K., Borah, N., Joseph, S., Chattopadhyay, R., Sharmila, S., et al. (2015). Improved spread–error relationship and probabilistic prediction from the CFS-based grand ensemble prediction system. *J. Appl. Meteorol. Climatol*. 54, 1569–1578. doi: 10.1175/JAMC-D-14-0200.1

Abhilash, S., Sahai, A. K., Pattnaik, S., and De, S. (2013). Predictability during active break phases of Indian summer monsoon in an ensemble prediction system using climate forecast system. *J. Atmos. Solar Terres. Phys*. 100–101, 13–23. doi: 10.1016/j.jastp.2013.03.017

Abhilash, S., Sahai, A. K., Pattnaik, S., Goswami, B. N., and Kumar, A. (2014c). Extended range prediction of active-break spells of Indian summer monsoon rainfall using an ensemble prediction system in NCEP climate forecast system. *Int. J. Climatol*. 34, 98–113. doi: 10.1002/joc.3668

Ajayamohan, R. S., and Rao, S. A. (2008). Indian Ocean dipole modulates the number of extreme rainfall events over India in a warming environment. *J. Meteorol. Soc. Jpn II* 86, 245–252. doi: 10.2151/jmsj.86.245

Annamalai, H. (2010). Moist dynamical linkage between the equatorial Indian Ocean and the South Asian monsoon trough. *J. Atmos. Sci*. 67, 589–610. doi: 10.1175/2009JAS2991.1

Barnston, A. G. (1992). Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke Score. *Weather Forecast*. 7, 699–709. doi: 10.1175/1520-0434(1992)007<0699:CATCRA>2.0.CO;2

Berner, J., Ha, S. Y., Hacker, J. P., Fournier, A., and Snyder, C. (2011). Model uncertainty in a mesoscale ensemble prediction system: stochastic versus multiphysics representations. *Mon. Weather Rev*. 139, 1972–1995. doi: 10.1175/2010MWR3595.1

Bjerknes, J., and Solberg, H. (1922). *Life Cycle of Cyclones and the Polar Front Theory of Atmospheric Circulation*. Geofysiske Publikationer. Cammermeyers Bogh.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Mon. Weather Rev*. 78, 1–3. doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Buizza, R., Bidlot, J. R., Wedi, N., Fuentes, M., Hamrud, M., Holt, G., et al. (2007). The new ECMWF VAREPS (variable resolution ensemble prediction system). *Q. J. R. Meteorol. Soc*. 133, 681–695. doi: 10.1002/qj.75

Buizza, R., Leutbecher, M., and Isaksen, L. (2008). Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc*. 134, 2051–2066. doi: 10.1002/qj.346

Buizza, R., and Palmer, T. N. (1995). The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci*. 52, 1434–1456. doi: 10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2

Buizza, R., and Palmer, T. N. (1998). Impact of ensemble size on ensemble prediction. *Mon. Weather Rev*. 126, 2503–2518. doi: 10.1175/1520-0493(1998)126<2503:IOESOE>2.0.CO;2

Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., et al. (2008). Forecast verification: current status and future directions. *Meteorol. Appl*. 15, 3–18. doi: 10.1002/met.52

Charney, J. G. (1947). The dynamics of long waves in a baroclinic westerly current. *J. Atmos. Sci*. 4, 136–162. doi: 10.1175/1520-0469(1947)004<0136:TDOLWI>2.0.CO;2

Chen, M., Wang, W., and Kumar, A. (2013). Lagged ensembles, forecast configuration, and seasonal predictions. *Mon. Weather Rev*. 141, 3477–3497. doi: 10.1175/MWR-D-12-00184.1

Eady, E. T. (1949). Long waves and cyclone waves. *Tellus* 1, 33–52. doi: 10.3402/tellusa.v1i3.8507

Ebert, E., Wilson, L., Weigel, A., Mittermaier, M., Nurmi, P., Gill, P., et al. (2013). Progress and challenges in forecast verification. *Meteorol. Appl*. 20, 130–139. doi: 10.1002/met.1392

Ferrier, B. S., Jin, Y., Lin, Y., Black, T., Rogers, E., and DiMego, G. (2002). "Implementation of a new grid-scale cloud and precipitation scheme in the NCEP Eta model," in *Conference on Weather Analysis and Forecasting*, Vol. 19 (AMS), 280–283.

Gadgil, S., and Gadgil, S. (2006). The Indian monsoon, GDP and agriculture. *Econ. Polit. Weekly* 41, 4887–4895.

Ghelli, A., and Ebert, E. (2008). Special issue on forecast verification. *Meteorol. Appl*. 15:1. doi: 10.1002/met.69

Goswami, B., Venugopal, V., Sengupta, D., Madhusoodanan, M., and Xavier, P. (2006). Earth by comets and meteorites. Further studies of these objects may elucidate whether their composition and membrane-like structures were important building blocks for the origin of life. *Science* 314, 1442–1445. doi: 10.1126/science.1132027

Goswami, B. N. (2005). "South Asian monsoon," in *Intraseasonal Variability in the Atmosphere-Ocean Climate System*, eds W. K. M. Lau and D. E. Waliser (Berlin; Heidelberg: Springer Berlin Heidelberg), 19–61. doi: 10.1007/3-540-27250-X_2

Goswami, B. N., Krishnamurthy, V., and Annmalai, H. (1999). A broad-scale circulation index for the interannual variability of the Indian summer monsoon. *Q. J. R. Meteorol. Soc.* 125, 611–633. doi: 10.1002/qj.49712555412

Goswami, B. N., Venugopal, V., and Chattopadhyay, R. (2019). "Chapter 2–South Asian monsoon extremes," in *Tropical Extremes:Natural Variability and Trends*, eds V. Venugopal, J. Sukhatme, R. Murtugudde, and R. B. T. T. E. Roca (Elsevier), 15–49. doi: 10.1016/B978-0-12-809248-4.00002-9

Goswami, B. N., and Xavier, P. K. (2003). Potential predictability and extended range prediction of Indian summer monsoon breaks. *Geophys. Res. Lett.* 30, 1–4. doi: 10.1029/2003GL017810

Green, B. W., Sun, S., Bleck, R., Benjamin, S. G., and Grell, G. A. (2017). Evaluation of MJO predictive skill in multiphysics and multimodel global ensembles. *Mon. Weather Rev.* 145, 2555–2574. doi: 10.1175/MWR-D-16-0419.1

Greybush, S. J., Saslo, S., and Grumm, R. (2017). Assessing the ensemble predictability of precipitation forecasts for the January 2015 and 2016 East Coast winter storms. *Weather Forecast.* 32, 1057–1078. doi: 10.1175/WAF-D-16-0153.1

Guhathakurta, P., and Rajeevan, M. (2008). Trends in the rainfall pattern over India. *Int. J. Climatol.* 28, 1453–1469. doi: 10.1002/joc.1640

Han, J., and Pan, H. L. (2011). Revision of convection and vertical diffusion schemes in the NCEP global forecast system. *Weather Forecast.* 26, 520–533. doi: 10.1175/WAF-D-10-05038.1

Harrison, M. S. J., Palmer, T. N., Richardson, D. S., and Buizza, R. (1999). Analysis and model dependencies in medium-range ensembles: two transplant case-studies. *Q. J. R. Meteorol. Soc.* 125, 2487–2515. doi: 10.1002/qj.49712555908

Hoffman, R. N., and Kalnay, E. (1983). Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A* 35, 100–118. doi: 10.3402/tellusa.v35i2.11425

Houze, R. A., McMurdie, L. A., Rasmussen, K. L., Kumar, A., and Chaplin, M. M. (2017). Multiscale aspects of the storm producing the June 2013 flooding in Uttarakhand, India. *Mon. Weather Rev.* 145, 4447–4466. doi: 10.1175/MWR-D-17-0004.1

Jiang, X., Li, T., and Wang, B. (2004). Structures and mechanisms of the northward propagating boreal summer intraseasonal oscillation. *J. Clim.* 17, 1022–1039. doi: 10.1175/1520-0442(2004)017<1022:SAMOTN>2.0.CO;2

Jolliffe, I. T., and Stephenson, D. B. (2011). *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons.

Joseph, P. V., and Sijikumar, S. (2004). Intraseasonal variability of the low-level jet stream of the Asian summer monsoon. *J. Clim.* 17, 1449–1458. doi: 10.1175/1520-0442(2004)017<1449:IVOTLJ>2.0.CO;2

Joseph, S., Sahai, A. K., Phani, R., Mandal, R., Dey, A., Chattopadhyay, R., et al. (2019). Skill evaluation of extended-range forecasts of rainfall and temperature over the meteorological subdivisions of India. *Weather Forecast.* 34, 81–101. doi: 10.1175/WAF-D-18-0055.1

Joseph, S., Sahai, A. K., Sharmila, S., Abhilash, S., Borah, N., Chattopadhyay, R., et al. (2015). North Indian heavy rainfall event during June 2013: diagnostics and extended range prediction. *Clim. Dyn.* 44, 2049–2065. doi: 10.1007/s00382-014-2291-5

Kalnay, E. (2019). Historical perspective: earlier ensembles and forecasting forecast skill. *Q. J. R. Meteorol. Soc.* 145, 25–34. doi: 10.1002/qj.3595

Kalnay, E., and Dalcher, A. (1987). Forecasting forecast skill. *Monthly Weather Rev.* 115, 349–356. doi: 10.1175/1520-0493(1987)115<0349:FFS>2.0.CO;2

Kaur, M., Krishna, R. P. M., Joseph, S., Dey, A., Mandal, R., Chattopadhyay, R., et al. (2020). Dynamical downscaling of a multimodel ensemble prediction system: application to tropical cyclones. *Atmos. Sci. Lett.* 21, 1–11. doi: 10.1002/asl.971

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., et al. (2000). Multimodel ensemble forecasts for weather and seasonal climate. *J. Clim.* 13, 4196–4216. doi: 10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2

Mandal, R., Joseph, S., Sahai, A. K., Phani, R., Dey, A., Chattopadhyay, R., et al. (2019). Real time extended range prediction of heat waves over India. *Sci. Rep.* 9:9008. doi: 10.1038/s41598-019-45430-6

Mitra, A. K., Bohra, A. K., Rajeevan, M. N., and Krishnamurti, T. N. (2009). Daily Indian precipitation analysis formed from a merge of rain-gauge data with the TRMM TMPA satellite-derived rainfall estimates. *J. Meteorol. Soc. Jpn. A* 87, 265–279. doi: 10.2151/jmsj.87A.265

Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T. (1996). The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.* 122, 73–119. doi: 10.1002/qj.49712252905

Murphy, A. H. (1991). Forecast verification: its complexity and dimensionality. *Mon. Weather Rev.* 119, 1590–1601. doi: 10.1175/1520-0493(1991)119<1590:FVICAD>2.0.CO;2

O'Malley, R. E. (1988). On nonlinear singularly perturbed initial value problems. *SIAM Rev.* 30, 193–212. doi: 10.1137/1030044

Orrell, D., Smith, L., Barkmeijer, J., and Palmer, T. N. (2001). Model error in weather forecasting. *Nonlin. Process. Geophys.* 8, 357–371. doi: 10.5194/npg-8-357-2001

Pai, D. S., Sridhar, L., Rajeevan, M., Sreejith, O. P., Satbhai, N. S., and Mukhopadhyay, B. (2014). Development of a new high spatial resolution ($0.25^\circ \times 0.25^\circ$) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam* 1, 1–18.

Pan, H. L., and Wu, W. S. (1995). Implementing a mass flux convective parameterization package for the NMC medium-range forecast model. *NMC Off. Note* 409:40.

Parker, D. J., Willetts, P., Birch, C., Turner, A. G., Marsham, J. H., Taylor, C. M., et al. (2016). The interaction of moist convection and mid-level dry air in the advance of the onset of the Indian monsoon. *Q. J. R. Meteorol. Soc.* 142, 2256–2272. doi: 10.1002/qj.2815

Parthasarathy, B., Kumar, K. R., and Kothawale, D. R. (1992). Indian summer monsoon rainfall indices: 1871–1990. *Meteorol. Mag.* 121, 174–186.

Pattanaik, D. R., Sahai, A. K., Mandal, R., Phani Muralikrishna, R., Dey, A., Chattopadhyay, R., et al. (2019). Evolution of operational extended range forecast system of IMD: Prospects of its applications in different sectors. *Mausam* 70, 233–264.

Raghavan, K. (1973). Tibetan anticyclone and tropical easterly jet. *Pure Appl. Geophys.* 110, 2130–2142. doi: 10.1007/BF00876576

Rai, P. K., Singh, G. P., and Dash, S. K. (2020). Projected changes in extreme precipitation events over various subdivisions of India using RegCM4. *Clim. Dyn.* 54, 247–272. doi: 10.1007/s00382-019-04997-6

Rajeevan, M., Bhate, J., and Jaswal, A. K. (2008). Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophys. Res. Lett.* 35, 1–6. doi: 10.1029/2008GL036105

Rajeevan, M., Gadgil, S., and Bhate, J. (2010). Active and break spells of the Indian summer monsoon. *J. Earth Syst. Sci.* 119, 229–247. doi: 10.1007/s12040-010-0019-4

Rao, S. A., Goswami, B. N., Sahai, A. K., Rajagopal, E. N., Mukhopadhyay, P., Rajeevan, M., et al. (2020). Monsoon mission: a targeted activity to improve monsoon prediction across scales. *Bull. Am. Meteorol. Soc.* 100, 2509–2532. doi: 10.1175/BAMS-D-17-0330.1

Rao, Y. P. (1976). *Southwest Monsoon*. Met. Monograph, India Meteorological Department, 367.

Rashid, H. A., Hendon, H. H., Wheeler, M. C., and Alves, O. (2011). Prediction of the Madden–Julian oscillation with the POAMA dynamical prediction system. *Clim. Dyn.* 36, 649–661. doi: 10.1007/s00382-010-0754-x

Richardson, D. (1997). "The relative effects of model and analysis differences on ECMWF and UKMO operational forecasts," in *Workshop on Predictability 20–22 October 1997* (Reading: ECMWF).

Robertson, A. W., Acharya, N., Goddard, L., Pattanaik, D. R., Sahai, A. K., Singh, K. K., et al. (2019). Subseasonal forecasts of the 2018 Indian summer monsoon over Bihar. *J. Geophys. Res. Atmos.* 124, 13861–13875. doi: 10.1029/2019JD031374

Roxy, M. K., Ghosh, S., Pathak, A., Athulya, R., Mujumdar, M., Murtugudde, R., et al. (2017). A threefold rise in widespread extreme rain events over central India. *Nat. Commun.* 8:708. doi: 10.1038/s41467-017-00744-9

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system version 2. *J. Clim.* 27, 2185–2208. doi: 10.1175/JCLI-D-12-00823.1

Sahai, A. K., Chattopadhyay, R., Joseph, S., R, P., and S, A. (2016). Extended range prediction system and its application. *Vayu Mandal* 42, 75–96.

Sahai, A. K., Sharmila, S., Abhilash, S., Chattopadhyay, R., Borah, N., Krishna, R. P., et al. (2013). Simulation and extended range prediction of monsoon intraseasonal oscillations in NCEP CFS/GFS version 2 framework. *Curr. Sci.* 104, 1394–1408.

Sikka, D. R., and Gadgil, S. (1980). On the maximum cloud zone and the ITCZ over Indian, longitudes during the southwest monsoon. *Mon. Weather Rev*. 108, 1840–1853. doi: 10.1175/1520-0493(1980)108<1840:OTMCZA>2.0.CO;2

Singh, R., Kishtawal, C. M., Pal, P. K., and Joshi, P. C. (2011). Assimilation of the multisatellite data into the WRF model for track and intensity simulation of the Indian Ocean tropical cyclones. *Meteorol. Atmos. Phys*. 111, 103–119. doi: 10.1007/s00703-011-0127-y

Sooraj, K. P., Terray, P., and Xavier, P. (2016). Sub-seasonal behaviour of Asian summer monsoon under a changing climate: assessments using CMIP5 models. *Clim. Dyn*. 46, 4003–4025. doi: 10.1007/s00382-015-2817-5

Stensrud, D. J., Bao, J. W., and Warner, T. T. (2000). Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Weather Rev*. 128, 2077–2107. doi: 10.1175/1520-0493(2000)128<2077:UICAMP>2.0.CO;2

Stensrud, D. J., and Fritsch, J. M. (1994). Mesoscale convective systems in weakly forced large-scale environments. Part III: numerical simulations and implications for operational forecasting. *Mon. Weather Rev*. 122, 2084–2104. doi: 10.1175/1520-0493(1994)122<2084:MCSIWF>2.0.CO;2

Suhas, E., Neena, J. M., and Goswami, B. N. (2013). An Indian monsoon intraseasonal oscillations (MISO) index for real time monitoring and forecast verification. *Clim. Dyn*. 40, 2605–2616. doi: 10.1007/s00382-012-1462-5

Tapiador, F. J., Tao, W. K., Shi, J. J., Angelis, C. F., Martinez, M. A., Marcos, C., et al. (2012). A comparison of perturbed initial conditions and multiphysics ensembles in a severe weather episode in Spain. *J. Appl. Meteorol. Climatol*. 51, 489–504. doi: 10.1175/JAMC-D-11-041.1

Toth, Z., and Kalnay, E. (1993). Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc*. 74, 2317–2330. doi: 10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2

Toth, Z., and Kalnay, E. (1997). Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev*. 125, 3297–3319. doi: 10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2

Vitart, F., and Molteni, F. (2009). Dynamical extended-range prediction of early monsoon rainfall over India. *Mon. Weather Rev*. 137, 1480–1492. doi: 10.1175/2008MWR2761.1

Vitart, F., and Robertson, A. W. (2018). The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *NPJ Clim. Atmos. Sci*. 1, 1–7. doi: 10.1038/s41612-018-0013-0

Wang, B., and Fan, Z. (1999). Choice of South Asian summer monsoon indices. *Bull. Am. Meteorol. Soc*. 80, 629–638. doi: 10.1175/1520-0477(1999)080<0629:COSASM>2.0.CO;2

Webster, P. J., and Yang, S. (1992). Monsoon and Enso: selectively interactive systems. *Q. J. R. Meteorol. Soc*. 118, 877–926. doi: 10.1002/qj.49711850705

Woo, S., Singh, G. P., Oh, J. H., and Lee, K. M. (2019). Projection of seasonal summer precipitation over Indian sub-continent with a high-resolution AGCM based on the RCP scenarios. *Meteorol. Atmos. Phys*. 131, 897–916. doi: 10.1007/s00703-018-0612-7

Xu, Z., Chen, J., Jin, Z., Li, H., and Chen, F. (2020). Assessment of the forecast skill of multiphysics and multistochastic methods within the GRAPES regional ensemble prediction system in the East Asian monsoon region. *Weather Forecast*. 35, 1145–1171. doi: 10.1175/WAF-D-19-0021.1

Zhao, Q., and Carr, F. H. (1997). A prognostic cloud scheme for operational NWP models. *Mon. Weather Rev*. 125, 1931–1953. doi: 10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2

**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 Climate

ORIGINAL RESEARCH
published: 09 June 2021
doi: 10.3389/fclim.2021.678109

# Toward Consistent Observational Constraints in Climate Predictions and Projections

*Gabriele C. Hegerl[1]*, Andrew P. Ballinger[1], Ben B. B. Booth[2], Leonard F. Borchert[3], Lukas Brunner[4], Markus G. Donat[5], Francisco J. Doblas-Reyes[5], Glen R. Harris[2], Jason Lowe[2], Rashed Mahmood[5], Juliette Mignot[3], James M. Murphy[2], Didier Swingedouw[6] and Antje Weisheimer[7]*

[1] School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom, [2] Met Office Hadley Centre, Exeter, United Kingdom, [3] Sorbonne Universités (SU/CNRS/IRD/MNHN), LOCEAN Laboratory, Institut Pierre Simon Laplace (IPSL), Paris, France, [4] Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland, [5] Barcelona Supercomputing Center (Centro Nacional de Supercomputación BSC-CNS), Barcelona, Spain, [6] EPOC, Université de Bordeaux, Pessac, France, [7] Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, United Kingdom

Observations facilitate model evaluation and provide constraints that are relevant to future predictions and projections. Constraints for uninitialized projections are generally based on model performance in simulating climatology and climate change. For initialized predictions, skill scores over the hindcast period provide insight into the relative performance of models, and the value of initialization as compared to projections. Predictions and projections combined can, in principle, provide seamless decadal to multi-decadal climate information. For that, though, the role of observations in skill estimates and constraints needs to be understood in order to use both consistently across the prediction and projection time horizons. This paper discusses the challenges in doing so, illustrated by examples of state-of-the-art methods for predicting and projecting changes in European climate. It discusses constraints across prediction and projection methods, their interpretation, and the metrics that drive them such as process accuracy, accurate trends or high signal-to-noise ratio. We also discuss the potential to combine constraints to arrive at more reliable climate prediction systems from years to decades. To illustrate constraints on projections, we discuss their use in the UK's climate prediction system UKCP18, the case of model performance weights obtained from the Climate model Weighting by Independence and Performance (ClimWIP) method, and the estimated magnitude of the forced signal in observations from detection and attribution. For initialized predictions, skill scores are used to evaluate which models perform well, what might contribute to this performance, and how skill may vary over time. Skill estimates also vary with different phases of climate variability and climatic conditions, and are influenced by the presence of external forcing. This complicates the systematic use of observational constraints. Furthermore, we illustrate that sub-selecting simulations from large ensembles based on reproduction of the observed evolution of climate variations is a good testbed for combining projections and predictions. Finally, the methods described in this paper potentially add value to projections and predictions for users, but must be used with caution.

**Keywords: climate change, climate predictions, future projections, observational constraints, model evaluation, climate modeling**

OPEN ACCESS

***Edited by:***
*Matthew Collins,*
*University of Exeter, United Kingdom*

***Reviewed by:***
*Benjamin Sanderson,*
*Centre Europeen De Recherche Et De Formation Avancee En Calcul Scientifique, France*
*Isla Simpson,*
*National Center for Atmospheric Research (UCAR), United States*

****Correspondence:***
*Gabriele C. Hegerl*
*gabi.hegerl@ed.ac.uk*

***Specialty section:***
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

***Received:*** *08 March 2021*
***Accepted:*** *26 April 2021*
***Published:*** *09 June 2021*

***Citation:***
*Hegerl GC, Ballinger AP, Booth BBB, Borchert LF, Brunner L, Donat MG, Doblas-Reyes FJ, Harris GR, Lowe J, Mahmood R, Mignot J, Murphy JM, Swingedouw D and Weisheimer A (2021) Toward Consistent Observational Constraints in Climate Predictions and Projections. Front. Clim. 3:678109. doi: 10.3389/fclim.2021.678109*

## INTRODUCTION

Information about future climate relies on climate model simulations. Given the uncertainty in the future climate's response to external forcings and climate models' persistent biases, there is a need for coordinated multi-model experiments. This need is addressed by the Coupled Model Intercomparison Project (CMIP), proposing a uniform protocol to evaluate the future climate. Currently, this protocol proposes to explore two future timescales separately: firstly the evolution of the climate toward the end of the century, and secondly the evolution of the climate within the first decade ahead (Eyring et al., 2016). Climate variations on the longer timescale are primarily driven by the climate responses to different scenarios of socio-economic development and resulting anthropogenic emissions of greenhouse gases and aerosols (Gidden et al., 2019; see also Forster et al., 2020). At decadal timescales on the other hand, the internal variability of the climate system is an important source of uncertainty, and part of the associated skill comes from successfully initializing models with the observed state of the climate. The two timescales are thus subject to different challenges and are therefore addressed by distinct experimental setups. In both cases, coordinated multi-model approaches are necessary to estimate uncertainty from model simulations.

To account for internal variability, the size of individual climate model ensembles has increased, so that there is a growing need to extract the maximum information from these ensembles and to grasp the opportunities associated with large ensembles (e.g., Kay et al., 2015). In particular, treating each model as equally likely (the so-called one-model-one-vote approach) may not provide the best information for climate decision making; This demonstrates the need for a well-informed decision on choice and processing of models for projections, while large ensembles may overcome, at least in part, concerns about signal-to-noise ratios in weighted ensembles (Weigel et al., 2010).

Furthermore, there is also a desire to provide decision makers with seamless information on the time-scale from a season to decades ahead. This involves the even more complex step of combining ensembles from initialized predictions started from observed conditions of near present-day with those from projections, the latter of which are typically started from conditions a century or more earlier. This paper discusses available methods using observations to evaluate and constrain ensemble predictions and projections, supporting the long-term goal of a consistent framework for their use in seamless predictions from years to decades.

Multiple techniques are available to constrain future projections drawing on different lines of evidence and considering different sources of uncertainty (e.g., Giorgi and Mearns, 2002; Knutti, 2010; Knutti et al., 2017; Sanderson et al., 2017; Lorenz et al., 2018; Brunner et al., 2020a,b; Ribes et al., 2021). Models that explore the full uncertainty in parameter space provide very wide uncertainty ranges (Stainforth, 2005), motivating the need to use observational constraints. Usually observational constraints are based on the assumption that there is a reliable link between model performance compared to observations over the historical era with future model behavior. This link is expressed using emergent constraints, weights, or other statistical approaches. For instance, this could mean excluding or downweighting models which are less successful in reproducing the climatological mean state or seasonal cycle. The constraint can also be based on the variability, representation of mechanisms or relationships between different variables, or changes in multi-model assessments of future changes (e.g., Hall and Qu, 2006; Sippel et al., 2017; Donat et al., 2018), which includes evaluation of the climate change magnitude in detection and attribution approaches (e.g., Stott and Kettleborough, 2002; Tokarska et al., 2020b). In a similar manner, the risk of experiencing an abrupt change in the subpolar North Atlantic gyre has been constrained by the capability of CMIP5 climate models to reproduce stratification in this region, which plays a key role in the dynamical behavior of the ocean (Sgubin et al., 2017). This is based on the fundamental idea that certain physical mechanisms of climate need to be appropriately simulated for the model to be "fit for purpose," and consistent with this thought, the Intergovernmental Panel on Climate Change (IPCC) reports have consistently dedicated a chapter to climate model evaluations. The IPCC has also drawn on observational constraints from attribution to arrive at uncertainty estimates in predictions both in assessment reports four (AR4) (Knutti et al., 2008) and AR5 (Collins et al., 2013).

Many methods constraining projections have been evaluated using model-as-truth approaches and several of them have been part of a recent method intercomparison based on a consistent framework (Brunner et al., 2020a). The authors found that there is a substantial diversity in the methods' underlying assumptions, uncertainties covered, and lines of evidence used. Therefore, it is maybe not surprising that the results of their application are not always consistent, and that they tend to be more consistent for the central estimate than the quantification of uncertainty. The latter is important, as reliable uncertainty ranges are often key to actionable climate information.

Emergent constraints is another highly visible research area that makes use of relationships between present day observable climate and projected future changes. Emergent constraints rely on statistical relationships between present day, observable, climate properties and the magnitude of future change. There is currently effort within this community to discriminate between those that are purely statistical from those where there is further confirmational evidence to support their usage (e.g., Caldwell et al., 2018, Hall et al., 2019). Efforts to identify consensus or consolidate constraints from multiple, often conflicting, emergent constraints have started to take place within the climate sensitivity context (Bretherton and Caldwell, 2020, Sherwood et al., 2020). However, these frameworks do not yet account for common model structural errors that will likely lead such assessments to an overly confident constraint (Sanderson et al., 2021). The reliability of emergent constraints for general climate projections is even less clear at this time (e.g., Brient, 2020), and therefore, we do not discuss such constraints further here as it is not clear how complete and reliable such constraints are.

For initialized predictions (Pohlmann et al., 2005; Meehl et al., 2009, 2021; Yeager and Robson, 2017; Merryfield et al., 2020; Smith et al., 2020), skill scores assess the model system's

performance in hindcasts compared to observations, allowing for a routine evaluation of the prediction system that is unavailable to projections. Multi-model studies on predictions have, however, only recently started to emerge as more sets of initialized decadal prediction simulations have become available as part of the CMIP6 Decadal Prediction Project (DCPP; Boer et al., 2016). Some studies merged CMIP5 and CMIP6 decadal prediction systems to maximize ensemble size for optimal filtering of the noise (e.g., Smith et al., 2020), or contrasted the multi-model means of CMIP5 and CMIP6 to pinpoint specific improvements in prediction skill from one CMIP iteration to the other (Borchert et al., 2021). Attempts to explicitly contrast and explain the decadal prediction skill of different model systems are yet very rare (Menary and Hermanson, 2018). There are therefore no methods of constraining or weighting multi-model ensembles of decadal prediction simulations in the literature which we could rely upon.

For these reasons, we provide in this paper a first exploration of discriminant features of multi-model decadal prediction ensembles with the aim of providing an indication which inherent model features benefit, and which degrade skill. We also discuss the contribution of forcing and internal variability to decadal prediction skill over time, and show how times of low and high skill (windows of opportunity; Mariotti et al., 2020) can be used to constrain sources of skill in space and time. We consider the cross-cutting relevance of observational constraints and reflect on their consistency across prediction and projection timescales and approaches. We also pilot opportunities for building upon multiple methods and investigate how observational constraints may be used in uncertainty characterization in a seamless prediction. Finally, we discuss the challenges in applying observational constraints to predictions, where skill varies over time and may therefore not be consistent across prediction timelines.

This paper examines the potential for observational constraints in the three European SREX regions Northern Europe (NEU), Central Europe (CEU) and Mediterranean (MED) [see, e.g., Brunner et al. (2020a)]. Many of our results will be transferable to other regions, although the signal-to-noise ratio as well as the skill of initialized predictions might be different for larger regions or lower latitude regions, with the potential for observational constraints being more powerful in some regions as a consequence. Hence our European example can be seen as a stress test for observational constraints in use.

We first illustrate examples of observational constraints for projections, identify contributing factors to model skill metrics, and explore the potential to use multiple constraints in sequence. We then illustrate, on the interface from projections to predictions, that the performance of a prediction system can be emulated by constraining a large ensemble to follow observational constraints on modes of sea surface temperature (SST). Lastly, the origin of skill and observational constraints in initialized predictions is illustrated across different models, different timelines and different regions as a first step toward consistently constraining predictions and projections for future merging applications. We draw lessons and recommendations for the use of observational constraints in the final section.

# CONSTRAINING PROJECTIONS

## Lessons Learned From the Use of Observational Constraints in Climate Projections in UKCP18

Observational constraints have played an important role in the latest generation of the UK climate projections (UKCP18; Murphy et al., 2018). UKCP18 includes sets of 28 global model simulations (∼60 km resolution), 12 regional (12 km resolution), and 12 local (2.2 km resolution) realizations of 21st century climate consisting of raw climate model data, for use in detailed analysis of climate impacts (Murphy et al., 2018; Kendon et al., 2019). Also provided is a set of probabilistic projections, the role of which is to provide more comprehensive estimates of uncertainty for use in risk assessments in their own right, and also as context for the realizations. The probabilistic climate projections are derived from a larger set of 360 model simulations, based on a combination of perturbed parameter ensembles with a single model, combined with simulations with different CMIP5 models. These have been combined to make probability density functions representing uncertainties due to internal variability and climate response, using a Bayesian framework that includes the formal application of observational constraints. The UKCP18 probabilistic approach is one of the methods covered in Brunner et al. (2020a). Key aspects include: (a) use of emulators to quantify parametric model uncertainties, by estimating results for parts of parameter space not directly sampled by a climate model simulation; (b) use of CMIP5 earth system models to estimate the additional contribution of structural model uncertainties (termed "discrepancy" in this framework) to the pdfs; (c) sampling of carbon cycle uncertainties alongside those due to physical climate feedbacks. The method, described in Murphy et al. (2018), is updated from earlier work by Sexton et al. (2012), Harris et al. (2013), Sexton and Harris (2015), and Booth et al. (2017). The climatological constraints are derived from seasonal spatial fields for 12 variables. These include latitude-longitude fields of surface temperature, precipitation, sea-level pressure, total cloud cover and energy exchanges at the surface and at the top of the atmosphere, plus the latitude-height distribution of relative humidity (denoted HIST in **Figure 1**). This amounts to 175,000 observables, reduced in dimensionality to six through eigenvector analysis (Sexton et al., 2012). Constraints from historical surface air temperature (SAT) change include the global average, plus three indices representing large-scale patterns (Braganza et al., 2003). The ocean heat content metric (OHC) is the global average in the top 700 m. The $CO_2$ constraint arises because the UKCP18 projections include results from earth system model simulations that predict the historical and future response of $CO_2$ concentration to carbon emissions, thus including uncertainties due to both carbon cycle and physical climate feedbacks. The observed trend in $CO_2$ concentration is therefore combined with the other metrics in the weighting methodology, to provide a multivariate set of constraints used to update joint prior probability distributions for a set of historical and future prediction variables (further details in Murphy et al., 2018).

**Figure 1** illustrates the impact of observational constraints on the UKCP18 pdfs for global mean temperature, and summer temperature and precipitation for Southern England; 2080–2099 relative to 1981–2000, under RCP8.5. Results show that as well as narrowing the range, specific constraints can also weight different parts of the pdf up or down, compared to the prior distribution. As an example, the chance of a summer drying is upweighted in the posterior, by the application of both the climatology and historical temperature trend constraints. Experiences with use of observational constraints in UKCP18 illustrate that considering multiple constraints can be powerful. This is shown in **Figure 1** by a sensitivity test, in which each pdf is modified by adding individual constraints in sequence. However, the impact of specific constraints can depend on the order in which they are applied. Here, e.g., the effect of historic changes in ocean heat content might appear larger, if applied as the first step in this illustration. This illustrates that there is plenty of scope to refine such constraint methods in the future. For example, metrics of climate variability are not yet considered in the set of historical climatology constraints.

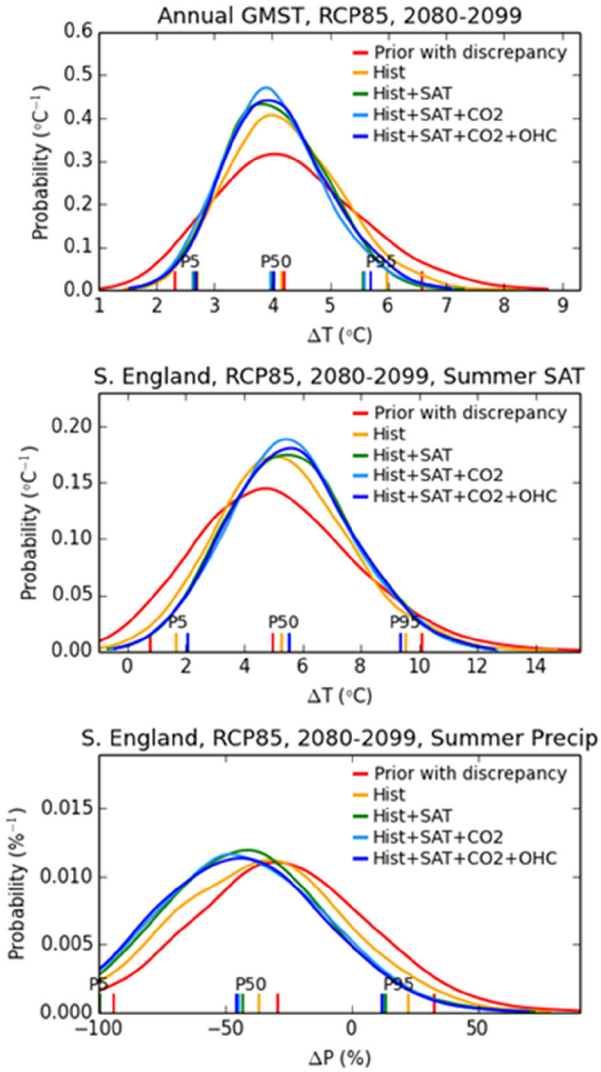



**FIGURE 1 |** Examples of the impact of constraints derived from historical climatology (Hist), added historical global surface air temperature trends (Hist+SAT), added historical trends in atmospheric $CO_2$ concentration (Hist+SAT+CO2), and added upper ocean heat content (Hist+SAT+CO2+OHC), in modifying the prior distribution to form the posterior. The 5th, 50th (median), and 95th percentiles are plotted, along with the pdfs. Results prior to the application of observational constraints (Prior with discrepancy) are also shown. Reproduced from Murphy et al. (2018).

## Examples of Methods for Observational Constraints on Projections

### Performance Weighting Methods (ClimWIP)

Methods using performance weighting evaluate if models are fit for purpose and weight them accordingly (see also UKCP18 Example discussed above). The fundamental idea is that projected climate change can only be realistic if the model simulates processes determining present day climate realistically as discussed e.g., in Knutti et al. (2017) for the case of Arctic sea ice. An updated version of the same method (termed Climate model Weighting by Independence and Performance—ClimWIP) was recently applied by Brunner et al. (2020b) to the case of global mean temperature change. Each model's weight is based on a range of performance predictors establishing its ability to reproduce observed climatology, variability and trend fields. These predictors are selected to be physically relevant and correlated to the target of prediction. Other approaches, such as emergent constraints, often use a single highly correlated metric, while ClimWIP draws on several such metrics. This can avoid giving heavy weight to a model which fits the observations well in one metric but is very far away in several others. In addition to that, they also include information about model dependencies within the multi-model ensemble (see Knutti et al., 2013), effectively downweighting model pairs which are similar to each other.

We show results from two applications of ClimWIP here: **Figure 2** illustrates the effect of ClimWIP, alongside other methods, compared to using unconstrained predictions from a set of CMIP5 models, illustrating that ClimWIP reduces spread in some seasons and regions, and also shifts the central tendency somewhat, depending on the case (Brunner et al., 2019), and in a similar manner as illustrated above for UKCP18. The CMIP6 weights used in the later part of the study are based on the latest version of ClimWIP described in Brunner et al. (2020b) and based on earlier work by Merrifield et al. (2020), Brunner et al. (2019), Lorenz et al. (2018), and Knutti et al. (2017). We used performance weights based on each model's generalized distance to reanalysis products (ERA5; Hersbach et al., 2020, and MERRA2, Gelaro et al., 2017) in five diagnostics evaluated from 1980 to 2014: global, spatially resolved fields of climatology and variability of near-surface air temperature and sea level pressure, as well as global, spatially resolved fields of near-surface air temperature trend (Brunner et al., 2020b). The weights were retrieved from the same setup as used in

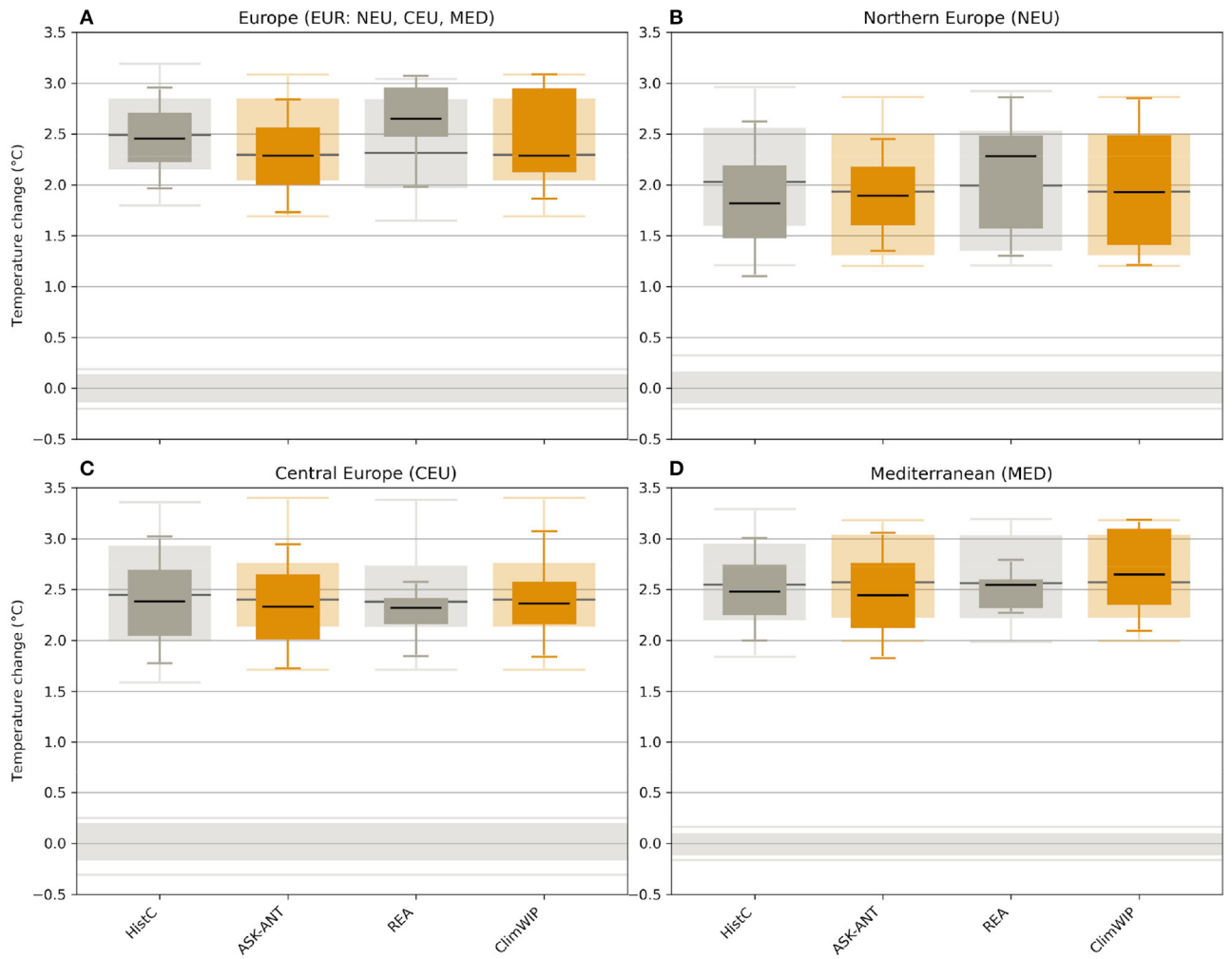



**FIGURE 2 |** Projected summer (June–August) temperature change 2041–2060 relative to 1995–2014 for **(A)** the combined European region as well as **(B–D)** the three European SREX regions (Northern Europe, Central Europe, and the Mediterranean) using RCP8.5. The lighter boxes give the unconstrained distributions originating from model simulations; the darker boxes give the observationally constrained distributions. Shown are median, 50% (bars) and 80% (whiskers) range. The gray box and lines centered around zero show the same percentiles of 20-years internal variability based on observations. Methods ASK-ANT and ClimWIP used in this paper are colored, additional methods HistC and REA shown in Brunner et al. (2020a), but not used in this paper, are in gray. After Figure 6 in Brunner et al. (2020a).

Brunner et al. (2020b) and it is important to note that they are optimized to constrain global mean temperature change in the second half of the 21st century for the full CMIP6 ensemble. Here we use them only to show the general applicability combining ClimWIP with the ASK approach (outlined next), which illustrates common inputs across constraints on projections and their relation to each other. We apply them to a subset of nine models for which Detection and Attribution Model Intercomparison Project (DAMIP simulations; Gillett et al., 2016) are available, and then focus on projections for Europe. For applications beyond the illustrative approach shown here, it is critical to retune the method for the chosen target and model subset.

## Trend and Attribution Based Methods (ASK Method)

Another widely used method for constraining projections focuses on the amplitude of forced changes (here referred to as "trend," although the constrained time-space pattern may be more complex than a simple trend). This method focuses on the performance of climate models in simulating externally forced climate change, with the idea that a model that responds too strongly or too weakly over the historical period may also do so in the future. Trend performance is included in ClimWIP, as its full implementation accounts for trends, and also in UKCP18, as illustrated above.

Trend based methods need to consider that the observed period is not only driven by greenhouse gas increases, but also

influenced by aerosol forcing, natural forcings (e.g., Bindoff et al., 2013) as well as internal variability[1], all of which impact on the magnitude of the observed climate change Since the future may show different combinations of external forcing than the past, including reducing aerosol forcing with increased pollution control, and different phases of natural forcing, non-discriminant use of trends may introduce errors. Two approaches have been used to circumvent this problem: one method is to use a period of globally flat aerosol forcing and argue that the largest contributor to trends is greenhouse gases over such periods, and then use trends as emergent constraints (Tokarska et al., 2020a). An alternative, the so-called Allen Stott Kettleborough "ASK method," introduced in the early 2000s (Allen et al., 2000; Stott and Kettleborough, 2002; Shiogama et al., 2016) uses results from detection and attribution of observed climate change to constrain projections. These methods seek to disentangle the role of different external forcings and internal variability in observed trends, and result in an estimate of the contribution by natural forcings, greenhouse gases, and other anthropogenic factors to recent warming. This allows us to estimate the observed greenhouse gas signal, and use it to constrain projections. This can be done by selecting climate models within the observed range of greenhouse gas response (Tokarska et al., 2020b) or by using the uncertainty range of greenhouse warming that is consistent with observations as an uncertainty range in future projections around the multi-model mean fingerprint (Kettleborough et al., 2007). The latter method has been included in assessed uncertainty ranges in projections in IPCC (see Knutti et al., 2008; Collins et al., 2013).

Here we illustrate the use of attribution based observational constraints. This method assumes that the true observed climate response, $y_{obs}$, to historical forcing is a linear combination of one or more (*n)* individual forcing fingerprints, $X_j$, scaled by adjustable scaling factors, $\beta_j$, to observations. We use the gridded observations E-OBS v19.0e dataset (Haylock et al., 2008), with monthly values computed from the daily data. Scaling factors are determined that optimize the fit to observations. Hence this method uses the response in observations to estimate the amplitude of a model-estimated space time pattern of response, with the rationale that uncertain feedbacks may lead to a larger or smaller response than anticipated in climate models (e.g., Hegerl and Zwiers, 2011). We use a total-least-squares (TLS) method to estimate the scaling factors, which accounts for noise in both the observations $\varepsilon_{obs}$, and in the modeled response to each of the forcings $\varepsilon_j$(see e.g., Schurer et al., 2018),

$$y_{obs} = \sum_{j=1}^{n} \beta_j (X_j - \varepsilon_j) + \varepsilon_{obs} \quad (1)$$

where the n fingerprints chosen may include the response to greenhouse gases only (GHG), natural forcings only (NAT), other anthropogenic forcings (OTH) or combinations thereof (ANT = GHG+OTH). A confidence interval for each of the scaling factors describes the range of magnitudes of the model response that are consistent with the observed signal. A forced model response is *detected* if the range of scaling factors are significantly >0, and can be described as being *consistent with observations* if the range of values contains the magnitude of one (=1). The uncertainty due to internal climate variability is here estimated by adding samples from the preindustrial Control simulations (of the same length) to the noise-reduced fingerprints and observations, and recomputing the TLS regression (10,000 times) in order to build a distribution of scaling factors, from which the 5th−95th percentile range can be computed. We have also explored confidence intervals based on bootstrapping (DelSole et al., 2019), and while there are slight differences in the spread, the two measures generally provide consistent and robust agreement.

CMIP6 model simulations (Eyring et al., 2016) run with historical forcings, and Detection and Attribution MIP (DAMIP) single-forcing simulations (Gillett et al., 2016) are used over the same period as E-OBS (1950–2014) to determine the fingerprints. Our analysis uses a set of nine models with 33 total ensemble members (**Table 1**), that were available in the Center for Environmental Data Analysis (CEDA) curated archive (retrieved in September 2020), common to the required set of simulations. For application of the ASK method, single forcing experiments are needed. Monthly surface air temperature fields from the observations and each of the CMIP6 model ensemble members were spatially regridded to a regular 2.5° × 2.5° latitude-longitude grid, with only the grid boxes over land (with no missing data throughout time) being retained in the analysis. The resulting masked fields (from observations and all individual model ensemble members) were spatially averaged over a European domain (EUR) and three sub-domains (NEU, CEU, and MED; as described in Brunner et al., 2020a). Fingerprints for each forcing are based on an unweighted, and in the example below (section Contrasting and Combining Constraints From Different Methodologies), weighted, average of each model's ensemble mean response to individual forcings. The total least squares approach requires an estimate of the signal-to-noise ratio of the fingerprint. This is calculated considering the noise reduction by averaging individual model ensemble averages, and assuming that the resulting variance adds in quadrature when averaging across ensembles. When weights are used, these are included in the calculation. Results from ASK are illustrated in **Figure 2**, again illustrating that the method reduces spread in some cases, and influences central tendency as well.

Whether this reduction in spread improves the reliability of projections is still uncertain, although some recent analysis supports these approaches: Gillett et al. (2021) applied an (im-) perfect model approach to estimate the attributable warming to CMIP6 models, and Schurer et al. (2018) an approach to estimate the transient climate sensitivity from individual simulations with withheld climate models for CMIP5. Gillett et al. (2021) found high reliability of the estimate of attributable warming, which increases confidence in its use for projections. Schurer et al. (2018) found that the method was somewhat overconfident for future warming if using the multi-model mean fingerprint,

[1]Bonnet R., Swingedouw D., Gastineau G., Boucher O., Deshayes J., Hourdin F., et al. (2021). Increased risk of near term global warming level due to a recent AMOC weakening. *Nat. Commun.* (in review).

**TABLE 1 |** List of the CMIP6 models used in the ASK-ClimWIP constraining intercomparison pilot study (restricting to models with individual forcing simulations available and normalizing weights to sum to unity for these relative to those shown in **Figure 3**).

| CMIP6 model name | Number of ensemble members included | ClimWIP weighting (no trend information) | ClimWIP weighting (with trend information) |
|---|---|---|---|
| ACCESS-ESM1-5 | 3 | 0.1627 | 0.1381 |
| BCC-CSM2-MR | 1 | 0.0132 | 0.0792 |
| CNRM-CM6-1 | 5 | 0.0772 | 0.0762 |
| CanESM5 | 10 | 0.0216 | 0.0049 |
| GFDL-ESM4 | 1 | 0.4051 | 0.5047 |
| HadGEM3-GC31-LL | 4 | 0.2582 | 0.0070 |
| IPSL-CM6A-LR | 5 | 0.0313 | 0.0639 |
| MIROC6 | 3 | 0.0052 | 0.0627 |
| MRI-EMS2-0 | 1 | 0.0256 | 0.0633 |
| Total | 33 | 1.0 | 1.0 |

but conservative if accounting fully for model uncertainty in a Bayesian approach, or if inflating residual variability.

## Contrasting and Combining Constraints From Different Methodologies

Eight different methods to arrive at weighted or constrained climate projections were recently compared for European regions in Brunner et al. (2020a). The study identified a lack of coordination across methods as a main obstacle for comparison since even studies which look at the same region in general might report results for slightly different domains, seasons, time periods or model subsets hindering a consistent comparison. Therefore, a common framework was developed to allow such a comparison between the different methods, including a set of European sub-regions. The results in Brunner et al. (2020a) focus on temperature and precipitation changes between 1995–2014 and 2041–2060 under RCP8.5 (i.e., using CMIP5) in the three European SREX regions. In addition, reasons for agreements and disagreements across these methods were also discussed.

**Figure 2** shows some results of the comparison illustrated in that review (for a detailed discussion of the results and the underlying methods see Brunner et al., 2020a). While all methods clearly show the anthropogenic warming signal, the comparison reveals different levels of agreement based on the region considered and the metric of interest (e.g., median vs. 80% range). In general, methods tend to agree better on the central estimate while uncertainty ranges can be fairly different, particularly for the more extreme percentiles (see, e.g., **Figure 2D**). However, for some regions also the median values can differ across methods and in isolated cases methods even disagree on the direction of the shift from the unconstrained distributions. Methods also constrain projections to different extents, with some methods leading to stronger constraints and others to weaker constraints. This can be due to using observations more or less completely and efficiently, but can also reflect differences in the underlying assumptions of the methods such as the statistical paradigm used. Some methods assume the models are exchangeable realization of the true observed response, while others assume that the models converge, as a sample, toward the truth, ranking models close to the model average as more likely to be correct.

For cases with such substantial differences, Brunner et al. (2020a) recommend careful evaluation of constraints projecting the future change in a withheld model based on each method. Full application of such withheld model approaches requires withholding a large number of simulations to ensure robust statistics, and is computationally expensive. Such work is ongoing in the community and will help resolve uncertainty across performance metrics. Brunner et al., also suggest attempting to merge methods, either based on their lines of evidence (before applying them) or based on their results (after applying them).

Here we pilot an example of combining two observational constraint methods. We do this to both illustrate what aspects of observed climate change influence performance metrics, and in order to test if a combined approach might harness the strengths of each paradigm. Results also illustrate the challenges and limitations involved in such an endeavor.

In order to do so, we limit the constraint used in ClimWIP to climatology and variance-based performance weights only. These can then be used to construct a weighted fingerprint (mentioned above) of the forced climate change that could be, arguably, more credible as a best estimate of the expected change than the simple one-model-one vote fingerprint generally used. It is also conceivable to combine both differently, e.g., by using the ASK constraint relative to a model's raw projection as weight in a ClimWIP weighted prediction. We chose the weighted ASK method for its ability to project changes outside the model range in cases where models over- or underestimate the actual climate change signal, but different choices are possible.

We use two different combinations of diagnostics to calculate the ClimWIP performance weights for this combination of constraints: one including temperature trends, and one without temperature trends (i.e., using only climatology and variability of temperature and sea level pressure). This is done to avoid accounting for trends twice when applying the constraints subsequently (**Table 1**), since the ASK method is strongly driven by temperature trends (while using also spatial information and the shape of the time series particularly to distinguish between the effects of different forcing and variability). Note that this modification of ClimWIP will most likely reduce its performance as a constraint on its own.

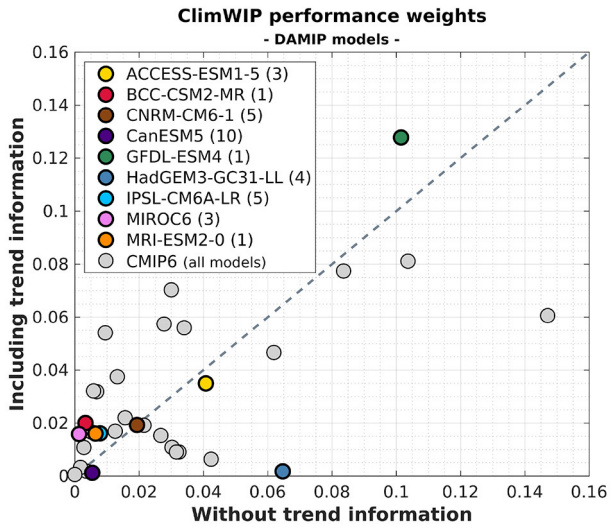



**FIGURE 3 |** Role of temperature trends in model performance weights, illustrating the relationship between the model-specific relative weights assigned by ClimWIP, computed using global fields excluding temperature trend information (along the x-axis), and including a temperature trend metric (along the y-axis). Climate models across CMIP6 are shown in gray; those also available in the DAMIP simulations (used in the ASK method) have been colored, with the number of ensemble members (n) shown in the legend.

The weights assigned to each of the 33 CMIP6 models (and the nine DAMIP models used in the ASK method) are shown in **Figure 3**, both when using all five diagnostics, and when not using the temperature trend. Results show that the performance weights from trends show a substantial influence on ClimWIP weights compared to the variant without trends, with largest differences for models with unusually strong trends, such as HadGEM3, which is almost disregarded in trend-based weighting but performs well on climatology. In contrast, trend information enhances the perceived value of a group of other models in the bottom left corner of the diagram, with very small weights in climatology-only cases compared to slightly larger ones in trend including cases. However, for many other models both metrics correlate (although their correlation is largely driven by a few highly weighted models). This illustrates that different information used can pull observational constraints in different directions.

There are suggestions that the role of trends in downweighting projections of higher end warming in both ClimWIP and ASK may be common across the wider set of projection methodologies. Historical trends in the UKCP18 methodology (**Figure 1**, labeled SAT) tend to reduce the upper tails of projected changes. Similarly, the HistC methodology (Brunner et al., 2020a section) is largely based on trend information, which also consistently downweights high end projected changes (Ribes et al., 2021), in response to too large change in such models over parts of the historical period.

We now use model performance weighting in constructing each of the multi-model mean fingerprints (**Figure 4**) that are subsequently used in the detection and attribution constraint. Thus, two sets of multi-model mean fingerprints are computed. Firstly, an equal-weighted set of multi-model fingerprints and, for comparison, a second set of multi-model fingerprints are computed as weighted average of each model's individual fingerprint in response to forcings. When combining the constraints in this way, we use the ClimWIP performance weights that were derived without temperature trend information (**Table 1**).

Annual surface temperature anomalies from 1950 to 2014 averaged over nine models (33 runs) are displayed in **Figure 4** with the upper left panel showing the equal-weighted time series, and the lower left panel showing the time series after applying the ClimWIP weights (without trend). The same observed annual time series (E-OBS, black line) has been plotted in each panel, along with the CMIP6 multi-model mean (of ensemble means) of the all-forcing historical simulations (brown line), the greenhouse gas single-forcing historical simulations (red line), and the natural single-forcing historical simulations (green line). A measure of the internal variability of the CMIP6 models is estimated by averaging the standard deviation (65-years samples) of the associated piControl simulations, and is indicated by the background shaded region.

The scaling factors were derived through a total least squares regression of the multi-model mean fingerprints onto the observations, estimating the amplitude of a single-fingerprint all forcing signal ("ALL"), and determining the separate amplitudes for combinations of fingerprints (for more details, see Brunner et al., 2020a; Ballinger et al., pers. com.; GHG, OTH, ANT, NAT). The scaling factors shown in **Figure 4** were derived using fingerprints comprising the conjoined annual time series of the three spatially-averaged European subregions (NEU, CEU, and MED; $3 \times 65 = 195$ years), each having first been normalized by a measure of that subregion's internal variability (using the standard deviation of equivalent piControl simulations). Hence the fingerprint used captures an element of the spatial signal in the three regions, and of their temporal evolution. The analysis was also performed using a single European average fingerprint, and separate single-subregion fingerprints (NEU/CEU/MED; not shown). As expected, the three-region fingerprint generally provides a tighter constraint because of the additional (spatial) information included, which strengthens the signal to noise ratio. However, the qualitative differences of using an equal-weighted or ClimWIP-weighted fingerprint, are found to be fairly robust irrespective of the particular fingerprint formulation.

**Figure 4** shows an overall narrowing of the uncertainty range in the scaling factors (providing a slightly tighter constraint) when using the ClimWIP-weighted model fingerprints, particularly for NAT, suggesting that at least in this case, the weighted fingerprints are more successful in identifying and separating responses to greenhouse gases from those to other forcings. The best-estimate magnitudes of the leading signal (ALL, GHG, ANTH) scaling factors remain reasonably robust. Results suggest that the weighted multi-model mean response to aerosols is larger than that in the observations, significantly so in the weighted case. Overall, the illustrated sensitivity of the estimated amplitude of natural and aerosol response probably

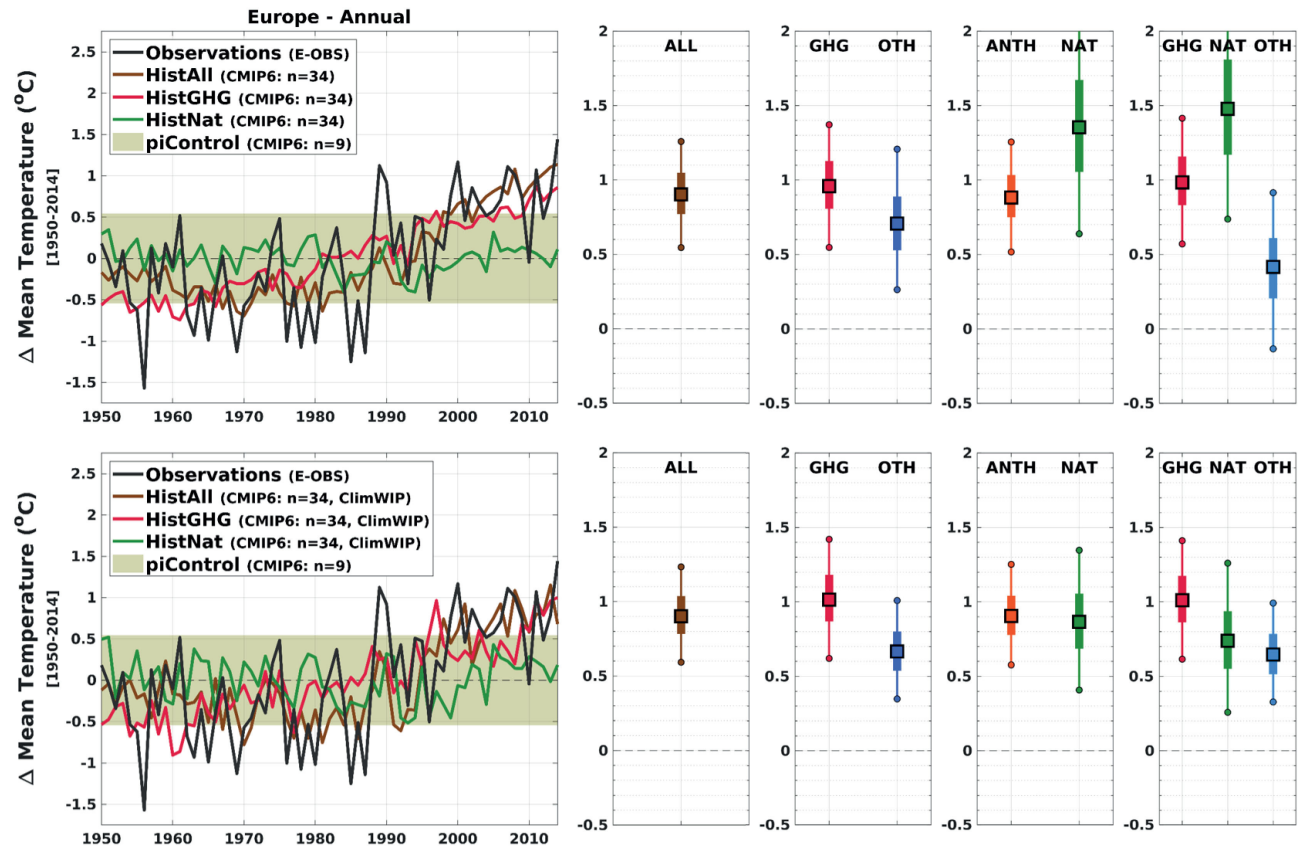



**FIGURE 4 |** Annual time series (left panels) of European surface air temperature anomalies (relative to 1950–2014) from observations (E-OBS v19, black line) and CMIP6 historical simulations (all forcings, brown line; GHG-only forcing, red line; and NAT-only forcing, green line). We display the multi-model mean of ensemble means from nine models, with the shaded region denoting the multi-model mean variability (±1 standard deviation) of the associated preindustrial Control simulations. The scaling factors (right panels) are derived by regression of the CMIP6 model fingerprints on the observations and indicate to what extent the multi-model mean fingerprint needs to be scaled to best match observations. Results show the 1-signal (ALL), two-signal (GHG & OTH; ANTH & NAT), and three-signal (GHG, NAT & OTH) scaling factors. The upper panels display the time-series fingerprints and associated scaling factors using an equal-weighted multi-model mean; the lower panels use ClimWIP performance weighting to weight the multi-model mean fingerprints for deriving the scaling factors. Confidence intervals show the 5th–95th (thin bars) and 25th–75th (thicker bars) percentile ranges of the resulting scaling factors.

reflects model differences in emphasis between ClimWIP weighted and unweighted cases.

We have further explored the robustness of results over different seasons (not shown). Results again suggest that the use of ClimWIP weights in the multi-model mean fingerprint yields stronger constraints when separating out the greenhouse gas signal (which is particularly useful for constraints). Also, the contribution by natural forcing, other anthropogenic and greenhouse gas forcing to winter temperature change is far less degenerate in the ClimWIP constrained case, although it needs to be better understood why this is the case. This illustrates some promise in combining constraints.

However, in order to evaluate if these narrowed uncertainty estimates reliably translate into better prediction skill, careful "perfect" and "imperfect" model studies will need to be performed, where single model simulations are withheld to predict their future evolution as a performance test, calculating performance weights relative to each withheld model (see e.g., Bo and Terray, 2015; Schurer et al., 2018; Brunner et al., 2020b). When doing so, it would be useful to consider forecast evaluation terminology used in predictions and to assess reliability (i.e., if model simulations that are synthetically predicted are within the uncertainty range of the prediction, given the statistical expectation; Schurer et al., 2018; Gillett et al., 2021), and if they show improved sharpness, i.e., their RMS error is smaller in order to avoid penalizing more confident methods unnecessarily). Another avenue is to draw perfect models from a different generation as explored, e.g., by Brunner et al. (2020b) where the skill of weighting CMIP6 was explored based on models from CMIP5 in order to provide an out-of-sample test to the extent that CMIP6 can be considered independent of CMIP5. A first pilot study using CMIP6 simulations to hindcast single CMIP5 simulations showed mixed results and no consistent preference for the ClimWIP vs. ASK vs. combined method in either metric (not shown; Ballinger et al., pers. com.).

In summary, there is an indication that the use of model weighting can potentially provide improved constraints on projections, fundamentally due to using fingerprints that rely

strongly on the most successful models. Europe as a target of reconstruction might be particularly tricky given high variability over a small continent, rendering more noisy fingerprints from weighted averages compared to straight multi-model averages, which can reduce the benefit in weighting approaches (Weigel et al., 2010).

However, climate variability can be considered as more than just "noise" in near term predictions, and hence the next method focuses on constraints for variability.

## Toward Seamless Predictions: Constraining Large Projection Ensembles to Match Recent Observed Variability

Above, observations were employed to evaluate projections in terms of processes, trends and climatology. Climate variability is considered in those previous analyses a random uncertainty that is separate from projections and adds uncertainty. This is in sharp contrast to initialized predictions, where one of the goals is to predict modes of variability. The forced signal is included in predictions, but skill initially originates largely from the initial condition and phasing in modes of climate variability. Observations are employed both to initialize the prediction and then to evaluate the hindcast.

In this section we illustrate the use of observations to align climate model projections with observed variability. The aim is to obtain improved information for predicting the climate of the following seasons and years, and to evaluate how such selected projections merge with the full ensemble as a case example for merging predictions and projections, as recommended in Befort et al. (2020). Sub-selecting ensemble members from a large ensemble that more closely resemble the observed climate state (e.g., Ding et al., 2018; Shin et al., 2020), is an attempt to try to align the internal climate variability of the sub-selected ensemble with the observed climate variability, similar to initialized climate prediction. We therefore also refer to these constraints relative to the observed anomalies as "pseudo-initialisation."

We use the Community Earth System Model (CESM) Large Ensemble (LENS; Kay et al., 2015) of historical climate simulations, extended with the RCP8.5 scenario after 2005. For each year (from 1961 to 2008) we select 10 ensemble members that most closely resemble the observed state of global SST anomaly patterns, as measured by pattern correlations. We then evaluate the skill of the sub-selected constrained ensembles in predicting the observed climate in the following months, years and decade, using anomaly correlation coefficient (ACC; Jolliffe and Stephenson, 2003). We also compare the skill of "un-initialised" (LENS40, the ensemble of all 40 LENS simulations) and "pseudo-initialised" (LENS10, the ensemble of the best 10 ensemble members identified in each year) simulations against "initialised" decadal predictions with the CESM Decadal Prediction Large Ensemble consisting of 40 initialized ensemble members (DPLE40; Yeager et al., 2018). The ocean and sea-ice initial conditions for DPLE40 are taken from an ocean/sea ice reconstruction forced by observation-based atmospheric fields from the Coordinated Ocean-Ice Reference Experiment forcing data, and the atmospheric initial conditions taken from LENS simulations. The anomalies are calculated based on lead-time dependent climatologies.

In this explorational study, the best 10 members of the LENS simulations are selected based on their pattern correlation of global SST anomalies with observed anomalies obtained from the Met Office Hadley Center's sea ice and sea surface temperature data (HadISST; Rayner et al., 2003). These pattern correlations are calculated using the average anomalies of the 5 months prior to 1st November of each year, for consistency of the 'pseudo-initialisation' with the initialized predictions (i.e., DPLE40), which are also initialized on 1st November of each year. We also tested ensemble selection based on the pattern correlation of different time periods (up to 10 years) prior to the 1st November initialization date, to better phase in low-frequency variability, but these tests did not provide clearly improved skill over the 5-months selection.

**Figure 5** compares the skill of different SST indices for the constrained pseudo-initialized ensemble (i.e., LENS10), the full LENS40, and the initialized prediction system. All three ensembles show very high skill ($R > 0.9$) in predicting global mean SSTs on inter-annual to decadal time-scales, primarily due to capturing the warming trend. For the first few months after initialization the constrained LENS10 ensemble shows skill that is comparable to the DPLE40 for global mean SSTs. Larger differences in the prediction skill between the three ensembles are apparent for indices of Pacific (El Nino Southern Oscillation, ENSO, and Interdecadal Pacific Oscillation, IPO) and Atlantic SST variability (Atlantic Multidecadal Variability, AMV). The constrained LENS10 ensemble shows significant skill in predicting the ENSO and IPO indices in the first ∼6–7 months after initialization, with correlations only about ∼0.1 lower compared to the initialized DPLE40 ensemble. LENS10 further shows improved skill over LENS40 during the first 2 forecast years for ENSO and IPO. For the AMV index, LENS10 shows increased skill over LENS40 for up to seven forecast years, while DPLE40 shows high skill ($R > 0.7$) for all forecast times up to one decade.

**Figure 6** shows that the spatial distribution of forecast skill of the LENS10 ensemble is often comparable to that of the DPLE40 for seasonal and annual mean forecasts. The skill of LENS40 is relatively lower than both the pseudo-initialized and the initialized predictions at least for the first few forecast months and the first forecast year. On longer time scales, LENS10 has some added skill in the North Atlantic, but decreased skill in other regions such as parts of the Pacific.

These analyses demonstrate the value of constraining large ensembles of climate simulations according to the phases of observed variability for predictions of the real-world climate. We find added value in comparison to the large (un-constrained) ensemble for up to 7 years in the Atlantic, and up to 2 years in indices of Pacific variability. It illustrates that using observational constraints by targeting modes of climate variability can produce skill that can approach that from initialization in some cases for large scale variability. **Figure 6** also illustrates that this skill is most pronounced in the first season in tropical regions and the Pacific, while added skill in near-term projections over the North

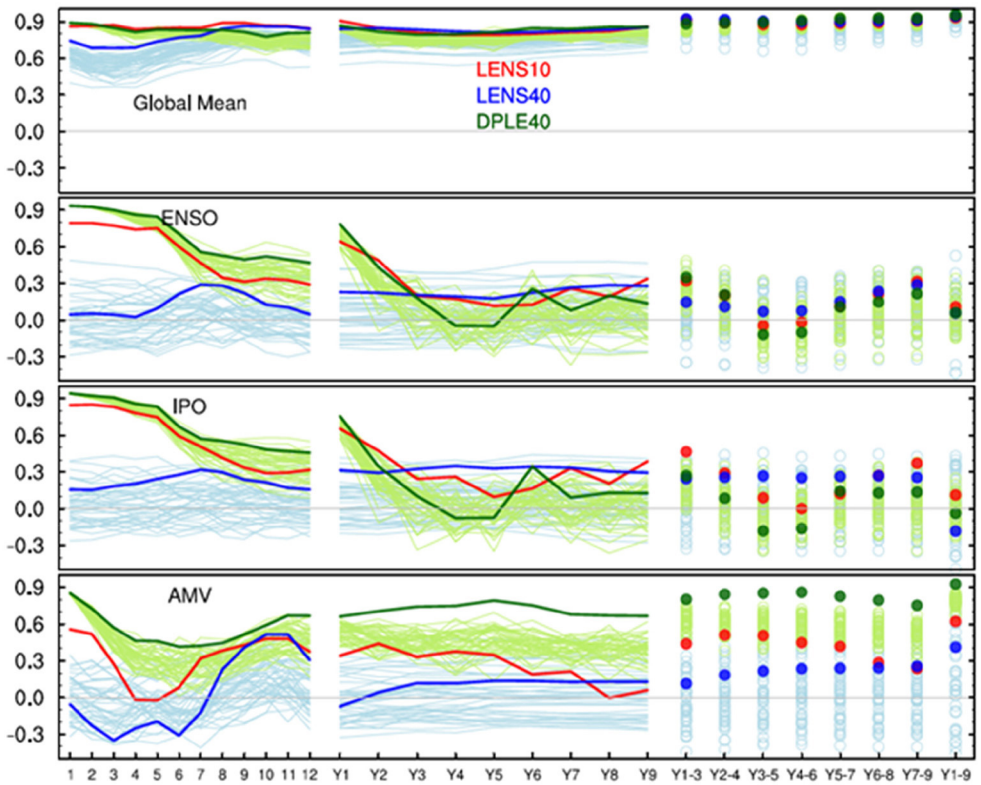



**FIGURE 5 |** Correlation skill for different forecast times: left lines represent skill for first 12 forecast months; center lines represent skill for first nine forecast years; dots represent skill for multi-annual mean forecasts. IPO is calculated as a tripole index (Henley et al., 2015) from SST anomalies, ENSO is based on area-weighted mean of SST anomalies at Nino3.4 region (i.e., 5°S–5°N, 170°W–120°W), and AMV is calculated as a weighted area average SST anomalies for 0–60°N of the North Atlantic ocean with global mean (60°S–60°N) SST removed.

Atlantic is more modest which could be related to weaker-than-observed variability in simulating North Atlantic Oscillation in LENS simulations (Kim et al., 2018). This work bridges between un-initialized and initialized predictions and their evaluation with observations, and illustrates how observational constraints can be used within a large ensemble of a single model to improve performance nearterm. The latter is the goal of initialized predictions, which we focus on in the subsequent sections.

## OBSERVATIONAL CONSTRAINTS ON INITIALIZED PREDICTIONS

In this section, several examples of observational constraints in initialized decadal climate prediction simulations are presented and tested for their potential. These examples include: identifying the level of agreement between model simulations and observations (predictive skill) that arises from the initialization process as well as different forcings of the climate system over time; identifying the predictive skill found in different initialized model systems using the models' inherent characteristics; and identifying the change of predictive skill over time, illustrated using predictions of North Atlantic sea surface temperature (SST). As with observational constraints for projections, we explore if there is potential to improve predictions by weighting or selecting prediction systems and chosen time horizons with the goal to improve performance.

We thus test decadal prediction experiments for observational constraints on the time dimension (exploring the changing importance of various forcings and internal variability over time) and the model dimension (a first step toward weighing initialized climate prediction ensembles). These analyses are both closely related to the approaches used for climate projections discussed above, and they will offer an indication about the degree to which the observational constraints that are applied to projections (see above) represent observed climatic variability. All of these explorational investigations will also pave the way toward eventually combining initialized and non-initialized climate predictions in order to tailor near-term climate prediction to individual users' needs.

The analyses we present rely on the following methods: We consider sea surface temperature (SST) and surface air temperature (SAT) for the period 1960–2014 in our analyses, based on simulations from the CMIP6 archive. We analyze initialized decadal hindcasts from the DCPP project (HC; Boer et al., 2016), as well as non-initialized historical simulations that are driven with reconstructed external forcing (HIST; Eyring

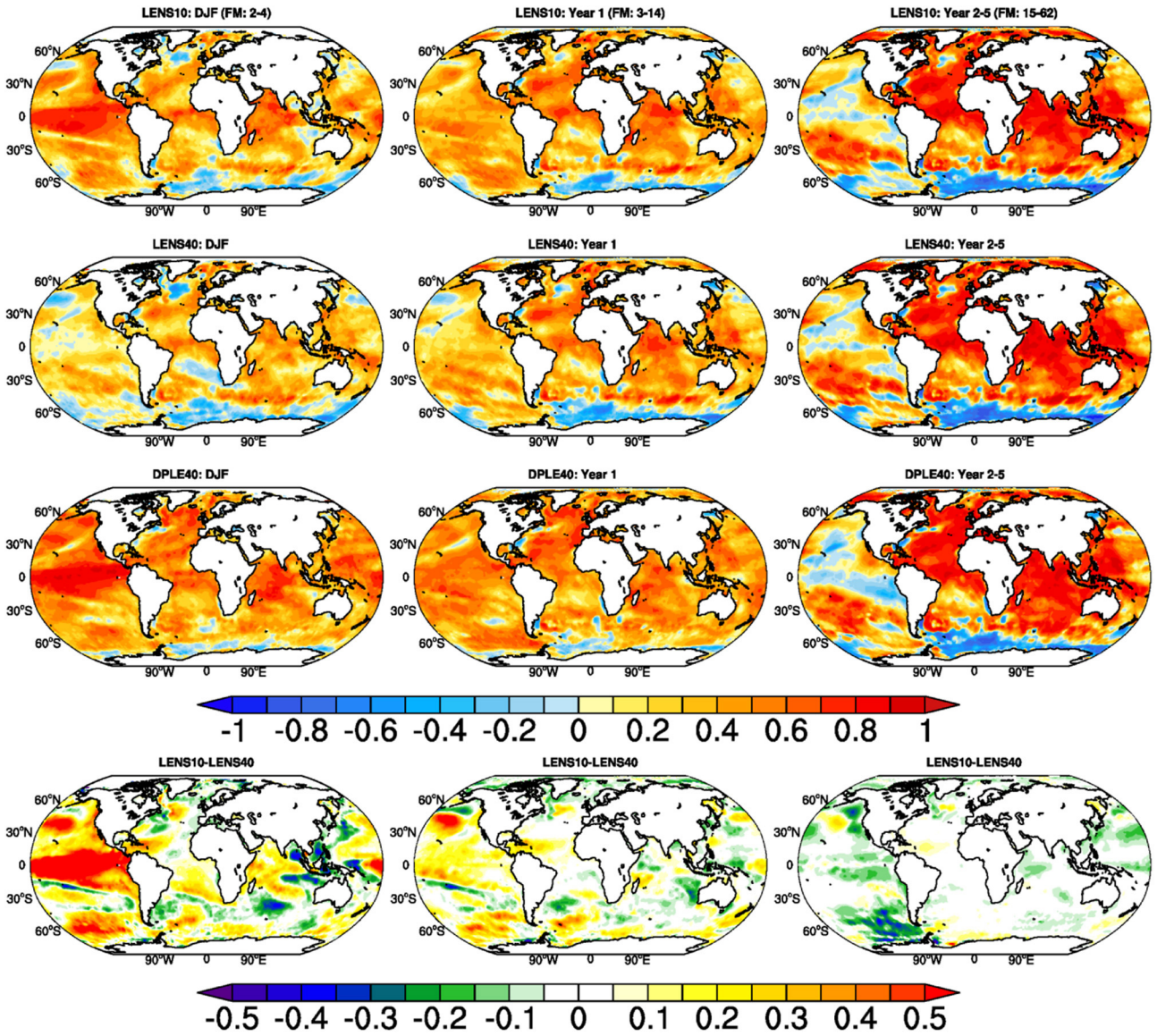



**FIGURE 6 |** Forecast skill, measured as anomaly correlation against HadISST, for LENS10 (top row), LENS40 (second row), DPLE40 (third row), and skill difference between LENS10 and LENS40 (bottom row). Left, center and right columns represent forecast skill of the mean of forecast months 2–4, forecast months 3–14, and forecast months 15–62, respectively.

et al., 2016). For comparison and to constrain predictions, SST from HadISST (Rayner et al., 2003) and SAT from the HadCRUTv4 gridded observational data set (Morice et al., 2012) are used. Agreement between model simulations and observations (prediction or hindcast skill) is quantified here as Pearson correlation between ensemble mean simulations and observations (Anomaly Correlation Coefficient, ACC) and mean squared skill score (MSSS; Smith et al., 2020). ACC tests whether a linear relationship exists between prediction and observation and quantifies the standardized variance explained by it (in its square), whereas MSSS quantifies the absolute difference between simulations and observations. Both ACC and MSSS indicate perfect agreement between prediction and observation at a value of 1 and decreasing agreement at decreasing values. Note that we do not compare these skill scores against a baseline (e.g., the uninitialized historical simulations); this was done and discussed extensively in Borchert et al. (2021). Instead much of our analysis focuses on detrended data to reduce the influence of anthropogenic forcing.

In all cases, anomalies against the mean state over the period 1970–2005 of the respective data set are formed; this equates to a lead time dependent mean bias correction in initialized hindcasts. We also subtract the linear trend from all time series prior to skill calculation to avoid the impact of

linear trends on the results. When focusing on the example of temperature in the subpolar gyre (SPG), we analyze area-weighted average SST in the region 45–60°N, 10–50°W. Surface temperature over Europe is represented by land grid-points in the NEU, CEU and MED SREX regions defined above. We examine summer (JJA) temperature over Europe. We also analyze how prediction skill changes over time (so-called windows of opportunity; Borchert et al., 2019; Christensen et al., 2020; Mariotti et al., 2020) to attribute changes in skill to specific climatic phases.

## Sources of Decadal Prediction Skill for North Atlantic SST

A recent paper detailed the influence of external forcing and internal variability on North Atlantic subpolar gyre region (SPG) SST variations and predictions (Borchert et al., 2021). The authors found North Atlantic SST to be significantly better predicted by CMIP6 models than by CMIP5 models, both in non-initialized historical simulations and initialized hindcasts. These findings indicated a larger role for forcing in influencing predictions of North Atlantic SST than previously thought. This work further showed that at times of strong forcing, predictions and projections of North Atlantic SST with CMIP6 multi-model averages exhibit high skill for predicting North Atlantic SST. Natural forcing, particularly major volcanic eruptions (Swingedouw et al., 2013; Hermanson et al., 2020; Borchert et al., 2021), plays a prominent role in influencing skill during the historical period, notably due to their impact on decadal variations of the oceanic circulation (e.g., Swingedouw et al., 2015). In the absence of strong forcing trends, initialization is needed to generate skill in decadal predictions of North Atlantic SST (Borchert et al., 2021; their Figure 2). Analyzing the contributions of forcing and internal variability to climate variations and their prediction is therefore an important step toward understanding observational constraints on initialized climate predictions. By examining the dominant factors governing the skill of predictions in the past, conclusions may be drawn for predictions of the future as well. This also illustrates that metrics for initialized model performance based on evaluating hindcasts are influenced not only by how well the method reproduces observed variability, but also by the response to forcing. Hence sources of skill in predictions (initialization) and projections (forcing) overlap, which is important to consider when comparing the role of observational constraints in both. This also needs to be considered when aiming to merge predictions and projections, which are driven by forcing only, and generally do not include volcanic forcing.

## Toward Performance Based Weighting for Initialized Predictions

Approaches discussed above, which identify the origin of skill among different external forcings and variability, could be seen as an observational constraint on predictions (section Sources of Decadal Prediction Skill for North Atlantic SST), constraining them based on the emerging importance of forcing and internal variability over time. This makes that technique similar to that used in ASK constraints (which, however, focuses on a different timescale). We now consider an approach similar to model-related weights used in the ClimWIP method. Instead of multi-model means, we here assess the seven individual CMIP6 DCPP decadal prediction systems (**Table 2**) with the aim of linking the skill in model systems to their inherent properties. We focus this analysis on North Atlantic subpolar gyre SST due to its high predictability (e.g., Marotzke et al., 2016; Brune and Baehr, 2020; Borchert et al., 2021) as well as its previously demonstrated ties to European summer SAT (Gastineau and Frankignoul, 2015; Mecking et al., 2019).

Initialized predictions from CMIP6 show broad agreement on ACC skill for SPG SST, with high initial skill and skill degradation over time (**Figure 7A**). The only prominent outlier to this is the CanESM5 model, which displays a strong initialization shock until approximately lead year 7 due to issues in the North Atlantic region with the direct initialization from the ORAS5 ocean reanalysis (Sospedra-Alfonso and Boer, 2020; Tietsche et al., 2020). For this reason, we will discuss CanESM5 as a special case whenever appropriate. The other six models generally agree on high skill in the initial years, which degrades over lead time (**Figure 7**), showing some degree of spread in ACC that could be linked to model or prediction system properties. This spread is found for both ACC (**Figure 7A**) and MSSS (**Figure 7B**), indicating the robustness of this result. Forming multi-model means results in comparatively high skill compared to individual models, evident in placement of the multi-model mean (black) at the upper half of ACC and MSSS skills. This is likely related to an improved filtering of the predictable signal from the noise (e.g., Smith et al., 2020), and the compensation of model errors between the systems; suggesting potential for further increased skill if using a weighted rather than simple average multimodel means. Skill degradation occurs at different rates in the different model systems, representing a possible angle at which to try and explain the skill differences.

### Constraint Based on SPG Stratification

Sgubin et al. (2017) showed that the representation of stratification over the recent period in the upper 2,000 m of the North Atlantic subpolar gyre in different models is a promising constraint on climate projections, impacting among other things the likelihood with which a sudden AMOC collapse is projected to happen in the future. Ocean stratification impacts North Atlantic climate variability not only on multidecadal time scales, but also locally on the (sub-)decadal time scale. It appears therefore appropriate to explore stratification as an observational constraint on the model dimension in predictions, and test whether models that show comparatively realistic SPG stratification also show higher SPG SST prediction skill and vice versa. To this end, we calculate a stratification indicator as in Sgubin et al. (2017) by integrating SPG density from the surface to 2,000 m depth for the period 1985–2014 in the different CMIP6 HIST models and EN4 reanalysis (Ingleby and Huddleston, 2007). We then calculate the root-mean square

**TABLE 2 |** Models used in the analysis presented in Section Observational Constraints on Initialized Predictions, based on availability at the time of analysis.

| Modeling Center | Model | Ensemble size | |
|---|---|---|---|
| | | Historical | Decadal Hindcasts |
| CCCma, Canada | CanESM5, (Sospedra-Alfonso and Boer, 2020) | 20 | 10 |
| IPSL, France | IPSL-CM6A-LR, (Boucher et al., 2020) | 30 | 10 |
| JAMSTEC, Japan | MIROC6, (Kataoka et al., 2020) | 10 | 10 |
| MOHC, UK | HadGEM3-GC31-MM, (Knight et al., 2014) | 4 | 10 |
| MPI-M, Germany | MPI-ESM1.2-HR, (Pohlmann et al., 2019) | 10 | 10 |
| NCAR, USA | CESM1.1-CAM5, (Yeager et al., 2018) | | 20 |
| | CESM2, (Danabasoglu et al., 2020) | 10 | |
| NCC, Norway | NorCPM1, (Counillon et al., 2021) | 30 | 10 |
| Total models (members) | 7 (114) | 7 (80) | |

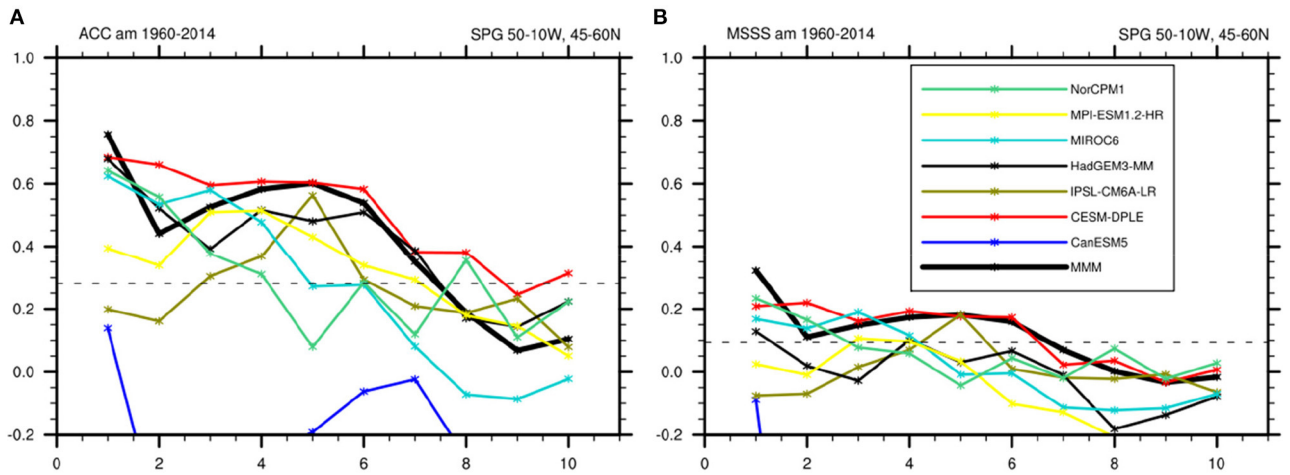



**FIGURE 7 | (A)** Anomaly correlation coefficient (ACC) for SPG SST in seven initialized CMIP6 model systems (y-axis) against lead time in years (x-axis). **(B)** as **(A)**, but showing mean squared skill score (MSSS) on the y-axis. Colors indicate the individual models, the thick black line the multi-model mean. The horizontal dashed line indicates the 95% confidence level compared to climatology based on a two-sided *t*-test.

difference between modeled and observed stratification (as in[2]). This index is then examined for a possible linear relationship to SPG SST prediction skill. Note that comparing stratification in the historical simulations to initialized predictions is not necessarily straight-forward as initialization may change the stratification.

For short lead times of up to 5 years, we find a strong, negative and linear relationship between SPG SST prediction skill in terms of ACC and MSSS in the seven prediction systems analyzed here, and mean SPG stratification bias in the corresponding historical simulation (**Figures 8A–C**). Models that simulate a more realistic SPG stratification show higher SST prediction skill than those that simulate less realistic stratification. At lead times longer than 6 years, this linear relationship is not as strong (**Figure 8C**), which also diminishes the skill-stratification relationship for the full 1–10 years lead time range (**Figure 8A**). Note that due to the initialization issue in CanESM5 discussed above, that model does not behave in line with the other models at short lead times.

These findings show that models that simulate a more realistic SPG stratification tend to predict SPG surface temperature for up to 5 years into the future more skillfully than models with a less realistic SPG stratification. The fact that mean stratification in the historical simulation seems to have an influence on the skill of initialized predictions at short lead time (although initialization has modified mean stratification) suggests that the modification of stratification through initialization is weak. Moreover, this finding hints at a reduced initialization-related shock in some models: models that simulate realistic SPG stratification without initialization experience less shock through initialization, while the shock itself may hamper the skill at long lead time. The role of physical realism of climatology vs. initialization shock should be explored further when analyzing performance of and constraints on prediction systems. While inspiring hope that SPG mean stratification state might be an accurate indicator of SPG SST hindcast skill, this result is based on a regression over 6 data points and therefore lacks robustness.

[2]Swingedouw, D., Bily, A., Esquerdo, C., Borchert, L. F., Sgubin, G., Mignot, M., et al. On the risk of abrupt changes in the North Atlantic subpolar gyre in CMIP6 models. Ann. NY Acad. Sci. (in review).

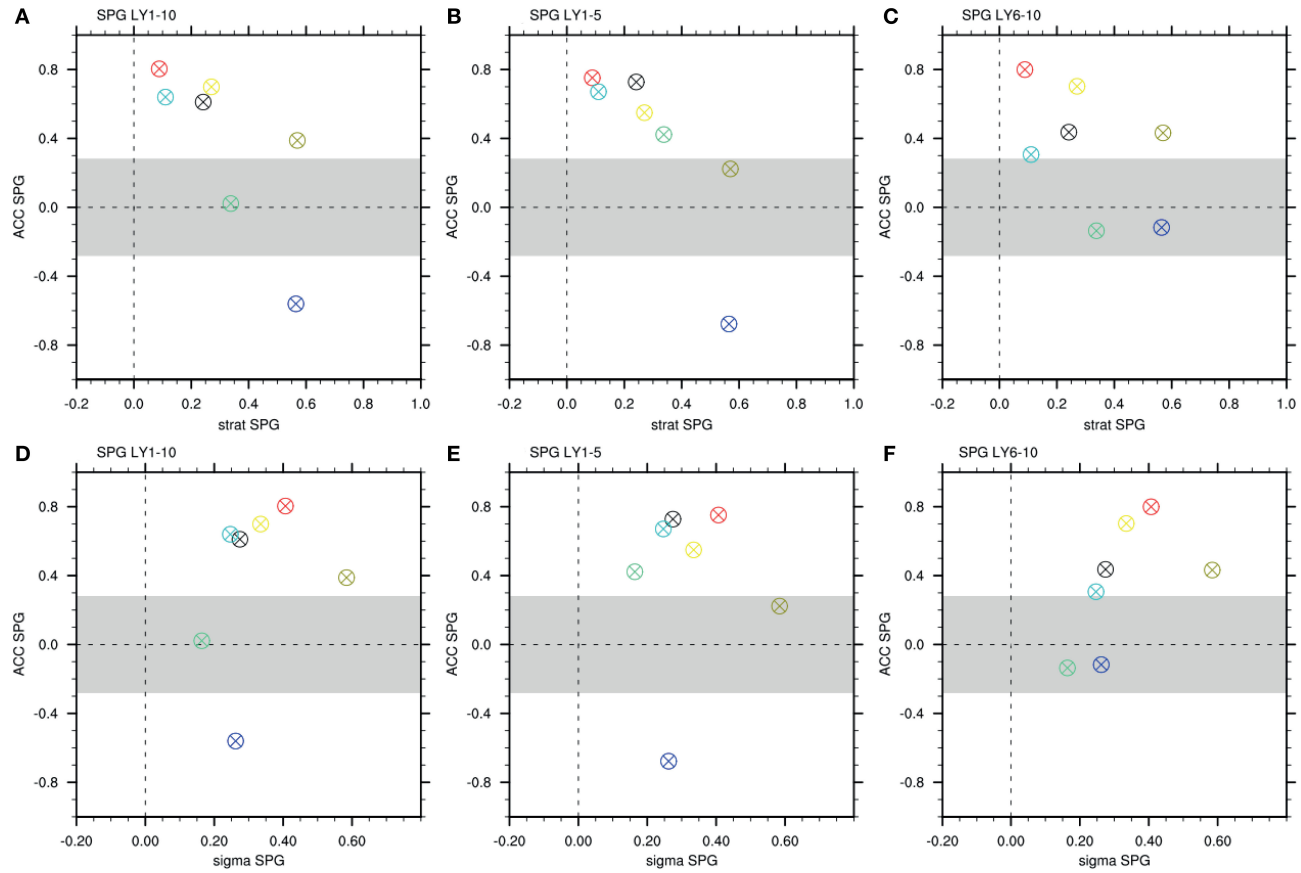



**FIGURE 8 |** ACC for SPG SST in seven initialized CMIP6 model systems at lead years **(A)** 1–10, **(B)** 1–5, and **(C)** 6–10 against RMSE between observed (vertical dashed line) and simulated SPG stratification in the period 1960–2014 from the corresponding historical simulations (x-axis). **(D–F)** as **(A–C)**, but showing the standard deviation of SPG SST in the first 500 years of the respective model's piControl simulation on the x-axis. Colors indicate the individual models: blue = CanESM5, red = CESM-DPLE, brown = IPSL-CM6A-LR, black = HadGEM3-GC31-MM, cyan = MIROC6, yellow = MPI-ESM1.2-HR, green = NorCPM1. The gray area indicates insignificant prediction skill at the 5% significance level based on 1,000 bootstraps.

## Constraint Based on Internal Climate Variability

The amount of climate variability inherently produced by the models might also relate to the skill of prediction systems. The assumption is that models that produce pronounced SPG SST variability by themselves reproduce strong observed decadal SPG SST changes more accurately than those that do not. This is a reasonable assumption since previous studies have shown that North Atlantic SST variability appears to be underestimated in climate models (Murphy et al., 2017; Kim et al., 2018). We represent this variability by standard deviation of SPG SST over 500 years in the pre-industrial control (piC) simulations of the seven different CMIP6 models (sigma SPG). Decadal SPG SST hindcast skill shows some increase with sigma SPG in the respective control simulations (**Figures 8D–F**), particularly at long lead time (**Figure 8F**). A possible cause for increased skill in models with higher sigma is a linear relationship of sigma SPG to the time lag at which autocorrelation becomes insignificant in the piControl simulations (not shown): higher variability implies longer decorrelation time scales which might indeed lead to longer predictability. Another hypothesis could be more robust variability produced by models with higher sigma, less perturbed by noise, associated with higher levels of variability. More work is needed to decipher the exact cause of this effect. The linear correlation between sigma SPG and SPG hindcast skill, however, is not robust across lead times. At long lead time of 6–10 years, the CanESM and IPSL models are outliers that potentially inhibit significant linear regression, due to the known initialization issue in CanESM (see above) and possible effect of the weak initialization in IPSL-CM6A (Estella-Perez et al., 2020). Again, this analysis is limited by the small number of models for which decadal hindcast simulations are currently available. Adding more models to this analysis could point toward other conclusions, or strengthen the results presented here. Additionally, extending this analysis to other regions in the piControl simulations (as in Menary and Hermanson, 2018) would provide valuable insights into the way that the representation of underlying dynamics in different models preconditions their skill for prediction of the SPG SST.

Other possible discriminant factors for decadal SPG SST prediction skill, such as equilibrium climate sensitivity (ECS), model initialization strategy (e.g., Smith et al., 2013) and resolution of the ocean model in the respective model

have been investigated, but with no firm conclusions at this point.

### Multi-Model Exploration of Windows of Opportunity

Analyses presented above analyze prediction skill for the period 1960–2014 as a whole, i.e., operate under the assumption that predictive skill is constant over time. In the North Atlantic region, however, the skill of decadal predictions was previously shown to change over time, forming windows of opportunity, with possible implications for the constraints discussed above. Here, we examine the model-dependency of windows of opportunity for decadal SPG SST prediction skill as a call for caution when applying observational constraints to predictions and projections.

Windows of opportunity for annual mean North Atlantic SPG SST across all models are presented for a lead time average over years 1–10 in **Figure 9**. These lead time averages bring out more skill due to temporal filtering of the time series, which is achieved by averaging all 10 predicted years that are predicted from a certain start year. The analysis of windows of opportunity enables an identification of model differences of prediction skill across time.

We find that there is general agreement among models on the approximate timing of windows of opportunity for SPG skill, where the general level of skill depends on the models' mean performance (**Figures 7**, **8**). Skill is high early and late in the analyzed time horizon, with a "skill hole" around the 1970s and 80s. These windows have been noted in earlier studies (e.g., Christensen et al., 2020), and interpreted by Borchert et al. (2018) to result from changes in oceanic heat transport. Since all examined models agree on this timing, it could be argued that windows of opportunity arise from the predictability of the climate system rather than the performance of individual climate models over time. This limits the applicability of observational constraints at times of low skill. Windows of opportunity should therefore be taken into account in observational constraints. While times of low skill appear to coincide with times of low trends for North Atlantic SST (e.g., Borchert et al., 2019; 2021), they are possibly caused by modes of climate variability that are mis-represented in the models, or times of small forced trends. As windows of opportunity found here are generally in line with those found for uninitialized historical simulations in Borchert et al. (2021), they appear to be at least partly a result of changes in forcing, e.g., natural forcing. This conclusion likely holds for observational constraints in predictions and projections alike.

While this assessment remains mainly qualitative, it highlights the potential for better estimating sources of prediction skill when combining observation-based skill metrics in time as well as between models. The presented analysis should thus be extended to include more models, and studying the underlying physics at work in-depth to produce actionable predictions for society.

## Potential for Constraining Decadal Prediction Skill of European Summer Temperature

Finally, we examine prediction skill for European surface air temperature, which is known to be difficult to predict due to small signal-to-noise ratio (e.g., Hanlon et al., 2013; Wu et al., 2019; Smith et al., 2020), yet one of our ultimate goals in terms of prediction. We contrast the skill of the different prediction systems currently available, and its change depending on time horizon, as a first step toward model weighing or selection. As in Section Constraining Projections, we use the SREX regions as a first step of homogenizing analysis methods between prediction and projection research.

We analyze the decadal prediction skill for SAT in SREX regions in CMIP6 models during summer (JJA) (**Figure 10**) after subtracting the linear trend. Summer temperature in SREX regions shows generally low hindcast skill for individual lead years. We find some differences between the different regions, with a tendency for higher skill toward the South. Forming the multi-model mean as a simple first step toward improving skill due to improved filtering of the signal from the noise (see above) does in fact lead to comparatively higher skill, but does not consistently elevate SREX summer temperature skill to significance at the 95% level (**Figure 10**). Because SPG SST was shown to impact European surface temperature during boreal summer (e.g., Gastineau and Frankignoul, 2015; Mecking et al., 2019) and hindcast skill for SREX SAT shows similar inter-model spread as for SPG SST, attempts to connect hindcast skill in individual models to inherent properties of the model (as above) is promising. The generally low (and mostly insignificant) level of skill for all models in the SREX regions indicates, however, that discriminatory features of skill between models would not enable the identification of skillful models without further treatment. Hence, further efforts such as an analysis of windows of opportunity (see Multi-Model Exploration of Windows of Opportunity) might reveal times of high prediction skill in the SREX regions in the future, indicating boundary conditions that benefit prediction skill.

## SPG Prediction Skill as Model Weights for Projections Over Europe

Finally, we experimented with using lead-year correlations from Subpolar gyre predictions, instead of the ClimWIP performance weights, for constructing weighted multi-model mean fingerprints of the response to external forcing, and estimating their contribution to observed change using the ASK method (Section **Contrasting and Combining Constraints From Different Methodologies**). While the estimate of the GHG signal against other forcings remained fairly robust, we found that when using those skill weighted fingerprints, the ASK method's ability to distinguish between contributions to JJA change from different forcings degenerated sharply compared to what is shown in **Figure 4**, top row (not shown). This is not very surprising as it is not clear that a model's prediction skill over the SPG would necessarily relate to its skill in simulating the response to forcing over European land, but an approach like this might prove more promising in other regions and seasons.

When considering combinations of performance-based weights derived from initialized model skill and long-term projection skill, different properties of models might come

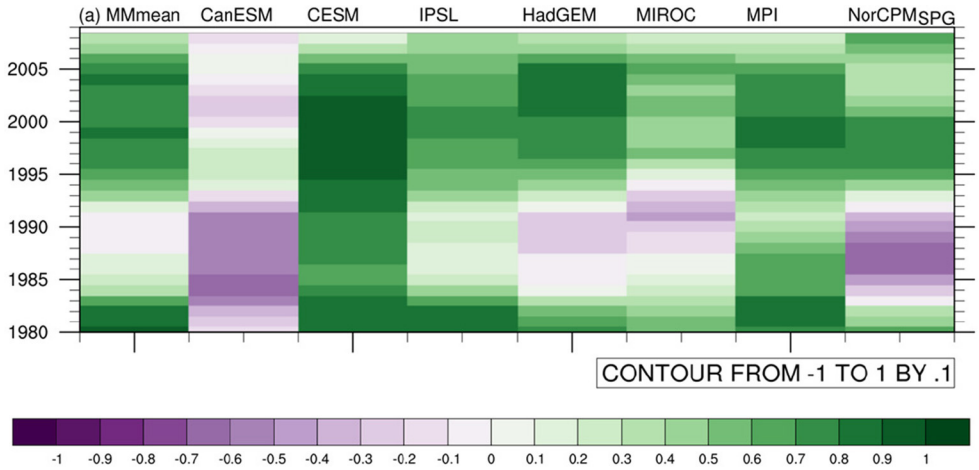



**FIGURE 9 |** Decadal ACC hindcast skill for detrended annual mean North Atlantic subpolar gyre SST for a sliding 20-years long window (the end year of which is shown on the y-axis) across multi-model mean and individual model means (x-axis). Colors indicate ACC (low to high = purple to green).

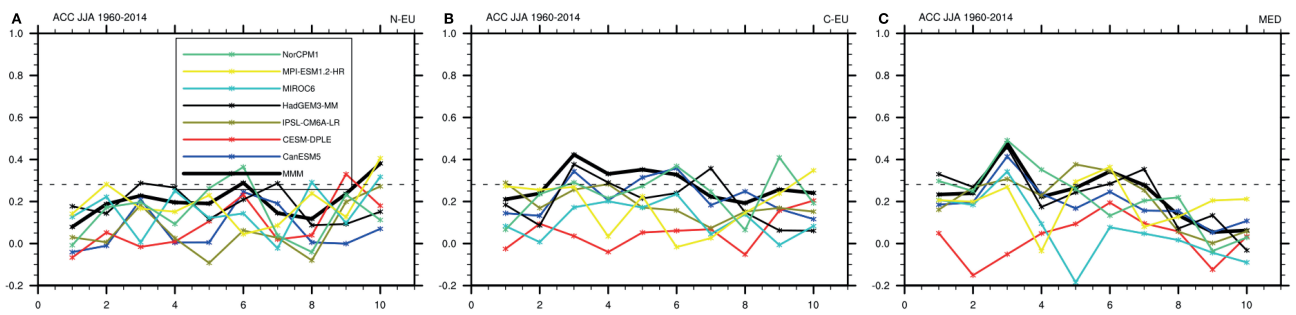



**FIGURE 10 |** ACC of detrended summer (JJA) SAT in the SREX regions **(A)** Northern Europe, **(B)** Central Europe, and **(C)** Mediterranean (y-axis) against lead time in years (x-axis). The horizontal dashed line represents the 95% confidence level compared to climatological prediction based on a two-sided *t*-test.

into play. For example, a model that shows strong response to forcings will show a higher signal-to-noise ratio in the response, which can drive up ACC skills influenced by forcing (although, potentially, at a cost of lower MSSS). Such a model may also show higher signal-to-noise ratio in fingerprints used for attribution which also improves the constraint from the forced response. On the other hand, a model that shows too strong trends (which may improve signal-to-noise ratios) would get penalized by ClimWIP for over-simulating trends, and hence rightly be identified as less reliable for predictions. Moreover, the skill of initialized predictions is heavily dependent on lead time and time-averaging windows, which requires careful consideration when applied as a weighting scheme to climate projections. This illustrates that different approaches to use observational constraints can pull a prediction system into different directions. It would be interesting to investigate links between the reliability of projections and the skill of predictions. Presently, the limited overlap between models providing individually forced simulations necessary for ASK, and being used in initialized predictions makes it difficult to pursue this further.

## CONCLUSIONS AND LESSONS LEARNT

Observational constraints for projections may both originate from weighting schemes that weight according to performance (Knutti et al., 2017; Sanderson et al., 2017; Lorenz et al., 2018; Brunner et al., 2020b), as well as from a binary decision which models are within an observational constraint and which outside (model selection methods; see e.g., discussion in Tokarska et al., 2020b; drawing on the ASK method; and Nijsse et al., 2020). Constraining projections based on the agreement with the observed climate state can phase in modes of climate variability and add skill, similar to initialization in decadal predictions. Retrospective initialized predictions (hindcasts) are evaluated against observations using skill scores that may also provide input for performance-based model weighting. This study illustrates that the prediction skill may vary strongly with lead time, climate model, in space, with climate state and over time (Borchert et al., 2019; Christensen et al., 2020; Yeager, 2020), suggesting a careful selection of cases to choose. Similarly, performance weighting varies depending on whether trends are included in

the analysis (i.e., if weights include evaluation of the forced response), or whetherthe weights are limited to performance in simulating mean climate. Constraints on future projections from attributed greenhouse warming (ASK, see Section Constraining Projections) show smaller uncertainties for targets of predictions where the signal-to-noise ratio is high compared to noisier variables, and correct implicitly for too strong or too weak a forced response compared to observations. In our view, these factors that control the skill of initialized predictions as well as the strength of observational constraints on projections need to be accounted for in upcoming attempts to combine projections and predictions, and this might complicate the seamless application of observational constraints in predictions and projections.

Overall, observational constraints on projections show substantial promise to correct for biases in the model ensemble of opportunity as illustrated from the UKCP18 example (**Figure 1**) as well as for other methods (**Figure 2**). However, several questions arise: If applied to similar model data, will different metrics for model performance favor similar traits and hence similar models? This is important when attempting to merge predictions and projections: If the choice of timescale strongly influences the model weights, or leads to selection of different models, the merged predictions might be inconsistent over time: in cases where the climate sensitivity deviates between up-weighted or selected models for projections and predictions, the merged predictions may have a discontinuous underlying climate change signal. Where high performance models show, possibly by chance, different variability dependends on choice of weights from projections or predictions, the merged predictions may also show different variability over time. It remains to be explored how detrimental a signal-strength discontinuity might be, as the signal during the initialized time horizon is still small compared to noise on all but global scales (Smith et al., 2020).

Finally, we found that observational constraints show promise and should be included in predictions and projections. This is a lesson also learned from UKCP18, where observational constraints have been used successfully in a major climate projections product, and thus influence planning and adaptation decisions based on the application. **Figures 1**, **2** both illustrate observational constraints and show potential to correct for model biases, such as too strong (or weak) response to forcing. However, we need to address the question to what extent observational constraints are reliable, even if they are intuitively well-justified (Weisheimer and Palmer, 2014). This is more straightforward for predictions than projections. For projections, such a 'perfect' or 'imperfect' model evaluation is not a trivial task, as it requires retuning the observational constraint for every model whose future performance is predicted in order to arrive at robust statistics. However, it provides powerful evaluation of observational constraints (see e.g., Schurer et al., 2018 or Brunner et al., 2020b among other recent examples).

Our analysis of initialized prediction simulations indicates that there are several dimensions (such as the change of skill over time) that add complexity when aiming to use observational constraints on climate projections combined with predictions. For an optimized constraining of merged, seamless climate prediction for the next 40 years, these dimensions need to be considered. Furthermore, initialization can introduce shocks that lead the initialized climate simulations into a different climate state to the uninitialised simulations (Bilbao et al., 2021), which can hamper attempts to merge predictions and projections, or to apply common constraints.

**Questions** that need to be considered when evaluating observational constraints and use them across prediction and projection timescales include:

- Is there potential to improve performance of prediction and projection systems by combining observed constraints? First results indicate that improvements may be possible, at least over projection timescales, particularly if combining climatological constraints and those based on the forced signal. This is unsurprising given that the one-model-one-vote system may well be suboptimal. However, any possible improvements need to be carefully evaluated, including in a perfect model setting. And the prediction-projection merging still requires some work.
- To what extent do weighting or model selection criteria used across projections and predictions favor similar model traits and to what extent are they uncorrelated or pull in different directions? Particularly:
  - do any of the weighting schemes preferentially select or highly weight models with stronger or weaker response to forcing? This can arise if correlation skills, e.g., are influenced by response to external forcing, such as volcanism (Borchert et al., 2021). Are there any other factors where different weighting schemes select differently?
  - relatedly, do any of the schemes which also draw on variability reward or penalize climate models with high or low internal climate variability, i.e., with low or high signal-to-noise ratio for external forcing or predictable signals? High variability may degrade the performance of the ASK system, but on the other hand appears to favor good performance over the subpolar gyre (**Figure 8**).
  - which dimensions beyond the model dimension show promise in the application of observational constraints? Is there a role for lead-time dependent skill, or skill depending on climatological conditions (see Section Observational Constraints on Initialized Predictions)?

**Overall, we recommend that:**

- Assumptions and mechanisms behind constraining methods need to be clear and transparent (this point was also made in Brunner et al., 2020a, but only for projections).
- It needs to be clear what main model characteristics explain a constraint. Skill/high model accuracy can originate from different model properties, e.g., from strong climatological performance of models, from the representation of physical mechanisms in specific models, from realistic representation of internal climate variability or from response to external forcing, or combinations of all.
- This is particularly important when considering observational constraints across the prediction/projection boundary. Furthermore, it needs to be considered that skill may vary over time, and short hindcast periods can be misleading. On

the other hand, we can increase performance by drawing on more lines of evidence but need to be wary of over fitting
- When evaluating/combining methods (for projections and across projections and predictions) we need a common and consistent test protocol for skill and reliance to ensure performance. This test protocol needs to consider all dimensions on which the skill of climate simulations varies. Combining different observational constraints may increase performance by drawing on more lines of evidence, but need to be wary of over fitting.
- It is important that consistent model versions and releases are used across the prediction/projection timeline to enable physically consistent merging for seamless prediction across the next 40 years. Carefully-designed simulations are thus required, using the same model versions for predictions and projections. Longer prediction lead times would be helpful to improve on some of the above robustness issues for bringing predictions and predictions closer together.

Last, but not least, it is important to also consider what our findings mean for users of climate information. Firstly, they illustrate that there may be a much greater information content in model prediction and projection ensembles than is first apparent when considering the raw ensemble alone to look at the spread in future conditions. Indeed, without applying constraints users may get a rather skewed view of the spread of future climate simulations. Secondly, the results show that whilst there are methods to extract this extra information they are currently affected by multiple choices around choice of constraint and how they are applied. It is recommended that studies include more focus on showing the effect of applying particular constraints, for instance using out of sample testing or the effect of windows of opportunity on constraining projections. This will help users avoid selecting approaches that provide over-confidence. Thirdly, although at an earlier stage in development, there is the growing potential for merging predictions and projections over their respective time-scales. Consideration of observational constraints is a vital part of this merging.

## DATA AVAILABILITY STATEMENT

The datasets analyzed for this study can be found in the CEDA archive, with derived products available on request.

## AUTHOR CONTRIBUTIONS

JMM and GRH drafted the first subsection of the section on constraints for projections, the next two sections are the result of a collaboration with LB, APB, and GCH, while the final subsection toward seamless predictions was drafted by RM, MGD, and FJD-R. LFB, JM, and DS drafted the section on initialized decadal predictions. GCH led the storyline across the manuscript and drafted the conclusions. All authors contributed to writing, and provided figures and text.

## FUNDING

All authors were supported by the EUCP project funded by the European Commission's Horizon 2020 programme, Grant Agreement number 776613. JM was also supported by the french ANR MOPGA project ARCHANGE and by the EU-H2020 Blue Action (GA 727852) and 4C projects (GA 821003). MGD also received funding by the Spanish Ministry for the Economy, Industry and Competitiveness grant reference RYC-2017-22964.

## REFERENCES

Allen, M. R., Stott, P. A., Mitchell, J. F., Schnur, R., and Delworth, T. L. (2000). Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature* 407, 617–620. doi: 10.1038/35036559

Befort, D. J., O'Reilly, C. H., and Weisheimer, A. (2020). Constraining projections using decadal predictions. *Geophys. Res. Lett*. 47:e2020GL087900. doi: 10.1029/2020GL087900

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, A. P., et al. (2021). Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth, Earth System *Dynamics*. 12, 173–196. doi: 10.5194/esd-12-173-2021

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., et al. (2013). Chapter 10 - Detection and attribution of climate change: From global to regional. In: *Climate Change 2013: The Physical Science Basis. IPCC Working Group I Contribution to AR5*. Cambridge: Cambridge University Press.

Bo,é, J., and Terray, L. (2015). Can metric-based approaches really improve multi-model climate projections? The case of summer temperature change in France. *Climate Dynamics*, 45(7–8), pp.1913–1928. doi: 10.1007/s00382-014-2445-5

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., et al. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci. Model Dev*. 9, 3751–3777. doi: 10.5194/gmd-9-3751-2016

Booth, B. B. B., Harris, G. R., Murphy, J. M., House, J. I., Jones, C. D., Sexton, D. M. H., et al. (2017). Narrowing the range of future climate projections using historical observations of atmospheric $CO_2$. *J. Clim*. 30, 3039–3053. doi: 10.1175/jcli-d-16-0178.1

Borchert, L. F., Düsterhus, A., Brune, S., Müller, W. A., and Baehr, J. (2019). Forecast-oriented assessment of decadal hindcast skill for North Atlantic SST. *Geophys. Res. Lett*. 46, 11444–11454. doi: 10.1029/2019GL084758

Borchert, L. F., Menary, M. B., Swingedouw, D., Sgubin, G., Hermanson, L., and Mignot, J. (2021). Improved decadal predictions of North Atlantic subpolar gyre SST in CMIP6. *Geophys. Res. Lett*. 48:e2020GL091307. doi: 10.1029/2020GL091307

Borchert, L. F., Müller, W., A., and Baehr, J. (2018). Atlantic ocean heat transport influences interannual-to-decadal surface temperature predictability in the north atlantic region. *Climate J*. 31, 6763–6782. doi: 10.1175/JCLI-D-17-0734.1

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., and Bastrikov, V., et al. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *J. Adv. Model. Earth Syst*. 12:e2019MS002010. doi: 10.1029/2019MS002010

Braganza, K., Karoly, D. J., Hirst, A. C., Mann, M. E., Stott, P., Stouffer, R. J., et al. (2003). Simple indices of global climate variability and change: part I — variability and correlation structure. *Clim. Dyn*. 20, 491–502. doi: 10.1007/s00382-002-0286-0

Bretherton, C. S., and Caldwell, P. M. (2020). Combining Emergent Constraints for Climate Sensitivity, Journal of *Climate*. 33, 7413–7430.

Brient, F. (2020). Reducing uncertainties in climate projections with emergent constraints: concepts, examples and prospects. *Adv. Atmos. Sci.* 37, 1–15. doi: 10.1007/s00376-019-9140-8

Brune, S., and Baehr, J. (2020). Preserving the coupled atmosphere–ocean feedback in initializations of decadal climate predictions. *WIREs Clim Change*. 11:e637. doi: 10.1002/wcc.637

Brunner, L., Lorenz, R., Zumwald, M., and Knutti, R. (2019). Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environ. Res. Lett*. 14:124010. doi: 10.1088/1748-9326/ab492f

Brunner, L., McSweeney, C., Ballinger, A. P., Befort, D. J., Benassi, M., Booth, B., et al. (2020a). Comparing methods to constrain future european climate projections using a consistent framework. *J. Climate*. 33, 8671–8692. doi: 10.1175/jcli-d-19-0953.1

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., and Knutti, R. (2020b). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth Syst. Dynam. Discuss*. 11, 995–1012. doi: 10.5194/esd-2020-23

Caldwell, P. M., Zelinka, M. D., and Klein, S. A. (2018). Evaluating emergent constraints on equilibrium climate sensitivity. *J. Climate* 31, 3921–3942. doi: 10.1175/JCLI-D-17-0631.1

Christensen, H. M., Berner, J., and Yeager, S. (2020). The value of initialization on decadal timescales: state-dependent predictability in the CESM decadal prediction large ensemble. *Climate J.* 33, 7353–7370. doi: 10.1175/JCLI-D-19-0571.1

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P. et al. (2013). "Long-term climate change: projections, commitments and irreversibility," in *Climate Change: 2013 Physical Science Basis*, ed T. Stocker et al. (Cambridge: Cambridge University Press), 1029–1136. doi: 10.1017/CBO9781107415324.024

Counillon, F., Keenlyside, N., Toniazzo, T., Koseki, S., Demissie, T., Bethke, I., et al. (2021). Relating model bias and prediction skill in the equatorial Atlantic. *Clim. Dyn.* 56, 2617–2630. doi: 10.1007/s00382-020-05605-8

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The community earth system model version 2 (CESM2). *J. Adv. Model. Earth Syst*. 12:e2019MS001916. doi: 10.1029/2019MS001916

DelSole, T., Trenary, L., Yan, X., et al. (2019). Confidence intervals in optimal fingerprinting. *Clim. Dyn.* 52, 4111–4126. doi: 10.1007/s00382-018-4356-3

Ding, H., Newman, M., Alexander, M. A., and Wittenberg, A. T. (2018). Skillful climate forecasts of the tropical Indo-Pacific Ocean using model-analogs. *J. Climate*. 31, 5437–5459. doi: 10.1175/JCLI-D-17-0661.1

Donat, M., Pitman, A. J., and Angélil, O. (2018). Understanding and reducing future uncertainty in midlatitude daily heat extremes via land surface feedback constraints. *Geophys. Res. Lett*. 45, 10627–10636. doi: 10.1029/2018GL079128

Estella-Perez, V., Mignot, J., Guilyardi, E., Swingedouw, D., and Reverdin, G. (2020). Advances in reconstructing the AMOC using sea surface observations of salinity. *Clim. Dyn*. 55, 975–992. doi: 10.1007/s00382-020-05304-4

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., et al. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* 9, 1937–1958. doi: 10.5194/gmd-9-1937-2016

Forster, P. M., Maycock, A. C., McKenna, C. M., and Smith, C. J. (2020). Latest climate models confirm need for urgent mitigation. *Nat. Clim. Change* 10, 7–10. doi: 10.1038/s41558-019-0660-0

Gastineau, G., and Frankignoul, C. (2015). Influence of the north atlantic sst variability on the atmospheric circulation during the twentieth century. *Clim. J.* 28, 1396–1416. doi: 10.1175/JCLI-D-14-00424.1

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Clim. J.* 30, 5419–5454. doi: 10.1175/JCLI-D-16- 0758.1

Gidden, M., Riahi, K., Smith, S., Fujimori, S., Luderer, G., Kriegler, E., et al. (2019). Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geosci. Model Dev. Discus.* 12, 1443–1475. doi: 10.5194/gmd-2018-266

Gillett, N. P., Kirchmeier-Young, M., Ribes, A., Shiogama, H., Hegerl, G. C., Knutti, R., et al. (2021). Constraining human contributions to observed warming since the pre-industrial period. *Nat. Clim. Chang.* 11, 207–212. doi: 10.1038/s41558-020-00965-9

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., et al. (2016). The detection and attribution model intercomparison project (DAMIP v1.0) contribution to CMIP6. *Geosci. Model Dev.* 9, 3685–3697. doi: 10.5194/gmd-9-3685-2016

Giorgi, F., and Mearns, L. O. (2002). Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "Reliability Ensemble Averaging" (REA) method. *J. Climate*. 15, 1141–1158.

Hall, A., Cox, P., Huntingford, C., and Klein, S. (2019). Progressing emergent constraints on future climate change. *Nat. Clim. Change.* 9, 269–278. doi: 10.1038/s41558-019-0436-6

Hall, A., and Qu, X. (2006). Using the current seasonal cycle to constrain snow albedo feedback in figure climate change. *Geophys. Res. Lett*. 33:L03502. doi: 10.1029/2005GL025127

Hanlon, H. M., Morak, S., and Hegerl, G. C. (2013). Detection and prediction of mean and extreme European summer temperatures with a multimodel ensemble. *J. Geophys. Res. Atmos*. 118, 9631–9641. doi: 10.1002/jgrd.50703

Harris, G. R., Sexton, D. M. H., Booth, B. B. B., Collins, M., and Murphy, J. M. (2013). Probabilistic projections of transient climate change. *Clim. Dyn.* 40:2937. doi: 10.1007/s00382-012-1647-y

Haylock, M., Hofstra, N., Klein Tank, A. M. G., and Klok, E. J. (2008). A European daily high-resolution gridded dataset of surface temperature, precipitation and sea-level pressure. *J. Geophys. Res*. 113:D20119. doi: 10.1029/2008JD010201

Hegerl, G., and Zwiers, F. (2011). Use of models in detection and attribution of climate change. *Wiley Interdisc. Rev. Clim. Change*. 2, 570–591. doi: 10.1002/wcc.121

Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., and Folland, C. K. (2015). Tripole index for the interdecadal pacific oscillation. *Clim. Dyn.* 45, 3077–3090. doi: 10.1007/s00382-015-2525-1

Hermanson, L., Bilbao, R., Dunstone, N., Ménégoz, M., Ortega, P., Pohlmann, H., et al. (2020). Robust multiyear climate impacts of volcanic eruptions in decadal prediction systems. *J. Geophys. Res. Atmosph*. 125:e2019JD031739. doi: 10.1029/2019JD031739

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc*. 146, 1999–2049. doi: 10.1002/qj.3803

Ingleby, B., and Huddleston, M. (2007). Quality control of ocean temperature and salinity profiles - historical and realtime data. *J. Mar. Syst*. 65, 158–175. doi: 10.1016/j.jmarsys.2005.11.019

Jolliffe, I. T., and Stephenson, D. B. (2003). *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. Hoboken: John Wiley & Sons Ltd., 240.

Kataoka, T., Tatebe, H., Koyama, H., Mochizuki, T., Ogochi, K., Naoe, H., et al. (2020). Seasonal to decadal predictions with MIROC6: description and basic evaluation. *J. Adv. Model. Earth Syst*. 12:e2019MS002035. doi: 10.1029/2019MS002035

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) Large Ensemble project: a community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc*. 96, 1333–1349. doi: 10.1175/BAMS-D-13-00255.1

Kendon, M., McCarthy, M., Jevrejeva, S., Matthews, A., and Legg, T. (2019). State of the UK climate 2018. *Int J Climatol*. 39 (Suppl. 1), 1–55. doi: 10.1002/joc.6213

Kettleborough, J. A., Booth, B., B., Stott, P. A., and Allen, M. R. (2007). Estimates of uncertainty in predictions of global mean surface temperature. *Clim. J.* 20, 843–855. doi: 10.1175/JCLI4012.1

Kim, W. M., Yeager, S., Chang, P., and Danabasoglu, G. (2018). Low-frequency north atlantic climate variability in the community earth system model large ensemble. *J. Clim*. 31, 787–813. doi: 10.1175/JCLI-D-17-0193.1

Knight, J. R., Andrews, M. B., Smith, D. M., Arribas, A., Colman, A. W., Dunstone, N. J., et al. (2014). Predictions of climate several years ahead using an improved decadal prediction system. *J. Clim*. 27, 7550–7567. doi: 10.1175/JCLI-D-14-00069.1

Knutti, R. (2010). The end of model democracy? *Clim. Change* 102, 395–404. doi: 10.1007/s10584-010-9800-2

Knutti, R., Allen, M. R., Friedlingstein, P., Gregory, J. M., Hegerl, G. C., Meehl, G. A., et al. (2008). A review of uncertainties in global

temperature projections over the twenty-first century. *J. Clim*. 21, 2651–2663. doi: 10.1175/2007JCLI2119.1

Knutti, R., Masson, D., and Gettelman, A. (2013). Climate model genealogy: CMIP5 and how we got there. *GRL* 40, 1194–1199. doi: 10.1002/grl.50256

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., and Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophys. Res. Lett*. 44, 1909–1918. doi: 10.1002/2016GL072012

Lorenz, R., Herger, N., Sedláček, J., Eyring, V., Fischer, E. M., and Knutti, R. (2018). Prospects and caveats of weighting climate models for summer maximum temperature projections over north america. *J. Geophys. Res. Atmosph*. 123, 4509–4526. doi: 10.1029/2017JD027992

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., et al. (2020). Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Amer. Meteor. Soc*. 101, E608–E625. doi: 10.1175/BAMS-D-18-0326.1

Marotzke, J., Müller, W. A., Vamborg, F. S. E., Becker, P., Cubasch, U., et al. (2016). MiKlip: a national research project on decadal climate prediction. *Bull. Am. Meteorol. Soc*. 97, 2379–2394. doi: 10.1175/BAMS-D-15-00184.1

Mecking, J. V., Drijfhout, S. S., J. J.-M., and Hirschi, Blaker, A. T. (2019). Ocean and atmosphere influence on the 2015 European heatwave. *Environ. Res. Lett*. 14:114035. doi: 10.1088/1748-9326/ab4d33

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., et al. (2009). Decadal prediction: can it be skillful? *BAMS* 2009, 1467–1486. doi: 10.1175/2009BAMS2778.1

Meehl, G. A., Richter, J. H., Teng, H., Capotondi A., Cobb, k., Doblas-Reyes, F., et al. (2021). Initialized Earth System prediction from subseasonal to decadal timescales. *Nat. Rev. Earth Environ*. 2, 340–357. doi: 10.1038/s43017-021-00155-x

Menary, M. B., and Hermanson, L. (2018). Limits on determining the skill of north atlantic ocean decadal predictions. *Nat. Commun*. 9:1694. doi: 10.1038/s41467-018-04043-9

Merrifield, A. L., Brunner, L., Lorenz, R., Medhaug, I., and Knutti, R. (2020). An investigation of weighting schemes suitable for incorporating large ensembles into multi-model ensembles. *Earth System Dyn*. 11, 807–834. doi: 10.5194/esd-11-807-2020

Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A. et al. (2020). Current and emerging developments in subseasonal to decadal prediction. *Bull. Am. Meteorol. Soc*. 101, E869–E896. doi: 10.1175/BAMS-D-19-0037.1

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set. *J. Geophys. Res*. 117:D08101. doi: 10.1029/2011JD017187

Murphy, J. M., Harris, G. R., Sexton, D. M. H., Kendon, E. J., Bett, P. E., Clark, R. T., et al. (2018). *UKCP18 Land Projections: Science Report*. Exeter: Met Office Hadley Centre. Available online at: https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report.pdf

Murphy, L. N., Bellomo, K., Cane, M., and Clement, A. (2017). The role of historical forcings in simulating the observed Atlantic multidecadal oscillation. *Geophys. Res. Lett*. 44, 2472–2480. doi: 10.1002/2016GL071337

Nijsse, F., Cox, P., and Williamson, M. (2020). An emergent constraint on Transient Climate Response from simulated historical warming in CMIP6 models. *Earth Syst. Dynamics* 1–14. doi: 10.5194/esd-2019-86

Pohlmann, H., Botzet, M., Latif, M., Roesch, A., Wild, M., and Tschuck, P. (2005). Estimating the decadal predictability of a coupled AOGCM. *J. Climate*. 17, 4463–4472. doi: 10.1175/3209.1

Pohlmann, H., Müller, W. A., Bittner, M., Hettrich, S., Modali, K., Pankatz, K., et al. (2019). Realistic quasi-biennial oscillation variability in historical and decadal hindcast simulations using CMIP6 forcing. *Geophys. Res. Lett*. 46, 14118–14125. doi: 10.1029/2019GL084878

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res*. 108, 4407. doi: 10.1029/2002JD002670, D14

Ribes, A., Qasmi, S., and Gillett, N. P. (2021). Making climate projections conditional on historical observations. *Sci. Adv*. 7:eabc0671. doi: 10.1126/sciadv.abc0671

Sanderson, B. M., Pendergrass, A., Koven, C. D., Brient, F., Booth, B. B. B., Fisher, R. A., et al. (2021). On structural errors in emergent constraints. *Earth Syst. Dynamics* (in review). doi: 10.5194/esd-2020-85

Sanderson, B. M., Wehner, M., and Knutti, R. (2017). Skill and independence weighting for multi-model assessments. *Geo. Sci. RModel Dev*. 10, 2379–2395. doi: 10.5194/gmd-10-2379-2017

Schurer, A., Hegerl, G., Ribes, A., Polson, D., Morice, C., and Tett, S. (2018). Estimating the transient climate response from observed warming. *J. Clim*. 31, 8645–8663. doi: 10.1175/JCLI-D-17-0717.1

Sexton, D. M., and Harris, G. R. (2015). The importance of including variability in climate change projections used for adaptation. *Nat. Clim. Change*. 5, 931–936. doi: 10.1038/nclimate2705

Sexton, D. M. H., Murphy, J. M., Collins, M., and Webb, M. J. (2012). Multivariate probabilistic projections using imperfect climate models, Part I: outline of methodology. *Clim. Dyn*. 38:2513. doi: 10.1007/s00382 011 1208 9

Sgubin, G., Swingedouw, D., Drijfhout, S., Mary, Y., and Bennabi, A. (2017). Abrupt cooling over the North Atlantic in modern climate models. *Nat. Commun*. 8:14375. doi: 10.1038/ncomms14375

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Rev. Geophys*. 58:e2019RG000678. doi: 10.1029/2019RG000678

Shin, J., Park, S., Shin, S.-I., Newman, M., and Alexander, M. (2020). Enhancing ENSO prediction skill by combining model analog and linear inverse models (MA-LIM). *Geophys. Res. Lett*. 47:e2019GL085914. doi: 10.1029/2019GL085914

Shiogama, H., and Stone, D., Emori, S., et al. (2016). Predicting future uncertainty constraints on global warming projections. *Sci. Rep*. 6, 18903. doi: 10.1038/srep18903

Sippel, S., Zscheischler, J., Mahecha, M., D., Orth R, Reichstein, M., Vogel, M., et al. (2017). Refining multi-model projections of temperature extremes by evaluation against land–atmosphere coupling diagnostics. *Earth Syst. Dynam*. 8, 387–403. doi: 10.5194/esd-8-387-2017

Smith, D., Eade, R., and Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Clim. Dyn*. 41, 3325–3338.

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., et al. (2020). North Atlantic climate far more predictable than models imply. *Nature* 583, 796–800. doi: 10.1038/s41586-020-2525-0

Sospedra-Alfonso, R., and Boer, G. J. (2020). Assessing the impact of initialization on decadal prediction skill. *Geophys. Res. Lett*. 47:e2019GL086361. doi: 10.1029/2019GL086361

Stainforth, D. A., et al. (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 433, 403–406. doi: 10.1038/nature03301

Stott, P. A., and Kettleborough, J. A. (2002). Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, 416, 723–726. doi: 10.1038/416723a

Swingedouw, D., Mignot, J., Labetoule, S., Guilyardi, E., and Madec, G. (2013). Initialisation and predictability of the AMOC over the last 50 years in a climate model. *Clim. Dyn*. 40, 2381–2399. doi: 10.1007/s00382-012-1516-8

Swingedouw, D., Ortega, P., Mignot, J., Guilyardi, E., Masson-Delmotte, V., Butler, P., et al. (2015). Bidecadal North Atlantic ocean circulation variability controlled by timing of volcanic eruptions. *Nat. Commun*. 6:6545. doi: 10.1038/ncomms7545

Tietsche, S., Balmaseda, M., Zuo, H., Roberts, C., Mayer, M., and Ferranti, L. (2020). The importance of North Atlantic Ocean transports for seasonal forecasts. *Clim. Dyn*. 55, 1995–2011. doi: 10.1007/s00382-020-05364-6

Tokarska, K., Hegerl, G. C., Schurer, A. P., Forster, P., and Marvel, K. (2020b). Observational Constraints on the effective climate sensitivity from the historical record. *Environ. Res. Lett*. 15:034043. doi: 10.1088/1748-9326/ab738f/pdf

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., et al. (2020a). Past warming trend constrains future

warming in CMIP6 models. *Sci. Adv.* 6:eaaz9549. doi: 10.1126/sciadv.aaz9549

Weigel, A. A. P., Knutti, R., Liniger, M. A., and Appenzeller, C. (2010). Risks of model weighting in multimodel climate projections. *J. Clim.* 23, 4175–4191. doi: 10.1175/2010jcli3594.1

Weisheimer, A., and Palmer, T. N. (2014). On the Reliability of Seasonal Climate Forecasts. *J. R. Soc. Interface.* 11:9620131162. doi: 10.1098/rsif.2013.1162

Wu, B., Zhou, T., Li, C., Müller, W. A., and Lin, J. (2019). Improved decadal prediction of Northern-Hemisphere summer land temperature. *Clim. Dyn.* 53, 1–13. doi: 10.1007/s00382-019-04658-8

Yeager, S. (2020). The abyssal origins of North Atlantic decadal predictability. *Clim. Dyn.* 55, 2253–2271. doi: 10.1007/s00382-020-05382-4

Yeager, S. G., Danabasoglu, G., Rosenbloom, N., A., Strand, W., Bates, S. C., and Meehl, G. A., et al. (2018). A large ensemble of initialized decadal prediction simulations using the community earth system model. *Bull. Amer. Meteor. Soc.* 99, 1867–1886. doi: 10.1175/BAMS-D-17-0098.2

Yeager, S. G., and Robson, J. I. (2017). Recent progress in understanding and predicting atlantic decadal climate variability. *Curr. Clim. Change Rep.* 3, 112–127. doi: 10.1007/s40641-017-0064-z

**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.

The handling Editor declared a past co-authorship with several of the authors GCH, BBBB, GRH, JL, JMM, DS.

frontiers in Climate

ORIGINAL RESEARCH
published: 16 June 2021
doi: 10.3389/fclim.2021.670740

# Uncertainty Quantification and Bayesian Inference of Cloud Parameterization in the NCAR Single Column Community Atmosphere Model (SCAM6)

*Raju Pathak[1,2], Hari Prasad Dasari[1], Samah El Mohtar[1], Aneesh C. Subramanian[3], Sandeep Sahany[2,4], Saroj Kanta Mishra[2], Omar Knio[1] and Ibrahim Hoteit[1]**

[1] Physical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [2] Center for Atmospheric Sciences, Indian Institute of Technology Delhi, New Delhi, India, [3] Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO, United States, [4] Center for Climate Research, Singapore, Singapore

Uncertainty quantification (UQ) in weather and climate models is required to assess the sensitivity of their outputs to various parameterization schemes and thereby improve their consistency with observations. Herein, we present an efficient UQ and Bayesian inference for the cloud parameters of the NCAR Single Column Atmosphere Model (SCAM6) using surrogate models based on a polynomial chaos expansion. The use of a surrogate model enables to efficiently propagate uncertainties in parameters into uncertainties in model outputs. We investigated eight uncertain parameters: the auto-conversion size threshold for ice to snow (*dcs*), the fall speed parameter for stratiform cloud ice (*ai*), the fall speed parameter for stratiform snow (*as*), the fall speed parameter for cloud water (*ac*), the collection efficiency of aggregation ice (*eii*), the efficiency factor of the Bergeron effect (*berg_eff*), the threshold maximum relative humidity for ice clouds (*rhmaxi*), and the threshold minimum relative humidity for ice clouds (*rhmini*). We built two surrogate models using two non-intrusive methods: spectral projection (SP) and basis pursuit denoising (BPDN). Our results suggest that BPDN performs better than SP as it enables to filter out internal noise during the process of fitting the surrogate model. Five out of the eight parameters (namely *dcs*, *ai*, *rhmaxi*, *rhmini*, and *eii*) account for most of the variance in predicted climate variables (e.g., total precipitation, cloud distribution, shortwave and longwave cloud radiative effect, ice, and liquid water path). A first-order sensitivity analysis reveals that *dcs* contributes ∼40–80% of the total variance of the climate variables, *ai* around 15–30%, and *rhmaxi, rhmini,* and *eii* around 5–15%. The second- and higher-order effects contribute ∼7 and 20%, respectively. The sensitivity of the model to these parameters was further explored using response curves. A Markov chain Monte Carlo (MCMC) sampling algorithm was also implemented for the Bayesian inference of *dcs, ai, as, rhmini,* and *berg_eff* using cloud distribution data collected at the Southern Great Plains (USA). The inferred parameters suggest improvements in the global Climate Earth System Model (CESM2) simulations of the tropics and sub-tropics.

**Keywords: climate modeling, uncertainty quantification, Bayesian inference, cloud parameters, parameterization schemes**

OPEN ACCESS

***Edited by:***
*Sarah Kang,*
*Ulsan National Institute of Science and Technology, South Korea*

***Reviewed by:***
*Jian Li,*
*Chinese Academy of Meteorological Sciences, China*
*Aiko Voigt,*
*University of Vienna, Austria*

***Correspondence:**
*Ibrahim Hoteit*
*ibrahim.hoteit@kaust.edu.sa*

***Specialty section:***
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

***Received:*** *22 February 2021*
***Accepted:*** *17 May 2021*
***Published:*** *16 June 2021*

***Citation:***
*Pathak R, Dasari HP, El Mohtar S, Subramanian AC, Sahany S, Mishra SK, Knio O and Hoteit I (2021) Uncertainty Quantification and Bayesian Inference of Cloud Parameterization in the NCAR Single Column Community Atmosphere Model (SCAM6). Front. Clim. 3:670740. doi: 10.3389/fclim.2021.670740*

## INTRODUCTION

Uncertainty quantification (UQ) in weather and climate models enables to evaluate model sensitivities and also to reduce inconsistencies between the model outputs and observations (e.g., Allen et al., 2000; Murphy et al., 2004; Stainforth et al., 2005; Lopez et al., 2006; Jackson et al., 2008; Le Maitre and Knio, 2010; Covey et al., 2013). In global climate models (GCMs), subgrid-scale processes (e.g., cloud characteristics and convection) are often parameterized using various schemes and assumptions depending on empirical parameters. These introduce different levels of uncertainty in the parameterization of subgrid-scales and, thus, in the eventual model simulations (e.g., Warren and Schneider, 1979). The UQ analysis requires a large number of model simulations in order to sample the probability distribution function (PDF) of a parameter, which increases exponentially with the number of uncertain parameters (Li et al., 2013). Parameter estimation can be undertaken through a series of model simulations that perturb the parameters individually and determine the predictive skill of the model for each simulation (e.g., Zaehle and Friend, 2010; Anand et al., 2018; Ricciuto et al., 2018). However, such methods cannot treat the non-linear interactions between the input parameters and model outputs (Tarantola, 2004; Hourdin et al., 2017). UQ has nonetheless been demonstrated as an effective method to determine the interactive effects of model parameters (e.g., Jackson et al., 2008; Collins et al., 2011; Yang et al., 2012, 2013; Covey et al., 2013; Zou et al., 2014; Guo et al., 2015; Qian et al., 2015; Sraj et al., 2016).

The multi-objective UQ framework is an advanced and robust approach to investigate interactions between model parameters and helps to identify the critical parameters to be tuned for best performances (e.g., Bastos and O'Hagan, 2009; Priess et al., 2011; Li et al., 2013; Yang et al., 2013; Zhao et al., 2013; Brown et al., 2014; Wang et al., 2014; Gong et al., 2015; Sraj et al., 2016; Qian et al., 2018). This framework consists of three principal components: (a) building an efficient surrogate model to quantify the sensitivity of the model's outputs to the input parameter, (b) identifying the most influential parameters, and (c) inferring optimized values for these parameters based on available data. The UQ machinery enables constructing surrogate models from a relatively reasonable number of model simulations, for a dozen of input parameters (e.g., Lee et al., 2011; Zhu et al., 2015). Surrogate models built using regression splines, Gaussian methods, generalized linear models, and polynomial chaos expansions (PCE) have been successfully employed in various regional and global applications (e.g., Lee et al., 2011; Carslaw et al., 2013; Guo et al., 2015; Zhu et al., 2015; Sraj et al., 2016).

Uncertainties in climate model simulations are mainly due to cloud parameterizations such as cloud distribution, cloud–aerosol interactions, cloud feedback, and the convective activity of clouds (e.g., Albrecht et al., 1988; Bony and Dufresne, 2005; Lee et al., 2011, 2012; Carslaw et al., 2013; Zelinka et al., 2013, 2020; Zhao et al., 2013; Bony et al., 2015; Pathak et al., 2020). Significant improvements in representing convection, clouds, and cloud–aerosol interactions in climate models have been achieved in the last two decades (e.g., Sanderson et al., 2008; Gettelman et al., 2010; Golaz et al., 2011; He and Posselt, 2015; Anand et al., 2018; Pathak et al., 2020). However, work is still required to minimize the uncertainties associated with cloud representation (e.g., Schwartz, 2004; Lohmann et al., 2007; Gettelman et al., 2012; Lee et al., 2012; Hazra et al., 2015).

The Community Atmosphere Model version-6 (CAM6) describes the atmospheric component of the National Center for Atmospheric Research (NCAR) Community Earth System Model version-2 (CESM2) (Danabasoglu et al., 2020). It uses substantially modified physical parameterizations relative to its predecessors' versions, except the radiative transfer scheme. CAM6 incorporates a single framework known as the Cloud Layer Unified by Binormals (CLUBB) scheme (Bogenschutz et al., 2013) to represent boundary layer turbulence, shallow convection, and cloud macrophysics. An improved two-moment prognostic cloud microphysics framework (also called MG2; Gettelman and Morrison, 2015; Gettelman et al., 2015) is also implemented in CAM6 to integrate prognostic precipitation species (e.g., rain and snow). MG2 interacts with advanced Modal Aerosol Module (MAM4) aerosol microphysics schemes to compute condensate mass fractions and number concentrations. Additional advances in CAM6 include the incorporation of topographic orientation (ridges) and blocking effects of low-level flows into the orographic gravity wave scheme.

An UQ analysis was performed on CLUBB parameters using the NCAR single-column atmospheric model (SCAM) version-5 by Guo et al. (2015), and Energy Exascale Earth System Model (E3SM) by Qian et al. (2018). The CLUBB scheme has significantly improved the simulation of the stratocumulus to cumulus transition through a substantial improvement in trade winds in the subtropical oceans and the simulation of coastal stratocumulus clouds (Bogenschutz et al., 2012, 2013). The single-column model simulates unresolved subgrid scale processes using parameterized physics (e.g., convection, clouds, turbulence, and radiation) by prescribing their dynamical state and tendencies, such as negligible dynamics-physics coupling (Jess et al., 2011; Guo et al., 2015; Zhang et al., 2016). In climate modeling, the single-column model has been a successful tool for developing, validating, and tuning physical parameterizations, since running a large number of model integrations from a GCM is time-consuming and computationally expensive (e.g., Lord et al., 1982; Betts and Miller, 1986; Guichard et al., 2004; Fridlind et al., 2012; Guo et al., 2015). Thus, we opt to use the NCAR SCAM-6 (Gettelman et al., 2019) to perform our UQ analysis of cloud parameterizations.

In this study, we attempt to understand the response of CAM6 to cloud micro- and macro-physics parameters. We also quantify the plausible physical mechanisms responsible for the sensitivity of model simulations to parametric uncertainties using cloud hydrometeor and cloud radiative effect distributions, as well as the convective instability. Our study will thus help to understand the uncertainties associated with model physics and enables a set of model parameters to be estimated and used for model calibration. We also highlight the use of PCE in climate UQ analysis and Basis Pursuit Denoising (BPDN) in alleviating internal noise inherent to model physics. We provide the model details, experimental setup, and parameters

in section Model Details and Methodology and introduce the framework for sensitivity analysis and Bayesian inference in section UQ Framework for Sensitivity Analysis and Bayesian Inference. We present the results in section Results and Discussion. In particular, we estimate the relative importance of sensitive parameters (Section Relative Importance of Sensitive Parameters), quantify the model responses to changes in these parameters for different relevant atmospheric variables (Section Response of Simulated QoI to Sensitive Parameters), and determine the posterior distributions of these parameters using Bayesian inference and available observations of cloud distribution (Section Bayesian Inference). Section Conclusions concludes the study with a summary of the results and findings.

# MODEL DETAILS AND METHODOLOGY

## Model Description

SCAM6 was used to investigate the sensitivity of the CAM6 physics package to multiple parameters. A large-scale flow and its related tendencies were prescribed in SCAM from observational data. CAM6 incorporates various physics packages, including (a) the CLUBB (Bogenschutz et al., 2012) scheme for the planetary boundary layer, shallow cumulus, and stratiform cloud macrophysics, (b) the MG2 cloud microphysics scheme (Gettelman and Morrison, 2015) for predicting the mass and number concentrations of falling condensed species (rain and snow), (c) the Modal Aerosol Model version-4 (MAM4; Liu et al., 2016) to account for the influence of aerosol on cloud microphysics, (d) subgrid orographic drag parameterization (Beljaars et al., 2004) to represent the turbulence from drag due to subgrid orography with the horizontal scale $<5$ km, and (e) the Zhang and McFarlane (1995) deep convection scheme, updated by Neale et al. (2008) to include the dilute CAPE computation and by Richter and Rasch (2008) to include convective momentum transport. More details on the model configuration can be found in Gettelman et al. (2019) and Danabasoglu et al. (2020).

## Experimental Setup

We configured SCAM6 with 32 vertical levels, the top-level at 2.26 hPa ($\sim$40 km), and forced it with the ARM97 observations (Gettelman et al., 2019) collected during the 30 day intensive observation period on June 1997 over the Southern Great Plains observatory (SGP: 36°N and 97°W; Zhang et al., 2016). The location of the SGP observatory has a significant impact due to the prevailing mid-latitude and mid-continent large-scale weather systems, the wide range of cloud and atmospheric conditions from migratory disturbances, and the presence of air masses with strong diurnal and annual cycles. ARM97 measurements are extensively used in single-column models (SCM) to understand convection, atmospheric radiation, cloud characteristics, and the interaction between radiation and clouds, aerosols, and gases (e.g., Guichard et al., 2004; Fridlind et al., 2012; Petch et al., 2014). In the SCM, the large-scale flow and its related tendencies are prescribed. The large-scale variables, such as the zonal and meridional components of the flow (U and V), temperature (T), moisture (Q) are measurable quantities, but the vertical advection tendencies of U, V, T, and Q are not, and are therefore computed using the dynamical core. ARM97 observations provide large-scale variables (i.e., U, V, T, and Q). The vertical advection tendencies of Q are computed using the Lagrangian dynamical core and the vertical advection tendencies of U, V, and T are computed using the Eulerian dynamical core.

## Investigated Parameters

We selected eight parameters to quantify the uncertainties in cloud microphysics and macrophysics (hereafter CMP) as outlined in **Table 1**. Although many uncertain parameters are involved in the CMP scheme, we considered only few parameters that have been reported as being sensitive in the previous version of this model (e.g., Covey et al., 2013; Yang et al., 2013; Qian et al., 2015, 2018; Pathak et al., 2020). The ice cloud fraction ($CF_i$) is calculated from relative humidity (RH) using the total ice water mixing ratio [i.e., the ice mass mixing ratio ($q_i$) plus water vapor mixing ratio ($q_v$)] and the saturated vapor mixing ratio over ice ($q_{sat}$), as follows:

$$\begin{aligned} RH_{ti} &= \frac{q_v + q_i}{q_{sat}}, \\ RH_d &= \left(0, \frac{RH_{ti} - RH_{i\_\min}}{RH_{i\_\max} - RH_{i\_\min}}\right), \\ CF_i &= min(1,\ RH_d^2), \end{aligned}$$

where $RH_{i\_\max}$ (*rhmaxi*) and $RH_{i\_\min}$ (*rhmini*) are the threshold relative humidity parameters with respect to the ice, reflecting high sensitivity to total ice super-saturation and ice cloud cover (Gettelman et al., 2010). The ice cloud fraction ($CF_i$) is greater than zero when $RH_{ti}$ reaches *rhmini*, and is equal to one (or 100%) when $RH_{ti}$ reaches *rhmaxi*. The mass- and number-weighted terminal fall speeds for all cloud and precipitation species (cloud water, cloud ice, rain, and snow) were obtained by performing an integration over particle size distributions with appropriate weighting by number concentration or mixing ratio as follows:

$$\begin{aligned} V_N &= \frac{\int_0^\infty \left(\frac{\rho_a}{\rho_{a0}}\right)^{0.54} aD^b\ \Phi\ (D)\ dD}{\int_0^\infty \Phi\ (D) dD}, \\ V_q &= \frac{\int_0^\infty \frac{\pi\rho}{6} \left(\frac{\rho_a}{\rho_{a0}}\right)^{0.54} aD^{b+3}\ \Phi\ (D)\ dD}{\int_0^\infty \frac{\pi\rho}{6} D^3 \Phi\ (D)\ dD}, \end{aligned}$$

where $\rho_{a0}$ is the reference air density at 850 hPa, and $a$ and $b$ are the empirical coefficients in the diameter-fall speed relationship ($V = aD^b$; where $V$ is the terminal fall speed for an individual particle of diameter $D$). The empirical coefficient $a$ for different hydrometeor species (e.g., *ai* for cloud ice, *as* for snow, and *ac* for cloud water) is another critical uncertain parameter. Since the auto-conversion of cloud ice to form snow is calculated by integrating cloud ice mass- and number-weighted size distributions larger than some threshold size, the resultant mixing ratio and number

**TABLE 1 |** Cloud microphysics and macrophysics parameters used in this study.

| Parameter name | Description | Range | | | Remarks |
|---|---|---|---|---|---|
| | | Low | Default | High | |
| dcs | Auto-conversion size threshold for ice to snow | 1e−4 | 5e−4 | 9e−4 | Mainly affects the distribution of high clouds; higher *dcs* values correspond to the lesser conversion of cloud ice to snow |
| ai | Fall speed parameter for stratiform cloud ice | 350 | 700 | 1,400 | Mainly affect ice water content |
| as | Fall speed parameter for stratiform snow | 5.860 | 11.72 | 23.44 | Mainly affect ice water content; larger *as* values correspond to higher cloud water terminal fall speeds |
| ac | Fall speed parameter for cloud water | 1.5e+7 | 3e+7 | 5e+7 | Mainly affect cloud water content |
| eii | Collection efficiency aggregation ice | 0.001 | 0.1 | 1.0 | Mainly affect ice water content |
| berg-eff | Efficiency factor for Bergeron effect | 0.5 | 1.0 | 1.5 | Mainly affect vapor deposition on cloud ice and liquid content |
| rhmaxi | Threshold maximum relative humidity for ice clouds | 0.90 | 1.0 | 1.10 | Mainly affect total ice supersaturation and ice cloud fractions |
| rhmini | Threshold minimum relative humidity for ice clouds | 0.70 | 0.80 | 0.90 | Mainly affect ice cloud fraction |

distribution into snow category are transferred over some specified time-scale ($\tau_{auto}$; Ferrier, 1994). Consequently, the grid-scale changes in $q_i$ and $N_i$ due to auto-conversion may be given as follows:

$$\left(\frac{\partial q_i}{\partial t}\right)_{auto} = -F\frac{\pi\rho_i N_{0i}}{6\tau_{auto}}\left[\frac{D_{cs}^3}{\lambda_i} + \frac{3D_{cs}^2}{\lambda_i^2} + \frac{6D_{cs}}{\lambda_i^3} + \frac{3D}{\lambda_i^4}\right]\exp^{-\lambda_i D_{cs}},$$
$$\left(\frac{\partial N_i}{\partial t}\right) = -F\frac{N_{0i}}{\lambda_i\tau_{auto}}\exp^{-\lambda_i D_{cs}},$$

where $D_{cs}$ (*or dcs*) is the threshold size parameter separating cloud ice from snow (used for UQ analysis), $\rho_i$ is the bulk density of cloud ice, and $\tau_{auto} = 3$ min. In addition, the parameter describing the efficiency factor for vapor deposition onto ice (*berg$_e$ff*) is also used. Korolev et al. (2016) found that the vapor deposition onto ice and the depletion of liquid (i.e., the Wegener-Bergeron-Findeisen process) is rarely equal to its theoretical efficiency due to inhomogeneities in humidity and updrafts, as well as the generation of supersaturation. The efficiency factor zero corresponds to the state in which no vapor is deposited, whereas an efficiency factor of one corresponds to a perfect deposition state. Gettelman et al. (2019) reported that the condition in which no vapor is deposited on ice corresponds to a higher fraction of liquid and supercooled liquid.

# UQ FRAMEWORK FOR SENSITIVITY ANALYSIS AND BAYESIAN INFERENCE

## Sensitivity Analysis

According to Sobol (1993), if a set of $m$-independent random parameters $\xi = (\xi_1, \ldots, \xi_m)$ produces response $f(\xi)$, it can be written in the form of an expansion as follows:

$$f(\xi) = f_0 + \sum_{i=1}^{m} f_i(\xi_i) + \sum_{i=1}^{m}\sum_{j>1}^{m} f_{ij}(\xi_i\xi_j) + \ldots + f_{1\ldots,m}(\xi_1, \ldots, \xi_m) \quad (1)$$

where $f_0$ is the expected value of $f(\xi)$, and $f_{i_1,\ldots,i_s}(\{i_1, \ldots, i_s\};\ s = 1, \ldots, m)$ are orthogonal functions. This representation is commonly referred to as the variance decomposition analysis. The expected square of this decomposition (Equation 1) leads to the following variance decomposition of $f(\xi)$:

$$V = \sum_{i=1}^{m} V_i + \sum_{i=1}^{m}\sum_{j>1}^{m} V_{ij} + \ldots + V_{1\ldots m} \quad (2)$$

where $V$ is the total variance of $f(\xi)$, $V_i$ is the partial variance due to the perturbation of input parameter $\xi_i$ alone, and $V_{i_1,\ldots,i_s}$ is the partial variance due to the interactions between perturbed input parameters $\{\xi_{i_1}, \ldots, ..\xi_{i_s}\}$. Sobol's variance-based sensitivity indices are:

$$S_{i_1,..,i_s} = \frac{V_{i_1,..,i_s}}{V} \quad (3)$$

The first-order sensitivity index (also called the main effect of $\xi_i$) is:

$$S_i = \frac{V_i}{V} \quad (4)$$

Polynomial Chaos (PC) has been suggested as an efficient method for describing the stochastic processes required to quantify uncertainties in a given system (Ghanem and Spanos, 1991; Le Maitre and Knio, 2010). PC is based on a probabilistic paradigm that reflects the stochastic quantities of interest as a truncated PC expansion, also termed PCE. The PCE of a particular quantity of interest (QoI; see **Table 2**) is written as a function of $m$ uncertain parameters $[\xi = (\xi_1, \xi_2, \ldots, \xi_m)]$ in the $[-1, +1]$ space as follows:

$$E(\xi) \approx \sum_{k=0}^{R} e_k\psi_k(\xi) \quad (5)$$

**TABLE 2 |** First-order sensitivity contribution of parameters for different variables.

| Variable | Major Contributing Parameter | Contribution (in percent) |
|---|---|---|
| Total Precipitation (PRECT) | *dcs*, *ai*, *rhmini*, *rhmaxi*, and *eii* | 32, 16, 7, 5, and 4 |
| Long-wave Cloud Radiative Effect (LWCF) | *dcs* and *ai* | 50 and 14 |
| Short-wave Cloud Radiative Effect (SWCF) | *dcs*, *ai*, *rhmaxi*, *eii*, and *rhmini* | 18, 19, 15, 4, and 2 |
| Surface Latent Heat Flux (LHFLX) | *dcs*, *ai*, *rhmaxi*, *rhmini*, and *eii* | 23, 19, 10, 4, and 4 |
| Liquid Water Path (LWP) | *dcs* and *ai* | 11 and 9 |
| Ice Water Path (IWP) | *dcs* and *ai* | 84 and 6 |
| Convective Available Potential Energy (CAPE) | *dcs*, *ai*, *rhmaxi*, and *rhmini* | 28, 14, 8, and 4 |
| Medium Cloud Cover | *rhmaxi*, *dcs*, and *rhmini* | 52, 17, and 11 |
| High Cloud Cover | *dcs* and *ai* | 42 and 5 |
| Total Cloud Cover | *dcs*, *ai*, and *rhmaxi* | 42, 5, and 2 |

where $E(\xi)$ is the desired QoI, $e_k$ are the PCE coefficients, and $\psi_k(\xi)$ are the multi-dimensional Legendre polynomials that form the following orthogonal basis:

$$< \psi_i, \psi_j > = \int \psi_i(\xi)\,\psi_j(\xi)\,\rho(\xi)\,d\xi = \delta_{ij} < \psi_i^2 >,$$

where $\rho(\xi)$ is the underlying uniform distribution.

The expanded form of the PCE in Equation (5) can be written as follows:

$$E(\xi) = e_0\psi_0 + \sum_{i=1}^{m} e_1\psi_1(\xi_i) + \sum_{i=1}^{m}\sum_{j\geq i}^{m} e_{ij}\psi_2(\xi_i,\xi_j) + \sum_{i=1}^{m}\sum_{j\geq i}^{m}\sum_{k\geq j}^{m} e_{ijk}\psi_3(\xi_i,\xi_j,\xi_k) + \ldots \quad (6)$$

The total number of expansion terms (R) in Equations 5 and 6 for $m$ parameters and a PC order $r$ is given by:

$$R = \frac{(m+r)!}{m!r!} \quad (7)$$

The PC expansion in Equation (6) is similar to the decomposition in Equation (1) and thus provides a surrogate for approximating model outputs for a given set of parameters. Transforming Equation (6) into Equation (2) enables the corresponding sensitivity indices to be estimated (Crestaux et al., 2009).

PCE coefficients can be computed using two approaches: intrusive and non-intrusive. In the intrusive methods, the model equations are reformulated through the substitution of stochastic or random variables such that the model deploys the stochastic Navier Stokes Equation (e.g., Kusch and Frank, 2018). It is often challenging to compute the PC coefficients from an intrusive method since it involves source code modification (Ghanem and Spanos, 1991). Non-intrusive methods, on the other hand, use samples of model simulations for different realizations of parameters to build a surrogate model (Peng et al., 2014; Sraj et al., 2016). Herein, we used a non-intrusive spectral projection (NISP) method (Reagan et al., 2003; Constantine et al., 2012) to compute the PCE coefficient as follows:

$$e_k = \frac{< E, \psi_k >}{< \psi_k\psi_k >} = \frac{1}{< \psi_k\psi_k >}\int E\psi_k(\xi)\,\rho(\xi)\,d\xi \quad (8)$$

The stochastic integral (Equation 8) is computed numerically using appropriate quadrature methods, as follows:

$$< E, \psi_k > \approx \sum_{q=1}^{Q} E(\xi_q)\psi_k(\xi_q)w_q \quad (9)$$

where $Q$ is the total number of quadrature points, $\xi_q$ denotes the vectors of the parameters at a quadrature point $q$, and $w_q$ is the corresponding weight. The Smolyak sparse nested grid-based quadrature method (Smolyak, 1963) was used to reduce the number of SCAM6 runs to build the surrogate model. In this study, for a PC order with $r = 4$ and $m = 8$ parameters (i.e., total truncated PCE terms $R = 495$), the total number of quadrature points created from the Smolyak level 5 was 3937 (see http://www.sparse-grids.de/ and Smolyak, 1963 for full details of the Smolyak quadrature).

**Figure 1** shows the spectrum of normalized PCE coefficients $\left(\frac{e_k}{e_0}\right)$ calculated from a non-intrusive spectral projection (NISP; see Equation 8), and the vertical solid lines separating PCE terms according to their PC order ($r = 4;\ 1,\ 2,\ 3,\ 4$). This indicates that PC does not converge properly since the estimated PCE coefficients increase instead of decaying with an increase in PC order, suggesting an overfitting of the model output. This is due to internal noise in model simulations that is not tolerated by the NISP method when calculating PCE coefficients (Peng et al., 2014; Sraj et al., 2016). We thus opted to use a non-intrusive technique based on compressed sensing (CS; Chen and Donoho, 1994; Van den Berg and Friedlander, 2007, 2009) to estate more suitable PCE coefficients.

CS estimates the noise in the data to be fitted and then approximates them using a PC representation that tolerates the corresponding noise level. In this case, CS solves Equation (5) by exploiting the approximate sparsity of its signal, which is set by the $l1$ norm. CS with the $l1$ norm is known as BPDN.

Thus, if we consider $E = [E(\xi_1), E(\xi_2), \ldots.., E(\xi_Q)]$ as the vector of model evaluations at different quadrature nodes ($\xi_q$), $e = (e_0, e_1, \ldots.., e_R)$ as the vector of PCE coefficients, and $\psi$ as the PC basis functions evaluated at each sampled $\xi_q$, then an equivalent of Equation (5) could be written as follow:

$$E = \psi e \quad (10)$$

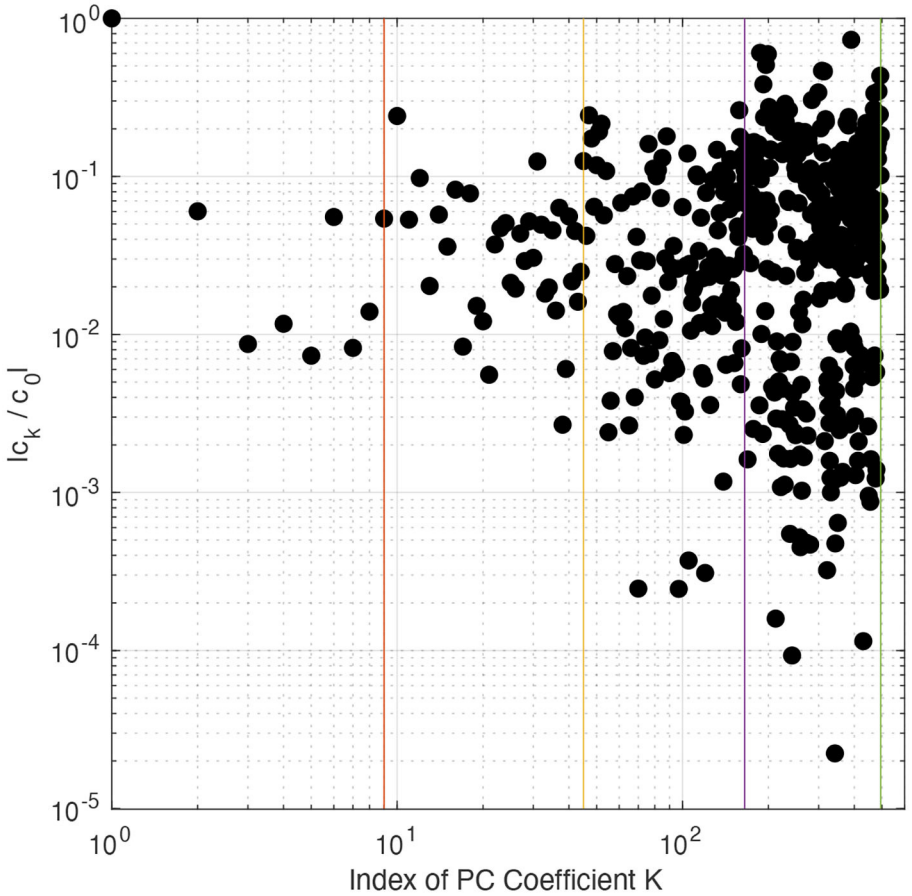



**FIGURE 1 |** PCE normalized coefficients $c_k/c_0$ for PC of order up to $r = 4$. Vertical lines separate the PCE terms into their different PC orders. PCE coefficients were calculated using the non-intrusive spectral projection (NISP) method.

The CS seeks a solution of Equation (10) with the minimum number of non-zero entries by solving the optimization problem:

$$\frac{minimize}{e} ||e||_1 : || E - \psi e| |_2 \leq \delta \tag{11}$$

where $\delta$ is noise, estimated by cross-validation (see Peng et al., 2014). Equation (11) was solved using a standard $l1$-minimization solver based on a spectral projection gradient algorithm (Van den Berg and Friedlander, 2009) provided in the SPGL1 MATLAB package by Van den Berg and Friedlander (2007) (see https://friedlander.io/spgl1/ for full details). The optimization problem (Equation 11) was solved using the previously simulated 3937 SCAM6 simulations. The resulting normalized PCE coefficients ($e_k/e_0$) are shown in **Figure 2**; their spectrum shows a decaying trend. In BPDN, the decrease in the PCE coefficients eventually reaches a plateau, while a further increase in the PC order leads to minimal improvement in the accuracy of the surrogate model. In light of these results, we used the BPDN approach to build the surrogate model (Equation 10).

**Figure 3** shows the scatter plot of SCAM6 simulations against results from the surrogate model built using PCE and BPDN for various variables or quantity of interest (QoI). The surrogate model and SCAM6 both show linear relationships for various QoIs and the constructed PCE model reproduces well the mean of the deterministic model signal. To quantify the level of agreement between SCAM6 and the surrogate model, we computed the $R^2$-value to estimate the fraction of total variance in SCAM6 results explained by the surrogate model. The highest $R^2$-value (0.99) were obtained for the longwave cloud radiative effect (LWCF) and the lowest value (0.79) for the convective available potential energy (CAPE) (**Figure 3**). The $R^2$-values of the remaining variables, including cloud water path (CWP), liquid water path (LWP), total cloud fraction (CLDTOT), surface latent heat flux (LHFLX), shortwave cloud radiative effect (SWCF), and total precipitation (PRECT) ranged between 0.79 and 0.99, indicating that BPDN enables to successfully build a surrogate model that is capable of describing the desired QoIs.

## Bayesian Inference

The posterior probability density function (PDF) of the CMP parameters can be calculated by updating the prior PDF according to Bayes' rule. Let $d = (d_1, d_2, \ldots d_n)^T$ be a vector of observation, $p = (p_1, p_2, \ldots p_n)^T$ a vector of uncertain input parameters, and $G$ a forward model such that $d \approx G(p)$. The prior PDF $[\pi(p)]$, which represents *a priori* information about $p$, is assumed to be uniform (non-informative, i.e., $\pi(p) =$

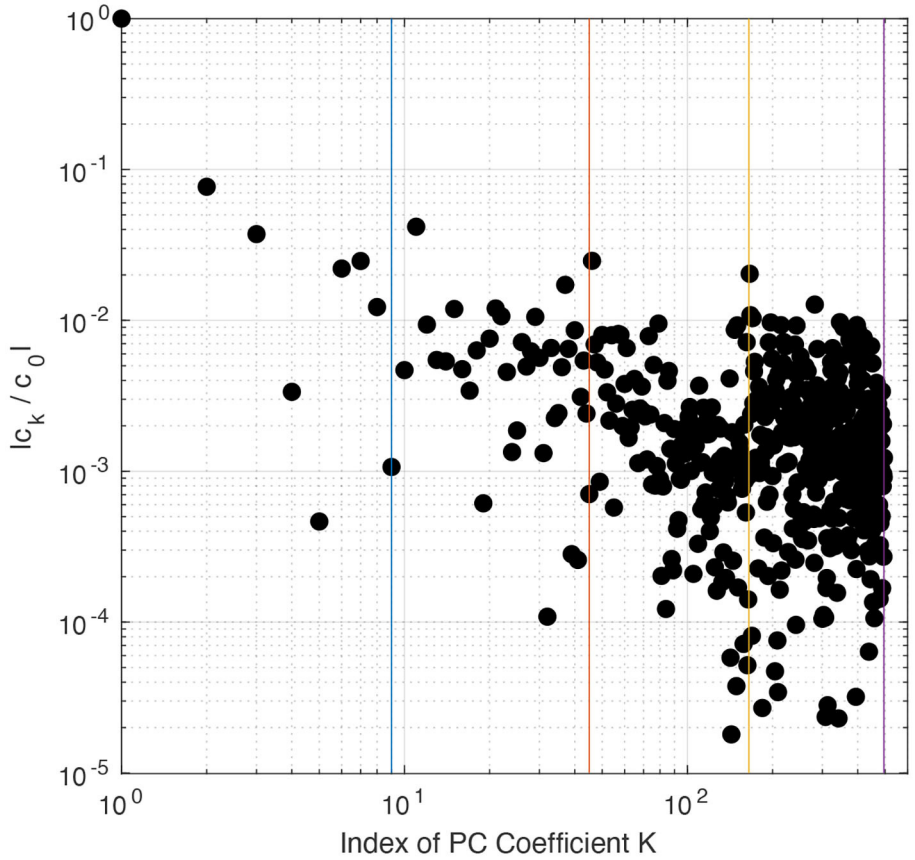



**FIGURE 2 |** PCE normalized coefficients $c_k/c_0$ for PC of order up to $r = 4$. Vertical lines separate the PCE terms into their different PC orders. PCE coefficients were calculated using the basis pursuit denoising (BPDN) method.

$\prod_{i=1}^{n_p} \frac{1}{(b_i - a_i)}$ , such that $a_i$ and $b_i$ are, respectively, the lower and upper bounds of parameter $p_i$; **Table 1**]. The input parameters $p$ are assumed to be independent with respect to one another. According to Bayes' rule, the posterior $[\pi(p|d)]$ is proportional to the product of the likelihood $[L(d|p)]$ and the prior as:

$$\pi\left(d\right) \propto L\left(p\right)\pi(p) \tag{12}$$

The likelihood is obtained from a cost function $E(p)$ using a Taylor score ($TS$, Taylor, 2001) and a scaling factor $S$ (used to normalize the cost function). Gaussian likelihoods based on misfits [i.e., the exponential form of the mean-square errors (MSE)] are widely adopted as cost functions (e.g., Sraj et al., 2016; Qian et al., 2018). Thus, to consider both the misfit in magnitude and the mismatch in the vertical cloud distribution, we used $[\ln(TS)]^2$ as the cost function (Guo et al., 2015; Qian et al., 2018) with a theoretical range of (0, *infinity*).

$$\pi\left(d\right) \propto exp\ exp\left[-E\left(p\right)\right]\ ^{*}\pi(p) \tag{13}$$

$$E\left(p\right) = S.\{\ln\ln\left[TS\left(p\right)\right]\}^2 \tag{14}$$

The TS used to evaluate model performance in terms of standard deviation and correlation with respect to observations is given as follows:

$$TS = \frac{(\frac{\sigma_{model}}{\sigma_{obs}} + \frac{\sigma_{obs}}{\sigma_{model}})^2 (1 + VCC_0)^k}{4\,(1 + VCC)^k} \tag{15}$$

$$VCC = \frac{\sum_{i=1}^{v} w_i \left(d_{i,model} - GM_{model}\right)\left(d_{i,obs} - GM_{obs}\right)}{\sigma_{model}.\sigma_{obs}\,.\,\sum_i^v w_i} \tag{15a}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{v} w_i (d_i - GM)^2}{\sum_{i=1}^{v} w_i}} \tag{15b}$$

$$GM = \frac{\sum_{i=1}^{v} w_i d_i}{\sum_{i=1}^{v} w_i} \tag{15c}$$

where subscripts "obs" and "model" denote the observed and simulated results, respectively. The standard deviation and vertical correlation coefficients between the observed and simulated results are denoted by $\sigma$ and VCC, respectively. $VCC_0$ denotes the maximum possible vertical correlation between the observations and the model outputs, and $k$ indicates a value that controls the correlation of the relative weight of the vertical level compared to the standard deviation in TS (Equation 15). $VCC_0$ and $k$ were set to 1 and 4,

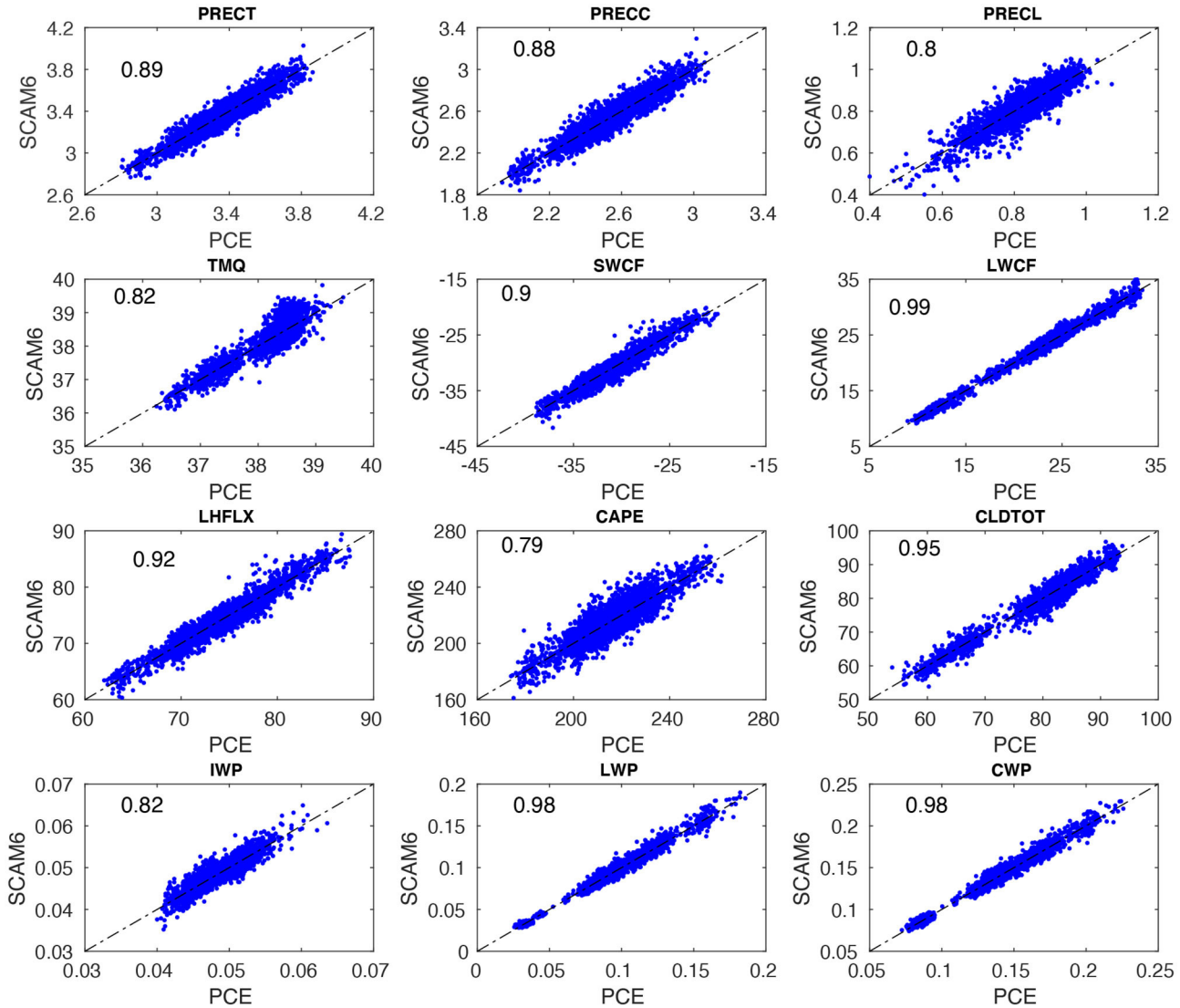



**FIGURE 3 |** Scatter plots of PC-fitted outputs vs. SCAM6 simulations for various climate variables. Row 1 corresponds to total precipitation (PRECT), convective precipitation (PRECC), and large-scale precipitation (PRECL). Row 2 corresponds to total precipitable water (TMQ), shortwave cloud radiative effect (SWCF), and longwave cloud radiative effect (LWCF). Row 3 corresponds to latent heat flux (LHFLX), convective available potential energy (CAPE), and total cloud cover (CLDTOT). Row 4 is similar to the previous rows but for ice water path (IWP), liquid water path (LWP), and cloud water path (CWP). The fraction (0–1) of the total variance of SCAM6 simulations explained by the PC-based surrogate model for various climatic variables is indicated by numbers in the upper left of each panel.

respectively, following Yang et al. (2013). A lower TS value is an indicator of better model performance, and model prediction and observation are considered identical when $TS = 1$. Vector $d$ represents the observed vertical cloud distribution from ARM97, $w_i$ is the vertical weight of a level, and $n$ is the number of vertical levels. The scaling factor was chosen as

$$S = \frac{2}{\Delta} \tag{16}$$

Jackson et al. (2004) noted that the spread of the cost function (i.e.,) due to natural variability could be used for its normalization (see Equation 14). In general, the natural variability is estimated *via* multiple model runs (Decremer et al., 2015). In this study, the average of all original simulations corresponding to the set of perturbed parameters was considered as an indicator of the spread of model bias due to natural variability (Qian et al., 2018). The Bayesian formulation thus requires the evaluation of the posterior distribution (Equation 13) to estimate the uncertain parameters. To this end, we employed a Markov Chain Monte Carlo (MCMC) technique; however, such an approach requires a large number of posterior evaluations in order to reach a meaningful solution. Each posterior evaluation requires a single model simulation, which is computationally prohibitive. Therefore, the PC-based surrogate model was used within the random-walk Metropolis MCMC algorithm to generate the 50,000 samples from posterior distributions for each parameter (Metropolis et al., 1953).

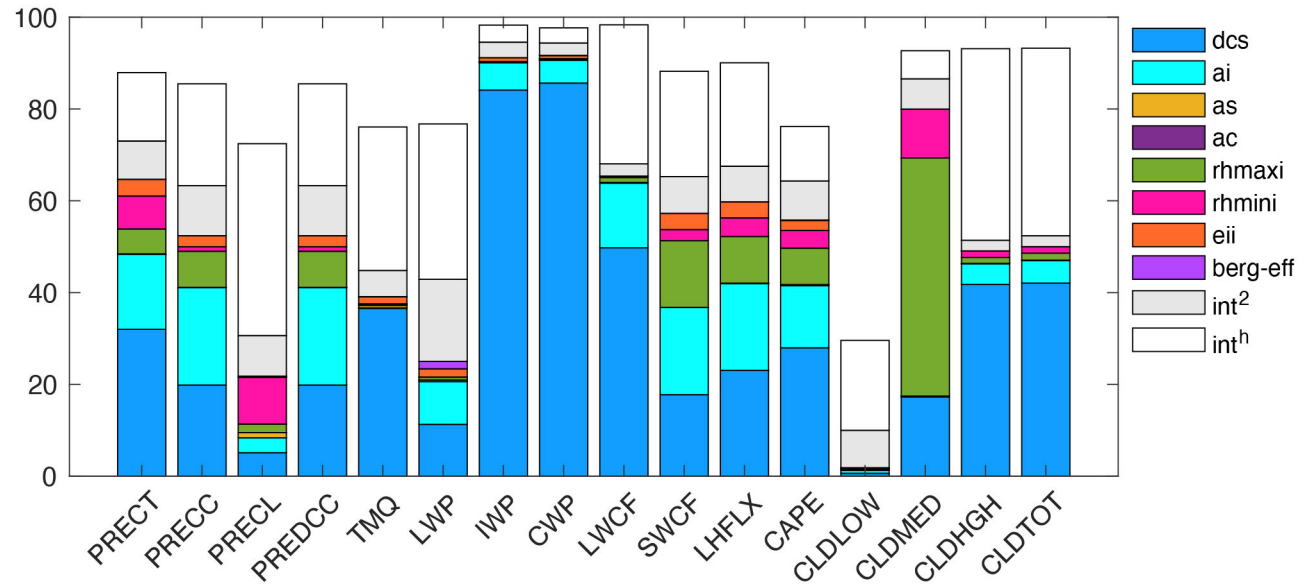



**FIGURE 4 |** Relative contributions of variance for individual parameters (main effect) to the overall variance of different climatic variables. The total contribution of variance from the second-order (two-way) interaction between parameters is indicated by the gray box in the bar plot ($Int^2$); the remainder of the contribution to variance from the higher-order (i.e., order > 2) interactions among other parameters is indicated by a white box for each variable.

# RESULTS AND DISCUSSION

## Relative Importance of Sensitive Parameters

**Figure 4** shows the contribution of first-order sensitivity for each parameter, the total second-order interactive effect ($int^2$) between any two-input parameters, and higher-order interactive effects ($int^h$) between more than two-input parameters, to the total explained variance for specific QoIs as resulting from the PCE surrogate model. The total variance explained by first-order and interactive effects ranges between 80 and 99% (**Figure 4**), as reflected by the $R^2$ values for different QoIs (**Figure 3**). In **Figure 4**, $int^2$ indicates the impact of a parameter on the model simulation depending on a second parameter, while $int^h$ corresponds to the impact of a parameter on the model simulations depending on more than two parameters. The contribution of first-order effects (in percent) to the total explained variance from different parameters are presented in **Table 2**. We found that *dcs* contribution is highest, about 32, 50, 18, 23, 11, 84, and 28% for PRECT, LWCF, SWCF, LHFLX, LWP, IWP, and CAPE, respectively. *dcs* also contributes about 17, 42, and 42% for the medium-level cloud (CLDMED), the high-level cloud (CLDHGH), and for the total cloud (CLDTOT), respectively. The total $int^2$ and $int^h$ effect contributed significantly to the total variance. For example, in PRECT, LWCF, SWCF, LHFLX, LWP, IWP, CAPE, CLDMED, CLDHGH, and CLDTOT, the $int^2$ was about 8, 3, 8, 7, 15, 4, 8, 7, 3, and 3%, respectively, while the corresponding $int^h$ values were about 15, 27, 20, 20, 30, 4, 10, 6, 43, and 43% (**Figure 4**).

In addition to the total $int^2$ effect, we also attempted to determine how effects induced by a single parameter may be amplified or suppressed by other parameters, compared to their OAT sensitivity (Zhao et al., 2013). As shown in **Figure 5**, we found that the interactive effects between any two-input parameters are not consequent (<3%) for most of the simulated variables, with the exception of the interactive effect between *dcs* and *ai*, which shows a prominent contribution of ∼4% to PRECT, 10% to LWP, and ∼4% to SWCF. Although the contributions of any two-parameter interactions are low relative to individual contributions, the large number of $int^2$ (28) makes their total $int^2$ contribution from all parameter pairs noteworthy (**Figure 4**).

In order to identify the overall most sensitive parameters, we quantified the total effects, including their first-order, $int^2$, and $int^h$ contributions (**Figure 6**). The total sensitivity for *dcs* ranges from 40 to 80%, *ai* from 15 to 30%, and the three-parameter combination (*rhmaxi*, *rhmini*, and *eii*) from 5 to 15% with respect to different simulated QoIs. The total sensitivity of some of these parameters is insensitive, for example, *as*, *ac*, and *berg_eff*, which contribute <3% to most simulated QoIs, with the exception of *as*, which contributed 2–5% to certain variables. Thus, *dcs*, *ai*, *rhmaxi*, *eii*, *rhmini*, and *as* were identified as the most influential parameters in CMP parameterization, arranged in decreasing order of their total sensitivity.

## Response of Simulated QoI to Sensitive Parameters

According to the parameter sensitivities results presented in section Relative Importance of Sensitive Parameters, we explain the response characteristics of different QoIs to different parameters based on the 3,937 SCAM6 simulations. The characteristics of the response of different QoIs to a parameter

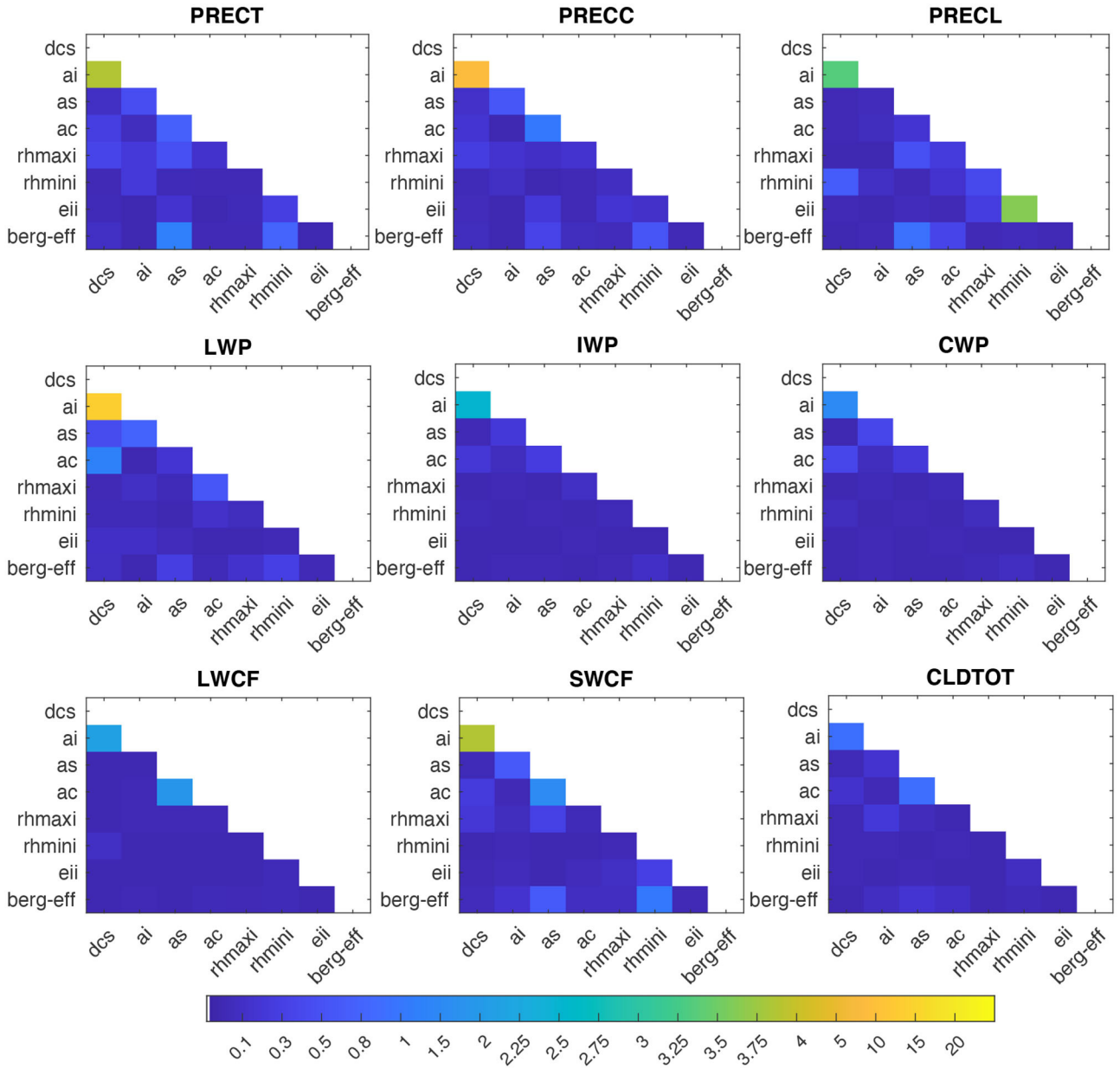



**FIGURE 5 |** Relative contributions (in %) from the interaction of a parameter with other parameters, to the total variance of PRECT, PRECC, and PRECL (in the first row), LWP, IWP, and CWP (in the second row), and LWCF, SWCF, and CLDTOT (in the third row), as estimated using the PC-fitted method.

are examined by analyzing the perturbation effect of a parameter while keeping the remaining parameter values fixed at their mean value (i.e., zero in a ±1 uncertainty range). We found that the substantial increase in CLDHGH (∼64 to 83%) and the partial increase in CLDMED (∼19 to 28%) following an increase in *dcs* (**Figure 7**) may have occurred because increasing *dcs* typically reduces the ice to snow conversion and thus increases the ice particle content of the upper atmosphere. Further, this increase in ice particles in the upper atmosphere following an increase in *dcs* leads to an increase in IWP (∼0.02 to 0.12 kg/m$^2$), LWCF (∼10 to 25 W/m$^2$), and SWCF (−26 to −33 W/m$^2$), and to a decrease in LWP (∼0.054 to 0.045 kg/m$^2$) (**Figure 8**). A decrease in PRECT (∼3.7 to 3.1 mm/day) in response to an increase in *dcs* is also noted (**Figure 9**), which indirectly alters (decreases) the stability of the atmosphere (CAPE; **Figure 7**) and thus the convective precipitation (PRECC; **Figure 9**), since changes occur in CMP but not in convection parameterizations. This phenomenon was also reported for different CMP parameters by Lin et al. (2016) and Pathak et al. (2020). In addition, the increase in *ai* caused more ice particles to fall, thereby decreasing CLDHGH (∼85 to 78%), LWCF (∼30 to 19 W/m$^2$) and SWCF (−35 to −29 W/m$^2$), and IWP (0.1 to 0.06 kg/m$^2$) (Mitchell et al., 2008), while causing

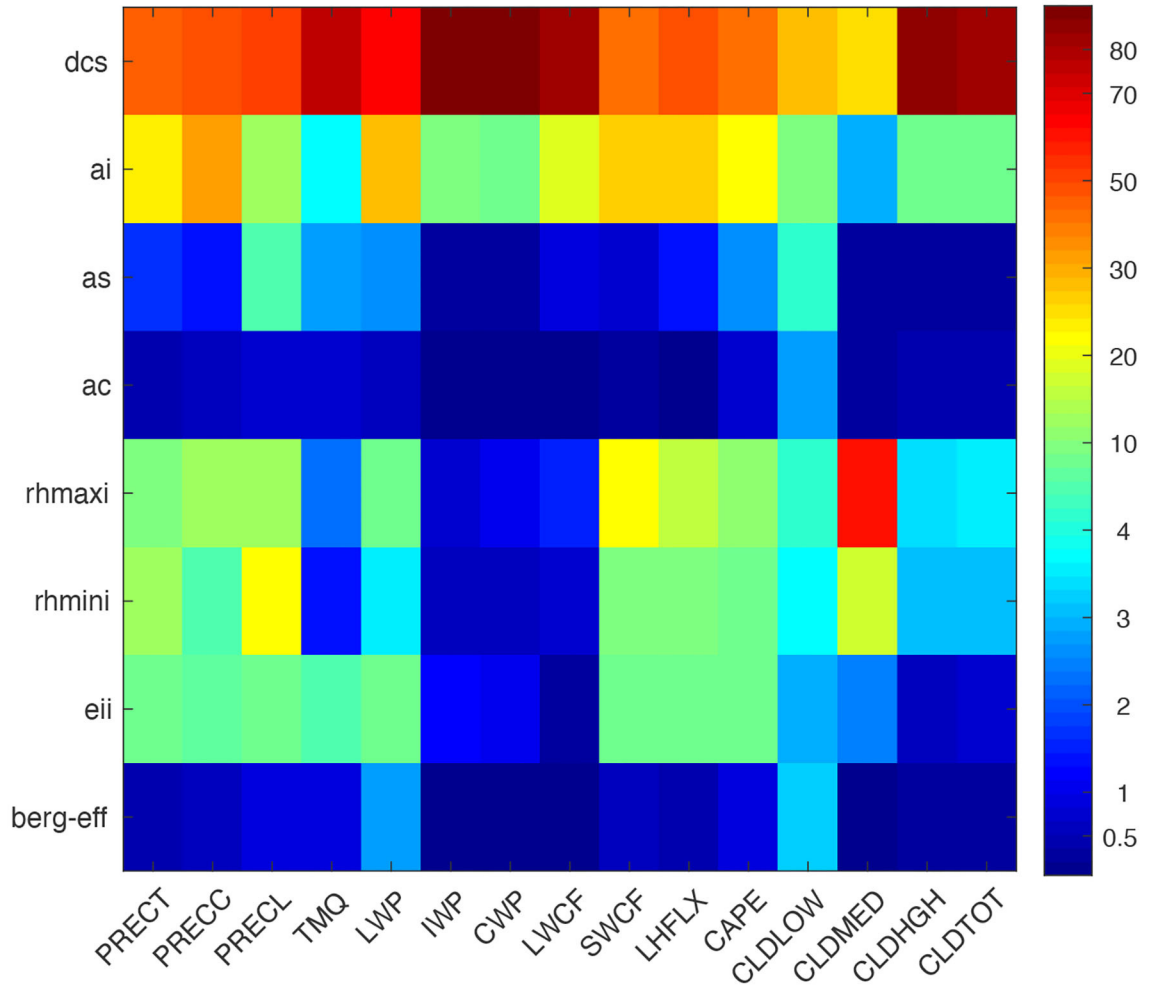



**FIGURE 6 |** Total contribution of each parameter to the total explained variance (in %) from PC-fitted outputs against SCAM6 simulations for various climatic variables. A larger number indicates greater significance.

an increase in LWP from 0.044 to 0.049 kg/m$^2$ (**Figures 7**, **8**). The *ai* also affects atmospheric instability, which typically increases with increasing *ai* (**Figure 7**), causing an increase in PRECT from ∼3.1 to 3.4 mm/day (**Figure 9**; Sun et al., 2011; Yang et al., 2013; Pistotnik et al., 2016). The effects of *ai* on LWCF are small when compared to the effect of *dcs*; this could reflect the small changes in CLDHGH that cause an increase in *ai*. In general, an increase in *dcs* increases the concentrations of ice clouds, whereas an increase in *ai* decrease ice cloud concentrations, hence the increased and decreased albedo effects leading, respectively, to increasing and decreasing SWCF.

Further, the increase in *rhmaxi* decreases both the grid box of full ice cloud cover (i.e., 100%) and clouds that reach supersaturation, leading to a large decrease in CLDMED (30 to 16%). A substantial decrease in CLDHGH was also found; however, due to lower temperatures in the upper atmosphere compared to the mid-level atmosphere, supersaturation was somewhat counterbalanced. The moderate increase in PRECT from 3.3 to 3.45 mm/day and decrease in SWCF from −31 to −29 W/m$^2$ were thus the main reasons behind the observed increase in *rhmaxi* (**Figures 7**–**9**; Xie et al., 2018). Additionally, the increase in *rhmini* reduced the fractional ice cloud cover formation and thus somewhat increased the presence of liquid clouds, as shown by the increase in CLDMED from 19 to 26% (**Figure 7**). We do not show here the response for CLDLOW since this was relatively insensitive to CMP parameters; its variation could be due to the control of frontal clouds and precipitation over the SGP site in addition to local convection (Qian et al., 2015).

Furthermore, we evaluated the response characteristics of the vertical distribution of different variables in **Figure 10** and the response characteristics of their vertical averages (from the surface to 100 hPa) in **Figure 11**. In **Figure 10**, the color bar values for each panel plot are not shown; however, the highest to lowest values are indicated by the dark yellow to the dark blue colors of the contour plot. This suggests that the average value of (a) snow concentration (maximum at ∼400 hPa with a spread between 600 and 200 hPa; **Figure 10**) decreases with increasing

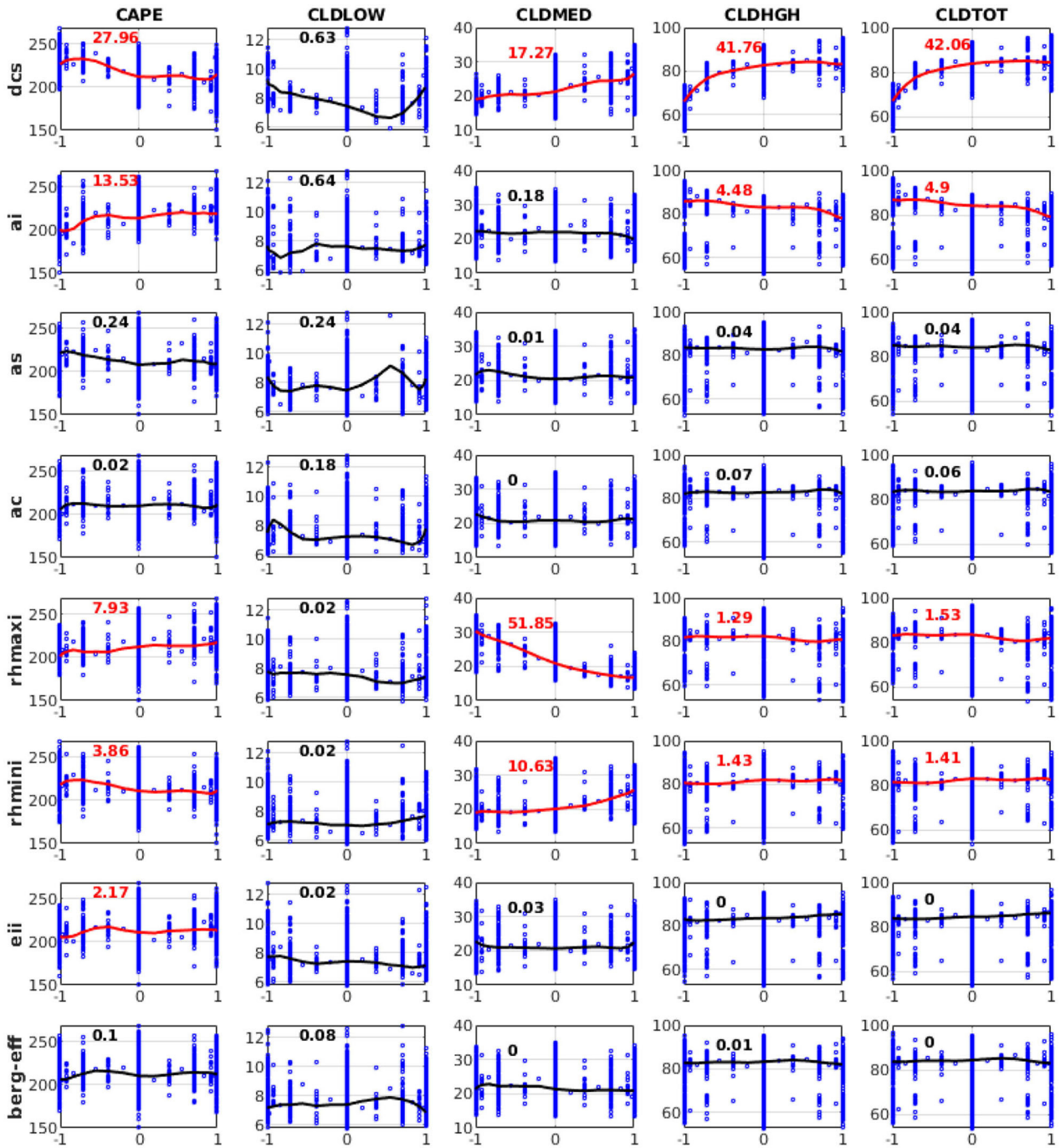



**FIGURE 7 |** Variation of convective available potential energy (CAPE), low-level cloud (CLDLOW), medium-level cloud (CLDMED), high-level cloud (CLDHGH), and total cloud cover (CLDTOT) in response to perturbations of eight different parameters from 3,937 SCAM6 simulations. The numbers in each plot represent the relative contributions (in %) of each input parameter perturbation to the overall variation of different variables. Red indicates that the contribution is significant (*F*-test) at the 95% significance level. The line in each plot corresponds to the response effect of the perturbation of a single parameter, keeping all other parameters at the zero (central) position of the parameter uncertainty range for the simulation of that variable. Vertical bars show the range in the variation of values of a particular variable in response to the perturbation of other parameters.

*dcs* for the first half of the parameter space, and (b) cloud liquid concentration (primarily concentrated between the surface and 600 hPa; **Figure 11**) significantly changes non-linearly relative to *dcs*; generally, it decreases for the first 45% of the parameter space and increases for the last 45% of the parameter space) and to *as* (sensitive only to the last 50% of the parameter

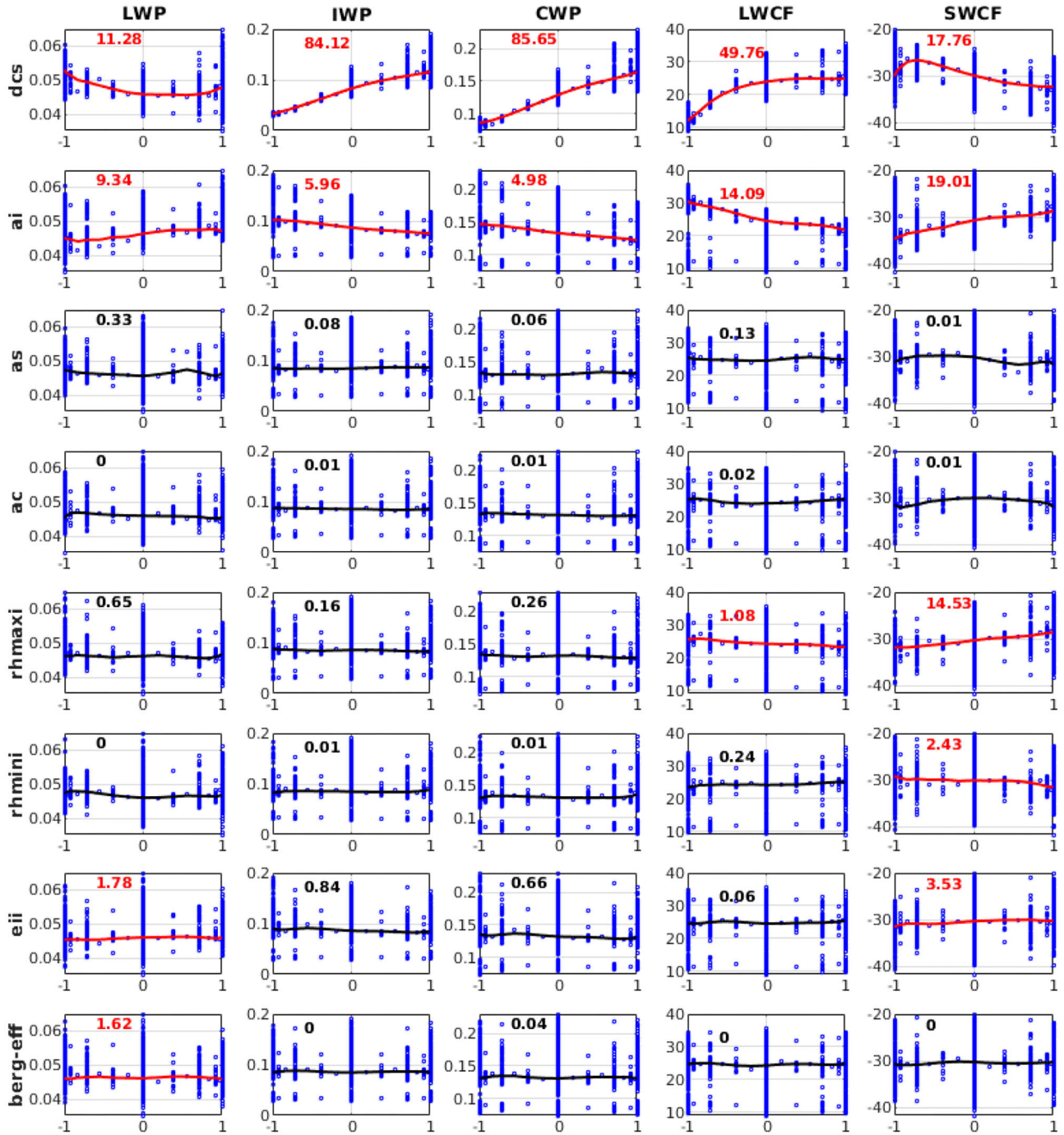



**FIGURE 8 |** Variation of liquid water path (LWP), ice water path (IWP), cloud water path (CWP), longwave cloud radiative effect (LWCF), and shortwave cloud radiative effect (SWCF) in response to perturbations of eight different parameters from 3,937 SCAM6 simulations. The numbers in each plot represent the relative contributions (in %) of each input parameter perturbation to the overall variation of different variables. Red indicates that the contribution is significant (*F*-test) at the 95% significance level. The line in each plot corresponds to the response effect of the perturbation of a single parameter when all other parameters at maintained at the zero (central) position of their uncertainty range for the simulation of that variable. Vertical bars denote the range of variation in values of a particular variable in response to the perturbation of other parameters.

space, initially increasing but later decreasing) (**Figure 11**). Furthermore, we note that the vertical average of cloud ice concentration (maximum at about 225 hPa with a spread from 300 to 150 hPa) increases significantly due to increasing *dcs* but decreases with increasing *ai*. The rate at which cloud condensate is converted to precipitation (maximum at ∼450 hPa with a

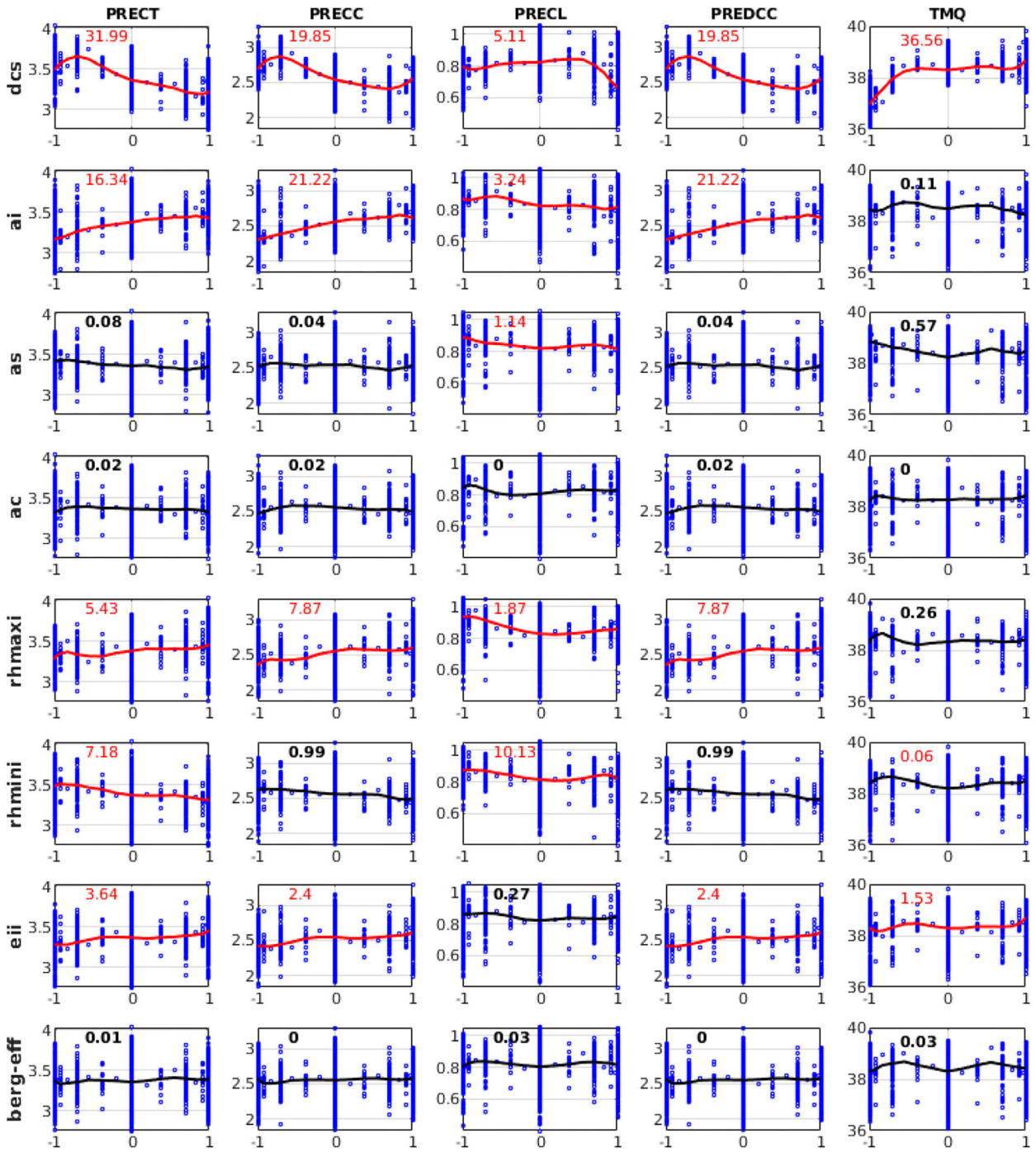



**FIGURE 9 |** Variation of PRECT, PRECC, PRECL, deep convective precipitation (PREDCC), and total precipitable water (TMQ) in response to perturbations of eight different parameters from 3,937 SCAM6 simulations. The numbers in each plot represent the relative contributions (in %) of each input parameter perturbation to the overall variation of different variables. Red denotes that the contribution is significant (*F*-test) at the 95% significance level. The line in each plot corresponds to the response effect of the perturbation of a single parameter when keeping all other parameters at the zero (central) position of their uncertainty ranges for the simulation of that variable. Vertical bars show the range of variation for values of a particular variable in response to the perturbation of other parameters.

spread from 300 to 700 hPa; **Figure 10**) decreases significantly with increasing *dcs* (**Figure 11**). The vertical averages of longwave cloud radiative effect (maximum at about 350 hPa and 100 hPa) and shortwave cloud radiative effect (centered at about 250 hPa) are primarily sensitive to *dcs* and *ai* in the first 60% of the parameter space. In general, shortwave and longwave cloud

radiative effect increase with increasing *dcs* and decrease with increasing *ai* (**Figure 11**). Additionally, the vertical average of relative humidity (maximum at about 300 hPa with a spread from 500 to 200 hPa) increases monotonically due to increasing *dcs*, *rhmaxi*, and *rhmini*, and decreases due to increasing *ai*. The vertical average of specific humidity, which is primarily distributed from the surface to 800 hPa, increases with increasing *dcs* and decreases with increasing *as* within the first 45% of parameter space; it also decreases monotonically with increasing *ai* in the last 60% of the parameter space (**Figure 11**). Thus, the increase in radiative effect with increasing *dcs* supports our finding of an increase in LWCF and SWCF, as well as the increase in ice cloud concentration with increasing *dcs*, which supports the substantial increase in IWP. Finally, the decrease in the rate of conversion of cloud condensate to precipitation with increasing *dcs* supports the corresponding reduction in PRECT through a reduction in its instability (**Figures 7**, **9**).

## Bayesian Inference

Finally, we analyze the Bayesian inference results for estimating CMP parameters. The MCMC algorithm (Metropolis et al., 1953) was used to generate sample chains for the different CMP parameters. We have run five times, each for 10,000 iterations (or samples), with different initial parameter values for efficient samples mixing and convergence. The PCE surrogate model was used to evaluate the cost function for each MCMC chain to reduce the computing requirements. **Figure 12** shows the trace plot of the CMP parameters (chains vs. MCMC iteration number), indicating a well-mixed chain. This implies that the distribution of the chains will remain unchanged with further sampling and converged to a stationary distribution. The average acceptance rate of MCMC samples was found to be 35%. From **Figure 12**, we find that the parameter samples of *dcs*, *ai*, *as*, *ac*, *rhmaxi*, *rhmini*, *berg_eff* move around their default CMP values. The parameter samples for *eii*, on the other hand, did not move around the default value, but rather skewed leftward to the default value to cover a larger part of the prior range, resulting in a wider posterior.

Further, we computed the marginalized posterior distribution from the MCMC chains by discarding the first 500 iterations (to begin with a good starting point of MCMC run) using kernel density estimates (KDE; Silverman, 1986). The marginalized posterior PDF for each parameter along the diagonal of the eight-by-eight matrix shows that the posterior PDFs of *dcs*, *ai*, *as*, *rhminl*, and *berg_eff* exhibit a sharp increase but are skewed leftward for *dcs* and rightward for *ai* and *as* relative to the default CMP value (**Figure 13**). Thus, with respect of the default values for *dcs*, *ai*, and *as* (0.0005, 700, and 11.72), their posterior mean estimates are found to be 0.0004, 903, 14.85, respectively, while those of *rhminl* and *berg_eff* are found to be 0.79 and 1.03, respectively, close to their default values of 0.80 and 1.0. The posterior mean estimates of these parameters are also found falling within the acceptable ranges (i.e., within the 95% of the intervals of high posterior probability). The posterior PDFs of *ac*, *rhmaxi*, and *eii* do not show a sharp peak and are mostly flat, consistent with our conclusions from their chains (**Figure 12**) and indicating less-informative posteriors, which is likely due to the lack of relevant observations.

The examination of 1D posterior PDFs provides only a measure of the integrated influence of parameters on model output; specific functional dependence is hidden. Thus, in addition to 1D posterior PDFs, we also show 2D PDFs, which facilitate parameter space visualization and the identification of optimized parameter sets. We found that the cost function for a parameter choice is strongly dependent on the other parameter chosen (Posselt, 2016). For example, the 2D PDF for *rhmaxi* and *as*, which exhibits a linear response, suggests an increase in cost function with an increase in both *rhmaxi* and *as*. We also found some parameters whose two-dimensional PDFs do not show any significant variations with changes in the other parameters, indicating that the cost function is functionally independent (e.g., the 2D PDF of *eii* does not show any significant changes when other parameters are changed).

To examine the extent to which the parameters posterior distributions inferred by the proposed UQ framework provide useful information, we assessed whether the posterior mean parameters can improve a global climate model performance. We therefore performed two sets of 7 year global climate simulations; one using the default parameters, and the other using the the posterior mean parameters. Both simulations are conducted using the CESM2 model with prescribed observed climatological sea surface and sea-ice temperatures. The simulation with posterior mean parameters suggests an overall improvement of ∼7% over the globe in annual means, with the largest improvement of ∼15% over the global land for December-February (**Figure 14**). The simulation of individual climate variables (e.g., total precipitation, cloud distribution, cloud radiative effect, and humidity) are improved by 5 to 50% over the globe, as well as over the tropical and sub-tropical regions, across the different periods (i.e., annual, June-August, and December-February), in comparison to the default model simulation (**Figure 14**). However, over the polar region, some of the simulated variables with the posterior mean parameters values have deteriorated by 5 to 20% compared to the default model simulation. In our setting, posterior parameters distributions were inferred from data collected in subtropics at the Southern Great Plains (36°N and 97°W; Zhang et al., 2016), and as such the resulting parameters are calibrated for such regions, which also explains the improvement we obtained in the subtropics and tropics. Recent results (Pathak et al., 2020) suggest a spatial dependency of some of the considered parameters despite being prescribed as constant values across the globe. Our results over the polar regions also support the use of spatially varying cloud parameters in climate models, which will be explored in depth in our future work.

# CONCLUSIONS

This study used an efficient multi-objective UQ framework to assess the sensitivity of NCAR SCAM6 outputs to cloud microphysics and macrophysics (CMP) parameterization schemes. The framework involved building a surrogate model

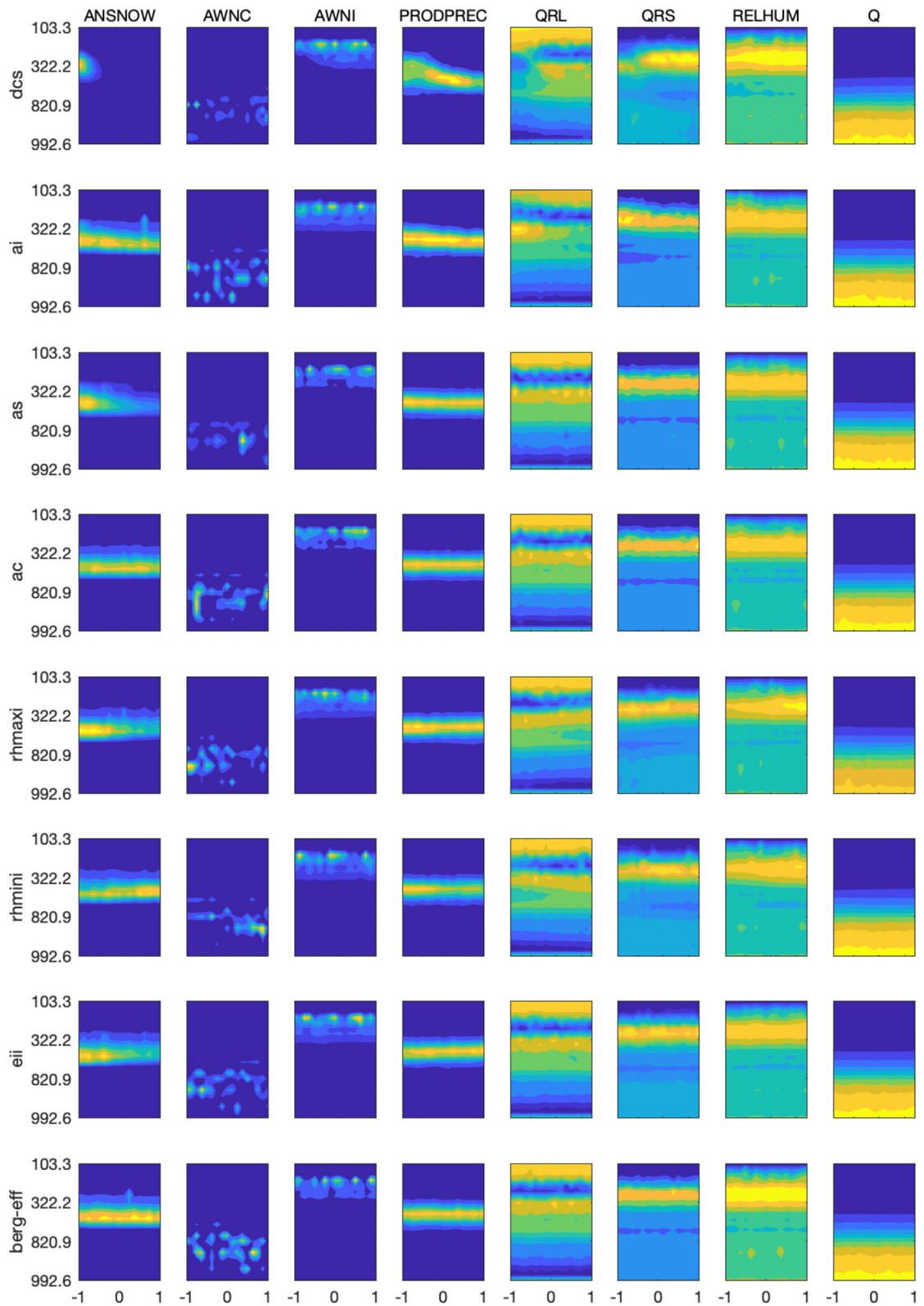



**FIGURE 10 |** Parameter perturbation response to the vertical distribution of snow concentration (ANSNOW), cloud water concentration (AWNC), ice concentration (AWNI), precipitation production (PRODPREC), longwave heating rate (QRL), shortwave heating rate (QRS), relative humidity (RELHUM), and specific humidity (Q).

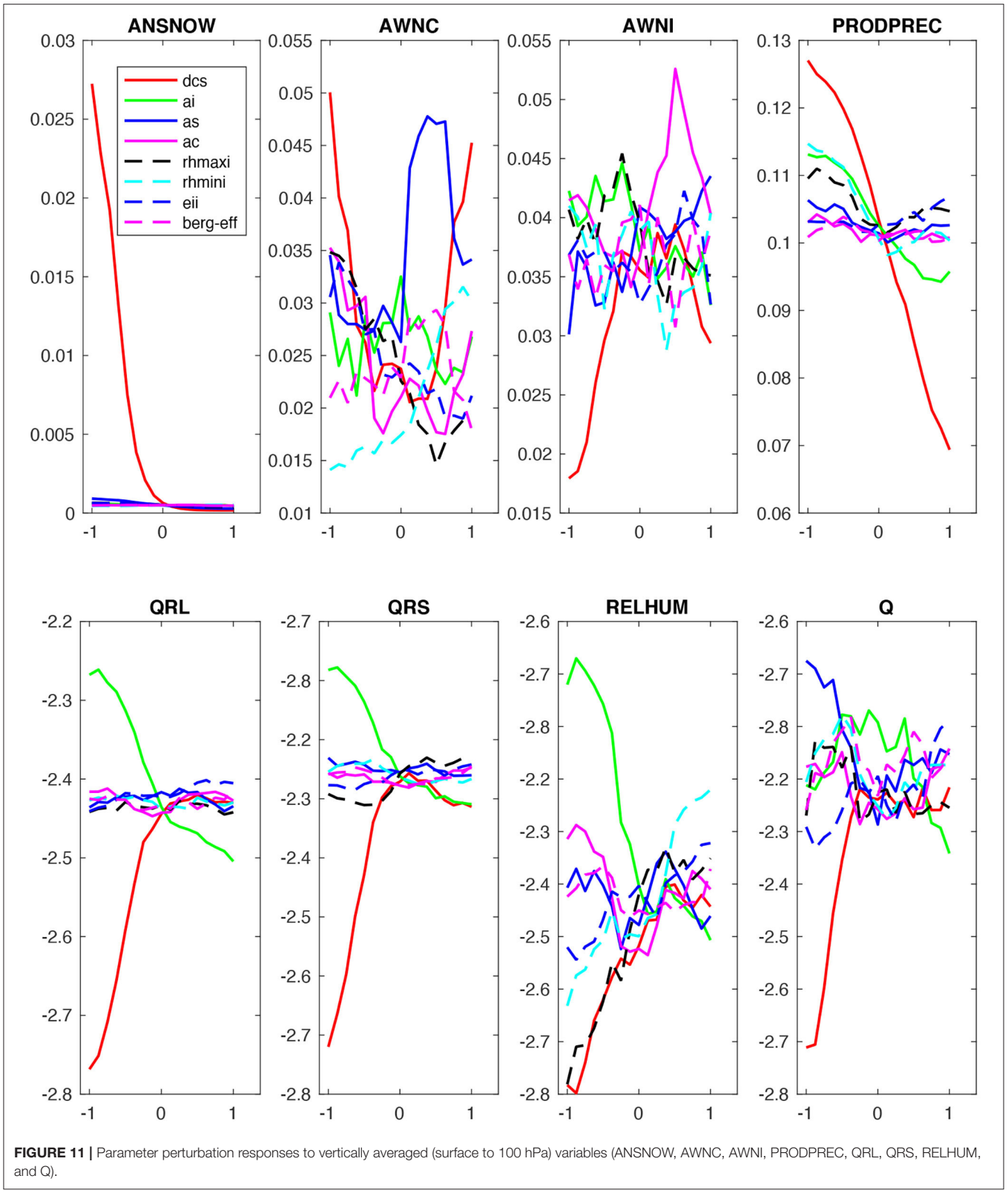



**FIGURE 11 |** Parameter perturbation responses to vertically averaged (surface to 100 hPa) variables (ANSNOW, AWNC, AWNI, PRODPREC, QRL, QRS, RELHUM, and Q).

using a polynomial chaos expansion, sensitivity analysis, and Bayesian inference to identify and quantify the uncertainties of various (quantities of interest) QoIs from the NCAR SCAM6 model associated with CMP parameterizations. The basis pursuit denoising (BPDN) approach was applied to build the polynomial chaos expansion (PCE) model to mitigate for internal noise

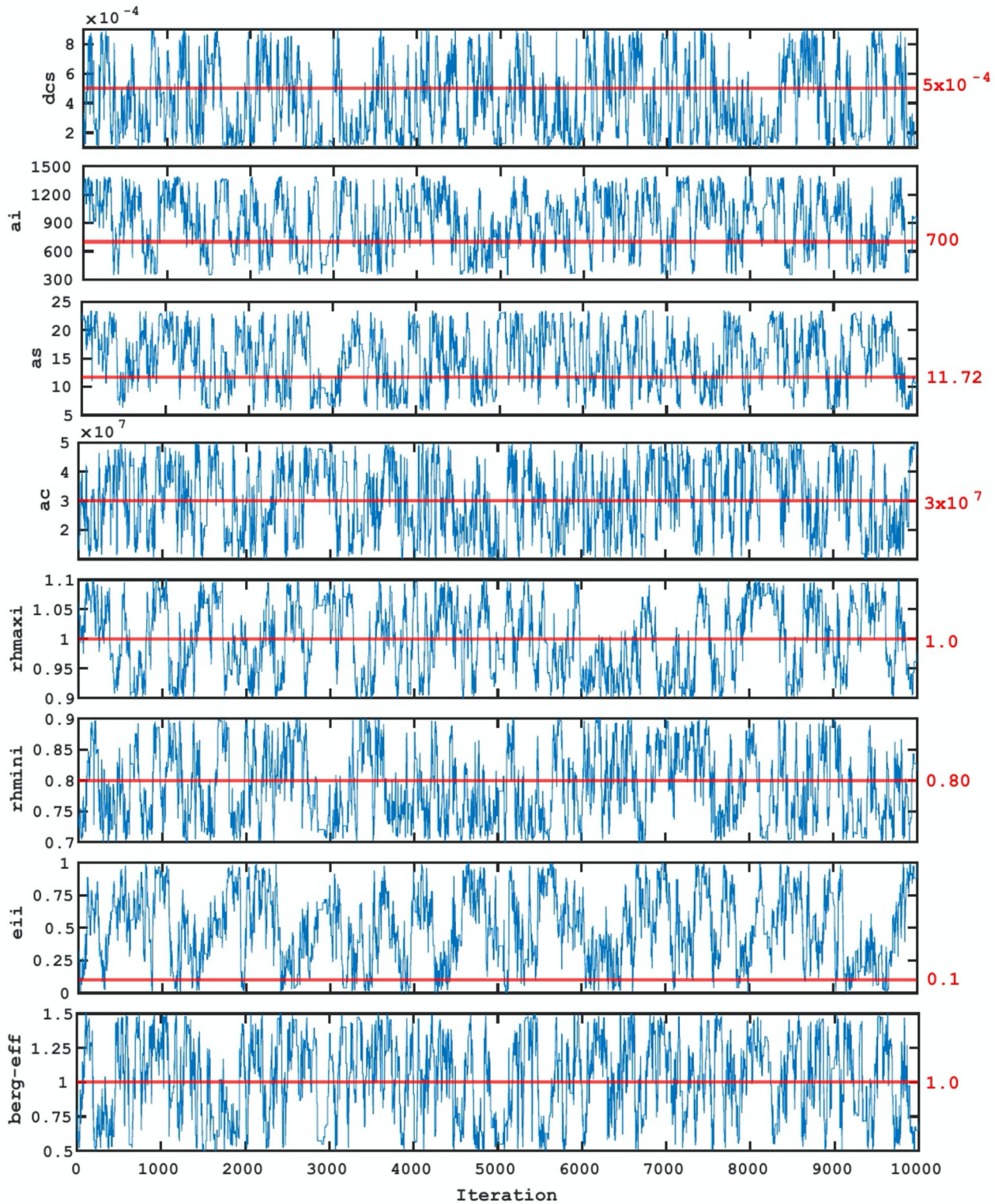



**FIGURE 12 |** Chains of parameters from MCMC samples. The horizontal red line (and the value shown in red) indicate the default parameter value used in NCAR SCAM6.

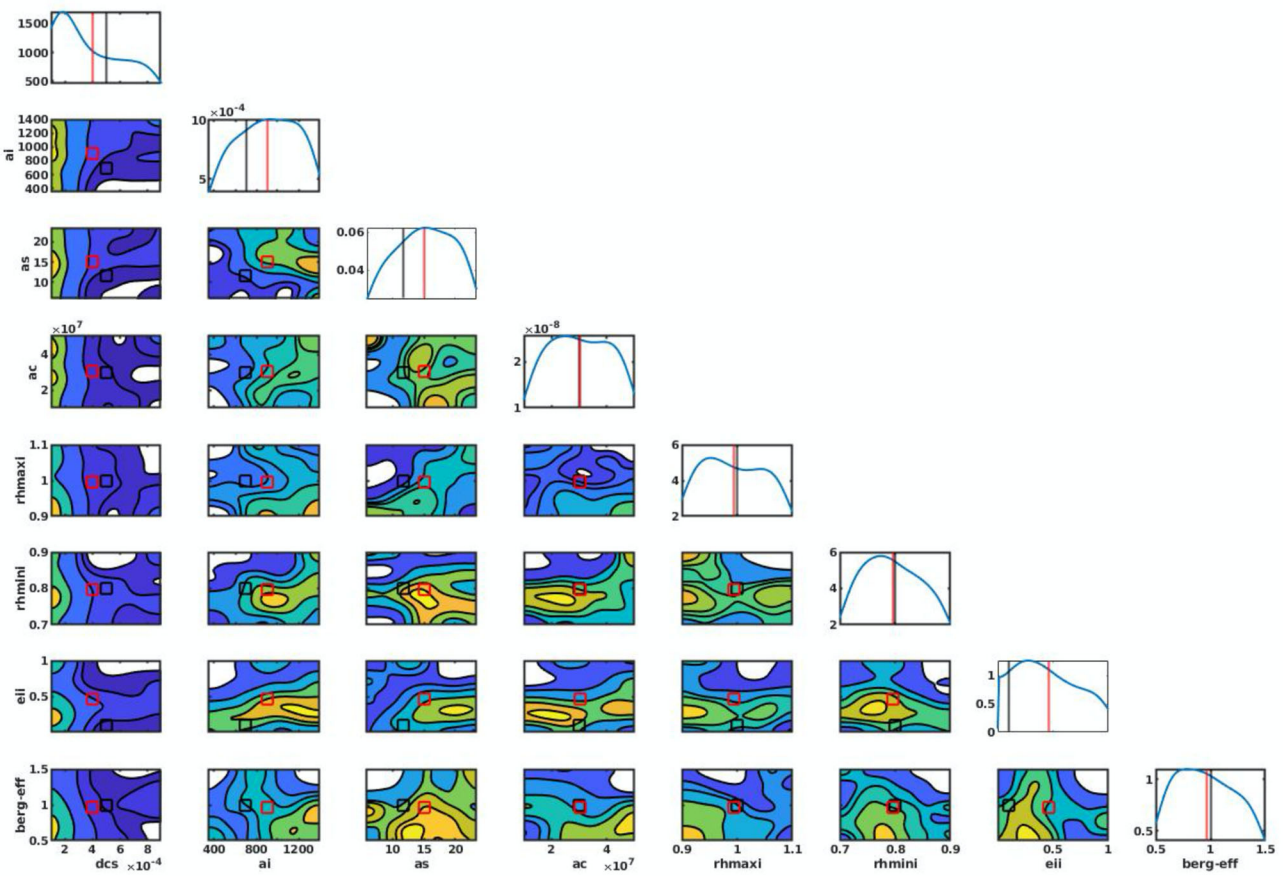



**FIGURE 13 |** Probability Distribution Functions (PDFs) of parameters using KDE along the diagonal of the matrix. The contour plots show the joint PDFs of two parameters. The black and red vertical lines show the default and posterior mean parameter values.

in model simulations. The surrogate model was shown to realistically approximate the model outputs, explaining most of the variance of different QoIs, with the highest explained variance (about 99%) for longwave cloud radiative effect (LWCF) and the lowest explained variance (about 79%) for convective available potential energy (CAPE).

The UQ analysis suggested that the simulated QoIs are most sensitive to five (*dcs*, *ai*, *rhmaxi*, *rhmini*, and *eii*) of the eight CMP parameters. *dcs* alone contributed 40–80% of the total variance of different simulated QoIs, *ai* 15–30%, and *rhmaxi*, *rhmini*, and *eii* each contributed 5–15%. Parameters *as*, *ac*, and *berg_eff* were found to be the least sensitive, each of which contributed <5% to the total variance. Further splitting the total sensitivity effect of a parameter into first-, second-, and higher-order sensitivities, we found that the first-order sensitivity of *dcs*, *ai*, *rhmini*, *rhmaxi*, and *eii* together contribute more than 60% of the total variance in different QoIs. While, the second- and higher-order sensitivity of *dcs*, *ai*, *rhmini*, *rhmaxi*, and *eii* together contribute about 7 and 20% of the total variance in different QoIs. Some previous studies using the predecessor version of this model have also argued that *dcs* and *ai* are the most sensitive parameters for the simulation of total precipitation over the tropical region (He and Posselt, 2015; Qian et al., 2015; Zhang, 2015; Pathak et al., 2020). Other studies that employed different models have also suggested *dcs* and *ai* to be highly sensitive to cloud distribution (e.g., Bony and Dufresne, 2005; Sanderson et al., 2008; Golaz et al., 2011; Gettelman et al., 2012). The sensitivity of *rhmaxi* to cloud distribution was also reported in the E3SM model (Qian et al., 2018; Xie et al., 2018). Uncertainties in these parameters are considered as important factors in climate model sensitivity (Zelinka et al., 2020).

We also found that an increase in *dcs* increases the concentration of ice between 300 and 150 hPa (and *vice versa* for snow concentrations) by decreasing the conversion rate of ice to snow, which causes a substantial increase in cloud cover. This leads to significant increase in ice water path (IWP), LWCF, and SWCF. Furthermore, the increase in longwave and shortwave heating rates due to increased *dcs* were shown to sustain the corresponding increases in LWCF and SWCF. A significant decrease in total precipitation (PRECT) occurred due to an increase in *dcs*, which indirectly affected (decreased) the conversion of cloud condensate to precipitation; this was clear from the reduced instability, i.e., the CAPE and reduced convective precipitation. The decrease in the conversion of cloud

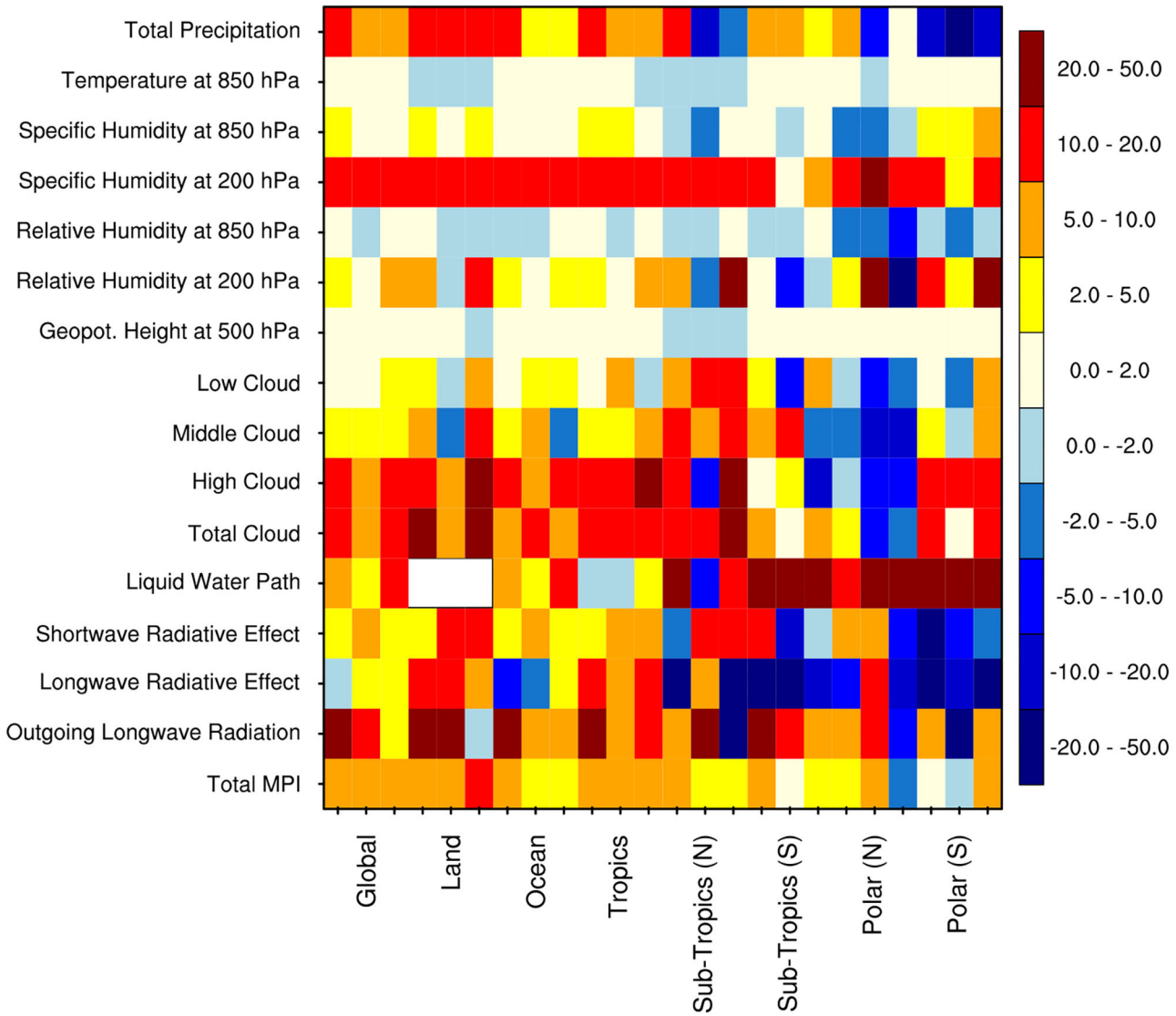



**FIGURE 14 |** The Model Performance Index (MPI in percent) resulting from simulations of various climate variables over the global, tropics (30°N−30°S), norther subtropics (30–60°N), southern sub-tropics (30–60°S), northern polar (60–90°N), and southern polar (60–90°S) regions, across different periods [i.e., Annual (ANN), June-August (JJA), December-February (DJF)]. The MPI is constructed using a metric developed by Zhang (2015). The contours in color show the percentage improvement (positive values) or deterioration (negative values) for different climate variables from the simulation with posterior mean parameters with respect to the default model simulation. Total MPI—model performance as averaged over all climate variables. Following observations and reanalysis data are used for construction of this metrics, the Global Precipitation Climatology Project (Adler et al., 2003) for total precipitation, the Clouds and Earth's Radiant Energy System-Energy Balanced and Filled (Loeb et al., 2009) project for shortwave and longwave radiative effect, the National Aeronautics and Space Administration Water Vapor Project (Randel et al., 1996) for liquid water path over ocean, and the International Satellite Cloud Climatology Project (Young et al., 2018) for low, middle, high, and total cloud, the ECMWF Interim reanalysis (Dee et al., 2011) for humidity, temperature, and geopotential height.

condensate to precipitation due to an increase in *dcs* also reduced the liquid water path (LWP), as noticed from reduced cloud water concentrations. The possible physical mechanism behind the sensitivity of QoIs with respect to *ai* values seems to work oppositely to *dcs*. In addition, an increase in relative humidity occurred in response to increased *dcs*, *rhmaxi*, and *rhmini*. A decrease (increase) in cloud water concentration in response to an increase in *rhmaxi* (*rhmini*) was also reported. The impact of *rhmaxi* and *rhmini* on PRECT was identified through studying its effects on middle clouds (CLDMED). Generally, CLDMED decreases (increases) due to increasing *rhmaxi* (*rhmini*).

The Bayesian inference results for vertical cloud distribution showed that the Markov chain Monte Carlo (MCMC) chains for *dcs*, *ai*, *as*, *ac*, *rhmaxi*, *rhmini*, and *berg_eff* fluctuated around their default CMP values. However, for *eii*, the MCMC chains did not fluctuate around the default value but were skewed

largely leftward; suggesting a non-informative cost function. The marginalized posterior distributions of the MCMC chains for *dcs*, *ai*, *as*, *rhminl*, and *berg_eff* showed well-defined peaks. The corresponding posterior mean estimate values were, respectively, found to be 0.0004, 903, 14.85, 0.79 and 1.03, compared to the corresponding default values of 0.0005, 700, 11.72, 0.80 and 1.0. The marginalized posterior distributions for *ac*, *rhmaxi*, and *eii* were mostly flat, indicating the probability of being nearly uniform across the parameter space; consequently, they do not provide meaningful information to estimate their posterior means. Global climate simulations performed using the posterior mean and default parameters have shown that the climate simulations using the posterior mean parameters values suggest an overall-improvement of ∼7% over the globe in annual means. The largest improvement of ∼15% was obtained over the global land for December-February. Specifically, the climate simulations using the posterior mean parameters values were improved over the tropical and sub-tropical regions, whereas it has deteriorated over the polar region. In line to the findings of Pathak et al. (2020), we anticipate that this deterioration in polar regions could be due to a spatial dependency of some of the considered parameters despite being prescribed as constant values across the globe in the climate models. Furthermore, the functional linear dependence of *rhmaxi* on *as* was also noted from the joint probability distribution of both parameters. This provides important information for understanding cloud processes and their associated physical processes and will assist developers in further improving climate models.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

RP, HD, AS, SS, SM, OK, and IH identified the problem and designed the work to meet the objective. RP, HD, SE, and AS conducted the analysis. RP, HD, SE, AS, SS, SM, OK, and IH wrote the manuscript using the analysis of RP, HD, SE, and AS. All authors contributed to the article and approved the submitted version.

## FUNDING

The research reported in this paper was supported by the office of Sponsor Research (OSR) at King Abdullah University of Science and Technology (KAUST) under the Virtual Red Sea Initiative (REP/1/3268-01-01) and the Saudi ARAMCO Marine Environmental Research Center at KAUST.

## ACKNOWLEDGMENTS

The DST Center of Excellence in Climate Modeling (RP03350) and Indian Institute of Technology Delhi, India, are acknowledged for their partial support. We thank NCAR for providing the Single-Column Community Atmosphere Model (SCAM).

## REFERENCES

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., et al. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeorol.* 4, 1147–67. doi: 10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2

Albrecht, B. A., Randall, D. A., and Nicholls, S. (1988). Observation of marine stratocumulus clouds during FIRE. *Bull. Amer. Meteor. Soc.* 69, 618–636. doi: 10.1175/1520-0477(1988)069<0618:OOMSCD>2.0.CO;2

Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R., and Delworth, T. L. (2000). Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature* 407, 617–620. doi: 10.1038/35036559

Anand, A., Mishra, S. K., Sahany, S., Bhowmick, M., Rawat, J. S., and Dash, S. K. (2018). Indian summer monsoon simulations: usefulness of increasing horizontal resolution, manual tuning, and semi-automatic tuning in reducing present-day model biases. *Sci. Rep.* 8:3522. doi: 10.1038/s41598-018-21865-1

Bastos, L. S., and O'Hagan, A. (2009). Diagnostics for Gaussian process emulators. *Technometrics* 51, 425–438. doi: 10.1198/TECH.2009.08019

Beljaars, A. C. M., Brown, A. R., and Wood, N. (2004). A new parametrization of turbulent orographic form drag. *Q. J. Roy. Meteor. Soc.* 130, 1327–1347. doi: 10.1256/qj.03.73

Betts, A. K., and Miller, M. J. (1986). A new convective adjustment scheme. Part II: SINGLE column tests using GATE wave, BOMEX, and arctic air-mass data sets. *Q. J. R. Meteor. Soc.* 112, 693–709. doi: 10.1002/qj.49711247308

Bogenschutz, P. A., Gettelman, A., Morrison, H., Larson, V. E., Craig, C., and Schanen, D. P. (2013). Higher-order turbulence closure and its impact on climate simulations in the community atmosphere model. *J. Clim.* 26, 9655–9676. doi: 10.1175/JCLI-D-13-00075.1

Bogenschutz, P. A., Gettelman, A., Morrison, H., Larson, V. E., Schanen, D. P., Meyer, N. R., et al. (2012). Unified parameterization of the planetary boundary layer and shallow convection with a higher-order turbulence closure in the Community Atmosphere Model: single-column experiments. *Geosci. Model Dev.* 5, 1407–1423. doi: 10.5194/gmd-5-1407-2012

Bony, S., and Dufresne, J. L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.* 32:L20806. doi: 10.1029/2005GL023851

Bony, S., Stevens, B., Frierson, M. W., Jakob, C., Kageyama, M., Pincus, R., et al. (2015). Clouds, circulation and climate sensitivity. *Nat. Publ. Gr.* 8, 261–268. doi: 10.1038/ngeo2398

Brown, S. J., Murphy, J. M., Sexton, D. M. H., and Harris, G. R. (2014). Climate projections of future extreme events accounting for modelling uncertainties and historical simulation biases. *Clim. Dyn.* 43, 2681–2705. doi: 10.1007/s00382-014-2080-1

Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., et al. (2013). Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature* 503, 67–71. doi: 10.1038/nature12674

Chen, S., and Donoho, D. (1994). "Basis pursuit," in *Proceedings of 1994 28th Asilomar Conference on Signals, Systems, and Computers* (Pacific Grove, CA). doi: 10.1109/ACSSC.1994.471413

Collins, M., Booth, B. B., Bhaskaran, B., Harris, G. R., Murphy, J. M., Sexton, D. M. H., et al. (2011). Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. *Clim. Dyn.* 36, 1737–1766. doi: 10.1007/s00382-010-0808-0

Constantine, P. G., Eldred, M. S., and Phipps, E. T. (2012). Sparse pseudospectral approximation method. *Comput. Methods Appl. Mech. Eng.* 229–232, 1–12. doi: 10.1016/j.cma.2012.03.019

Covey, C., Lucas, D. D., Tannahill, J., Garaizar, X., and Klein, R. (2013). Efficient screening of climate model sensitivity to a large number of perturbed input parameters. *J. Adv. Model. Earth Syst*. 5, 598–610. doi: 10.1002/jame.20040

Crestaux, T., Le Maitre, O. P., and Martinez, J.-M. (2009). Polynomial chaos expansion for sensitivity analysis. *Reliab. Eng. Syst. Saf*. 94, 1161–1172. doi: 10.1016/j.ress.2008.10.008

Danabasoglu, G., Lamarque, J. -F., Bacmeister, J., Bailey, D. A., and DuVivier, A. K., Edwards, J., et al. (2020*)*. The Community Earth System Model Version 2 (CESM2). *J. Adv. Model. Earth Syst*. 12, 1–35. doi: 10.1029/2019MS001916

Decremer, D., Chung, C. E., and Raisanen, P. (2015). Strategies for reducing the climate noise in model simulations: ensemble runs versus a long continuous run. *Clim. Dyn*. 44, 1367–1379. doi: 10.1007/s00382-014-2161-1

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. Meteor. Soc*. 137, 553–597. doi: 10.1002/qj.828

Ferrier, B. S. (1994). A double-moment multiple-phase four-class bulk ice scheme. Part I: description. *J. Atmos. Sci*. 51, 249–280.

Fridlind, A. M., Ackerman, A. S., Chaboureau, J.-P., Fan, J., Grabowski, W. W., Hill, A. A., et al. (2012). A comparison of TWP-ICE observational data with cloud-resolving model results. *J. Geophys. Res*. 117:D05204. doi: 10.1029/2011JD016595

Gettelman, A., Kay, J. E., and Shell, K. M. (2012). The evolution of climate sensitivity and climate feedbacks in the community atmosphere model. *J. Clim*. 25, 1453–1469. doi: 10.1175/JCLI-D-11-00197.1

Gettelman, A., Liu, X., Ghan, S. J., Morrison, H., Park, S., Conley, A. J., et al. (2010*)*. Global simulations of ice nucleation and ice supersaturation with an improved cloud scheme in the Community Atmosphere Model. *J. Geophys. Res. Atmos*. 115, 1–19. doi: 10.1029/2009JD013797

Gettelman, A., and Morrison, H. (2015). Advanced two-moment bulk microphysics for global models. Part I: off-line tests and comparison with other schemes. *J. Clim*. 28, 1268–1287. doi: 10.1175/JCLI-D-14-00102.1

Gettelman, A., Morrison, H., Santos, S., Bogenschutz, P., and Caldwell, P. M. (2015). Advanced two-moment bulk microphysics for global models. Part II: global model solutions and aerosol-cloud interactions. *J. Clim*. 28, 1288–1307. doi: 10.1175/JCLI-D-14-00103.1

Gettelman, A., Truesdale, J. E., Bacmeister, T., Caldwell, P. M., Neale, R. B., Bogenschutz, P. A., et al. (2019). The Single Column Atmosphere Model Version 6 (SCAM6): not a scam but a tool for model evaluation and development. *J. Adv. Model. Earth Syst*. 11, 1381–1401. doi: 10.1029/2018MS001578

Ghanem, R. G., and Spanos, P. D. (1991). *Stochastic Finite Elements: A Spectral Approach*. New York, NY: Springer-Verlag, 214. doi: 10.107/978-1-4612-3094-6

Golaz, J.-C., Salzmann, M., Donner, L. J., Horowitz, L. W., Ming, Y., and Zhao, M. (2011). Sensitivity of the aerosol indirect effect to sub-grid variability in the cloud parameterization of the GFDL atmosphere general circulation model AM3. *J. Clim*. 24, 3145–3160. doi: 10.1175/2010JCLI3945.1

Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Dai, Y., et al. (2015). Multi-objective parameter optimization of common land model using adaptive surrogate modeling. *Hydrol. Earth Syst. Sci*. 19, 2409–2425. doi: 10.5194/hess-19-2409-2015

Guichard, F., Petch, J. C., Redelsperger, J.-L., Bechtold, P., Chaboureau, J.-P., Cheinet, S., et al. (2004). Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models. *Q. J. R. Meteor. Soc*. 130, 3139–3172. doi: 10.1256/qj.03.145

Guo, Z., Wang, M., Qian, Y., Larson, V. E., Ghan, S., Ovchinnikov, M., et al. (2015). A sensitivity analysis of cloud properties to CLUBB parameters in the single-column Community Atmosphere Model (SCAM5). *J. Adv. Model. Earth Syst*. 6, 829–858. doi: 10.1002/2014MS000315

Hazra, A., Chaudhari, H. S., Rao, S. A., Goswami, B. N., Dhakate, A., Pokhrel, S., et al. (2015). Impact of revised cloud microphysical scheme in CFSv2 on the simulation of the Indian summer monsoon. *Int. J. Climatol*. 35, 4738–4755. doi: 10.1002/joc.4320

He, F., and Posselt, D. J. (2015). Impact of parameterized physical processes on simulated tropical cyclone characteristics in the community atmosphere model. *J. Clim*. 28, 9857–9872. doi: 10.1175/JCLI-D-15-0255.1

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J. C., Balaji, V., Duan, Q., et al. (2017). The art and science of climate model tuning. *Bull. Am. Meteor. Soc*. 98, 589–602. doi: 10.1175/BAMS-D-15-00135.1

Jackson, C., Sen, M., and Stoffa, P. (2004). An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. *J. Clim*. 17, 2828–2841. doi: 10.1175/1520-0442(2004)017andlt

Jackson, C. S., Sen, M. K., Huerta, G., Deng, Y., and Bowman, K. P. (2008). Error reduction and convergence in climate prediction. *J. Clim*. 21, 6698–6709. doi: 10.1175/2008JCLI2112.1

Jess, S., Spichtinger, P., and Lohmann, U. (2011). A statistical subgrid-scale algorithm for precipitation formation in stratiform clouds in the ECHAM5 single column model. *Atmos. Chem. Phys. Discuss*. 11, 9335–9374. doi: 10.5194/acpd-11-9335-2011

Korolev, A., Khain, A., Pinsky, M., and French, J. (2016). Theoretical study of mixing in liquid clouds—Part 1: classical concepts. *Atmos. Chem. Phys*. 16, 9235–9254. doi: 10.5194/acp-16-9235-2016

Kusch, J., and Frank, M. (2018). Intrusive methods in uncertainty quantification and their connection to kinetic theory. *Int. J. Adv. Eng. Sci. Appl. Math*. 10, 54–69. doi: 10.1007/s12572-018-0211-3

Le Maitre, O. P., and Knio, O. M. (2010). *Spectral Methods for Uncertainty Quantification*. Dordrecht: Springer-Verlag, 536. doi: 10.1007/978-90-481-3520-2

Lee, L. A., Carslaw, K. S., Pringle, K., Mann, G. W., and Spracklen, D. V. (2011). Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmos. Chem. Phys. Discuss*. 11, 12253–12272. doi: 10.5194/acp-11-12253-2011

Lee, L. A., Carslaw, K. S., Pringle, K. J., and Mann, G. W. (2012). Mapping the uncertainty in global CCN using emulation. *Atmos. Chem. Phys*. 12, 9739–9751. doi: 10.5194/acp-12-9739-2012

Li, J. D., Duan, Q. Y., Gong, W., ye, A. Z., Dai, Y. J., Miao, C. Y., et al. (2013). Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis. *Hydrol. Earth Syst. Sci. Discuss*. 10, 2243–2286. doi: 10.5194/hessd-10-2243-2013

Lin, G., Wan, H., Zhang, K., Qian, Y., and Ghan, S. J. (2016). Can nudging be used to quantify model sensitivities in precipitation and cloud forcing?. *J. Adv. Model. Earth Syst*. 8, 1073–1109. doi: 10.1002/2016MS000659

Liu, X., Ma, P. L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., et al. (2016). Description and evaluation of a new four-mode version of the modal aerosol module (MAM4) within version 5.3 of the community atmosphere model. *Geosci. Model Dev*. 9, 505–522. doi: 10.5194/gmd-9-505-2016

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., et al. (2009). Toward optimal closure of the earth's top-of-atmosphere radiation budget. *J. Clim*. 22, 748–766. doi: 10.1175/2008JCLI2637.1

Lohmann, U., Stier, P., Hoose, C., Ferrachat, S., Kloster, S., Roeckner, E., et al. (2007). Cloud microphysics and aerosol indirect effects in the global climate model ECHAM5-HAM. *Atmos. Chem. Phys. Discuss*. 7, 3719–3761. doi: 10.5194/acpd-7-3719-2007

Lopez, A., Tebaldi, C., New, M., Stainforth, D., Allen, M., Kettleborough, J., et al. (2006). *Two* approaches to quantifying uncertainty in global temperature changes. *J. Clim*. 19, 4785–4796. doi: 10.1175/JCLI3895.1

Lord, S. J., Chao, W. C., and Arakawa, A. (1982). Interaction of a cumulus cloud ensemble with the large-scale environment. Part IV: the discrete model. *J. Atmos. Sci*. 39, 104–113.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys*. 21, 1087–1092. doi: 10.1063/1.1699114

Mitchell, D. L., Rasch, P., Ivanova, D., McFarquhar, G., and Nousiainen, T. (2008). Impact of small ice crystal assumptions on ice sedimentation rates in cirrus clouds and GCM simulations. *Geophys. Res. Lett*. 35, 1–5. doi: 10.1029/2008GL033552

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., et al. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* 430, 768–772. doi: 10.1038/nature02771

Neale, R. B., Richter, J. H., and Jochum, M. (2008). The impact of convection on ENSO: from a delayed oscillator to a series of events. *J. Clim*. 21, 5904–5924. doi: 10.1175/2008JCLI2244.1

Pathak, R., Sahany, S., and Mishra, S. K. (2020). Uncertainty quantification based cloud parameterization sensitivity analysis in the NCAR community atmosphere model. *Sci. Rep*. 10, 1–17. doi: 10.1038/s41598-020-74441-x

Peng, J., Hampton, J., and Doostan, A. (2014). A weighted $\ell$1-minimization approach for sparse polynomial chaos expansions. *J. Comput. Phys*. 267, 92–111. doi: 10.1016/j.jcp.2014.02.024

Petch, J. C., Hill, A., Davies, L., Fridlind, A., Jakob, C., Lin, Y., Xie, S., and Zhu, P. (2014). Evaluation of intercomparisons of four different types of model simulating TWP-ICE. *Quart. J. Roy. Meteor. Soc.* 140, 826–837. doi: 10.1002/qj.2192

Pistotnik, G., Groenemeijer, P., and Sausen, R. (2016). Validation of convective parameters in MPI-ESM decadal hindcasts (1971-2012) against ERA-interim reanalyses. *Meteorol. Zeitschr.* 25, 631–643. doi: 10.1127/metz/2016/0649

Posselt, D. J. (2016). A Bayesian examination of deep convective squall line sensitivity to changes in cloud microphysical parameters. *J. Atmos. Sci.* 73, 637–665. doi: 10.1175/JAS-D-15-0159.1

Priess, M., Koziel, S., and Slawig, T. (2011). Surrogate-based optimization of climate model parameters using response correction. *J. Computer Sci.* 2, 335–344. doi: 10.1016/j.jocs.2011.08.004

Qian, Y., Wan, H., Yang, B., Golaz, J.-C., Harrop, B., Hou, Z., et al. (2018). Parametric sensitivity and uncertainty quantification in the version 1 of E3SM atmosphere model based on short perturbed parameter ensemble simulations. *J. Geophys. Res. Atmos.* 123, 13046–13073. doi: 10.1029/2018JD028927

Qian, Y., Yan, H., Hou, A., Johannesson, G., and Klein, S. (2015). Parametric sensitivity analysis of precipitation at global and local scales in the Community Atmosphere Model CAM5. *J. Adv. Model. Earth Syst.* 6, 513–526. doi: 10.1002/2014MS000354

Randel, D. L., Vonder Haar, T. H., Ringerud, M. A., Stephens, G. L., Greenwald, T. H., and Combs, C. L. (1996). A new global water vapor dataset. *Bureau Am. Meteor. Soc.* 77, 1233–1246.

Reagan, M. T., Najm, H. N., Ghanem, R. G., and Knio, O. M. (2003). Uncertainty quantification in reacting flow simulations through non-intrusive spectral projection. *Combust. Flame* 132, 545–555. doi: 10.1016/S0010-2180(02)00503-5

Ricciuto, D., Sargsyan, K., and Thornton, P. (2018). The impact of parametric uncertainties on biogeochemistry in the E3SM land model. *J. Adv. Model. Earth Syst.* 10, 297–319. doi: 10.1002/2017MS000962

Richter, J. H., and Rasch, P. J. (2008). Effects of convective momentum transport on the atmospheric circulation in the community atmospheric model, Version 3. *J. Clim.* 21, 1487–1499. doi: 10.1175/2007JCLI1789.1

Sanderson, B. M., Piani, C., Ingram, W. J., Stone, D. A., and Allen, M. R. (2008). Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. *Clim. Dyn.* 30, 175–190. doi: 10.1007/s00382-007-0280-7

Schwartz, S. E. (2004). Uncertainty requirements in radiative forcing of climate change. *J. Air Waste Manag. Assoc.* 54, 1351–1359. doi: 10.1080/10473289.2004.10471006

Silverman, B. W. (1986). *Density Estimation: For Statistics and Data Analysis*. London: Chapman and Hall, 175.

Smolyak, S. A. (1963). Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR*. 4, 240–243.

Sobol, I. (1993). Sensitivity analysis or nonlinear mathematical models. *Math. Model Comput. Exp.* 1, 407–414.

Sraj, I., Zedler, S. E., Knio, O. M., Jackson, C. S., and Hoteit, I. (2016). Polynomial chaos-based Bayesian inference of K-profile parameterization in a general circulation model of the tropical pacific. *Mon. Weather Rev.* 144, 4621–4640. doi: 10.1175/MWR-D-15-0394.1

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., and faull, N., Frame, D. J., et al. (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 433, 403–406. doi: 10.1038/nature03301

Sun, F., Hall, A., and Qu, X. (2011). On the relationship between low cloud variability and lower tropospheric stability in the Southeast Pacific. *Atmos. Chem. Phys.* 11, 9053–9065. doi: 10.5194/acp-11-9053-2011

Tarantola, A. (2004). *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 342.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* 106, 7183–7192. doi: 10.1029/2000JD900719

Van den Berg, E., and Friedlander, M. P. (2007). *SPGL1: A Solver for Large-Scale Sparse Reconstruction*. Available online at: https://www.cs.ubc.ca/~mpf/spgl1/index.html (accessed January 10, 2020).

Van den Berg, E., and Friedlander, M. P. (2009). Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.* 31, 890–912. doi: 10.1137/080714488

Wang, C., Duan, Q. Y., Gong, W., Ye, A. Z., Di, Z. H., and Miao, C. Y. (2014). An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. *Environ. Model. Softw.* 60, 167–179. doi: 10.1016/j.envsoft.2014.05.026

Warren, S. G., and Schneider, S. H. (1979). Seasonal simulation as a test for uncertainties in the parameterizations of a Budyko-Sellers zonal climate model. *J. Atmos. Sci.* 36, 1377–1397.

Xie, S., Lin, W., Rasch, P. J., Ma, P.-L., Neale, R., Larson, V. E., et al. (2018*)*. Understanding cloud and convective characteristics in version 1 of the E3SM atmosphere model. *J. Adv. Model. Earth Syst.* 10, 2618–2644. doi: 10.1029/2018MS001350

Yang, B., Qian, Y., Lin, G., Leung, L. R., Rasch, P. J., Zhang, G. J., et al. (2013). Uncertainty quantification and parameter tuning in the cam5 zhang-mcfarlane convection scheme and impact of improved convection on the global circulation and climate. *J. Geophys. Res. Atmos.* 118, 395–415. doi: 10.1029/2012JD018213

Yang, B., Qian, Y., Lin, G., Leung, R., and Zhang, Y. (2012). Some issues in uncertainty quantification and parameter tuning: a case study of convective parameterization scheme in the WRF regional climate model. *Atmos. Chem. Phys.* 12, 2409–2427. doi: 10.5194/acp-12-2409-2012

Young, A., Knapp, K. R., Inamdar, A., Hankins, W., and Rossow, W. B. (2018). The International Satellite Cloud Climatology Project H-Series climate data record product. *Earth Syst. Sci. Data* 10, 583–593. doi: 10.5194/essd-10-583-2018

Zaehle, S., and Friend, A. D. (2010). Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. *Glob. Biogeochem. Cycles* 24:GB1005. doi: 10.1029/2009GB003521

Zelinka, M. D., Klein, S. A., and Taylor, K. E. (2013). Contributions of different cloud types to feedbacks and rapid adjustments in CMIP5. *J. Clim.* 26, 5007–5027. doi: 10.1175/JCLI-D-12-00555.1

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.* 47, 1–12. doi: 10.1029/2019GL085782

Zhang, G. J., and McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the canadian climate centre general circulation model. *Atmosph. Ocean* 33, 407–446. doi: 10.1080/07055900.1995.9649539

Zhang, M., Somerville, R. C. J., and Xie, S. (2016). The SCM concept and creation of ARM forcing datasets. *Meteorol. Monogr.* 57, 24.1–24.12. doi: 10.1175/AMSMONOGRAPHS-D-15-0040.1

Zhang, T. (2015). An automatic and effective parameter optimization method for model tuning. *Geosci. Model Dev.* 8, 3579–3591. doi: 10.5194/gmd-8-3579-2015

Zhao, C., Liu, X., Qian, Y., Yoon, J., Hou, Z., Lin, G., et al. (2013*)*. A sensitivity study of radiative fluxes at the top of atmosphere to cloud-microphysics and aerosol parameters in the community atmosphere model CAM5. *Atmos. Chem. Phys.* 13, 10969–10987. doi: 10.5194/acp-13-10969-2013

Zhu, Q., Xu, X., Gao, C., Ran, Q.-H., and Xu, Y.-P. (2015). Qualitative and quantitative uncertainties in regional rainfall frequency analysis. *J. Zhejiang Univ. Sci. A* 16, 194–203. doi: 10.1631/jzus.A1400123

Zou, L. W., Qian, Y., Zhou, T. J., and Yang, B. (2014). Parameter tuning and calibration of RegCM3 with MIT-Emanuel cumulus parameterization scheme over CORDEX East Asia domain. *J. Clim.* 27, 7687–7701. doi: 10.1175/JCLI-D-14-00229.1

**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 Climate

ORIGINAL RESEARCH
published: 17 June 2021
doi: 10.3389/fclim.2021.681127

# A Novel Initialization Technique for Decadal Climate Predictions

*Danila Volpi[1]*, Virna L. Meccia[1], Virginie Guemas[2], Pablo Ortega[3], Roberto Bilbao[3], Francisco J. Doblas-Reyes[3,4], Arthur Amaral[3], Pablo Echevarria[3], Rashed Mahmood[3] and Susanna Corti[1]*

[1] National Research Council, Institute of Atmospheric Sciences and Climate, Bologna, Italy, [2] Centre National de Recherches Météorologiques, Météo-France, UMR3589 CNRS, Toulouse, France, [3] Earth Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain, [4] Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

Model initialization is a matter of transferring the observed information available at the start of a forecast to the model. An optimal initialization is generally recognized to be able to improve climate predictions up to a few years ahead. However, systematic errors in models make the initialization process challenging. When the observed information is transferred to the model at the initialization time, the discrepancy between the observed and model mean climate causes the drift of the prediction toward the model-biased attractor. Although such drifts can be generally accounted for with a posteriori bias correction techniques, the bias evolving along the prediction might affect the variability that we aim at predicting, and disentangling the small magnitude of the climate signal from the initial drift to be removed represents a challenge. In this study, we present an innovative initialization technique that aims at reducing the initial drift by performing a quantile matching between the observed state at the initialization time and the model state distribution. The adjusted initial state belongs to the model attractor and the observed variability amplitude is scaled toward the model one. Multi-annual climate predictions integrated for 5 years and run with the EC-Earth3 Global Coupled Model have been initialized with this novel methodology, and their prediction skill has been compared with the non-initialized historical simulations from CMIP6 and with the same decadal prediction system but based on full-field initialization. We perform a skill assessment of the surface temperature, the heat content in the ocean upper layers, the sea level pressure, and the barotropic ocean circulation. The added value of the quantile matching initialization is shown in the North Atlantic subpolar region and over the North Pacific surface temperature as well as for the ocean heat content up to 5 years. Improvements are also found in the predictive skill of the Atlantic Meridional Overturning Circulation and the barotropic stream function in the Labrador Sea throughout the 5 forecast years when compared to the full field method.

**Keywords: decadal climate prediction, initialization, drift, quantile matching, full field initialization**

OPEN ACCESS

***Edited by:***
*Roxy Mathew Koll, Indian Institute of Tropical Meteorology (IITM), India*

***Reviewed by:***
*Leonard Borchert, Sorbonne Universités, France*
*Vimal Koul, University of Hamburg, Germany*

****Correspondence:***
*Danila Volpi*
*d.volpi@isac.cnr.it*

***Specialty section:***
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

***Received:*** *15 March 2021*
***Accepted:*** *26 April 2021*
***Published:*** *17 June 2021*

***Citation:***
*Volpi D, Meccia VL, Guemas V, Ortega P, Bilbao R, Doblas-Reyes FJ, Amaral A, Echevarria P, Mahmood R and Corti S (2021) A Novel Initialization Technique for Decadal Climate Predictions. Front. Clim. 3:681127. doi: 10.3389/fclim.2021.681127*

## 1. INTRODUCTION

Providing reliable climate information for the near-term future is of paramount importance for many socioeconomic sectors, such as agriculture, energy, health, and insurance. Incorporating this information in the decision-making process is a key goal for the climate service community (Goddard, 2016; Otto et al., 2016). Decadal predictions cover 1-year to 10-year timescales.

The information extracted from this timescale is involved in the process of planning for adaptation strategies (Goddard et al., 2012; Brasseur and Gallardo, 2016). There is an internationally coordinated effort to produce and study predictions that cover such a timescale known as the Decadal Climate Prediction Project (DCPP, Boer et al., 2016), which contributes to the Coupled Model Intercomparison Project Phase 6 (CMIP6, Eyring et al., 2016).

Decadal predictability can arise from two main sources, namely the radiative forcings and the internal variability. Analogously to climate projections, decadal predictions are partly considered a boundary value problem whose crucial goal is to estimate the response of the system to variations in external forcings (Meehl et al., 2009). Apart from the anthropogenic contribution to changes in the external forcing (Booth et al., 2012), there also exist changes of natural origins, such as solar variability and volcanic eruptions, which are known to strongly affect the natural variability in the North Atlantic (Borchert et al., 2021; Mann et al., 2021).

The sources of interannual to decadal predictability originating from the slow components of the internal climate variability are associated mainly with the sea surface temperature (SST) state and ocean heat content (Doblas-Reyes et al., 2013a; Guemas et al., 2013; Merryfield et al., 2020). El Niño Southern Oscillation (ENSO) is the main process that contributes to the forecast quality in the Tropics as well as in large parts of the world, thanks to its expanded remote impacts, known as teleconnections (Doblas-Reyes et al., 2013b; Beverley et al., 2019). Another source of predictability is the sea ice, which is predictable 3 years ahead (Tietsche et al., 2014; Day et al., 2015) and whose main source of predictability is given by its persistence (Blanchard-Wrigglesworth et al., 2011). Climate predictability over the extratropics is controlled by teleconnections with the Tropics and the Arctic (Jung et al., 2015), as well as with soil moisture and land snow (Bellucci et al., 2015a).

The leading modes of decadal variability that dominate the Atlantic and the Pacific oceans are, respectively, the Atlantic multidecadal variability (AMV) and the Pacific decadal variability (PDV) (Kushnir et al., 2019). Although the processes responsible for the AMV and its predictability are not fully understood (Latif and Keenlyside, 2011; Cassou et al., 2018), climate models suggest that this mode of variability is linked to the variations in strength of the Atlantic Meridional Overturning Circulation (AMOC) (Zhang and Wang, 2013), which in turn is driven mainly by the convection activity that takes place in the northern high latitudes (Ortega et al., 2015; Robson et al., 2016), by a wind-dependent contribution (Mignot et al., 2006) and also by volcanic eruptions (Borchert et al., 2021; Mann et al., 2021). Since the AMOC exerts significant influences on the European climate through its net northward heat transport in the Atlantic, having good levels of prediction skill over the North Atlantic is of crucial importance. The North Atlantic upper-ocean heat content has an impact on Atlantic hurricanes and inland temperature and precipitation (Dunstone et al., 2011; Gastineau and Frankignoul, 2015; Buckley et al., 2019).

In order to capture the oscillations and the impacts of these variability modes, an effective initialization is crucial. Initialized predictions generally show improved forecast skill with respect to historical simulations, which only respond to changes in radiative forcings (Corti et al., 2015; Boer et al., 2016). Initializing a climate prediction consists of incorporating into the model the observed state information at the initial time of the forecast. However, models have systematic errors, which are associated with misrepresentation of key processes that are unresolved at the particular model grid and need to be parameterized (Nadiga et al., 2019). The main consequence of model errors is the difference between the model and observed mean state. Such a difference complicates the initialization task and one open issue is how to provide the model with the best estimate of the real initial state, without introducing inconsistencies that could compromise the prediction quality (Brune and Baehr, 2020).

One of the common initialization strategies is the full-field initialization (FFI), where the initial state is the best estimate of the observed climate state—the reanalysis (Pohlmann et al., 2009). After initialization, the prediction drifts away from the real-world attractor toward the mean model-biased state. To account for such a bias, a posteriori bias correction needs to be applied; various techniques that take into account the forecast time, start date, or initial condition dependence of the bias have been designed and implemented (Kharin et al., 2012; Fučkar et al., 2014). The correction of the bias in interannual predictions takes up the challenge of disentangling the small magnitude of climate signal to be predicted from the initial drift to be removed (Smith et al., 2013).

An alternative technique to limit the drift is the anomaly initialization (AI) that aims at phasing the model variability with the observed one by assimilating the observed anomaly onto the model mean state (Smith et al., 2008; Pohlmann et al., 2013). Previous studies have applied these initialization techniques to different models to highlight the relative strengths and limitations (Hazeleger et al., 2013; Smith et al., 2013; Bellucci et al., 2015b; Marotzke et al., 2016). The results show that in their standard implementation, at interannual time scales, the differences in skill between these techniques are small and limited to specific regions (Smith et al., 2013). The best strategy has been suggested to be model dependent because models have different biases (Magnusson et al., 2013; Polkova et al., 2014).

In this work, we present a new initialization method, the quantile matching (QM) technique that aims at reducing the drift and at limiting inconsistencies coming from the differences between the model and the observed variability amplitude. After performing the decadal hindcasts with EC-Earth3, we explore the impact of the new initialization method on the forecast skill by comparing the predictions initialized with QM, with a set of predictions initialized with FFI and a set of historical simulations. Section 2 introduces the initialization method, its implementation, and describes the model set-up in use. Section 3 is organized as follows: section 3.1 provides an overview of the behavior of the decadal predictions based on the new initialization method in terms of mean bias and drift. The prediction skill of the surface and sub-surface fields is presented in section 3.2, and the skill in the North Atlantic region is explored in section 3.3. Finally, the main findings are summarized in section 4.

# 2. METHODS AND DATA

## 2.1. Quantile Matching as Initialization Method

Predictions after initialization can experience two potential problems that can affect the forecast skill. On the one hand, rapid model adjustments known as initial shocks can occur (He et al., 2017). They can be generated from different mechanisms, one of them being the imbalance caused by the use of inconsistent atmosphere and ocean reanalyses as initial states. Previous studies suggested that the impact of initial shocks on the forecast skill is negligible at seasonal timescale (Mulholland et al., 2015). Nevertheless, this is not the case at longer timescales, where initial shock plays an important role in sensitive areas such as the North Atlantic subpolar region (Kröger et al., 2018; Bilbao et al., 2021). On the other hand, the model tends to adjust toward its biased mean climate after being initialized. Such a drift, which is a common feature of predictions initialized with FFI, is known to affect the forecast quality (Hazeleger et al., 2013).

While initial shocks can occur with both FFI and AI, the drift is expected to be largely reduced in predictions initialized with AI, as such technique employs an initial state that belongs to the model attractor and only imposes the observed variability (i.e., the initial state is the sum of the model mean climate and the observed anomalies). However, the use of AI can introduce observed anomalies whose amplitude does not belong to the range of the internal variability generated by the model. Volpi et al. (2017b) addressed this issue by weighting the observed anomalies with the ratio between the model and observed standard deviations. This technique of weighting, together with the anomaly initialization of the ocean density, showed improved skill in predicting the sea-ice variability, the AMV and the SST in the Labrador Sea, in parts of the North Pacific and Southern Ocean.

However, such refinement employed the standard deviation to characterize the variability amplitude, although the statistical distribution of variability might be skewed. The QM introduced here, therefore, expands over the idea of weighting the observed anomalies by respecting the distribution of the model variability. The QM consists of initializing the prediction with the model state whose percentile in the model distribution is the same as the percentile of the observed state in the observed distribution at the initialization time. The added value of this method is as follows:

- The initial state belongs to the model attractor (as any other AI technique),
- By matching the cumulative distributions of the model and observations, the observed initial state is effectively scaled with respect to the model variability.

Since with the current version of the model we do not have available an analogous set of decadal predictions as in Volpi et al. (2017b), we cannot evaluate the impact of respecting the distribution of the model variability. Therefore, the objective of this study is to assess the relative benefits and drawbacks of the novel method with respect to the state-of-the-art decadal predictions initialized with FFI. The QM is applied to all the grid-points of all ocean prognostic variables, which are the ones that are directly predicted by the model. For the implementation of the technique, the ocean reanalysis NEMOVAR-ORAS4 (Mogensen et al., 2012) is taken as the observational truth. Although NEMOVAR-ORAS4 is subject to some uncertainties, it has the advantage of providing observationally constrained and physically consistent values for all the prognostic ocean variables. At the time of this study, there was only one EC-Earth3 historical simulation available that stored the initial conditions for November. Therefore, the model distribution of each variable and grid point is computed using that historical simulation (r4i1p1f1). **Figure 1** illustrates an example of the implementation of the QM method for sea surface temperature (SST) at one grid point. The blue curve represents the cumulative distribution function (defined as the probability of a variable to take a value smaller than the value given in the x-axis) of SST calculated with one member of the ocean reanalysis NEMOVAR-ORAS4, over the period 1960–2014, for the grid point considered. Similarly, the SST cumulative distribution function calculated with the historical simulation of EC-Earth3 is shown in red, for the same grid point. The circles in the NEMOVAR-ORAS4 distribution indicate the value taken by the reanalysis on the $1^{st}$ of November of the years marked in the figure. Assuming November 1960 as the target initial date, the model is initialized with the model value (marked with a yellow star) whose cumulative distribution function matches the observed one at the initialization time.

The ratio between the model and the reanalysis SST variance is shown in **Figure 2**. Regions with the highest variance mismatch (darkest colors in **Figure 2**) are the areas where the QM method applies larger corrections to scale the observed variability onto the model amplitude.

## 2.2. Model Description and Experimental Set-Up

The model in use for this study is the CMIP6 version of EC-Earth3 GCM (Döscher et al., 2021). Its atmospheric component is the European Centre for Medium-Range Weather Forecasts Integrated Forecasting System (IFS cycle cy36r4) in its standard resolution, with 91 vertical levels and a T255 horizontal resolution. The ocean component is the NEMO model version 3.6 (Madec and NEMO System Team, 2015), with ORCA1 configuration (about 1 degree with enhanced tropical resolution) and 75 vertical levels. The sea-ice component is LIM3 (Rousset et al., 2015) directly embedded into NEMO. The atmospheric and ocean components are coupled via OASIS3 (Craig et al., 2017). Information on the dynamic vegetation is prescribed from a previous simulation with LPJ-GUESS (Smith et al., 2014).

The benchmark hindcasts are a full field initialized experiment and an ensemble of 15 uninitialized CMIP6 historical simulations (Histo), assessed by Bilbao et al. (2021). All the experiments, including our QM, are carried out with the same model version. In the FFI experiment, all the variables from each model component are initialized with observational estimates (reanalysis). The atmosphere and land surface initial conditions are taken from the ERA-40 reanalysis (Uppala et al., 2005) for start dates before 1979 and ERA-Interim (Dee et al., 2011)

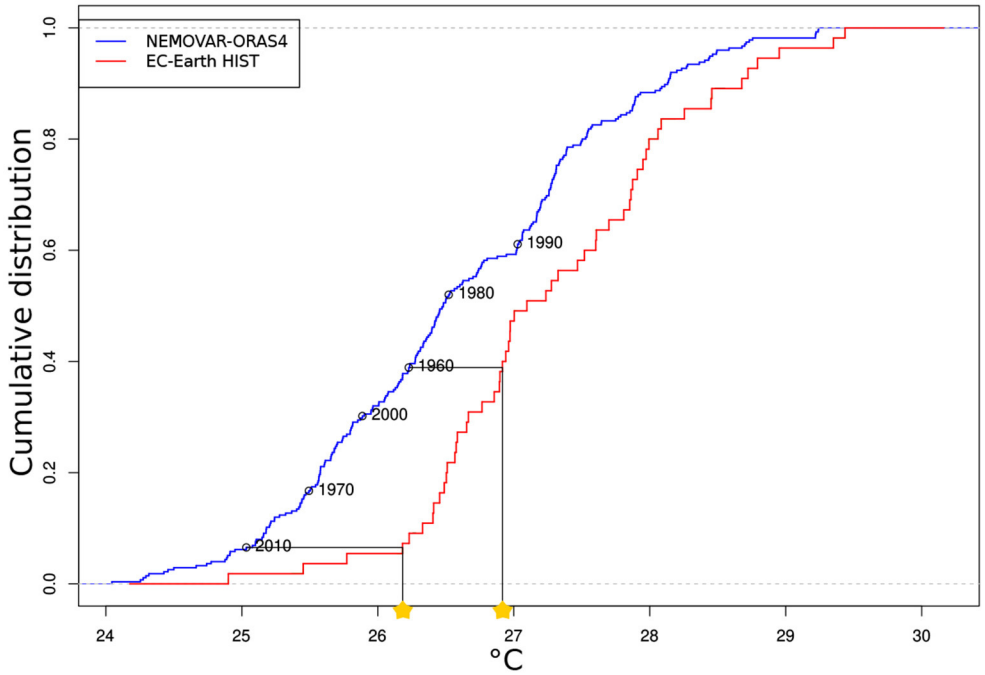



**FIGURE 1 |** Example of the implementation of the quantile matching method: cumulative distribution function for the SST in one grid point of the Tropical Pacific in blue for NEMOVAR-ORAS4 and in red for a historical simulation of EC-Earth3. The circles in the reanalysis distribution indicate the value on November 1 of the year indicated in the figure. The prediction at that grid point will be initialized with the value from the historical simulation, which has the same cumulative distribution as NEMOVAR-ORAS4 at the initialization time (the yellow stars indicate the value for the start dates of 1960 and 2010). This calculation is made for all the grid points and all the ocean prognostic variables.

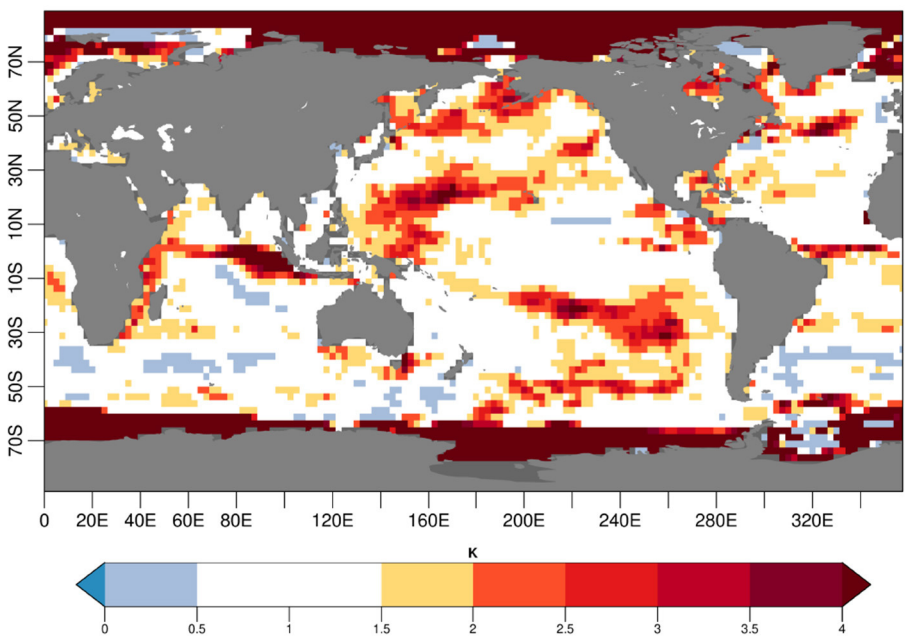



**FIGURE 2 |** Ratio of SST model variance over the NEMOVAR-ORAS4 variance calculated over the November 1960–2014 start dates. The NEMOVAR-ORAS4 variance is calculated concatenating the 5 ensemble members available, while for the model one EC-Earth3 historical simulation is used.

afterwards. The ocean initial conditions are taken from the 3D-Var five-member ocean reanalysis NEMOVAR-ORAS4, while the sea-ice initial conditions are produced with a NEMO-LIM simulation driven by DFS5.2 forcing fluxes (Brodeau et al., 2009). The DFS5.2 forcing data corresponds to a corrected version of ERA-interim accounting for radiative and precipitation observations in the Arctic. The solar irradiance, volcanic and anthropogenic aerosol load, and greenhouse gas concentrations are prescribed using the CMIP6 radiative forcing estimates up to 2014. After that date, the SSP2-4.5 scenario (O'Neill et al., 2016) is used.

Both initialized experiments (FFI and QM) are composed by 10 ensemble members, initialized every November from 1960 until 2014 and running for 5 years. The ensemble is generated using perturbations in the 3-dimensional temperature field in the atmospheric component, and using the 5-member ensemble from the NEMOVAR-ORAS4 reanalysis in the ocean.

The QM is applied to the ocean component and is performed for all the ocean prognostic variables and grid points, by matching the 5 members of NEMOVAR-ORAS4 separately, in order to obtain 5 different initial conditions. The atmospheric and sea-ice components of the QM experiment have identical initial condition as the FFI.

## 2.3. Bias and Skill Estimation

The forecast time-dependent climatologies are computed for the longest common period (1965–2014). For each forecast year (from 1 to 5), a set of predictions is available within the 1965–2014 period (e.g., 1965 is predicted at forecast year 1 for the prediction starting in November 1964, and also as forecast year 5 for the prediction starting in November 1960). The forecast time-dependent bias is calculated as the difference between the model and the observed climatologies. The drift, i.e., the evolution of the bias with forecast time, is estimated as the difference between the first and the last forecast year climatologies for the model. This estimate of the drift provides a simplistic picture but it already highlights the main features of the model drift.

To measure the forecast quality, we use the anomaly correlation (AC) and the root mean square error (RMSE). The confidence interval is calculated with a t-distribution for the AC and with a $\chi^2$ distribution for the RMSE. The serial dependence between the hindcasts is accounted for in the computation of

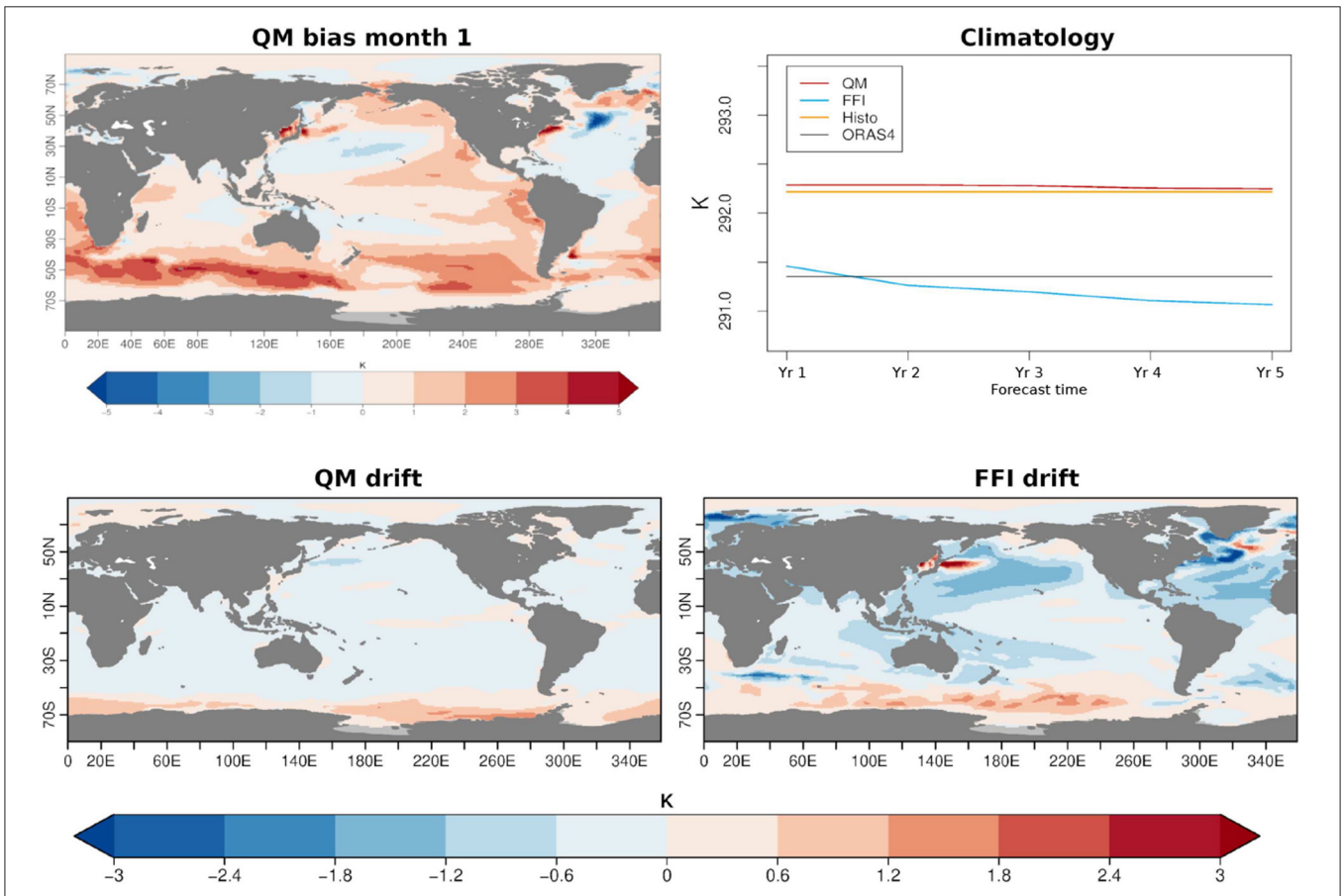



**FIGURE 3 | Top left**: SST bias of the QM experiment for the first forecast month calculated against NEMOVAR-ORAS4 reanalysis. **Top right**: forecast time-dependent climatologies of the annual mean global SST, calculated as explained in section 2.3. The ensemble mean climatology is shown in red for QM, in blue for FFI, in yellow for Histo, and in black for NEMOVAR-ORAS4. **Bottom**: drift calculated as the difference between the first and the last forecast year climatologies of QM (left panel) and FFI (right panel).

the confidence interval using the Von Storch and Zwiers (2001) formula. The confidence interval also takes into account the trend, that is not removed in the computation of the skill. The skill scores are computed either on annual mean values, or after smoothing the timeseries with a 1-year running mean in order to filter out seasonal climate variability and focus on interannual prediction skill. To measure the impact of the QM on the forecast quality, we compute the differences of ACs between the QM and FFI, and between QM and Histo. We use the method by Siegert et al. (2017) to identify the statistically significant differences in skill between the various experiments.

We have used several observational datasets for verification purposes: the GISTEMPv4 (Lenssen et al., 2019) NASA gridded surface temperature anomaly, the EN4 (Good et al., 2013) for the ocean heat content, and HadISST (Rayner et al., 2003) for the computation of the AMV. For the study of the initial SST bias and drift, we have used the NEMOVAR-ORAS4 reanalysis (as it represents the initial reference state for QM and the initial condition for FFI). Since observations of the full AMOC cell and the barotropic stream function are not available, they are validated against the NEMOVAR-ORAS4 reanalysis. The sea level pressure is validated against the ERA-40 and ERA-Interim reanalyses.

# 3. RESULTS

## 3.1. Forecast Drift and Mean State Biases

Initializing the predictions with a state that belongs to the model attractor implies that the mean state of the predictions is biased with respect to the observations since the beginning of the forecast. The top left panel of **Figure 3** shows the SST model bias with respect to the NEMOVAR-ORAS4 reanalysis in the first forecast month. The model displays a widespread warm bias over the Tropical Pacific extending toward the extratropics in the eastern basin. An even more pronounced warm bias appears in the Southern Ocean, while the Subpolar North Atlantic region presents an intense cold bias. These features are consistent with the ones found in the bias of Histo (**Supplementary Figure 1**). The top right panel of **Figure 3** shows the annual mean climatologies of the global SST as a function of the forecast time. Consistently with the bias map, the

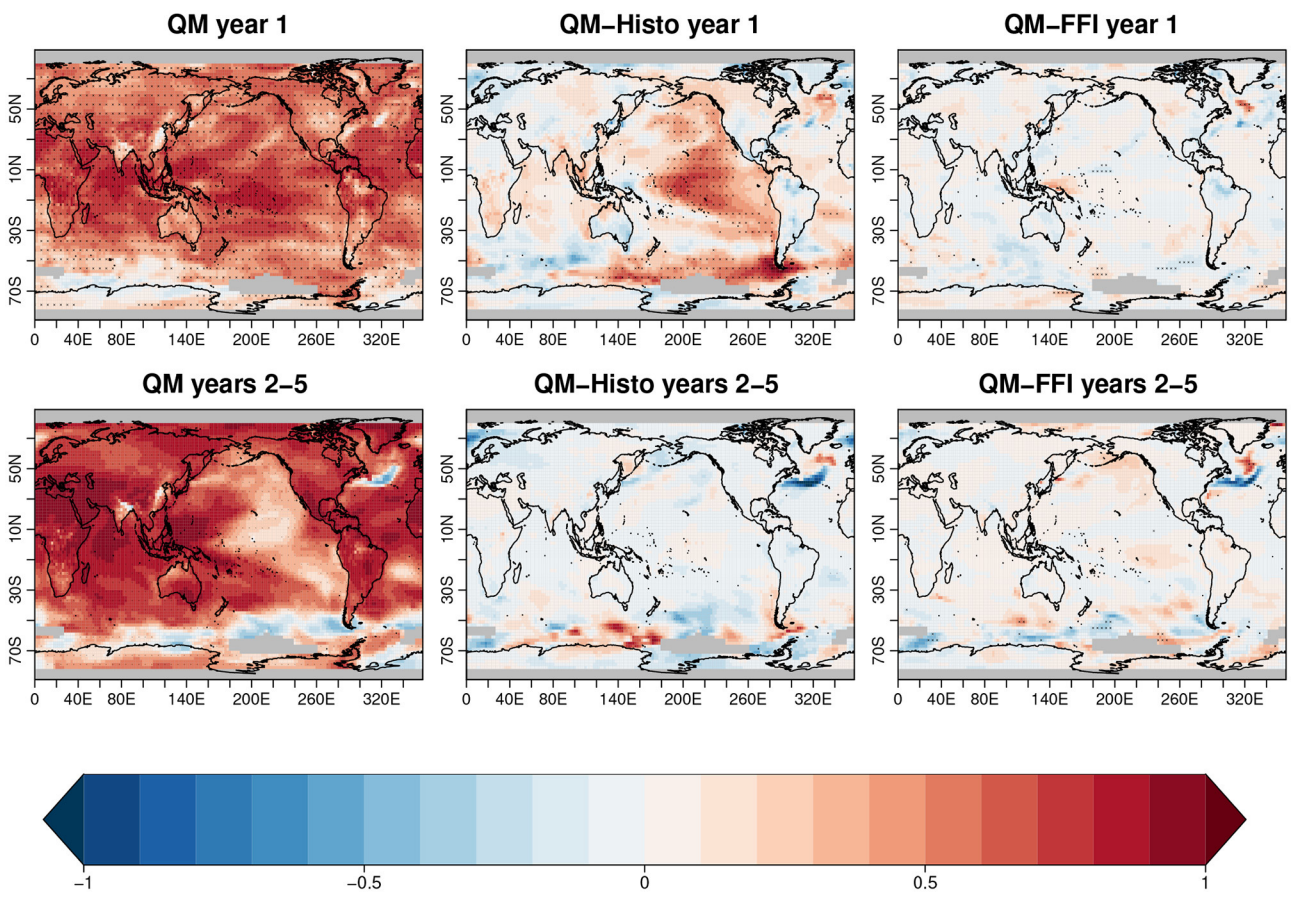



**FIGURE 4 |** Anomaly correlation for the surface temperature calculated against the GISTEMPv4 dataset. The first column represents the quantile matching (QM) skill in the first forecast year **(top)** and the 2–5 forecast years **(bottom)**. The middle column shows the difference between QM and Histo skill, and the third column the difference between QM and full-field initialization (FFI) skill. The black dots indicate regions of significant anomaly correlation (AC) (first column), and regions where the difference in AC is statistically significant with 95% confident level (second and third column).

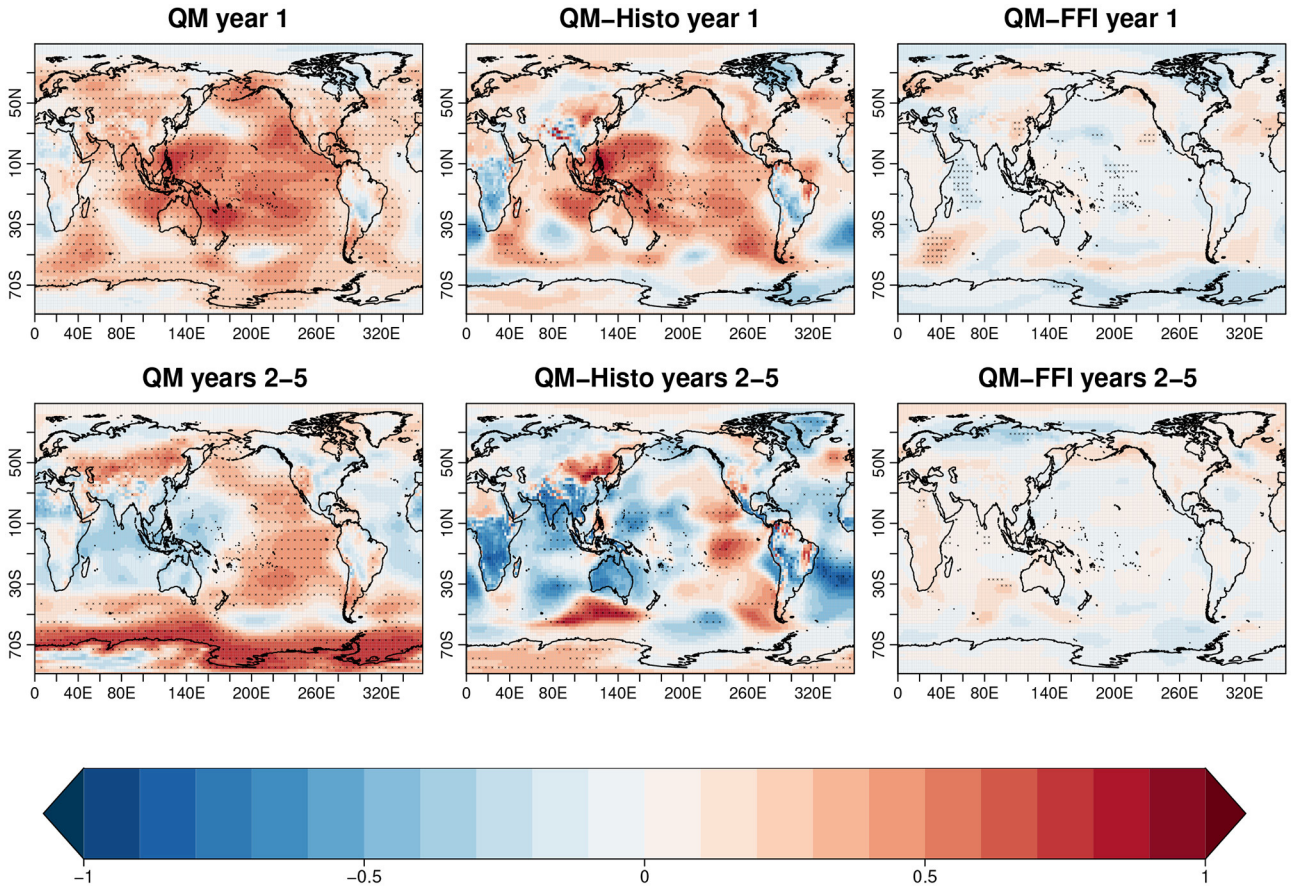



**FIGURE 5 |** The same as in **Figure 4**, but for the sea level pressure. The anomaly correlation (AC) is calculated against ERA-40 and Era-Interim reanalysis.

QM climatology (in red) results warmer than the NEMOVAR-ORAS4 climatology (black line), and remains close to the Histo climatology (in yellow) throughout the whole forecast period. In contrast, the climatology of the FFI experiment starts closer to NEMOVAR-ORAS4 but it drifts after the first forecast year. The reason for which the FFI experiment drifts away from the Histo climatologies is extensively investigated in Bilbao et al. (2021), where it is suggested to be an effect of the initial shock that leads to a collapse in Labrador Sea convection. This issue will be further discussed in section 3.3. However, the global drift toward colder temperatures is not occurring all around the globe as it is shown in the bottom right panel of **Figure 3**. Although a positive drift in SST is present in the Southern Ocean, the largest negative ones occur in the North Atlantic, particularly in the subpolar gyre region. Conversely, and as expected, the QM largely prevents the model drifts, with values of even an order of magnitude lower with respect to the FFI (bottom left panel of **Figure 3**).

## 3.2. Global Skill

We first evaluate the global skill in predicting the surface fields. The QM experiment exhibits high forecast quality in surface temperature (first column **Figure 4**) and sea level pressure (first column **Figure 5**) as indicated by the AC against the GISTEMPv4 dataset and ERA-40-ERA-Interim, shown in the first columns of **Figures 4**, **5**, respectively. The skill in surface temperature in the first forecast year is globally significant, the main exceptions being a few regions over Asia and the Southern Ocean and a small region in the North Atlantic. At longer forecast times (2–5 forecast years), the region of non-significant skill in the North Atlantic expands, as well as in the Southern Ocean, and a new region in the central Tropical Pacific loses significance. Also the sea level pressure partly loses skill after the first forecast year and the regions with significant skill remain the subtropical gyre region in the Pacific and the Southern Ocean, which are the regions with the highest model biases during the first forecast month (**Figure 3**). When compared to Histo, the QM predictions show significant improvements over a predominant part of the Pacific and over the North Atlantic subpolar region (NASP) as well as the Southern Ocean, during the first forecast year for both the surface temperature (middle column **Figure 4**) and the sea level pressure fields (middle column **Figure 5**). The significant added-value with respect to Histo in the NASP surface temperature is maintained at longer forecast years (bottom panels of **Figure 4**). However, the improvement is partially lost at forecast years 2–5 in the zone of the inter-gyre position, where the subpolar and subtropical

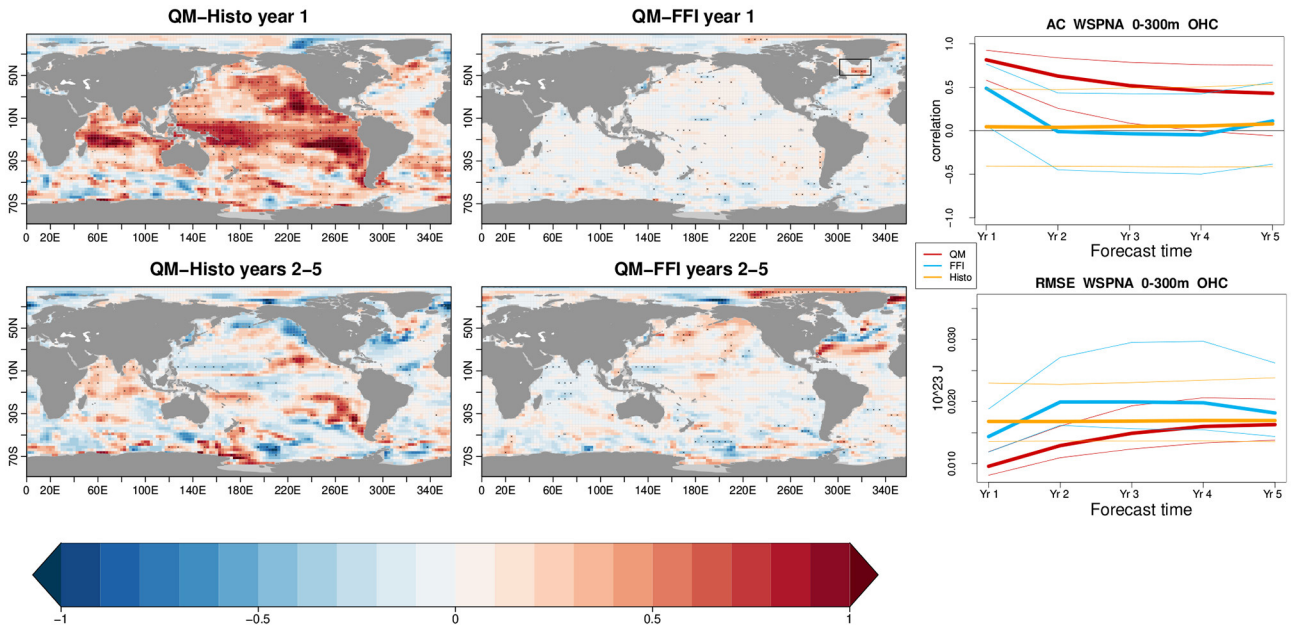



**FIGURE 6 |** Skill difference in the upper 300 m ocean heat content, computed against EN4 observational dataset. The first column shows the anomaly correlation (AC) difference between quantile matching (QM) and Histo for the first forecast year **(top)** and the 2–5 forecast years **(bottom)**. The second column shows the difference between QM and full-field initialization (FFI). The third column is a zoom on the regional mean of the western subpolar North Atlantic sector (WSPNA; 50–65° N, 60–30° W) as indicated by the black box. The AC and root mean square error (RMSE) are calculated with yearly mean data along the forecast time and are shown, respectively, in the **top** and **bottom** panel. In red is shown the QM experiment, in blue the FFI, and in yellow the Histo. The thin lines represent the 95% confidence interval obtained with a t-distribution for the correlation and a $\chi^2$ distribution for the RMSE.

gyres meet. The QM also outperforms the FFI in predicting the NASP surface temperature and, more importantly, the improved skill maintains significance throughout the whole forecast time (third column **Figure 4**). However, analogously to what is shown in the comparison with Histo, there is a degradation of skill along the Gulf stream for the forecast years 2–5, which could potentially be due to a wrong positioning of the current in the QM predictions. Differences between QM and FFI or Histo inland temperature are marginal. The AC results are broadly consistent with the residual correlations computed according to Smith et al. (2019), which are shown in **Supplementary Figure 2** of the **Supplementary Material**.

The comparison of QM with FFI for the sea level pressure (third column **Figure 5**) highlights a statistically significant skill improvement in the Antarctic circumpolar current and a degradation of skill in the Tropical Pacific and the Indian Ocean during the first forecast year, while significance in the skill difference is lost at forecast years 2–5.

We now focus on the upper layer ocean heat content (0–300 m) because of its crucial role in the ocean meridional transport: **Figure 6** shows the skill comparison between QM and Histo/FFI. Similarly to what is shown for the surface temperature, the QM significantly improves the skill in the Pacific and the NASP region with respect to Histo, although there is not a clear improvement over the Southern Ocean. Besides, skill over the Indian Ocean is improved. Part of those improvements are also maintained at longer timescales. When comparing with the FFI prediction, we find that the added value of the QM is mainly located over the NASP region, consistently with the surface temperature results.

We now zoom over the western subpolar North Atlantic region (50–65° N, 60–30° W, WSPNA), as this is a region in which SST and ocean heat content have been shown to influence the temperature and precipitation over the neighboring continents (Buckley et al., 2019). The third column of **Figure 6** shows the AC (first row) and the RMSE (second row) over the WSPNA region as a function of the forecast time. The highest values of ACs and the lowest values of RMSE over the WSPNA region and during the entire forecast time are obtained with the QM (red lines). Conversely, the Histo simulations (yellow lines) do not show significant AC in any of the forecast years, whereas the FFI loses its skill after the first forecast year.

## 3.3. Skill in the North Atlantic

In the previous section, we have shown significant improvements in the NASP region, specifically in the Western sector of QM with respect to FFI. Bilbao et al. (2021) performed an assessment for an analogous decadal prediction experiment based on FFI, and highlighted the collapse of deep convection in the Labrador Sea with a consequent weakening of the AMOC, subsequent to initialization. This led to a lack of prediction skill for the upper ocean heat content and the surface temperature in the NASP. We therefore look at the deep convection activity in the Labrador Sea in an attempt to explain the origins of the QM skill improvements presented in the previous section. **Figure 7** shows the ensemble

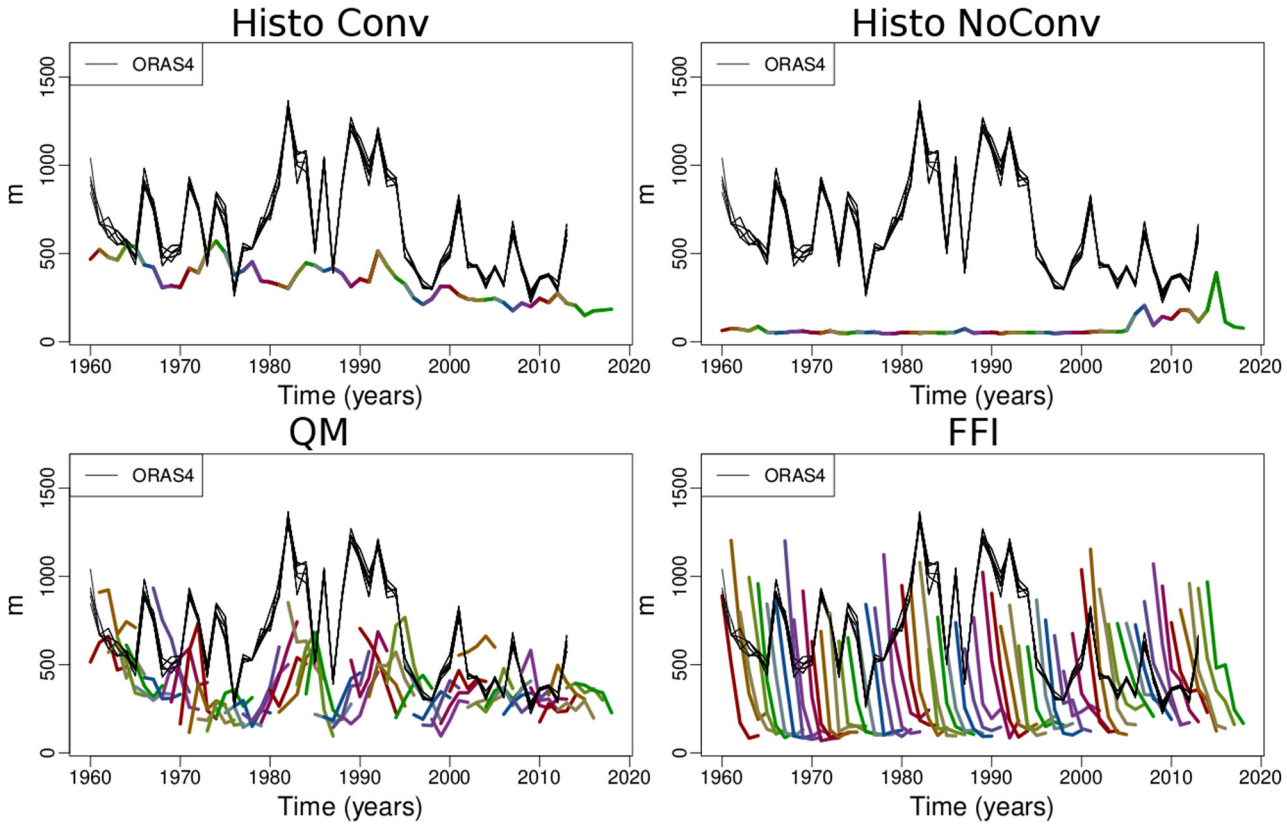



**FIGURE 7 |** Evolution of the mixed layer depth ensemble mean for the February–March–April average in the Labrador Sea. The top row shows the Histo ensemble splitted into members that exhibit a regular convection activity **(left)**, from the three which show no convection throughout most of the historical period **(right)**. The second row shows, respectively, the QM **(left)** and the FFI **(right)** evolution. The different start dates are represented with different colors. The NEMOVAR-ORAS4 is shown in black.

mean of the mixed layer depth in the Labrador Sea (55–65° N, 56–46° W), as a proxy for the deep convection intensity, for the February–March–April mean, which are the months of deepest mixing. In order to better understand the behavior of the Histo simulations, we have isolated three members that show no deep convection throughout most of the historical period (top right panel of **Figure 7**), from the rest of the Histo ensemble (top left panel). The Histo ensemble mean with regular convection is representative of the response to external forcings only and shows lower values than the NEMOVAR-ORAS4 reference. In the experiments with FFI, deep convection collapses very quickly with forecast time (bottom right panel of **Figure 7**) preventing FFI to follow the NEMOVAR-ORAS4 variability, consistently with what was found by Bilbao et al. (2021). In contrast, QM (bottom left panel of **Figure 7**) is able to partly reproduce the variability present in NEMOVAR-ORAS4: after initialization, it shows a slight increase in convective activity during the two peaks observed in the 1980s and in the 1990s. However, this does not directly translate into an increase of forecast skill, as the QM does not improve with respect to the Histo ensemble that preserves the convection. While Histo has a constant skill throughout the whole forecast time, QM presents statistically significant AC only during the first year (top left panel of **Figure 8**). This is probably due to the fact that the skill in predicting the mixed layer depth is dominated by the external forcing changes that have induced an increase in stratification hindering convection activity. The additional model variability that attempts at capturing the observed one rather constitutes an additional noise, which is detrimental to the forecast performance. The results in terms of RMSE (bottom left panel of **Figure 8**) are consistent with what is shown by the AC.

We assess the skill in predicting the barotropic circulation by looking at the barotropic stream function, which characterizes the horizontal water transport integrated vertically in the Labrador Sea (right column of **Figure 8**). We find that Histo do not show any skill. The initialized predictions, on the other hand, have similar skill at the beginning of the forecast. However, the skill of FFI deteriorates with forecast time and is lost at forecast year 3, while the skill of QM is maintained throughout the whole forecast time.

In summary, we have shown that the QM avoids the Labrador Sea deep convection collapse that occurs in the FFI experiment, and improves the prediction skill of the barotropic stream function with respect to both the FFI and the Histo simulations.

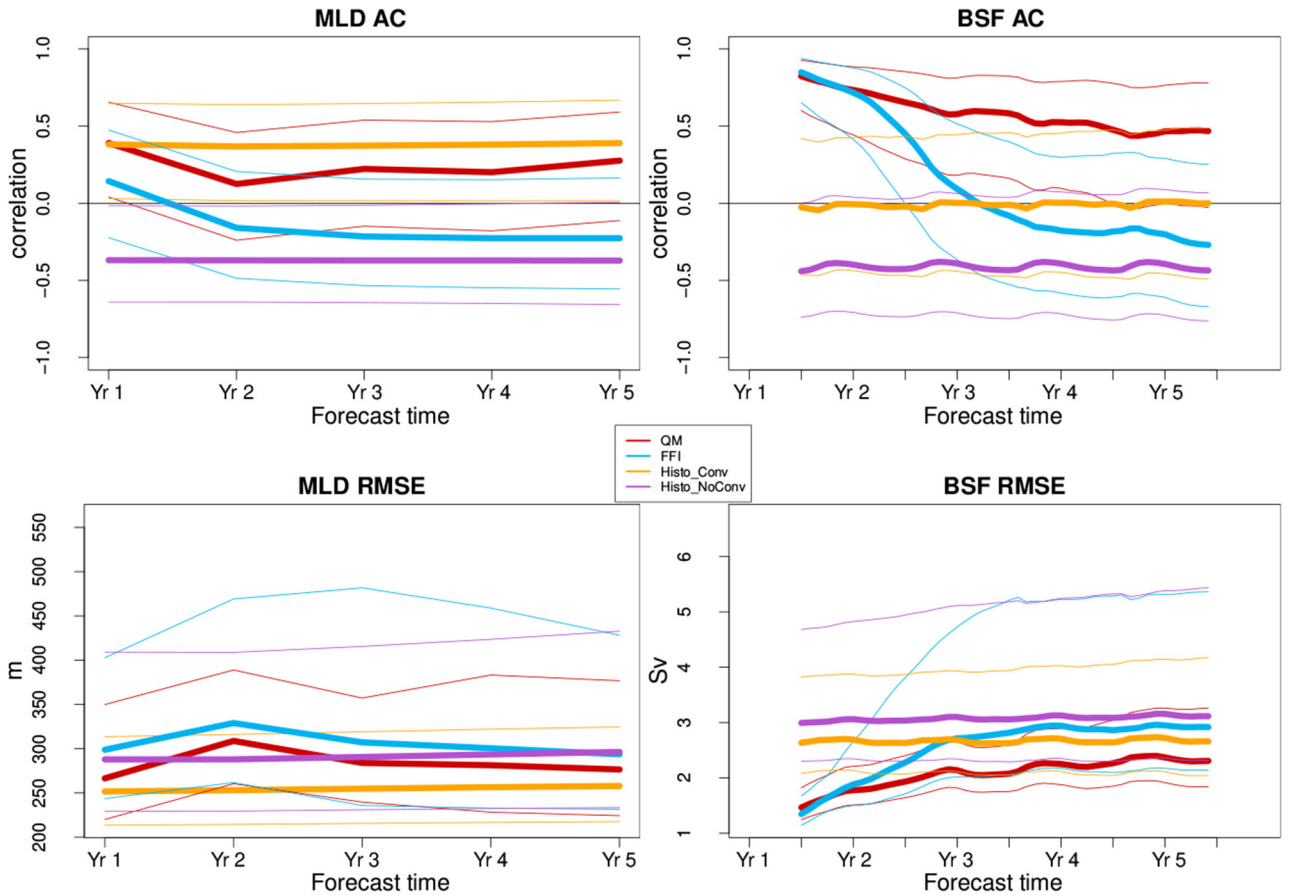



**FIGURE 8 |** Anomaly correlation of the mixed layer depth (**top left** panel) and the barotropic stream function (**top right** panel) in the Labrador Sea (55–65°N, 56–46°W). The second row shows the respective root mean square error (RMSE). Quantile matching (QM) is shown in red, the full-field initialization (FFI) in blue, the Histo with convection is shown in yellow, and in purple the Histo members with no convection. The skill are computed against NEMOVAR-ORAS4, using the February–March–April mean for the mixed layer depth, and a running mean of 12 months for the barotropic stream function. The thin lines represent the 95% confidence interval obtained with a t-distribution for the correlation and a $\chi^2$ distribution for the RMSE.

We are now interested in investigating the effect of these improvements on the AMOC. Since the AMOC skill depends on the latitude range at which it is considered, we show the results in a Hovmoller diagram of the AC (**Figure 9**). We have computed the AMOC index as the maximum of the Atlantic meridional overturning stream function in adjacent latitude-depth boxes of latitude range of 2 degrees (from 20° to 60° N) and depth range of 900–3,000 m depth. If we focus north of 40°N, the FFI starts with high skill that progressively deteriorates until the third forecast year down to an AC smaller than 0.1, probably due to the collapse in deep convection. This is largely improved by the QM that shows high skill in predicting the AMOC, particularly at subpolar latitudes, for the entire forecast time. The skill of FFI at lower latitudes is marginally higher than the QM one throughout the whole forecast time. One reason to explain this is that the detrimental effect of the convection collapse might need longer time to propagate at lower latitudes than the 5 years that we have available. The skill of the Histo simulations is constant with forecast time. This is consistent with the results in Borchert et al. (2021), where the good representation of the NASP SST in the CMIP6 historical simulations is attributed to the AMOC-related response to the forcings (volcanic eruptions and partly solar forcing).

Finally, we complement the North Atlantic analysis with the assessment of the skill in predicting the AMV. The AMV index is calculated as the difference between the regional SST anomalies in the North Atlantic (0° to 60° N and 80° to 0° W) and the global mean SST anomalies (between 60° S and 60° N), following the definition by Trenberth and Shea (2006). The ensemble QM predictions successfully capture the observed warming trend (left panel of **Figure 10**). In terms of skill, both the initialized predictions and the Histo simulations succeed in predicting the AMV index with skill significantly different than 0 along the whole forecast period as shown by the confidence intervals (thin lines in **Figure 10**). The prediction skill of FFI starts close to the skill of Histo and it might be generated

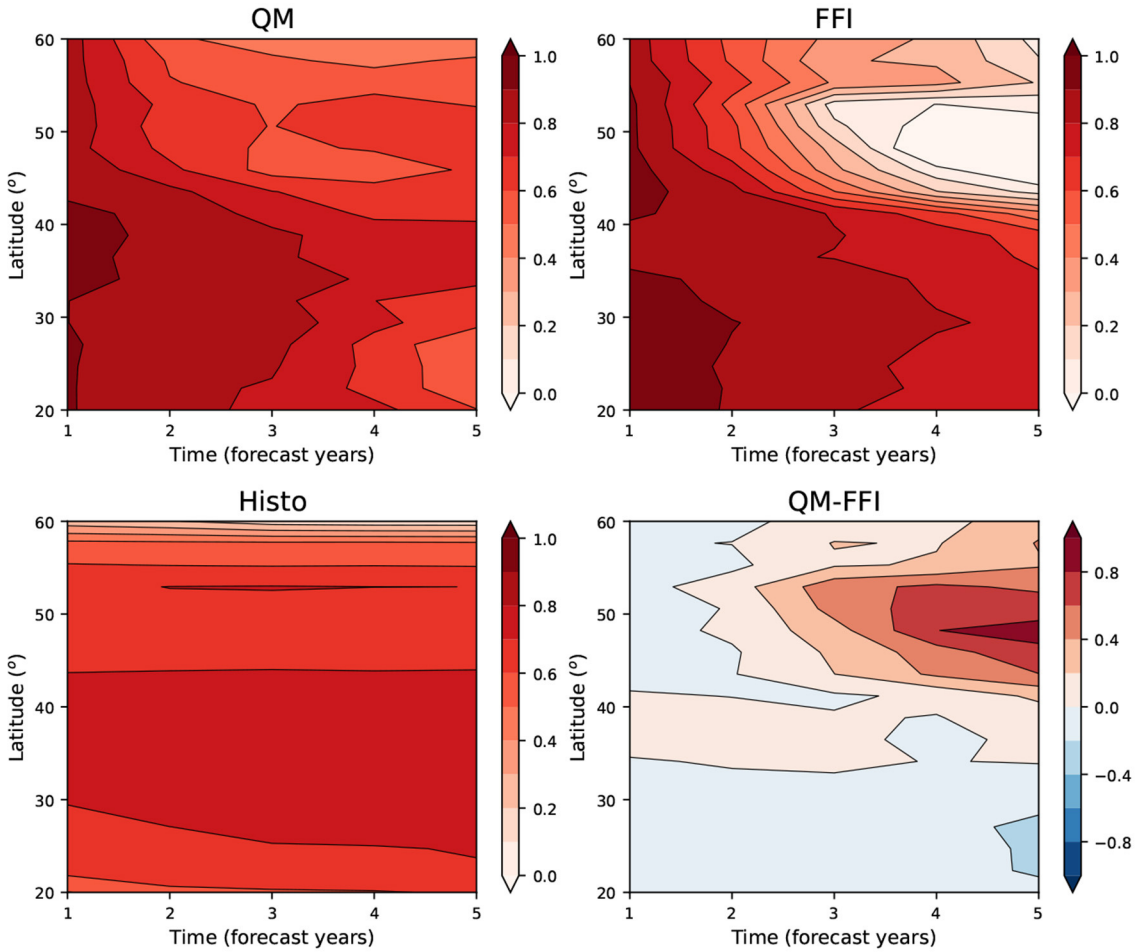



**FIGURE 9 |** Hovmoller diagrams of the AMOC anomaly correlation with NEMOVAR-ORAS4. The Atlantic Meridional Overturning Circulation (AMOC) is calculated as the maximum of the Atlantic meridional overturning stream function over adjacent boxes of 2° latitude (from 20° to 60° N) and 900–3,000 m depth. The panels show, respectively, the quantile matching (QM) **(top left)**, the full-field initialization (FFI) **(top right)**, the Histo **(bottom left)**, and the difference between the QM and the FFI skill **(bottom right)**.

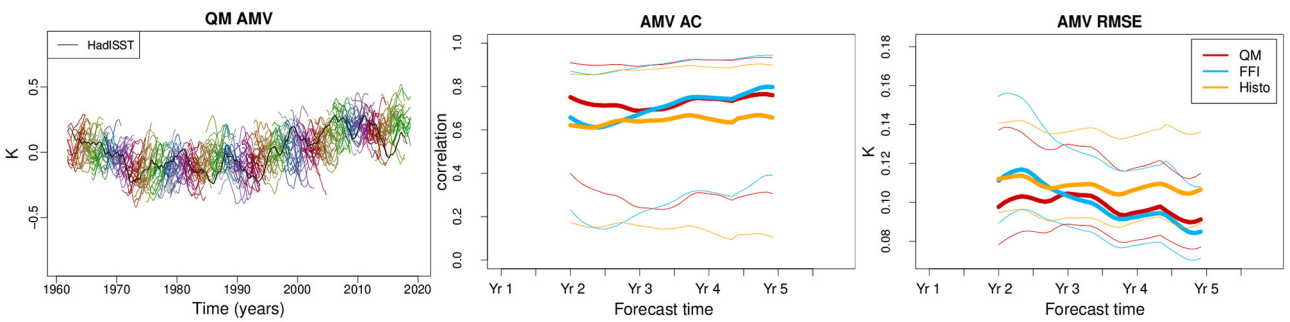



**FIGURE 10 |** Atlantic multidecadal variability (AMV) index for the quantile matching (QM) predictions **(left)**. The index computed from HadISST is shown in black. The different colors represent the different start dates. The anomaly correlation (AC) and the root mean square error (RMSE) of the AMV index are shown, respectively, in the **central** and **right** panels. The skill scores are calculated against the HadISST dataset, applying a 2-year running mean to the data. As in previous figures, the thin lines represent the 95% confidence interval obtained with a t-distribution for the correlation and a $\chi^2$ distribution for the RMSE.

by the AMV persistence. During the first two forecast years, QM outperforms FFI, both in terms of AC and RMS (second and third panel of **Figure 10**), although the difference is not statistically significant.

## 4. DISCUSSION AND CONCLUSIONS

In this study, we have presented a novel initialization method for decadal climate predictions, namely the quantile matching. This method aims at tackling two major issues:

- The drift coming from initializing the predictions away from the model preferred trajectories, similarly to any other anomaly initialization method;
- The potential inconsistencies between the observed/model distribution of variability.

The quantile matching method selects the initial condition of the prediction as the model state, which is identically located in the model distribution as the observed initial state in the observed distribution. Therefore, the initial state belongs to the model attractor, reducing the drift. Moreover, matching in the observed and model statistical distributions scales the observed variability toward the model one, correcting any potential amplitude incompatibilities. The forecast skill assessment of a decadal prediction with EC-Earth3, initialized with the quantile matching method, has led to the following findings:

- Surface temperature has high significant skill throughout the forecast years, in agreement with other studies (e.g., Smith et al., 2019).
- In the North Atlantic subpolar region, the quantile matching enhances the surface temperature and ocean heat content skill, compared to the historical simulations and the full field initialized predictions.
- The quantile matching avoids a collapse of deep convection in the Labrador Sea that occurred in the full field initialized predictions. This could be due to the positive effect of initializing the predictions on the model attractor. However, initialization does not seem to have a direct impact on the skill of the mixed layer depth, as no significant skill improvements are detected compared to the historical simulations. This is probably due to the fact that the skill in predicting the mixed layer depth is dominated by the trend that the historical simulations is able to capture.
- The skill of the barotropic stream function in the Western subpolar North Atlantic sector for the quantile matching predictions is significant throughout the whole forecast period and is the highest compared to both the full field initialized predictions and the historical simulations.
- The quantile matching has a positive impact on the Atlantic Meridional Overturning Circulation, which is skillfully predicted throughout the whole forecast time. The Atlantic Multidecadal Variability is well-forecasted along the whole forecast time by both the initialized predictions and the historical simulations.

The effect of the reduced prediction drift translates into an improved forecast skill of the meridional and barotropic ocean transports in the North Atlantic.

Considering that the ocean holds most of the memory in the climate system, we have performed the quantile matching on the ocean model component only. However, previous studies have highlighted the importance of having a sea-ice initial state consistent with the ocean state (Volpi et al., 2017a; Tian et al., 2020). Therefore, the use of different initialization techniques for the ocean and sea-ice components could have generated some instabilities in the quantile matching decadal predictions. A step further for this work would be performing the quantile matching also in the sea-ice model component. Besides, the use of more sophisticated atmosphere-ocean coupled initialization technique (Counillon et al., 2014; Brune and Baehr, 2020) would address the issue of the initial shock originated by the use of disjoint reanalyses for the initialization of the model components (Mulholland et al., 2015).

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

DV, VG, PO, and FD-R designed the quantile matching experiment. DV performed the decadal climate predictions and performed the analysis with the contributions of RB, VM, and RM. PE assisted with the settings of the workflow manager. AA contributed to the cmorizations of the outputs. RM and AA ran the benchmark simulations. PO, VG, FD-R, VM, and SC contributed to the discussion and the interpretation of the results. DV and VM prepared the manuscript with the contribution of all co-authors.

## FUNDING

This study was supported by the project LISTEN funded by the European Commission Horizon 2020 Marie Skłodowska-Curie Actions - IF (GA 799930).

## ACKNOWLEDGMENTS

The authors thankfully acknowledge the computer resources from the ECMWF special project INCIPIT (spitvolp) and the technical assistance provided by ECMWF and BSC. The climate simulations have been performed using Autosubmit workflow manager (Manubens-Gil et al., 2016). The authors thank Paolo Davini for providing the restart files of the historical simulation used to implement the quantile matching. We acknowledge Saskia Loosveldt and the earthdiags and ESMValTool suite developers, as well as the startR and s2dverification (Manubens

et al., 2018) software packages developers, as these tools were used to postprocess, analyze, and visualize the results presented in this work. PO acknowledges support by the Spanish Ministry of Economy, Industry and Competitiveness through the Ramon y Cajal grant (RYC-2017-22772).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fclim.2021.681127/full#supplementary-material

## REFERENCES

Bellucci, A., Haarsma, R., Bellouin, N., Booth, B., Cagnazzo, C., van den Hurk, B., et al. (2015a). Advancements in decadal climate predictability: the role of nonoceanic drivers. *Rev. Geophys*. 53, 165–202. doi: 10.1002/2014RG000473

Bellucci, A., Haarsma, R., Gualdi, S., Athanasiadis, P. J., Caian, M., Cassou, C., et al. (2015b). An assessment of a multi-model ensemble of decadal climate predictions. *Clim. Dyn*. 44, 2787–2806. doi: 10.1007/s00382-014-2164-y

Beverley, J. D., Woolnough, S. J., Baker, L. H., Johnson, S. J., and Weisheimer, A. (2019). The northern hemisphere circumglobal teleconnection in a seasonal forecast model and its relationship to european summer forecast skill. *Clim. Dyn*. 52, 3759–3771. doi: 10.1007/s00382-018-4371-4

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonniére, P.-A., et al. (2021). Assessment of a full-field initialized decadal climate prediction system with the cmip6 version of ec-earth. *Earth Syst. Dyn*. 12, 173–196. doi: 10.5194/esd-12-173-2021

Blanchard-Wrigglesworth, E., Bitz, C. M., and Holland, M. M. (2011). Influence of initial conditions and climate forcing on predicting Arctic sea ice. *Geophys. Res. Lett*. 38:L18503. doi: 10.1029/2011GL048807

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., et al. (2016). The decadal climate prediction project (DCPP) contribution to CMIP6. *Geosci. Model Dev*. 9, 3751–3777. doi: 10.5194/gmd-9-3751-2016

Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T., and Bellouin, N. (2012). Aerosols implicated as a prime driver of twentieth-century north atlantic climate variability. *Nature* 484, 228–232. doi: 10.1038/nature10946

Borchert, L. F., Menary, M. B., Swingedouw, D., Sgubin, G., Hermanson, L., and Mignot, J. (2021). Improved decadal predictions of north Atlantic subpolar gyre SST in CMIP6. *Geophys. Res. Lett*. 48:e2020GL091307. doi: 10.1029/2020GL091307

Brasseur, G. P., and Gallardo, L. (2016). Climate services: lessons learned and future prospects. *Earths Future* 4, 79–89. doi: 10.1002/2015EF000338

Brodeau, L., Barnier, B., Treguier, A., Penduff, T., and Gulev, S. (2009). An ERA40-based atmospheric forcing for global ocean circulation models. *Ocean Model* 31, 88–104. doi: 10.1016/j.ocemod.2009.10.005

Brune, S., and Baehr, J. (2020). Preserving the coupled atmosphere-ocean feedback in initializations of decadal climate predictions. *WIREs Climate Change* 11:e637. doi: 10.1002/wcc.637

Buckley, M. W., DelSole, T., Lozier, M. S., and Li, L. (2019). Predictability of north Atlantic sea surface temperature and upper-ocean heat content. *J. Clim*. 32, 3005–3023. doi: 10.1175/JCLI-D-18-0509.1

Cassou, C., Kushnir, Y., Hawkins, E., Pirani, A., Kucharski, F., Kang, I.-S., et al. (2018). Decadal climate variability and predictability: challenges and opportunities. *Bull. Am. Meteorol. Soc*. 99, 479–490. doi: 10.1175/BAMS-D-16-0286.1

Corti, S., Palmer, T., Balmaseda, M., Weisheimer, A., Drijfhout, S., Dunstone, N., et al. (2015). Impact of initial conditions versus external forcing in decadal climate predictions: a sensitivity experiment. *J. Clim*. 28, 4454–4470. doi: 10.1175/JCLI-D-14-00671.1

Counillon, F., Bethke, I., Keenlyside, N., Bentsen, M., Bertino, L., and Zheng, F. (2014). Seasonal-to-decadal predictions with the ensemble filter and the Norwegian earth system model: a twin experiment. *Tellus A Dyn. Meteorol. Oceanogr*. 66:21074. doi: 10.3402/tellusa.v66.21074

Craig, A., Valcke, S., and Coquart, L. (2017). Development and performance of a new version of the oasis coupler, OASIS3-MCT_3.0. *Geosci. Model Dev*. 10, 3297–3308. doi: 10.5194/gmd-10-3297-2017

Day, J., Tietsche, S., Collins, M., Goessling, H., V., Guemas, Guillory, A., et al. (2015). The arctic predictability and prediction on seasonal-to-interannual timescales (apposite) data set. *Geosci. Model Dev. Discuss*. 10, 8809–8833. doi: 10.5194/gmdd-8-8809-2015

Dee, D. P., Uppala, S. M., Simmons, A. J., P., Berrisford, P. 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

Doblas-Reyes, F. J., Andreu-Burillo, I., Chikamoto, Y., Garcia-Serrano, J., Guemas, V., Kimoto, M., et al. (2013a). Initialized near-term regional climate change prediction. *Nat. Commun*. 4:1715. doi: 10.1038/ncomms2704

Doblas-Reyes, F. J., Garcia-Serrano, J., Lienert, F., Biescas, A. P., and Rodrigues, L. R. L. (2013b). Seasonal climate predictability and forecasting: status and prospects. *WIREs Climate Change* 4, 245–268. doi: 10.1002/wcc.217

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., et al. (2021). The ec-earth3 earth system model for the climate model intercomparison project 6. *Geosci. Model Dev. Discuss*. 2021, 1–90. doi: 10.5194/gmd-2020-446

Dunstone, N. J., Smith, D. M., and Eade, R. (2011). Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude north Atlantic ocean. *Geophys. Res. Lett*. 38:L14701. doi: 10.1029/2011GL047949

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., et al. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev*. 9, 1937–1958. doi: 10.5194/gmd-9-1937-2016

Fučkar, N. S., Volpi, D., Guemas, V., and Doblas-Reyes, F. J. (2014). A posteriori adjustment of near-term climate predictions: accounting for the drift dependence on the initial conditions. *Geophys. Res. Lett*. 41, 5200–5207. doi: 10.1002/2014GL060815

Gastineau, G., and Frankignoul, C. (2015). Influence of the north Atlantic SST variability on the atmospheric circulation during the twentieth century. *J. Clim*. 28, 1396–1416. doi: 10.1175/JCLI-D-14-00424.1

Goddard, L. (2016). From science to service. *Science* 353, 1366–1367. doi: 10.1126/science.aag3087

Goddard, L., Hurrell, J. W., Kirtman, B. P., Murphy, J., Stockdale, T., and Vera, C. (2012). Two time scales for the price of one (almost). *Bull. Am. Meteorol. Soc*. 93, 621–629. doi: 10.1175/BAMS-D-11-00220.1

Good, S. A., Martin, M. J., and Rayner, N. A. (2013). EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Oceans* 118, 6704–6716. doi: 10.1002/2013JC009067

Guemas, V., Doblas-Reyes, F. J., Andreu-Burillo, I., and Asif, M. (2013). Retrospective prediction of the global warming slowdown in the past decade. *Nat. Clim. Change* 3, 649–653. doi: 10.1038/nclimate1863

Hazeleger, W., Guemas, V., Wouters, B., Corti, S., Andreu-Burillo, I., Doblas-Reyes, F. J., et al. (2013). Multiyear climate predictions using two initialization strategies. *Geophys. Res. Lett*. 40, 1794–1798. doi: 10.1002/grl.50355

He, Y., Wang, B., Liu, M., Liu, L., Yu, Y., Liu, J., et al. (2017). Reduction of initial shock in decadal predictions using a new initialization strategy. *Geophys. Res. Lett*. 44, 8538–8547. doi: 10.1002/2017GL074028

Jung, T., Doblas-Reyes, F., Goessling, H., Guemas, V., Bitz, C., Buontempo, C., et al. (2015). Polar lower-latitude linkages and their role in weather and climate prediction. *Bull. Am. Meteorol. Soc*. 96, ES197–ES200. doi: 10.1175/BAMS-D-15-00121.1

Kharin, V., Boer, G., Merryfield, W., Scinocca, J., and Lee, W. (2012). Statistical adjustment of decadal predictions in a changing climate. *Geophys. Res. Lett*. 39:L19705. doi: 10.1029/2012GL052647

Kröger, J., Pohlmann, H., Sienz, F., Marotzke, J., Baehr, J., Köhl, A., et al. (2018). Full-field initialized decadal predictions with the mpi earth system model: an initial shock in the north atlantic. *Clim. Dyn*. 51, 2593–2608. doi: 10.1007/s00382-017-4030-1

Kushnir, Y., Scaife, A. A., Arritt, R., Balsamo, G., Boer, G., Doblas-Reyes, F., et al. (2019). Towards operational predictions of the near-term climate. *Nat. Clim. Change* 9, 94–101. doi: 10.1038/s41558-018-0359-7

Latif, M., and Keenlyside, N. S. (2011). A perspective on decadal climate variability and predictability. *Deep Sea Res. II Top. Stud. Oceanogr*. 58, 1880–1894. doi: 10.1016/j.dsr2.2010.10.066

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., et al. (2019). Improvements in the gistemp uncertainty model. *J. Geophys. Res. Atmos*. 124, 6307–6326. doi: 10.1029/2018JD029522

Madec, G., and NEMO System Team (2015). NEMO ocean engine.

Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., and Stockdale, T. (2013). Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Clim. Dyn*. 41, 2393–2409. doi: 10.1007/s00382-012-1599-2

Mann, M. E., Steinman, B. A., Brouillette, D. J., and Miller, S. K. (2021). Multidecadal climate oscillations during the past millennium driven by volcanic forcing. *Science* 371, 1014–1019. doi: 10.1126/science.abc5810

Manubens, N., Caron, L.-P., Hunter, A., Bellprat, O., Exarchou, E., Fučkar, N. S., et al. (2018). An R package for climate forecast verification. *Environ. Modell. Softw*. 103, 29–42. doi: 10.1016/j.envsoft.2018.01.018

Manubens-Gil, D., Vegas-Regidor, J., Prodhomme, C., Mula-Valls, O., and Doblas-Reyes, F. J. (2016). "Seamless management of ensemble climate prediction experiments on hpc platforms," in *2016 International Conference on High Performance Computing Simulation (HPCS)*, Innsbruck, Austria, 895–900. doi: 10.1109/HPCSim.2016.7568429

Marotzke, J., Müller, W. A., Vamborg, F. S. E., Becker, P., Cubasch, U., Feldmann, H., et al. (2016). MiKlip: a national research project on decadal climate prediction. *Bull. Am. Meteorol. Soc*. 97, 2379–2394. doi: 10.1175/BAMS-D-15-00184.1

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., et al. (2009). Decadal prediction: can it be skillful? *Bull. Am. Meteorol. Soc*. 90, 1467–1486. doi: 10.1175/2009BAMS2778.1

Merryfield, W. J., Baehr, J., Batte, L., Becker, E. J., Butler, A. H., Coelho, C. A. S., et al. (2020). Current and emerging developments in subseasonal to decadal prediction. *Bull. Am. Meteorol. Soc*. 101, E869–E896. doi: 10.1175/BAMS-D-19-0037.1

Mignot, J., Levermann, A., and Griesel, A. (2006). A decomposition of the atlantic meridional overturning circulation into physical components using its sensitivity to vertical diffusivity. *J. Phys. Oceanogr*. 36, 636–650. doi: 10.1175/JPO2891.1

Mogensen, K. S., Balmaseda, M. A., and Weaver, A. (2012). *The Nemovar Ocean Data Assimilation as Implemented in the ECMWF Ocean Analysis for System4*. ECMWF Technical Memorandum. Reading, UK.

Mulholland, D. P., Laloyaux, P., Haines, K., and Balmaseda, M. A. (2015). Origin and impact of initialization shocks in coupled atmosphere-ocean forecasts. *Mnthly Weather Rev*. 143, 4631–4644. doi: 10.1175/MWR-D-15-0076.1

Nadiga, B. T., Verma, T., Weijer, W., and Urban, N. M. (2019). Enhancing skill of initialized decadal predictions using a dynamic model of drift. *Geophys. Res. Lett*. 46, 9991–9999. doi: 10.1029/2019GL084223

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model intercomparison project (scenarioMIP) for CMIP6. *Geosci. Model Dev*. 9, 3461–3482. doi: 10.5194/gmd-9-3461-2016

Ortega, P., Mignot, J., Swingedouw, D., Sevellec, F., and Guilyardi, E. (2015). Reconciling two alternative mechanisms behind bi-decadal variability in the north Atlantic. *Prog. Oceanogr*. 137, 237–249. doi: 10.1016/j.pocean.2015.06.009

Otto, J., Brown, C., Buontempo, C., Doblas-Reyes, F., Jacob, D., Juckes, M., et al. (2016). Uncertainty: lessons learned for climate services. *Bull. Am. Meteorol. Soc*. 97, ES265–ES269. doi: 10.1175/BAMS-D-16-0173.1

Pohlmann, H., Jungclaus, J. H., Köhl, A., Stammer, D., and Marotzke, J. (2009). Initializing decadal climate predictions with the gecco oceanic synthesis: effects on the north atlantic. *J. Clim*. 22, 3926–3938. doi: 10.1175/2009JCLI2535.1

Pohlmann, H., Müller, W. A., Kulkarni, K., Kameswarrao, M., Matei, D., Vamborg, F. S. E., et al. (2013). Improved forecast skill in the tropics in the new miklip decadal climate predictions. *Geophys. Res. Lett*. 40, 5798–5802. doi: 10.1002/2013GL058051

Polkova, I., Köhl, A., and Stammer, D. (2014). Impact of initialization procedures on the predictive skill of a coupled ocean-atmosphere model. *Clim. Dyn*. 42, 3151–3169. doi: 10.1007/s00382-013-1969-4

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos*. 108:D14. doi: 10.1029/2002JD002670

Robson, J., Ortega, P., and Sutton, R. (2016). A reversal of climatic trends in the north Atlantic since 2005. *Nat. Geosci*. 9, 513–517. doi: 10.1038/ngeo2727

Rousset, C., Vancoppenolle, M., Madec, G., Fichefet, T., Flavoni, S., Barthélemy, A., et al. (2015). The louvain-la-neuve sea ice model lim3.6: global and regional capabilities. *Geosci. Model Dev*. 8, 2991–3005. doi: 10.5194/gmd-8-2991-2015

Siegert, S., Bellprat, O., Menegoz, M., Stephenson, D. B., and Doblas-Reyes, F. J. (2017). Detecting improvements in forecast correlation skill: Statistical testing and power analysis. *Mnthly Weather Rev*. 145, 437–450. doi: 10.1175/MWR-D-16-0037.1

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., et al. (2014). Implications of incorporating n cycling and n limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences* 11, 2027–2054. doi: 10.5194/bg-11-2027-2014

Smith, D. M., Eade, R., and Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Clim Dyn*. 41, 3325–3338. doi: 10.1007/s00382-013-1683-2

Smith, D. M., Eade, R., Scaife, A. A., Caron, L.-P., Danabasoglu, G., DelSole, T. M., et al. (2019). Robust skill of decadal climate predictions. *Clim. Atmos. Sci*. 2:13. doi: 10.1038/s41612-019-0071-y

Smith, T., Reynolds, R., Peterson, T., and Lawrimore, J. (2008). Improvements to noaa's historical merged land-ocean surface temperature analysis (1880-2006). *J. Clim*. 21, 2283–2296. doi: 10.1175/2007JCLI2100.1

Tian, T., Yang, S., Karami, M. P., Massonnet, F., Kruschke, T., and Koenigk, T. (2020). Benefits of sea ice thickness initialization for the arctic decadal climate prediction skill in ec-earth3. *Geosci. Model Dev. Discuss*. 2020, 1–29. doi: 10.5194/gmd-2020-331-supplement

Tietsche, S., Day, J., Guemas, V., Hurlin, W., Keeley, S., Matei, D., et al. (2014). Seasonal to interannual arctic sea ice predictability in current global climate models. *Geophys. Res. Lett*. 3, 1035–1043. doi: 10.1002/2013GL058755

Trenberth, K. E., and Shea, D. J. (2006). Atlantic hurricanes and natural variability in 2005. *Geophys. Res. Lett*. 33:L12704. doi: 10.1029/2006GL026894

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., et al. (2005). The ERA-40 reanalysis. *Q. J. R. Meteorol. Soc*. 131, 2961–3012. doi: 10.1256/qj.04.176

Volpi, D., Guemas, V., and Doblas-Reyes (2017a). Comparison of full field and anomaly initialisation for decadal climate prediction: towards an optimal consistency between the ocean and sea-ice anomaly initialisation state. *Clim. Dyn*. 49, 1181–1195. doi: 10.1007/s00382-016-3373-3

Volpi, D., Guemas, V., Doblas-Reyes, F. J., Hawkins, E., and Nichols, N. (2017b). Decadal climate prediction with a refined anomaly initialisation approach. *Clim. Dyn*. 48, 1841–1853. doi: 10.1007/s00382-016-3176-6

Von Storch, H., and Zwiers, F. (2001). *Statistical Analysis in Climate Research*. Cambridge: Cambridge University Press.

Zhang, L., and Wang, C. (2013). Multidecadal north atlantic sea surface temperature and atlantic meridional overturning circulation variability in CMIP5 historical simulations. *J. Geophys. Res. Oceans* 118, 5772–5791. doi: 10.1002/jgrc.20390

**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 Climate

ORIGINAL RESEARCH
published: 12 July 2021
doi: 10.3389/fclim.2021.686058

# Estimating Remaining Carbon Budgets Using Temperature Responses Informed by CMIP6

*Martin Rypdal[1]*, Niklas Boers[2,3,4], Hege-Beate Fredriksen[5], Kai-Uwe Eiselt[5], Andreas Johansen[1], Andreas Martinsen[1], Endre Falck Mentzoni[1], Rune G. Graversen[5] and Kristoffer Rypdal[1]*

[1] Department of Mathematics and Statistics, UiT – The Arctic University of Norway, Tromsø, Norway, [2] Department of Mathematics and Computer Science, Free University of Berlin, Berlin, Germany, [3] Potsdam Institute for Climate Impact Research, Potsdam, Germany, [4] Department of Mathematics and Global Systems Institute, University of Exeter, Exeter, United Kingdom, [5] Department of Physics and Technology, UiT – The Arctic University of Norway, Tromsø, Norway

OPEN ACCESS

**Edited by:**
*Thomas Frölicher, University of Bern, Switzerland*

**Reviewed by:**
*Kevin Grise, University of Virginia, United States*
*Kazuaki Nishii, Mie University, Japan*
*Chris Jones, Met Office Hadley Centre (MOHC), United Kingdom*
*Philip Goodwin, University of Southampton, United Kingdom*

***Correspondence:**
*Martin Rypdal*
*martin.rypdal@uit.no*

**Specialty section:**
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

**Received:** *26 March 2021*
**Accepted:** *18 June 2021*
**Published:** *12 July 2021*

**Citation:**
*Rypdal M, Boers N, Fredriksen H-B, Eiselt K-U, Johansen A, Martinsen A, Falck Mentzoni E, Graversen RG and Rypdal K (2021) Estimating Remaining Carbon Budgets Using Temperature Responses Informed by CMIP6. Front. Clim. 3:686058. doi: 10.3389/fclim.2021.686058*

A remaining carbon budget (RCB) estimates how much $CO_2$ we can emit and still reach a specific temperature target. The RCB concept is attractive since it easily communicates to the public and policymakers, but RCBs are also subject to uncertainties. The expected warming levels for a given carbon budget has a wide uncertainty range, which increases with less ambitious targets, i.e., with higher $CO_2$ emissions and temperatures. Leading causes of RCB uncertainty are the future non-$CO_2$ emissions, Earth system feedbacks, and the spread in the climate sensitivity among climate models. The latter is investigated in this paper, using a simple carbon cycle model and emulators of the temperature responses of the Earth System Models in the Coupled Model Intercomparison Project Phase 6 (CMIP6) ensemble. Driving 41 CMIP6 emulators with 127 different emission scenarios for the 21st century, we find almost perfect linear relationship between maximum global surface air temperature and cumulative carbon emissions, allowing unambiguous estimates of RCB for each CMIP6 model. The range of these estimates over the model ensemble is a measure of the uncertainty in the RCB arising from the range in climate sensitivity over this ensemble, and it is suggested that observational constraints imposed on the transient climate response in the model ensemble can reduce uncertainty in RCB estimates.

**Keywords: remaining carbon budget, climate model emulator, climate sensitivity, CMIP6, transient climate response, integrated assessment model**

## 1. INTRODUCTION

The concept of remaining carbon budgets (RCBs) is appealing and highly applicable to climate mitigation policy (Zickfeld et al., 2009). It allows us to relate a specific climate target to the remaining greenhouse gases humans can release into the atmosphere and still comply with this target. However, like all simple ideas in climate science, it demonstrates ambiguities and uncertainties. Ambiguities arise as the number of specific definitions of temperature targets and RCBs increase during efforts to make concepts and procedures precise. More important, however, are the uncertainties arising from the large spread in model projections, including those of state-of-the-art climate and Earth system models (ESMs). Notably, the spread across the ensemble of

models and the corresponding uncertainties in their equilibrium climate sensitivity (ECS) and RCB do not seem to diminish with increasing model complexity. The state-of-the-art versions of ESMs included in the Coupled Model Intercomparison Project Phase 6 (CMIP6) show a span in ECS of 1.6–5.6 K with 10 models out of 27 exceeding 4.5 K (Zelinka et al., 2020). The increase in ECS is primarily linked to a stronger positive cloud feedback in some of the models, although this is still under investigation. The transient climate response (TCR), defined as the mean global temperature anomaly in a 20-year period centered on year 70 in a model experiment where $CO_2$ concentrations increases by 1% per year, shows a span of 1.3–3.0°C in the CMIP6 experiments shown in **Table 1**. In the remainder of this paper these are referred to as 1% per year experiments.

It is generally accepted that there is an approximate scenario independence in the relationship between the cumulative $CO_2$ emissions and the global mean surface air temperature (GSAT) over a considerable range of realistic mitigation scenarios (Allen et al., 2009; Gregory et al., 2009; Matthews et al., 2009; Meinshausen et al., 2009; Gillett et al., 2013; Goodwin et al., 2015; MacDougall and Friedlingstein, 2015; MacDougall, 2016; Rogelj et al., 2016, 2019). More precisely, there is an approximately linear relationship between the GSAT a given year and the cumulative emissions up to that year. Moreover, it turns out that in scenarios for which the emissions drop to zero at a given year, the GSAT will peak approximately that year, and hence the peak GSAT and the cumulative emissions up to the year of zero annual emissions satisfy the same linear relationship. The increase in GSAT per unit of emitted $CO_2$ given by this linear relation is called the transient climate response to cumulative emissions of carbon (TCRE) (Gregory et al., 2009; Stocker et al., 2013).

In this paper, we compare the cumulative emissions after 2018 in emission scenarios from the Integrated Assessment Modeling Consortium & International Institute for Applied Systems Analysis (IIASA) (Huppmann et al., 2018). Details are given in **Figure 1** and **Table 2**. For those scenarios where annual $CO_2$ emissions have dropped to zero a year in this century, we compute the cumulative emissions up to that year. For those scenarios where annual emissions are still positive in year 2100, we compute the cumulative emissions up to year 2100. The corresponding GSAT values are evaluated for those years by means of a simple impulse-response model, similar to the FaIR model (Smith et al., 2018; Leach et al., 2020). A linear relationship between GSAT and cumulative emissions computed this way is estimated using linear regression, and the slope of the regression line serves as an estimate of TCRE. We define a climate target as a particular GSAT-value, e.g., 2.0°C above the pre-industrial baseline, and the estimated RCB for this target is obtained by the estimated linear relationship.

The transient climate response obtained by this procedure is the so-called *effective* transient climate response to cumulative emissions of carbon (ETCRE), since the emission scenarios contain other anthropogenic emissions than $CO_2$ (Matthews et al., 2017). The ETCRE includes warming from other greenhouse gases than $CO_2$, most importantly methane, and for cooling effects due to atmospheric aerosols. In contrast, the $CO_2$-only TCRE is defined as the warming attributable to $CO_2$ forcing alone. One can estimate the $CO_2$-only TCRE from ESM experiments, driven by atmospheric $CO_2$ concentration increases by 1% per year. The $CO_2$ emissions can be derived from the specified atmospheric $CO_2$ concentrations and the modeled atmosphere-ocean and atmosphere-land $CO_2$-fluxes, and hence the $CO_2$-only TCRE can be computed by dividing the GSAT increase by the cumulative emissions. Using 15 CMIP5 models, Gillett et al. (2013) find $CO_2$-only TCRE in the range 0.22 − −0.65°C per 1,000 Gt $CO_2$, with a mean of 0.44°C per 1,000 Gt $CO_2$. Analyzing 11 CMIP6 models, Arora et al. (2020) found $CO_2$-only TCRE in the range 0.33 − 0.58°C per 1,000 Gt $CO_2$, with a mean of 0.44°C per 1,000 Gt $CO_2$.

The basis of these estimates are scenarios where atmospheric $CO_2$ concentration increases by 1% per year, and not scenarios where we reduce emissions to mitigate climate change. The reason why this does not pose a problem is the above mentioned scenario-independence of the relation between the GSAT and the cumulative emissions. The physical mechanism behind this scenario-independence is a subtle balance between a delayed warming of earlier emissions and a cooling associated with a negative forcing due to $CO_2$ uptake by oceans and land. If all emitted $CO_2$ would have remained in the atmosphere (no sinks) the warming would be delayed due to the thermal inertia of the ocean, and more so in scenarios with high emissions. However, the net $CO_2$ take-up by the ocean and land biosphere will increase as atmospheric concentration increases, and the ESMs indicate that the reduced $CO_2$ forcing due to this uptake approximately offsets the additional forcing represented by the radiation imbalance due to the delayed warming of the ocean surface. It also turns out that the warming is approximately proportional to the size of the emission increment and not strongly dependent on the background $CO_2$ concentration. The implication is the linear dependence of GSAT on cumulative emissions, and hence the GSAT will not increase if the emissions stop; the temperature maximum will coincide with the time the annual emissions drop to zero (Matthews et al., 2017; MacDougall et al., 2020).

There are several ways of adjusting $CO_2$-only TCREs and RCBs to obtain their effective counterparts. One method is to estimate the fraction of the total radiative forcing attributable to anthropogenic $CO_2$-emissions. In the CMIP5 ensemble, the multi-model mean ratio of $CO_2$ forcing to total anthropogenic forcing has been estimated to be 0.86 (Meinshausen et al., 2011; Matthews et al., 2017), which yields a multi-model mean ETCRE of 0.51°C per 1,000 Gt $CO_2$ based on the $CO_2$-only estimate of 0.44°C per 1,000 Gt $CO_2$ (Gillett et al., 2013).

Another approach (Matthews et al., 2017) is to estimate the ETCRE by dividing the observed 1861–2015 GSAT increase of 0.99°C by the 1870–2015 cumulative $CO_2$ emissions of 2,035 Gt $CO_2$ to obtain ETCRE = 0.49°C per 1,000 Gt $CO_2$. However, in ambitious yet realistic future mitigation scenarios, where emissions are brought rapidly to zero in this century, the ratio of $CO_2$ forcing to total anthropogenic forcing may deviate from the historical estimates. The method applied in this paper is to analyze open-source scenarios constructed using integrated assessment models (IAMs) (Huppmann et al., 2018) (**Figure 1** and **Table 2**). In these scenarios, the total emissions of various

**TABLE 1 |** The parameters $d_1, d_2, d_3$ are estimated weights for the three temperature responses with time constants 0.5, 10, 100 yrs, respectively, for the 41 CMIP6 models.

| Earth system model | $d_1$ (K m$^2$/W yrs) | $d_2$ (K m$^2$/W yrs) | $d_3$ (K m$^2$/W yrs) | $F_{2\times CO_2}$ (W m$^{-2}$) | ECS (K) | TCR (K) |
|---|---|---|---|---|---|---|
| ACCESS-CM2 | 0.34 | 0.040 | 0.006 | 3.4 | 4.7 | 2.1 |
| ACCESS-ESM1-5 | 0.49 | 0.035 | 0.005 | 2.8 | 3.9 | 1.9 |
| AWI-CM-1-1-MR | 0.45 | 0.031 | 0.003 | 3.6 | 3.2 | 2.1 |
| BCC-CSM2-MR | 0.60 | 0.020 | 0.004 | 3.1 | 3.0 | 1.7 |
| BCC-ESM1 | 0.35 | 0.037 | 0.004 | 3.0 | 3.3 | 1.8 |
| CAMS-CSM1-0 | 0.39 | 0.023 | 0.001 | 4.2 | 2.3 | 1.7 |
| CanESM5 | 0.22 | 0.054 | 0.006 | 3.7 | 5.6 | 2.5 |
| CAS-ESM2-0 | 0.62 | 0.029 | 0.003 | 3.3 | 3.5 | 2.1 |
| CESM2 | 0.46 | 0.023 | 0.008 | 3.3 | 5.2 | 2.1 |
| CESM2-FV2 | 0.61 | 0.027 | 0.007 | 2.9 | 5.2 | 2.0 |
| CESM2-WACCM | 0.65 | 0.026 | 0.005 | 3.3 | 4.7 | 2.3 |
| CESM2-WACCM-FV2 | 0.27 | 0.037 | 0.007 | 2.9 | 4.8 | 1.7 |
| CMCC-CM2-SR5 | 0.38 | 0.037 | 0.003 | 3.8 | 3.5 | 2.1 |
| CNRM-CM6-1 | 0.17 | 0.051 | 0.004 | 3.6 | 4.9 | 2.2 |
| CNRM-CM6-1-HR | 0.50 | 0.038 | 0.003 | 4.0 | 4.3 | 2.4 |
| E3SM-1-0 | 0.26 | 0.070 | 0.006 | 3.3 | 5.3 | 3.0 |
| EC-Earth3-Veg | 0.33 | 0.050 | 0.006 | 3.4 | 4.3 | 2.6 |
| FGOALS-f3-L | 0.47 | 0.022 | 0.002 | 4.1 | 3.0 | 1.9 |
| FGOALS-g3 | 0.42 | 0.017 | 0.003 | 3.7 | 2.8 | 1.5 |
| GFDL-CM4 | 0.59 | 0.028 | 0.004 | 3.2 | 3.9 | 2.0 |
| GFDL-ESM4 | 0.39 | 0.026 | 0.001 | 3.8 | 2.6 | 1.6 |
| GISS-E2-1-G | 0.74 | 0.021 | 0.000 | 3.6 | 2.8 | 2.2 |
| GISS-E2-1-H | 0.39 | 0.033 | 0.003 | 3.5 | 3.1 | 1.9 |
| GISS-E2-2-G | 0.45 | 0.028 | 0.000 | 3.7 | 2.4 | 1.7 |
| IITM-ESM | 0.41 | 0.015 | 0.001 | 4.6 | 2.4 | 1.7 |
| INM-CM4-8 | 0.53 | 0.022 | 0.002 | 2.7 | 1.8 | 1.3 |
| INM-CM5 | 0.40 | 0.026 | 0.002 | 2.9 | 1.9 | 1.4 |
| IPSL-CM6A-LR | 0.32 | 0.046 | 0.006 | 3.4 | 4.6 | 2.3 |
| KACE-1-0-G | 0.01 | 0.037 | 0.006 | 3.3 | 4.4 | 1.4 |
| MIROC-ES2L | 0.41 | 0.024 | 0.001 | 4.1 | 2.7 | 1.8 |
| MIROC6 | 0.23 | 0.028 | 0.002 | 3.7 | 2.6 | 1.3 |
| MPI-ESM1-2-HR | 0.43 | 0.021 | 0.003 | 3.6 | 3.0 | 1.7 |
| MPI-ESM1-2-LR | 0.39 | 0.024 | 0.002 | 4.2 | 3.0 | 1.8 |
| MRI-ESM2-0 | 0.50 | 0.016 | 0.003 | 3.5 | 3.1 | 1.6 |
| NESM3 | 0.70 | 0.034 | 0.003 | 3.7 | 4.8 | 2.7 |
| NorCPM1 | 0.41 | 0.024 | 0.003 | 3.3 | 3.0 | 1.6 |
| NorESM2-LM | 0.67 | 0.003 | 0.003 | 3.4 | 2.6 | 1.5 |
| NorESM2-MM | 0.52 | 0.009 | 0.003 | 3.8 | 2.5 | 1.3 |
| SAM0-UNICON | 0.49 | 0.030 | 0.002 | 3.9 | 3.7 | 2.3 |
| TaiESM1 | 0.44 | 0.037 | 0.003 | 4.0 | 4.3 | 2.4 |
| UKESM1-0-LL | 0.41 | 0.052 | 0.005 | 3.6 | 5.4 | 2.8 |

*The table also contains the $4\times CO_2$ forcing, ECS, and TCR derived from $4\times CO_2$ and 1% experiments in each model.*

greenhouse gasses and aerosols emissions are known, and we can obtain corresponding temperatures using a simplified version of the FaIR model (Smith et al., 2018; Leach et al., 2020). To assess the uncertainties in RCBs, one should ideally explore an ensemble of realistic mitigation scenarios using the full set of ESMs in the CMIP6 ensemble, which is not feasible due to the computational costs. In this study, we parameterize the temperature response module in our simple model by fitting those model parameters to the temperature response in two standard $CO_2$-forcing scenarios in each of the ESMs in the CMIP6 ensemble. Each of these simple response models emulates the corresponding temperature response to total forcing in the ESM. Combining this temperature module with the greenhouse gas and aerosol forcing module in the FaIR model we compute a temperature response to each of

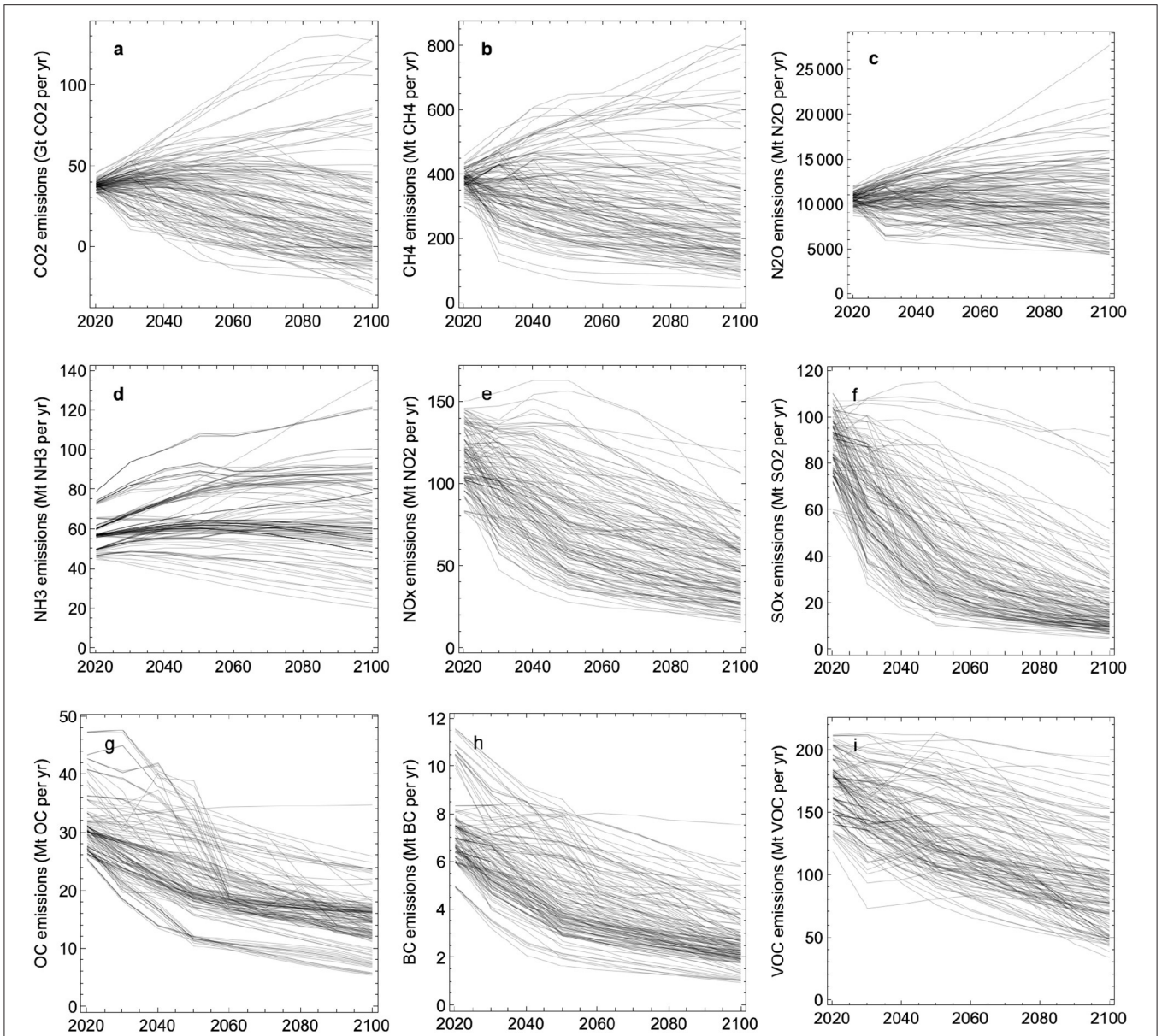



**FIGURE 1 |** Emissions of greenhouse gasses and aerosols in the scenarios listed in **Table 2**. **(a)**: Carbon dioxide ($CO_2$). **(b)**: Methane ($CH_4$). **(c)**: Nitrous oxide ($N_2O$). **(d)**: Ammonia ($NH_3$). **(e)**: Nitrogen oxide ($NO_x$). **(f)**: Sulfur oxide ($SO_x$). **(g)**: Organic carbon (OC). **(h)**: Black carbon (BC). **(i)**: Volatile organic compounds (VOC).

the emission scenarios, and the resulting GSAT time series and $CO_2$ emission time series in each of these model runs allows us to analyze the relationship between cumulative emissions and peak temperatures, and estimate ETCRE and RCBs. Our simple modeling set-up, described in section 2, is based on generally accepted results from the climate modeling literature, while keeping them operational and straightforward.

The philosophy of our approach has similarities to that of MacDougall et al. (2017). They emulated an ensemble of CMIP5 models by means of a climate model of intermediate complexity parameterized to have the climate sensitivity, radiative forcing, and ocean heat uptake efficiency as diagnosed from each CMIP5 model. However, their modeling framework was restricted to 1% per year experiments in the CMIP5 models which imposes carbon fluxes between the atmosphere and ocean and the atmosphere and terrestrial biosphere that may not be consistent with a fully coupled system. Apart from using a different emulator model and emulating a more recent generation of CMIP models, the main novelty in our approach is the application of the emulator model fitted to 41 CMIP6 model versions to an ensemble of 127 emission scenarios for the twenty-first century. The statistics of the estimates of ETRCE and RCB are therefore based on 5,207 distinct simulations of the emulator model.

**TABLE 2 |** The Shared Socioeconomic Pathways (SSPs) and integrated assessment models (IAMs) that form the 127 emission scenarios shown in **Figure 1**.

| | AIM/CGE | GCAM4 | IMAGE | MESSAGE-GLOBIOM | REMIND-MAGPIE | WITCH-GLOBIOM |
|---|---|---|---|---|---|---|
| SSP1-19 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SSP1-26 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SSP1-34 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SSP1-45 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SSP1-60 | | | | | | ✓ |
| SSP1-Baseline | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SSP2-19 | ✓ | ✓ | | ✓ | ✓ | |
| SSP2-26 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SSP2-34 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SSP2-45 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SSP2-60 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SSP2-Baseline | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SSP3-34 | ✓ | | ✓ | ✓ | | ✓ |
| SSP3-45 | ✓ | | ✓ | ✓ | | ✓ |
| SSP3-60 | ✓ | | ✓ | ✓ | | ✓ |
| SSP3-Baseline | ✓ | ✓ | ✓ | ✓ | | ✓ |
| SSP4-19 | | | | | | ✓ |
| SSP4-26 | ✓ | ✓ | ✓ | | | ✓ |
| SSP4-34 | ✓ | ✓ | ✓ | | | ✓ |
| SSP4-45 | ✓ | ✓ | ✓ | | | ✓ |
| SSP4-60 | | ✓ | ✓ | | | ✓ |
| SSP4-Baseline | ✓ | ✓ | ✓ | | | ✓ |
| SSP5-19 | | ✓ | | | ✓ | |
| SSP5-26 | ✓ | ✓ | | | ✓ | |
| SSP5-34 | ✓ | ✓ | ✓ | | ✓ | ✓ |
| SSP5-45 | ✓ | ✓ | ✓ | | ✓ | ✓ |
| SSP5-60 | ✓ | ✓ | ✓ | | ✓ | ✓ |
| SSP5-Baseline | ✓ | ✓ | ✓ | | ✓ | ✓ |

*Note that the emission time series given in **Figure 1** that can be used to drive ESMs for each SSP come in different versions from each IAM used to generate these time series. These IAMs are indicated in the upper row in the table.*

## 2. MODELING SET-UP

We use a simple modeling set-up where atmospheric $CO_2$ concentrations are computed from the emissions, $E_{CO_2}(t)$, using the approach of Leach et al. (2020) which builds on Smith et al. (2018). Details are explained in those papers. The FAIR model uses anthropogenic fossil fuel and land use $CO_2$ emissions as input and partitions them into four pools $R_i$;

$$C_{\mathrm{CO}_2}(t) = C_{\mathrm{CO}_2,\mathrm{PI}} + \sum_{i=1}^{4} R_i(t)\,,$$

where $C_{\mathrm{CO}_2,\mathrm{PI}} = 280\,\mathrm{ppm}$ is the pre-industrial concentration. The pools represent differing time scales of carbon uptake. Here $i = 1$ represents uptake by geological processes, $i = 2$ the deep ocean, $i = 3$ the biosphere, and $i = 4$ the ocean mixed layer. The concentration in each pool varies according to the equation,

$$\frac{dR_i}{dt} = a_i E_{\mathrm{CO}_2}(t) - \frac{1}{\tau_{\mathrm{CO}_2,i}\alpha} R_i\,,$$

where $E_{\mathrm{CO}_2}(t)$ is the $CO_2$ emission rate, $a_i$ is the partition fraction ($\sum_{i=1}^{4} a_i = 1$), and $\tau_{\mathrm{CO}_2,i}\alpha$ is the characteristic time scale of the $i$'th pool, where the state-dependence is built into the model by letting $\alpha$ depend on the global temperature $T(t)$ and the cumulative uptake $G_u$ of agent $u$ since initialization of the model;

$$G_u(t) = \sum_{i=1}^{4}\left[a_i \int_{t_0}^{t} E_{\mathrm{CO}_2}(s)ds - R_i(t)\right].$$

The time $t_0$ refers to the year 1750. The model for $\alpha$ is

$$\alpha(T, G_u) = g_0 \exp\left(\frac{r_0 + r_u G_u + r_T T}{g_1}\right),$$

where $r_0$ is the strength of pre-industrial uptake from the atmosphere, $r_u$ is sensitivity of uptake from atmosphere to cumulative uptake of agent since model initialization, and $r_T$ is such sensitivity to model temperature. The parameters $g_0$ and $g_1$ are determined by $a_i$ and $\tau_{\mathrm{CO}_2,i}$, $i = 1,\ldots,4$, and are not independent parameters. The equations determining them are given and explained in Leach et al. (2020).

We model the concentrations of methane and nitrous oxide as linear responses of scenario data for emissions:

$$C_{\mathrm{CH_4}}(t) = C_{\mathrm{CH_4,\,PI}} + \int_{t_0}^{t} G_{\mathrm{CH_4}}(t-s)E_{\mathrm{CH_4}}(s)\,ds\,.$$

and

$$C_{\mathrm{N_2O}}(t) = C_{\mathrm{N_2O,\,PI}} + \int_{t_0}^{t} G_{\mathrm{N_2O}}(t-s)E_{\mathrm{N_2O}}(s)\,ds\,.$$

with $G_{\mathrm{CH_4}}(t) = c_{\mathrm{CH_4}}\, e^{-t/\tau_{\mathrm{CH_4}}}$, and similarly for $N_2O$. The factors $c_{\mathrm{CH_4}}$ and $c_{\mathrm{N_2O}}$ are chosen to yield the current atmospheric methane and nitrous oxide concentrations based on the emissions since 1750 (Boden et al., 2017; Saunois et al., 2020). The pre-industrial concentrations are set to $C_{\mathrm{CH_4,\,PI}} = 700$ ppb and $C_{\mathrm{N_2O,\,PI}} = 270$ ppb.

The radiative forcing associated with greenhouse gas concentrations is computed using Equations (7)– (9) in Smith et al. (2018) with parameters presented in **Table 3**:

$$F_{\mathrm{CO_2}} = \Big[\xi_1(C_{\mathrm{CO_2}} - C_{\mathrm{CO_2,\,PI}})^2 + \xi_2|C_{\mathrm{CO_2}} - C_{\mathrm{CO_2,\,PI}}| + \xi_3(C_{\mathrm{N_2O}} + C_{\mathrm{N_2O,\,PI}}) + \frac{F_{2\times\mathrm{CO_2}}}{\ln(2)}\Big]\ln\left(\frac{C_{\mathrm{CO_2}}}{C_{\mathrm{CO_2,\,PI}}}\right)$$

$$F_{\mathrm{CH_4}} = \Big[\xi_4(C_{\mathrm{CH_4}} + C_{\mathrm{CH_4,\,PI}}) + \xi_5(C_{\mathrm{N_2O}} + C_{\mathrm{N_2O,\,PI}}) + \xi_6\Big]\left(\sqrt{\frac{C_{\mathrm{CH_4}}}{\mu}} - \sqrt{\frac{C_{\mathrm{CH_4,\,PI}}}{\mu}}\right)$$

$$F_{\mathrm{N_2O}} = \Big[\xi_7(C_{\mathrm{CO_2}} + C_{\mathrm{CO_2,\,PI}}) + \xi_8(C_{\mathrm{N_2O}} + C_{\mathrm{N_2O,\,PI}}) + \xi_9(C_{\mathrm{CH_4}} + C_{\mathrm{CH_4,\,PI}}) + \xi_{10}\Big]\left(\sqrt{\frac{C_{\mathrm{N_2O}}}{\mu}} - \sqrt{\frac{C_{\mathrm{N_2O,\,PI}}}{\mu}}\right),$$

where $\mu = 1$ ppm. The number $F_{2\times\mathrm{CO_2}}$ is the forcing associated with a $CO_2$-doubling. This number is model-dependent and obtained from the Gregory plots for the abrupt $4\times CO_2$ experiments in the CMIP6 ensemble (Gregory et al., 2004). Aerosol forcing is modeled to be proportional to aerosol emissions:

$$F_{\mathrm{aero}} = \gamma_{\mathrm{NH_3}}E_{\mathrm{NH_3}} + \gamma_{\mathrm{BC}}E_{\mathrm{BC}} + \gamma_{\mathrm{OC}}E_{\mathrm{OC}} + \gamma_{\mathrm{NOX}}E_{\mathrm{NOX}} + \gamma_{\mathrm{VOC}}E_{\mathrm{VOC}} + \gamma_{\mathrm{SOX}}E_{\mathrm{SOX}} + F_{\mathrm{aero,\,cloud}}(E_{\mathrm{BC}}, E_{\mathrm{OC}}, E_{\mathrm{SOX}}),$$

where the additional term

$$F_{\mathrm{aero,\,cloud}}(E_{\mathrm{BC}}, E_{\mathrm{OC}}, E_{\mathrm{SOX}}) = F_0\frac{f(E_{\mathrm{BC}}, E_{\mathrm{OC}}, E_{\mathrm{SOX}}) - f(E_{\mathrm{BC}}^{(1765)}, E_{\mathrm{OC}}^{(1765)}, E_{\mathrm{SOX}}^{(1765)})}{f(E_{\mathrm{BC}}^{(2011)}, E_{\mathrm{OC}}^{(2011)}, E_{\mathrm{SOX}}^{(2011)}) - f(E_{\mathrm{BC}}^{(1765)}, E_{\mathrm{OC}}^{(1765)}, E_{\mathrm{SOX}}^{(1765)})}$$

**TABLE 3 |** Overview of the model parameters used to compute greenhouse gas concentrations, greenhouse forcing, and aerosol forcing, following the approaches in Smith et al. (2018) and Leach et al. (2020).

| Parameter | Unit | Value |
|---|---|---|
| $C_{CO_2,PI}$ | ppm | 280 |
| $C_{CH_4,PI}$ | ppb | 700 |
| $C_{N_2O,PI}$ | ppb | 270 |
| $a_1$ | ppm/Gt $CO_2$ | 0.059 |
| $a_2$ | ppm/Gt $CO_2$ | 0.061 |
| $a_3$ | ppm/Gt $CO_2$ | 0.077 |
| $a_4$ | ppm/Gt $CO_2$ | 0.075 |
| $\tau_{CO_2,1}$ | yrs | $1\times10^5$ |
| $\tau_{CO_2,2}$ | yrs | 394 |
| $\tau_{CO_2,3}$ | yrs | 36.5 |
| $\tau_{CO_2,4}$ | yrs | 4.3 |
| $\tau_{CH_4}$ | yrs | 12.3 |
| $\tau_{N_2O}$ | yrs | 110 |
| $r_0$ | | 30.4 |
| $r_u$ | 1/(Gt $CO_2$) | $4.8\times10^{-3}$ |
| $r_T$ | $K^{-1}$ | 2.64 |
| $g_0$ | | 0.01 |
| $g_1$ | | 11.4 |
| $c_{CH_4}$ | ppb/(Mt $CH_4$ $yr^{-1}$) | 0.34 |
| $c_{N_2O}$ | ppb/(Mt $N_2O$ $yr^{-1}$) | $2\times10^{-4}$ |
| $\xi_1$ | $Wm^{-2}/ppm^2$ | $-2.4\times10^{-7}$ |
| $\xi_2$ | $Wm^{-2}$/ppm | $7.2\times10^{-4}$ |
| $\xi_3$ | $Wm^{-2}$/ppb | $-1.05\times10^{-4}$ |
| $\xi_4$ | $Wm^{-2}$/ppb | $-6.5\times10^{-7}$ |
| $\xi_5$ | $Wm^{-2}$/ppb | $-4.1\times10^{-6}$ |
| $\xi_6$ | $Wm^{-2}$ | 0.043 |
| $\xi_7$ | $Wm^{-2}$/ppm | $-4.0\times10^{-6}$ |
| $\xi_8$ | $Wm^{-2}$/ppb | $2.1\times10^{-6}$ |
| $\xi_9$ | $Wm^{-2}$/ppb | $-2.45\times10^{-6}$ |
| $\xi_{10}$ | $Wm^{-2}$ | 0.117 |
| $\gamma_{NH_3}$ | $Wm^{-2}$/(Mt $NH_3$ $yr^{-1}$) | $-1.56\times10^{-3}$ |
| $\gamma_{BC}$ | $Wm^{-2}$/(Mt BC $yr^{-1}$) | $16\times10^{-3}$ |
| $\gamma_{OC}$ | $Wm^{-2}$/(Mt OC $yr^{-1}$) | $-1.45\times10^{-3}$ |
| $\gamma_{NOx}$ | $Wm^{-2}$/(Mt $NO_2$ $yr^{-1}$) | $-3.6\times10^{-4}$ |
| $\gamma_{VOC}$ | $Wm^{-2}$/(Mt VOC $yr^{-1}$) | $-3.8\times10^{-4}$ |
| $\gamma_{SOx}$ | $Wm^{-2}$/(Mt $SO_2$ $yr^{-1}$) | $-2.07\times10^{-3}$ |
| $F_0$ | $Wm^{-2}$ | −0.45 |
| $\beta_1$ | | −1.95 |
| $\beta_2$ | 1/(Mt $SO_2$ $yr^{-1}$) | $5.55\times10^{-3}$ |
| $\beta_3$ | 1/(Mt BC $yr^{-1}$) | $13.9\times10^{-3}$ |

accounts for aerosol-cloud indirect effect. Here $f(E_{\mathrm{BC}}, E_{\mathrm{OC}}, E_{\mathrm{SOX}}) = \beta_1\ln(1+\beta_2E_{\mathrm{SOX}} + \beta_3(E_{\mathrm{BC}} + E_{\mathrm{OC}}))$. All parameter values are listed in **Table 3**.

Our model for the temperature response is

$$T(t) = \int_{t_0}^{t} G_T(t-s)F_{\mathrm{tot}}(s)ds\,, \qquad (1)$$

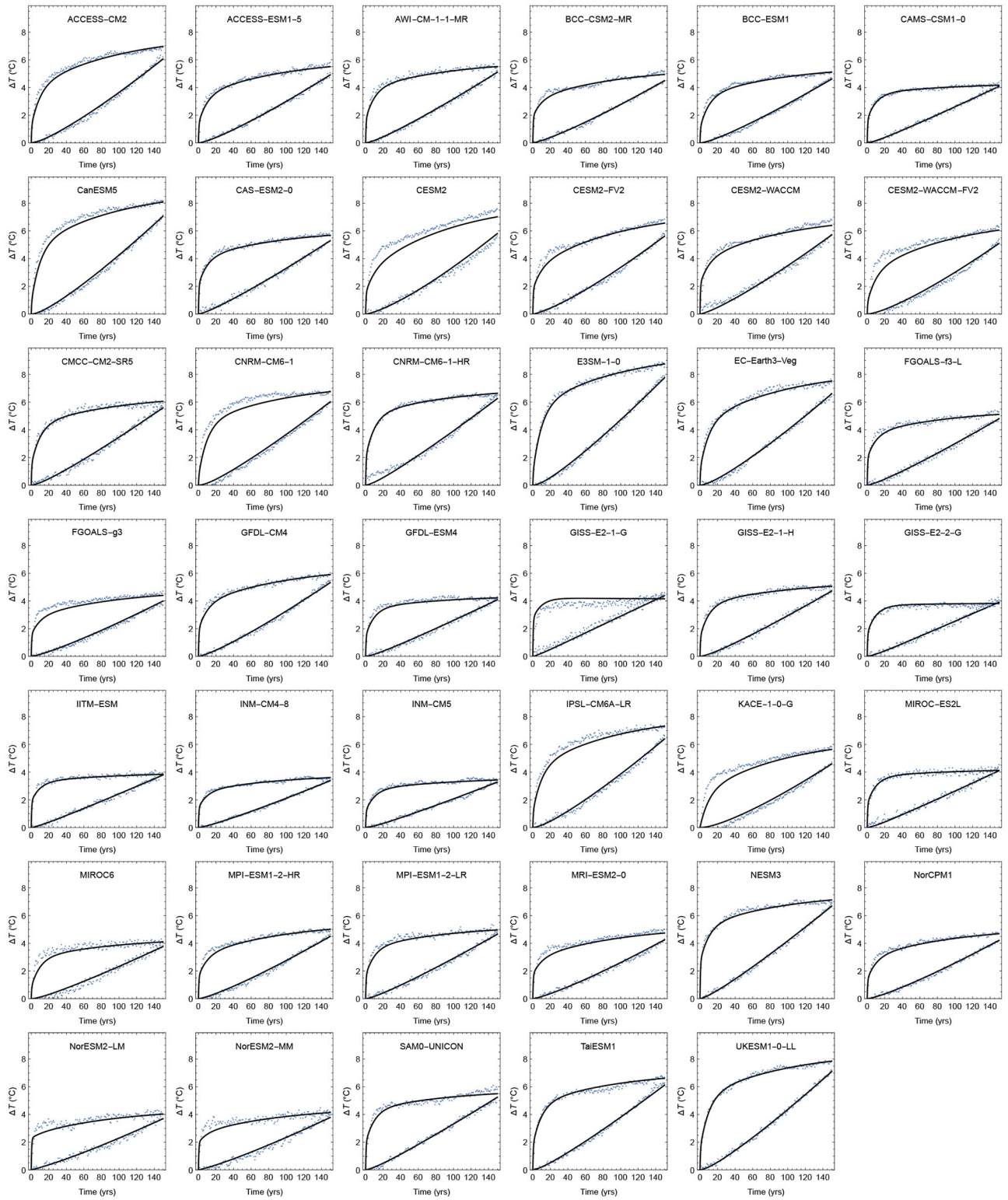



**FIGURE 2 |** The points are the GSAT from the first 150 yrs in $4\times CO_2$ experiments and 1%-per-yr experiments in CMIP6. The solid curves are the simultaneous least-squares estimates to the two time series of a linear response to the forcings $F(t) = F_{4\times CO_2}\ \Theta(t)$ and $F(t) = F_{2\times CO_2}(\log(1.01)/\log(2))t$.

with $F_{\rm tot} = F_{\rm CO_2} + F_{\rm CH_4} + F_{\rm N_2O} + F_{\rm aero}$ and

$$G_T(t) = \sum_{i=1}^{3} d_i e^{-t/\tau_i}\,.$$

To prevent statistical overfitting we use fixed, but well-separated time scales $\tau_i$, chosen to be 0.5, 10, and 100 yrs (Fredriksen and Rypdal, 2017). The factors $d_i$ are estimated simultaneously from the first 150 yrs in $4\times CO_2$ experiments in CMIP6, and the first 150 yrs in experiments where the $CO_2$ concentration is increased by 1% per yr (**Figure 2**). The time series are drift-adjusted using control runs of the CMIP6 models. The method for estimation is linear regression and the forcings used are $F(t) = F_{4\times CO_2}\Theta(t)$, where $\Theta(t)$ is the unit step function, and $F(t) = F_{2\times CO_2}(\ln(1.01)/\ln(2))t$, for the two experiments, respectively. The slow climate response, in this case the parameter $d_3$, is not well-constrained by 150-yr runs (Sanderson, 2020). However, the analyses in presented in this paper only concern GSAT up to the year 2100, and are insensitive to this uncertainty. **Table 1** shows the estimated parameters $d_1$, $d_2$, and $d_3$ for the 41 models in the CMIP6 ensemble. The table also shows the TCR, ECS, and $F_{2\times CO_2}$ of each climate model. The ECS-values are estimated using the standard Gregory-plot technique and the TCR-values are obtained from the CMIP6 runs where $CO_2$ concentrations are increased by 1% per year. Using the updated HadCrut data set we set the present-day GSAT at 1.1°C above the 1850–1900 baseline (Morice et al., 2012). Historical $CO_2$ and methane emissions are obtained from Hoesly et al. (2018).

The integral in Equation (1) is computed as a discrete sum

$$\sum_{s=t_0}^{t} G_T(t-s+\delta)F_{\rm tot}(s)\Delta s\,, \qquad (2)$$

where $F_{\rm tot}(s)$ are annual forcing values and $\Delta s = 1$ yr. We use $\delta = 0.5$ yrs, which corresponds to the midpoint rule in the approximation of the integral. Using $\delta = 0$ will lead to over-estimation of the temperature response compared to the exact integrals used in the parameter estimation. **Figure 3** shows that $\delta = 0.5$ yrs gives agreement between the TCRs estimated directly from the ESMs and the TCRs estimated from the discrete-time emulators.

## 3. RESULTS

Our results show that the linear relationship between total emissions and maximum GSAT is an excellent approximation for each temperature-response model for cumulative emissions up to 5,000 Gt $CO_2$ after 2018, but that the ETCRE varies considerably over the ensemble of different temperature responses (**Figure 4**). Over the ensemble we find a mean ETCRE of 0.42°C per 1,000 Gt CO2, with a 66% confidence range of 0.35–0.47°C per 1,000 Gt CO2 (**Figure 5A**). Here 66% confidence range means the range between the 17 and 83% percentiles for the ensemble of 41 ETCREs estimated as the slope of the regression lines shown in **Figure 4**. Throughout this paper, 66% confidence range always refers to a range over a specified ensemble of models.

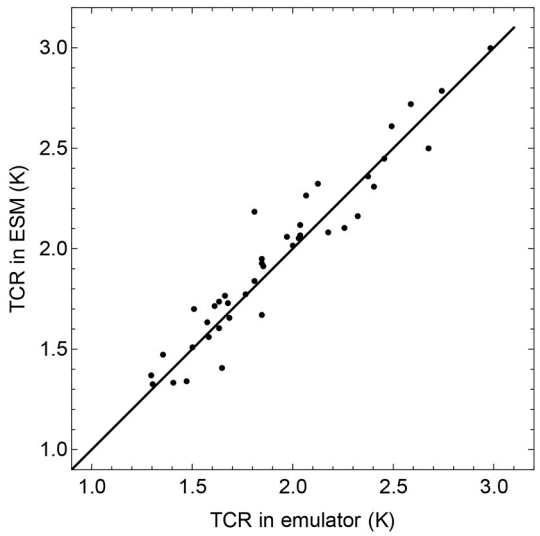



**FIGURE 3 |** Comparison of TCR estimated from the 1% per year experiments in 41 ESMs and the TCR estimated in the corresponding emulators. The figure shows that the computed climate response using the discretization in Equation (2) is unbiased when using the midpoint rule, i.e., $\delta = 0.5$ yrs.

Nijsse et al. (2020) have recently constrained TCR to the range 1.3–2.1 K by leaving out models with TCR$\geq$ 2.2 from the ensemble. Restricting to this sub-ensemble, the 66% confidence range for ETCRE is lowered to 0.33–0.40°C per 1,000 Gt CO2 (**Figure 5B**).

The cumulative emissions (the RCB) for a given peak temperature target, computed for each ESM, is estimated from the regression line for that ESM in **Figure 4**. It allows us to construct the histograms shown in **Figures 6A–D**. They show how the cumulative emissions are distributed over the 41 ESMs for four different temperature targets. We note that the RCB varies by a factor of two over the model ensemble. Restricting to the sub-ensemble of models with TCR$<$ 2.2, we find the histograms shown in **Figures 6E–H**. This restriction imposes a significant constraint on the lower end of the RCB range for each target, ruling out the more pessimistic estimates for the remaining carbon budget.

The differences in ETCRE between the high- and low-sensitivity models are illustrated in **Figure 7** For the sub-ensemble of climate models with TCR $\geq$ 2.2 K, the mean ETCRE is 0.52°C per 1,000 Gt CO2, with a 66% confidence range of 0.45–0.57°C per 1,000 Gt CO2. For those climate models with TCR $<$ 2.2 K, the mean ETCRE is 0.37°C per 1,000 Gt CO2, with a 66% range of 0.30–0.43°C per 1,000 Gt CO2. In **Figure 8** the maximum GSAT is plotted against the cumulative $CO_2$ emission. This cumulative emission is different for every emission scenario, while the maximum GSAT varies across models for the same scenario. Thus, there are 127 columns of points, one for each scenario, and each point in the same column gives the GSAT for a specific model driven by that scenario. Hence, each column contains 41 points, where the blue points represents models in the TCR$<$ 2.2 K sub-ensemble, and the red points the models in

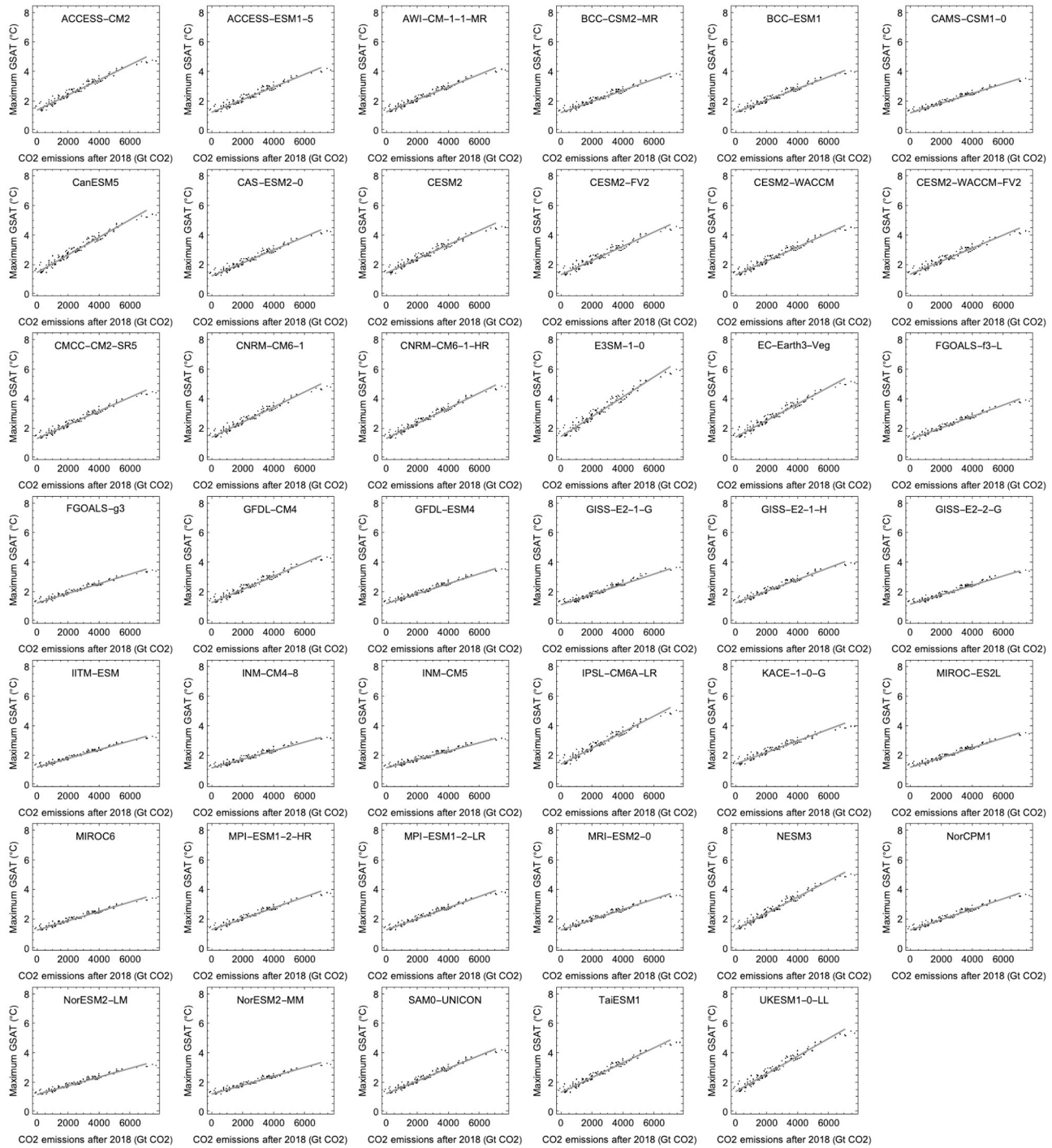



**FIGURE 4 |** Estimates of ETRCE. Each panel contains 127 points representing maximum GSAT vs. cumulative $CO_2$ emissions obtained from our simple emulator for the indicated CMIP6 ESM, i.e., each point represents one of the 127 emission scenarios. The parameters $d_1, d_2, d_3$ estimated for each ESM are shown in **Table 1**. The regression lines demonstrate approximately linear relationships between total positive $CO_2$ emissions between 2018 and 2100 and the maximum GSAT for the ensemble of emission scenarios for each of the 41 different climate models in the CMIP6 ensemble. ETCRE estimates are obtained from the slopes of regression lines.

the TCR$\geq$ 2.2 K sub-ensemble. The blue- and red-shaded areas depict the 66% ranges of GSAT in the two sub-ensembles for each scenario. We observe that the corresponding difference in RCB between the two sub-ensembles grows approximately linearly with increasing temperature target above 1.5 °C, and that model uncertainty in RCB grows linearly with the temperature target.

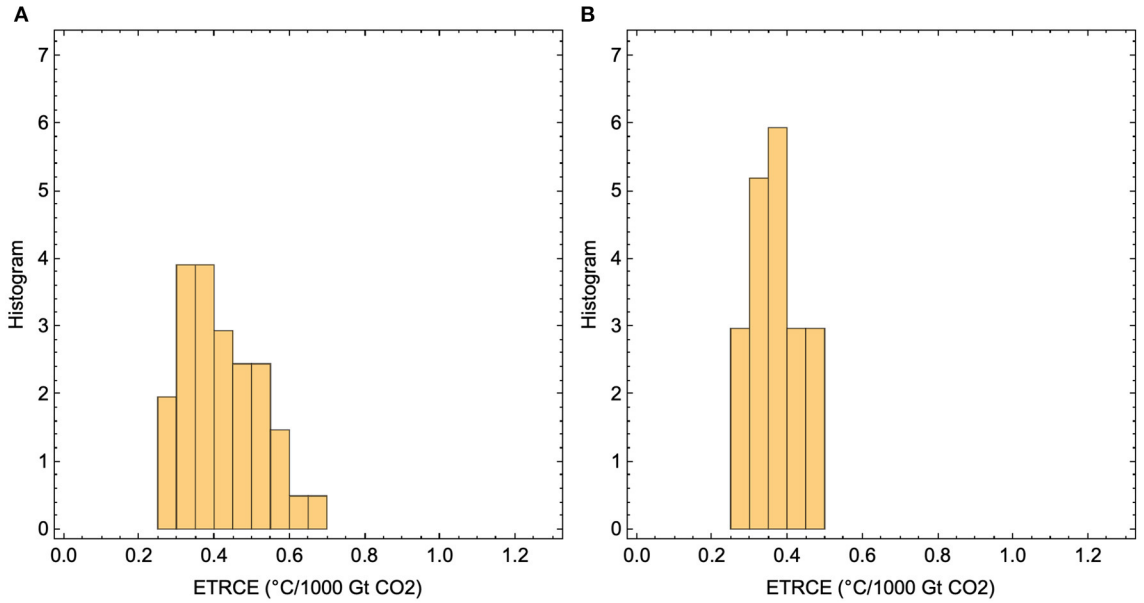



**FIGURE 5 | (A)** The histogram of ETCRE-values obtained by using the full set of temperature-response models. **(B)** The histogram of ETCRE-values obtained using only temperature response models informed by CMIP6 models with TCR in the range 1.3–2.2 K.

The width of the 66% confidence range for RCB increases from 800 to 2,500 GtCO2 as the target increases from 1.5 to 3.0 °C.

**Figure 8** shows that the estimated ETCRE scales linearly with the TCR of the ESMs estimated from 150 yrs 1% per year experiments. The estimated relationship is

$$\mathrm{ETCRE} = a + b\,\mathrm{TCR}\,,$$

with $a = 0.21$°C per 1,000 Gt CO2 and $b = 0.01$ per 1,000 Gt CO2. The TCR ranges from 1.3 to 3.0°C in the CMIP6 ensemble, corresponding to a range of 0.29 to 0.64°C per 1,000 Gt CO2 in ETCRE.

## 4. DISCUSSION

The results shown in **Figure 4** demonstrate the linearity of the maximal GSAT response to cumulative emissions over the ensemble of 127 SSP scenarios in all emulated CMIP6 models. It allows accurate estimates (small spread over the ensemble of scenarios) of the ETCRE associated with each emulated CMIP6 model. There is, however, a large spread in this model-specific ETRCE over the CMIP6 ensemble as shown in **Figures 5A,B**. The importance and novelty of these results are that the main uncertainty of the ETRCE and the associated RCB is not due to the spread of realistic emission scenarios, but rather the spread of sensitivities over the CMIP6 model ensemble. **Figure 8** also demonstrates the close correlation between the ETRCE and the transient climate response TCR over the CMIP6 ensemble, which suggests that constraints obtained on the climate sensitivity, leading for instance to removal of hyper-sensitive models from the ensemble, will reduce the uncertainty in the estimates of the ETRCE and the RCB.

The proportionality between TCRE and TCR is not new, this was discussed in Gillett et al. (2013), and more recently in Jones and Friedlingstein (2020). It is shown in this paper, however, that it also holds for the ETCRE, i.e., as the non-$CO_2$ emissions are taken into account. This may not come entirely as a surprise, since studies based on the standard Representative Concentration Pathways Scenarios RCP2.6, RCP4.5, RCP6.0, and RCP8.5 show consistent dependence between non-$CO_2$ and $CO_2$ forcing throughout the twenty-first century (e.g., Williams et al., 2017). The RCPs set pathways for greenhouse gas concentrations from which the emission pathways are derived, and hence do not represent realistic socioeconomic scenarios. The SSPs, on the other hand, are based on narratives describing broad socioeconomic trends that could shape future society. These are intended to span the range of plausible futures, so we believe that the confirmation of the proportionality for this ensemble of scenarios strengthens the prospects of using the TCR to constrain TCRE and RCB.

The design of our study precludes explicit study of uncertainty due to model variation in the sensitivity of radiative forcing from $CO_2$ emission. This is because the parameters of module of the emulator that computes forcing from emissions are fixed and not fitted to each CMIP6 model. Our rationale for not fitting all the coefficients of the module that calculates forcing based on the input from emission scenarios is two-fold: First, we would need to have available results from at least one model run forced by such full emission scenarios for all the 41 CMIP6 models in order to make such a calibration. Second, **Table 3** shows the additional 40 parameters that would

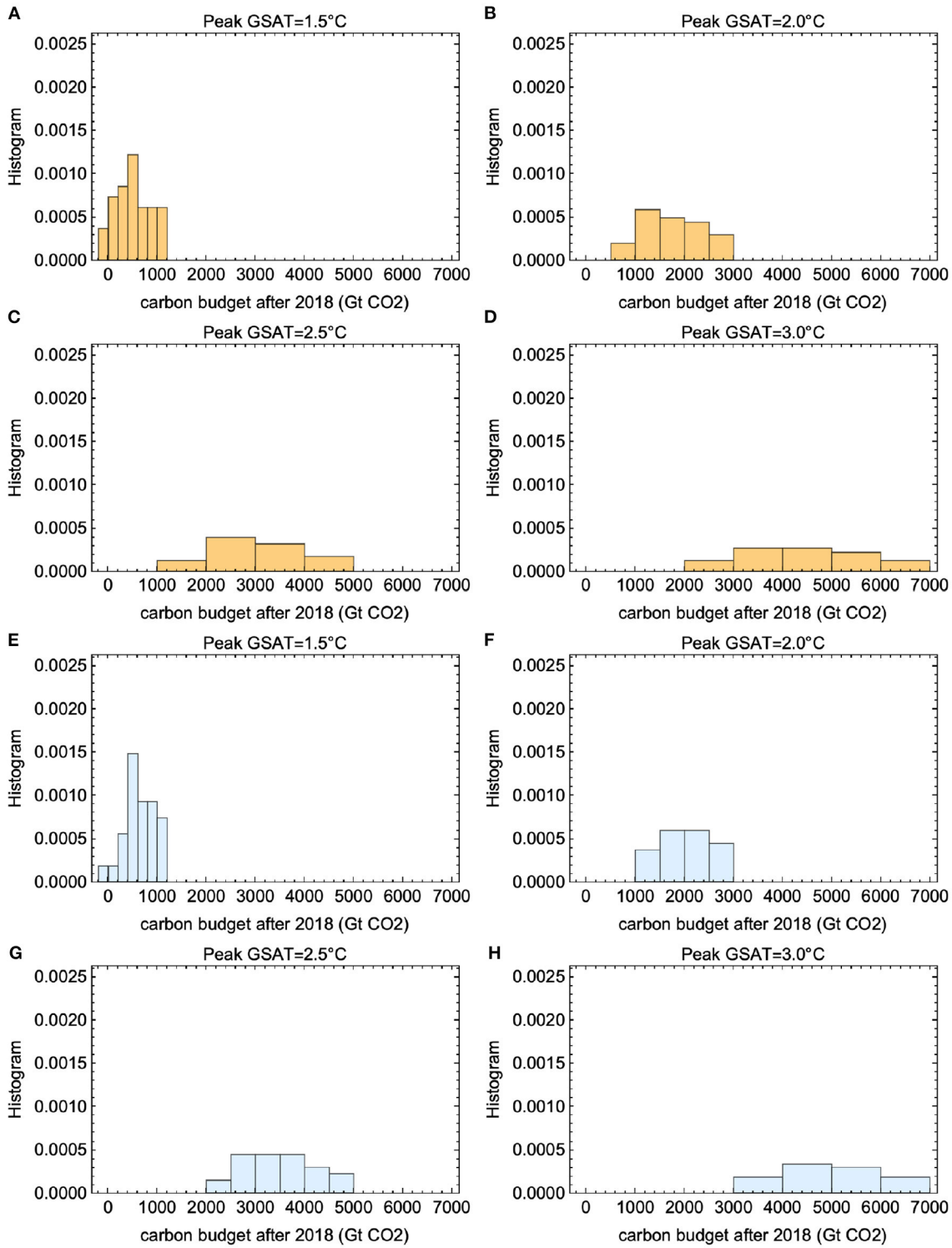



**FIGURE 6 |** Estimates of RCB for different temperature targets. Panels **(A–D)** show histograms for the 1.5, 2.0, 2.5, and 3.0°C-targets, respectively. **(E–H)** The same estimates, but based only on the temperature response models informed by CMIP6 models with TCR in the range 1.3–2.2 K.

be fitted to each of these model runs. Even if only a subset of the parameters were subject to fitting, the risk of statistical overfitting would be unavoidable. It seems that one is left with the choice between using a reasonably complex module with fixed coefficients for computation of forcing from emission input, or a very simple model with a few fitting parameters.

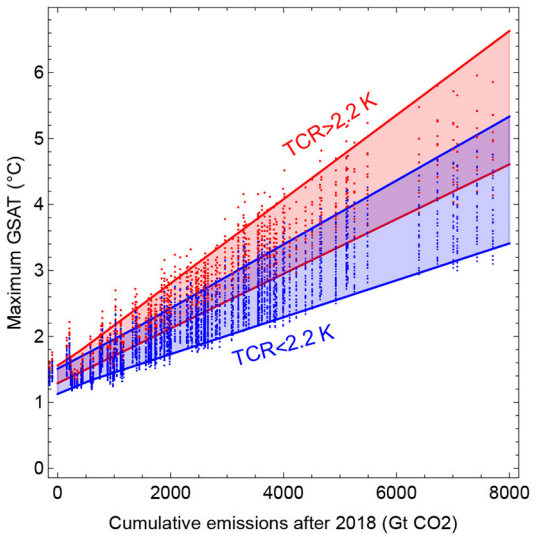



**FIGURE 7 |** Each column of points shows maximum GSAT for a given emission scenario, so their spread indicates the variance over the ensemble of ESMs. The red points are for ESMs with TCR≥2.2 K, and the blue points for ESMs with TCR<2.2 K.

In the former case, one will miss the variability among the CMIP6 models when it comes to forcing calculations. In the latter case, one may miss important mechanisms. We have chosen the former option, and unfortunately that precludes explicit evaluation of the contribution of some aspects of the CMIP6 variability to the uncertainty in RCB. Thus, the real ESM model uncertainty is probably greater than estimated in this paper, but at present, we have no means of quantifying this additional uncertainty.

The performance of the emulator model could be tested if we had available CMIP6 model runs forced by the 127 emission scenarios, or at least, by a selected few. In the CMIP6 database, we find a few runs driven by selected SSPs, but we have not been able to identify exactly the emission data input used in these runs. As more data from CMIP6 runs becomes available, we hope more comprehensive testing and refinement of the emulator will be possible.

Our analyses show that estimates of RCBs are associated with considerable uncertainty related to the global temperature response to radiative forcing, quantified for example as the spread over different members of the CMIP6 model ensemble. We further show that model estimates of ETCRE correlate strongly with TCR across models, which is convenient since much effort is being made to use observations to constrain the TCR and ECS. Cox et al. (2018) used the instrumental temperature record to constrain ECS in the CMIP5 ensemble to a 66% confidence interval of 2.2–3.4 K. This approach was based on an assumed theoretical relation between ECS and unforced temperature fluctuations, whereas the analysis reflected the forced temperature responses (Brown et al., 2018; Po-Chedley et al., 2018; Rypdal et al., 2018). To circumvent this issue, Jiménez-de-la Cuesta and Mauritsen (2019) used observational data of post 1970 warming to constrain ECS in the CMIP5 ensemble to a 95% confidence interval of 1.72–4.12 K. This result is roughly consistent with the recent results of Sherwood et al. (2020), who used multiple lines of evidence to argue that ECS above 4.5 K is unlikely. The results of this paper suggests that ruling-out the ESMs with the highest climate sensitivity would narrow the uncertainty in ETCRE. An alternative, but related, approach is to tune emulators to observational data (Smith et al., 2018). The estimated uncertainty in ETCRE corresponds directly to the uncertainty in RCB, which we find to depend linearly on the temperature target. Hence, the less ambitious the temperature target, the higher the uncertainty in the corresponding RCB.

Since we use a relatively simple carbon cycle model, there is an additional source of uncertainty that is not accounted for, induced by potentially changing feedbacks in the dynamics of the Earth system, which have been shown to be a significant source of uncertainty for RCBs (Jones and Friedlingstein, 2020). Permafrost thawing in response to rising surface temperatures leads to the release of greenhouse gases stored in high-latitude soils. The release of these additional greenhouse gases will in turn accelerate global warming.

The Amazon rainforest is another example of such a positive feedback. It has been argued and observed in climate model projections that the Amazon ecosystem might transition from its current rainforest state to a state dominated by grassland and savanna vegetation (Cox et al., 2004; Hirota et al., 2011; Lovejoy and Nobre, 2018, 2019) which would be accompanied by the release of large amounts of carbon dioxide to the atmosphere. Carbon-cycle feedbacks have an overall accelerating effect on global warming (Cox et al., 2000), and the situation seems particularly evident for the Amazon. Increasing tree mortality during a transition from rainforest to Savanna will cause the rainforest to turn from a global carbon sink to a global source of carbon (Brienen et al., 2015), as has already happened temporarily during the severe droughts of 2005 and 2010 (Feldpausch et al., 2016). Climate-change-induced dieback of the Amazon would lead to the release of additional greenhouse gases, which would further accelerate global temperature rise.

The Amazon rainforest also provides an example of how anthropogenic forcing other than greenhouse gas release can affect the climate system. Modeling evidence suggests that only partial deforestation of the Amazon rainforest might—through intricate couplings between evapotranspiration, condensational latent heating, and the South American low-level circulation system—lead to a collapse of the South American monsoon system and thus, ultimately, of the Amazon rainforest (Boers et al., 2017).

As a third example, the ice-albedo feedback implies rising temperatures in the Arctic, leading to accelerating sea ice retreat, lowering albedo, and effectively increasing mean surface temperatures regionally. This positive feedback contributes to uncertainty in ETCRE, which translates to even more considerable uncertainty in the amount of greenhouse gas

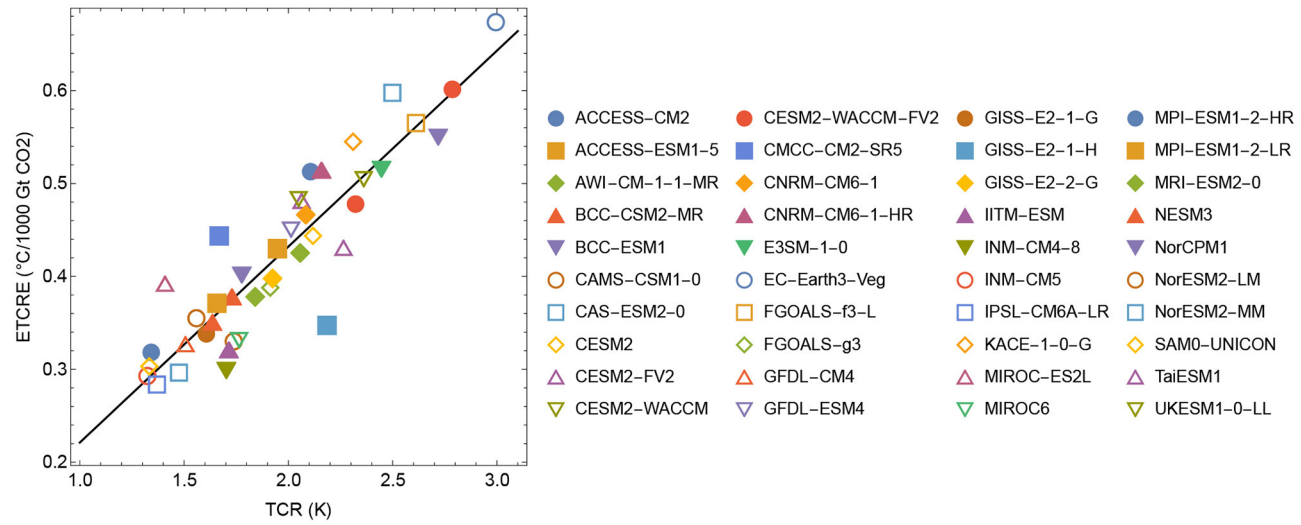



**FIGURE 8 |** The approximately linear relationship between the estimated ETCRE and TCR of the CMIP6 model used to inform the temperature response model.

emissions we can allow to still limit peak temperature to specified targets.

These three examples of positive Earth system feedbacks are all—in some form—implemented in state-of-the-art models such as the ones from the CMIP6 suite (Eyring et al., 2016), and systematic searches have revealed many abrupt transitions related to such positive feedbacks in model projections (Drijfhout et al., 2015). Nevertheless, it is still assumed that state-of-the-art models remain too stable (Valdes, 2011). The presence of positive feedbacks and potential tipping points within the Earth system adds a layer of uncertainty to RCBs that is extremely difficult to quantify.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

MR, AJ, AM, EF, NB, and KR designed the study with input from all authors. K-UE, H-BF, AM, and MR processed and analyzed the CMIP6 data. MR, AJ, AM, and EF carried out the analyses. MR, NB, RG, and KR wrote the original manuscript with input from all authors, while KR wrote the revised version with input from MR.

## FUNDING

This was TiPES contribution #66; the TiPES 399 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 401 No. 820970. This work was supported by the UiT Aurora Centre Program, UiT The Arctic University of Norway (2020), and the Research Council of Norway (project number 314570). NB acknowledges funding by the Volkswagen foundation.

## REFERENCES

Allen, M. R., Frame, D. J., Huntingford, C., Jones, C. D., Lowe, J. A., Meinshausen, M., et al. (2009). Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* 458, 1163–1166. doi: 10.1038/nature08019

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., et al. (2020). Carbon-concentration and carbon-climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences* 17, 4173–4222. doi: 10.5194/bg-17-4173-2020

Boden, T., Marland, G., and Andres, R. J. (2017). *Global, Regional, and National Fossil-Fuel Co2 Emissions (1751 - 2014)*. Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL).

Boers, N., Marwan, N., Barbosa, H. M. J., and Kurths, J. (2017). A deforestation-induced tipping point for the south American monsoon system. *Sci. Rep.* 7:41489. doi: 10.1038/srep41489

Brienen, R. J. W., Phillips, O. L., Feldpausch, T. R., Gloor, E., Baker, T. R., Lloyd, J., et al. (2015). Long-term decline of the amazon carbon sink. *Nature* 519, 344–348. doi: 10.1038/nature14283

Brown, P. T., Stolpe, M. B., and Caldeira, K. (2018). Assumptions for emergent constraints. *Nature* 563, E1–E3. doi: 10.1038/s41586-018-0638-5

Cox, P. M., Betts, R. A., Collins, M., Harris, P. P., Huntingford, C., and Jones, C. D. (2004). Amazonian forest dieback under climate-carbon cycle projections for the 21st century. *Theoret. Appl. Climatol.* 78, 137–156. doi: 10.1007/s00704-004-0049-4

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., and Totterdell, I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* 408, 184–187. doi: 10.1038/35041539

Cox, P. M., Huntingford, C., and Williamson, M. S. (2018). Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature* 553, 319–322. doi: 10.1038/nature25450

Drijfhout, S., Bathiany, S., Beaulieu, C., Brovkin, V., Claussen, M., Huntingford, C., et al. (2015). Catalogue of abrupt shifts in intergovernmental panel on

climate change climate models. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5777–E5786. doi: 10.1073/pnas.1511451112

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., et al. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* 9, 1937–1958. doi: 10.5194/gmd-9-1937-2016

Feldpausch, T. R., Phillips, O. L., Brienen, R. J. W., Gloor, E., Lloyd, J., Lopez-Gonzalez, G., et al. (2016). Amazon forest response to repeated droughts. *Glob. Biogeochem. Cycles* 30, 964–982. doi: 10.1002/2015GB005133

Fredriksen, H.-B., and Rypdal, M. (2017). Long-range persistence in global surface temperatures explained by linear multibox energy balance models. *J. Clim.* 30, 7157–7168. doi: 10.1175/JCLI-D-16-0877.1

Gillett, N. P., Arora, V. K., Matthews, D., and Allen, M. R. (2013). Constraining the ratio of global warming to cumulative CO2 emissions using CMIP5 simulations*. *J. Clim.* 26, 6844–6858. doi: 10.1175/JCLI-D-12-00476.1

Goodwin, P., Williams, R. G., and Ridgwell, A. (2015). Sensitivity of climate to cumulative carbon emissions due to compensation of ocean heat and carbon uptake. *Nat. Geosci.* 8, 29–34. doi: 10.1038/ngeo2304

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., et al. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.* 31:L03205. doi: 10.1029/2003GL018747

Gregory, J. M., Jones, C. D., Cadule, P., and Friedlingstein, P. (2009). Quantifying carbon cycle feedbacks. *J. Clim.* 22, 5232–5250. doi: 10.1175/2009JCLI2949.1

Hirota, M., Holmgren, M., Van Nes, E. H., and Scheffer, M. (2011). Global resilience of tropical forest and savanna to critical transitions. *Science* 334, 232–235. doi: 10.1126/science.1210657

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2018). Historical (1750-2014) anthropogenic emissions of reactive gases and aerosols from the community emissions data system (CEDS). *Geosci. Model Dev.* 11, 369–408. doi: 10.5194/gmd-11-369-2018

Huppmann, D., Kriegler, E., Krey, V., Riahi, K., Rogelj, J., Calvin, K., et al. (2018). *IAMC 1.5°C Scenario Explorer and Data Hosted by IIASA*. Geneva: Intergovernmental Panel on Climate Change.

Jiménez-de-la Cuesta, D., and Mauritsen, T. (2019). Emergent constraints on earth's transient and equilibrium response to doubled Co2 from post-1970s global warming. *Nat. Geosci.* 12, 902–905. doi: 10.1038/s41561-019-0463-y

Jones, C. D., and Friedlingstein, P. (2020). Quantifying process-level uncertainty contributions to TCRE and carbon budgets for meeting paris agreement climate targets. *Environ. Res. Lett.* 15:074019. doi: 10.1088/1748-9326/ab858a

Leach, N. J., Jenkins, S., Nicholls, Z., Smith, C. J., Lynch, J., Cain, M., et al. (2020). Fairv2.0.0: a generalised impulse-response model for climate uncertainty and future scenario exploration. *Geosci. Model Dev. Discuss.* 2020, 1–48. doi: 10.5194/gmd-14-3007-2021

Lovejoy, T. E., and Nobre, C. (2018). Amazon tipping point. *Sci. Adv.* 4:eaat2340. doi: 10.1126/sciadv.aat2340

Lovejoy, T. E., and Nobre, C. (2019). Amazon tipping point: last chance for action. *Sci. Adv.* 5:eaba2949. doi: 10.1126/sciadv.aba2949

MacDougall, A. H. (2016). The transient response to cumulative Co2 emissions: a review. *Curr. Clim. Change Rep.* 2, 39–47. doi: 10.1007/s40641-015-0030-6

MacDougall, A. H., and Friedlingstein, P. (2015). The origin and limits of the near proportionality between climate warming and cumulative CO2 emissions. *J. Clim.* 28, 4217–4230. doi: 10.1175/JCLI-D-14-00036.1

MacDougall, A. H., Frölicher, T. L., Jones, C. D., Rogelj, J., Matthews, H. D., Zickfeld, K., et al. (2020). Is there warming in the pipeline? A multi-model analysis of the zero emissions commitment from Co2. *Biogeosciences* 17, 2987–3016. doi: 10.5194/bg-17-2987-2020

MacDougall, A. H., Swart, N. C., and Knutti, R. (2017). The uncertainty in the transient climate response to cumulative Co2 emissions arising from the uncertainty in physical climate parameters. *J. Clim.* 30, 813–827. doi: 10.1175/JCLI-D-16-0205.1

Matthews, H. D., Gillett, N. P., Stott, P. A., and Zickfeld, K. (2009). The proportionality of global warming to cumulative carbon emissions. *Nature* 459, 829–832. doi: 10.1038/nature08047

Matthews, H. D., Landry, J.-S., Partanen, A.-I., Allen, M., Eby, M., Forster, P. M., et al. (2017). Estimating carbon budgets for ambitious climate targets. *Curr. Clim. Change Rep.* 3, 69–77. doi: 10.1007/s40641-017-0055-0

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C. B., Frieler, K., Knutti, R., et al. (2009). Greenhouse-gas emission targets for limiting global warming to 2 °c. *Nature* 458, 1158–1162. doi: 10.1038/nature08017

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., et al. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* 109:213. doi: 10.1007/s10584-011-0156-z

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set. *J. Geophys. Res. Atmos.* 117:D08101. doi: 10.1029/2011JD017187

Nijsse, F. J. M. M., Cox, P. M., and Williamson, M. S. (2020). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth Syst. Dyn.* 11, 737–750. doi: 10.5194/esd-11-737-2020

Po-Chedley, S., Proistosescu, C., Armour, K. C., and Santer, B. D. (2018). Climate constraint reflects forced signal. *Nature* 563, E6–E9. doi: 10.1038/s41586-018-0640-y

Rogelj, J., Forster, P. M., Kriegler, E., Smith, C. J., and Séférian, R. (2019). Estimating and tracking the remaining carbon budget for stringent climate targets. *Nature* 571, 335–342. doi: 10.1038/s41586-019-1368-z

Rogelj, J., Schaeffer, M., Friedlingstein, P., Gillett, N. P., van Vuuren, D. P., Riahi, K., et al. (2016). Differences between carbon budget estimates unravelled. *Nat. Clim. Change* 6:245. doi: 10.1038/nclimate2868

Rypdal, M., Fredriksen, H.-B., Rypdal, K., and Steene, R. J. (2018). Emergent constraints on climate sensitivity. *Nature* 563, E4–E5. doi: 10.1038/s41586-018-0639-4

Sanderson, B. (2020). Relating climate sensitivity indices to projection uncertainty. *Earth System Dyn.* 11, 721–735. doi: 10.5194/esd-11-721-2020

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., et al. (2020). The global methane budget 2000-2017. *Earth Syst. Sci. Data* 12, 1561–1623. doi: 10.5194/essd-12-1561-2020

Sherwood, S., Webb,M. J., Annan, J. D., Armour, K. C., Forster, P.M., Hargreaves, J. C., et al. (2020). An assessment of earth's climate sensitivity using multiple lines of evidence. *Rev. Geophys.* 58:e2019RG000678. doi: 10.1029/2019RG000678

Smith, C. J., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., et al. (2018). Fair v1.3: a simple emissions-based impulse response and carbon cycle model. *Geosci. Model Dev.* 11, 2273–2297. doi: 10.5194/gmd-11-2273-2018

Stocker, T., Qin, D., Plattner, G.-K., Alexander, L., Allen, S., Bindoff, N., et al. (2013). *Technical Summary*. Cambridge, UK; New York, NY: Cambridge University Press.

Valdes, P. (2011). Built for stability. *Nat. Geosci.* 4, 414–416. doi: 10.1038/ngeo1200

Williams, R. G., Roussenov, V., Goodwin, P., Resplandy, L., and Bopp, L. (2017). Sensitivity of global warming to carbon emissions: effects of heat and carbon uptake in a suite of earth system models. *J. Clim.* 30, 9343–9363. doi: 10.1175/JCLI-D-16-0468.1

Zelinka,M. D.,Myers, T. A.,McCoy, D. T., Po-Chedley, S., Caldwell, P.M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.* 47:e2019GL085782. doi: 10.1029/2019GL085782

Zickfeld, K., Eby, M., Matthews, H. D., and Weaver, A. J. (2009). Setting cumulative emissions targets to reduce the risk of dangerous climate change. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16129–16134. doi: 10.1073/pnas.0805800106

**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 Climate

ORIGINAL RESEARCH
published: 27 September 2021
doi: 10.3389/fclim.2021.736759

# Emerging Skill in Multi-Year Prediction of the Indian Ocean Dipole

*F. Feba*[1]*, *Karumuri Ashok*[1]*, *Matthew Collins*[2] and *Satish R. Shetye*[1]

[1] Centre for Earth, Ocean and Atmospheric Sciences, University of Hyderabad, Hyderabad, India, [2] College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, United Kingdom

The Indian Ocean Dipole is a leading phenomenon of climate variability in the tropics, which affects the global climate. However, the best lead prediction skill for the Indian Ocean Dipole, until recently, has been limited to ∼6 months before the occurrence of the event. Here, we show that multi-year prediction has made considerable advancement such that, for the first time, two general circulation models have significant prediction skills for the Indian Ocean Dipole for at least 2 years after initialization. This skill is present despite ENSO having a lead prediction skill of only 1 year. Our analysis of observed/reanalyzed ocean datasets shows that the source of this multi-year predictability lies in sub-surface signals that propagate from the Southern Ocean into the Indian Ocean. Prediction skill for a prominent climate driver like the Indian Ocean Dipole has wide-ranging benefits for climate science and society.

**Keywords: Indian Ocean (Dipole), decadal prediction, CanCM4, MIROC5, Southern Ocean, IOD and Southern Ocean, Antarctic Circumpolar Current (ACC), decadal prediction in tropics**

OPEN ACCESS

***Edited by:***
*Swadhin Kumar Behera, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan*

***Reviewed by:***
*Tomoki Tozuka, The University of Tokyo, Japan*
*Benjamin Ng, CSIRO Oceans and Atmosphere, Australia*

***Correspondence:**
*F. Feba*
*feba.francis@outlook.com*
*Karumuri Ashok*
*ashokkarumuri@uohyd.ac.in*

***Specialty section:***
*This article was submitted to Predictions and Projections, a section of the journal Frontiers in Climate*

***Received:*** *05 July 2021*
***Accepted:*** *31 August 2021*
***Published:*** *27 September 2021*

***Citation:***
*Feba F, Ashok K, Collins M and Shetye SR (2021) Emerging Skill in Multi-Year Prediction of the Indian Ocean Dipole. Front. Clim. 3:736759. doi: 10.3389/fclim.2021.736759*

## INTRODUCTION

The Indian Ocean Dipole (IOD) is an inter-annual coupled ocean–atmosphere phenomenon in the tropical Indian Ocean that peaks during the boreal fall season (September to November; SON). Positive IODs are associated with anomalously warmer western Indian Ocean and anomalously cooler eastern Indian Ocean. Negative IOD events are associated with opposite anomalous sea surface temperatures (SSTs) across the Indian Ocean (Saji et al., 1999; Webster et al., 1999; Murtugudde et al., 2000). The IOD is a well-known driver of global climate (Saji and Yamagata, 2003). In addition to affecting the neighboring Maritime Continent to the east and East Africa to the west (Behera et al., 2005), the positive IOD events, for example, have been associated with reduced rainfall over western and southern Australia (Ashok et al., 2003; Ashok and Saji, 2007; Cai et al., 2011), enhanced seasonal Indian summer monsoon (ISM) rainfall (Ashok et al., 2001; Ashok and Saji, 2007), and climates of even distant regions of South America (Chan et al., 2008) and Europe (Hardiman et al., 2020). In addition to its own coupled dynamics (Gualdi et al., 2003; Yamagata et al., 2004; Behera et al., 2006; Ha et al., 2017; Tanizaki et al., 2017; Saji, 2018; Marathe et al., 2021), the IOD is suggested to be triggered by other inter-annual processes such as the El Niño–Southern Oscillation (ENSO), the dominant climate driver from the tropics.

Attempts to predict the IOD on a seasonal scale have been on-going for about two decades (Iizuka et al., 2000; Shinoda et al., 2004; Luo et al., 2007; Doi et al., 2016) and have met with relatively shorter lead time skills, the highest being of 4 months. This is in contrast to the 12–17 months of lead prediction skill for the ENSO as shown in recent studies (Barnston et al., 2012; Park et al., 2018; Tang et al., 2018). Several papers claim that prediction skills for the IOD and ENSO are linked and intertwined (Wajsowicz, 2005; Izumo et al., 2010; Luo et al., 2010).

The emerging discipline of decadal prediction, i.e., prediction of climate information for the near future, shows great promise for societal needs and economic policymakers. Decadal prediction lies between weather prediction and climate projection. Therefore, decadal prediction has to resolve initial value problems like in weather prediction and seasonal to inter-annual forecasts and climate variability and trends effective from boundary conditions, including slow varying oceanic processes, snow cover, and anticipated changes in anthropogenic greenhouse gases and aerosols (Meehl et al., 2009, 2014; Smith et al., 2013; Boer et al., 2016). Interestingly, the inter-annual ENSO and the IOD can also be modulated by decadal processes (Ashok et al., 2004; Tozuka et al., 2007). Such processes should enhance the decadal prediction skill of these inter-annual events.

Given the limited lead skill of the IOD on seasonal scales, its prediction at multi-year lead times is exceptionally challenging. Remarkably, in this study, we show that two model hindcasts from the CMIP5 (Coupled Model Intercomparison Project—Phase 5) decadal prediction datasets are found to show successful prediction skill for IOD events. These sets give us an opportunity to track the lead prediction skill of an event at a lead of up to 2 years.

# DATA AND METHODS

## Data

We have mainly used a sub-set of the decadal hindcast outputs from the CMIP5 decadal runs (Meehl et al., 2009, 2014; Taylor et al., 2012; Smith et al., 2013; Boer et al., 2016). These runs have anthropogenic and natural forcing in them, and comprise several independent decade-long ensemble-runs by each model. For example, the four members of the first ensemble are initialized in 1960 and run up to 1970 (**Figure 1**). Such runs are available with initial conditions from 1960 to 2011. Together, we have 51 ensemble members of decadal hindcasts for each model, each differing from the other in the initialization year (from 1960 to 2011). For each model, we average multiple ensembles with same initialization dates. These averages are referred to as the model hindcasts for the initial conditions of that particular year. The models, whose hindcasts we consider, are the CanCM4 (Merryfield et al., 2013), MIROC5 (Watanabe et al., 2010), BCC-CSM (Wu et al., 2013), and the GFDL (Delworth et al., 2006; Zhang et al., 2007) models. The output data from all the models are divided into 10 groups depending on the year after initialization (**Figure 1**). Yr1, yr2, yr3… indicate 1, 2, 3 … year(s) after initialization.

We used Hadley Center Sea Ice and Sea Surface Temperature (HadISST; Rayner et al., 2003) and Ocean Reanalysis data from the European Center for Medium-Range Weather Forecasts (ORAS4; Balmaseda et al., 2013) for ocean temperatures. As for ENSO indices, we calculate the NINO3 and NINO4 indices as anomalies of SSTs in area averaged over the boxes 150–90°W, 5°S−5°N and 160°E−150°W, 5°S−5°N, respectively. We calculate the Indian Ocean Dipole Mode Index (IODMI) as the SST gradient between the western equatorial Indian Ocean (50–70°E and 10°S−10°N) and the south-eastern equatorial Indian Ocean (90–110°E and 10°S−0°). We take into consideration the IODMI only during the months of September to November as the IOD peaks during this season.

## Statistical Methods

We have carried out correlation and regression analysis. We have used the two-tailed Student's *t*-test to determine the significance of the correlation coefficients. We ascertained the robustness of the skills by applying a boot-strapping test for 1,000 simulations (using NCL) and found our results to be significant at 90% for up to 4 years. For verifying the skill of persistence for both the models and observations, we use the auto-correlation (**Supplementary Figure 1**). To illustrate briefly, say in the case of CanCM4, we correlate the predicted IODMI at lead 1 with that at lead 2 to get the persistence skill at lead 1. Similarly, the persistence skill at lead 2 is obtained by correlating the predicted IODMI at lead 0 with that at lead 2, and so on. We have used the Empirical Orthogonal Function (EOF) analysis technique (von Storch and Zwiers, 1999) to determine the dominant patterns of spatio-temporal variability in the equatorial Indian Ocean, as an additional analysis to compare the simulated IODs with the observations. The EOF analysis is a multivariate statistical technique to calculate the eigenvalues and eigenvectors of a spatially weighted anomaly covariance matrix. The corresponding eigenvalues quantify the variance percent explained by each pattern, which is, by definition, orthogonal to the other patterns.

To have an estimate of the inter-annual variability of IOD that could be attributed to the Southern Ocean (20°W:50°E; 60°S:40°S), we projected the IODMI on to the corresponding area-averaged subsurface temperatures (vertically averaged over 300–800 m depth), i.e., through a regression analysis, and measured the goodness-of-fit (figure not shown).

# RESULTS

## Lead Prediction Skills for ENSO and IOD

ENSO events peak during the boreal winter (through December), while the IOD events peak during boreal fall (September–November). Our analysis shows that the hindcasts analyzed in this study (**Figure 2A**) predict the peak NINO3 index (Trenberth, 1997) significantly at leads up to 12–13 months, in agreement with earlier studies (Sun et al., 2018; Pal et al., 2020).

Surprisingly, the hindcasts from the MIROC5 and CanCM4 predict the IODMI during the fall season with significant lead prediction skill for at least 2–3 years (**Figure 2B**). The skill levels for most of the lead times are not only statistically significant at the 95% confidence level from a two-tailed Student's *t*-test but also are better than persistence (**Supplementary Figure 1**). While the lead prediction skills of the IODMI from the CanCM4 fall below 95% confidence level at 4–5-year leads, these skill levels are still significant at a 90% confidence level. The MIROC5 also shows a weak skill for the eighth year, still significant at 80% confidence level.

It is noteworthy that this skill is present despite the maximum lead predictability of only 1 year for the NINO3 index in all the models. We also ascertain that the maximum lead time skill for the NINO4 is similar to those for the NINO3 index. The ISM

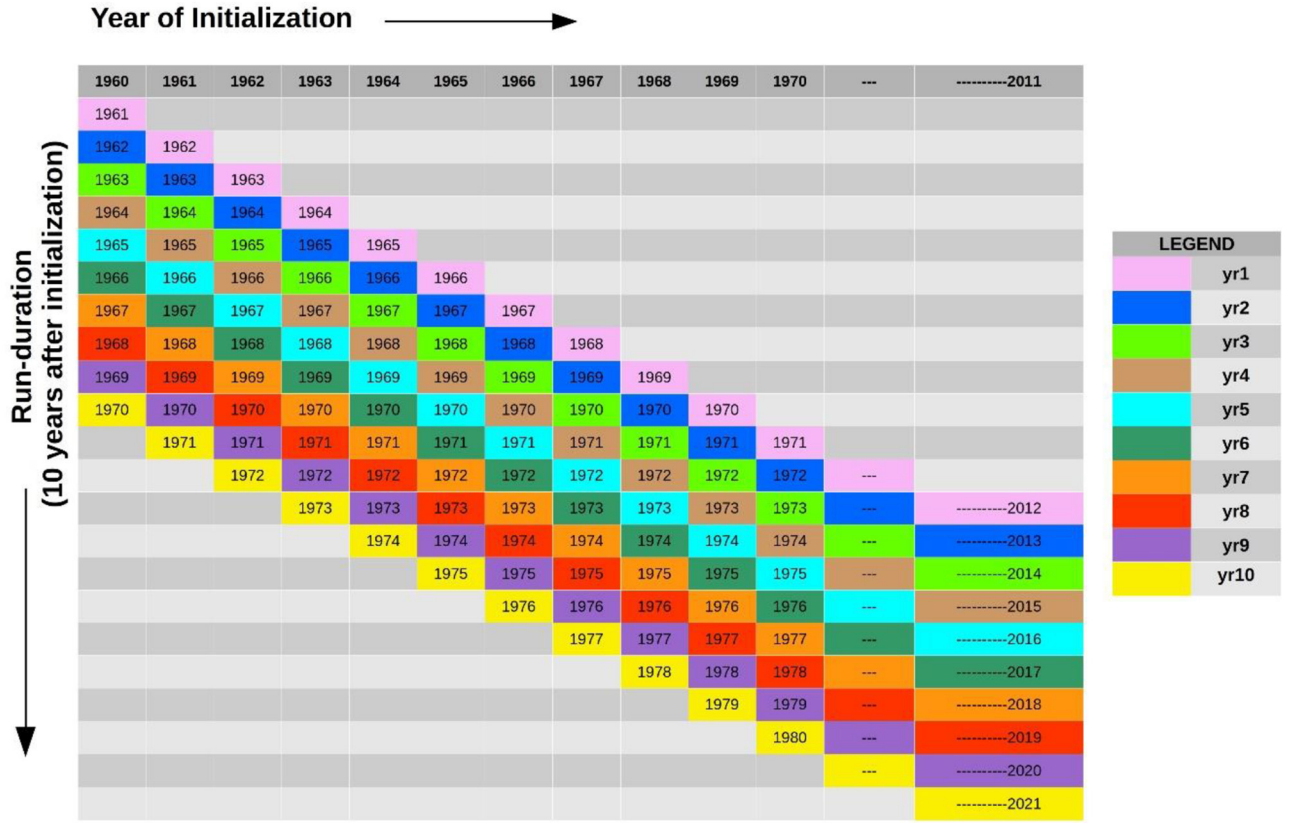



**FIGURE 1 |** The output data from all the models are divided into 10 groups depending on the year after initialization. Yr1, yr2, yr3… indicates 1, 2, 3 … year(s) after initialization, which is represented by individual colors.

rainfall is an ocean–atmosphere coupled phenomenon occurring every year from June to September bringing copious amount of rainfall over the Indian subcontinent. ISM is strongly influenced by ENSO and IOD, along with other factors. Our study shows that ISM has no predictability in the analyzed models, presumably due to model inadequacies in representing monsoon processes (figure not shown). Nevertheless, rainfall predictions could be improved by exploiting known statistical relationships between ENSO-ISM or IOD-ISM (Jourdain et al., 2013; Swapna et al., 2015; Dutta and Maity, 2018).

## Fidelity of the Simulated IOD

To further ascertain whether the IODMI computed from the two hindcast datasets represents some of the observed features of the IOD (Yamagata et al., 2003), we compare the second leading simulated EOF modes of the SST from the hindcasts of MIROC5 and CanCM4 with that from the corresponding observations. **Figure 3**, derived from the EOF analysis at 1-year lead, shows that the models indeed capture the observed dominant IOD variance pattern associated with the EOF2 in terms of the location of the two centers of action. The simulated variance explained by this statistical mode from the MIROC5 is 12.2%, comparable with the corresponding observed value of 12% (Saji et al., 1999; Ashok et al., 2004). EOF2 of CanCM4 is 27.4%, which is higher than the expected value as the dipole-like structure shows more spread in the eastern Indian Ocean (**Figure 3**). The simulated EOF1, associated with Indian Ocean Basin mode, and the corresponding variance explained are also realistic.

## Source of the IOD Prediction Skill

A question arises as to what processes are responsible for the IOD predictability. Earlier studies have shown that the key component of improved decadal prediction lies in ocean physical processes, which internally generate decadal climate variability (Meehl et al., 2014; Farneti, 2017). Thus, the memory of the upper ocean (∼800 m) provides an improved predictability of SST variability in models on decadal timescales (Yeager et al., 2012; Wang et al., 2014). Motivated by these, we looked for a source of predictability of IOD at several levels in the subsurface temperature (every 50 m). We found the maximum signal in the Southern Ocean at 300 to 800 m depth 7–10 years before the occurrence of the IOD event. From an analysis of the ORAS4 and SODA3 reanalysis datasets, we find a signal for the IODMI in the Southern Ocean. This is evidenced by significant correlations between depth-averaged 300–800 m temperatures in the Southern Ocean, which lead the IODMI by 10–6 years, respectively (**Figure 4**). The significant positive correlations in the Southern Ocean indicate that a positive temperature anomaly in the Southern Ocean leads

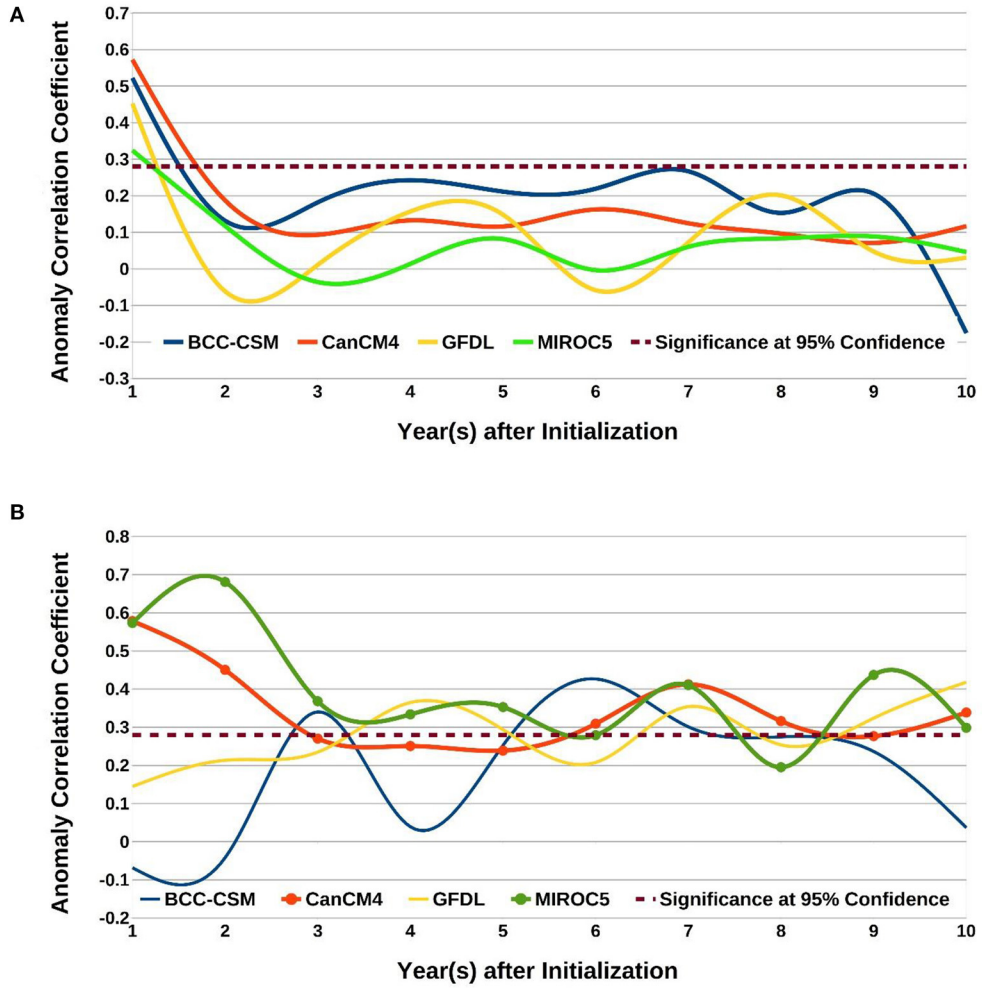



**FIGURE 2 | (A)** Anomaly correlations between observed NINO3 index with the corresponding predicted NINO3 index at different lead years for various models, shown in different colors. The significant correlation value at 95% confidence level is 0.27 and is shown by dashed lines. Details of how the correlations at different leads were obtained, from the sets of ∼51 year-long decadal prediction runs with different initial years, are available from section Statistical Methods and **Figure 1**. **(B)** Same as **(A)** but for the IODMI. The lines with circles in **(B)** are the two models of interest for this study.

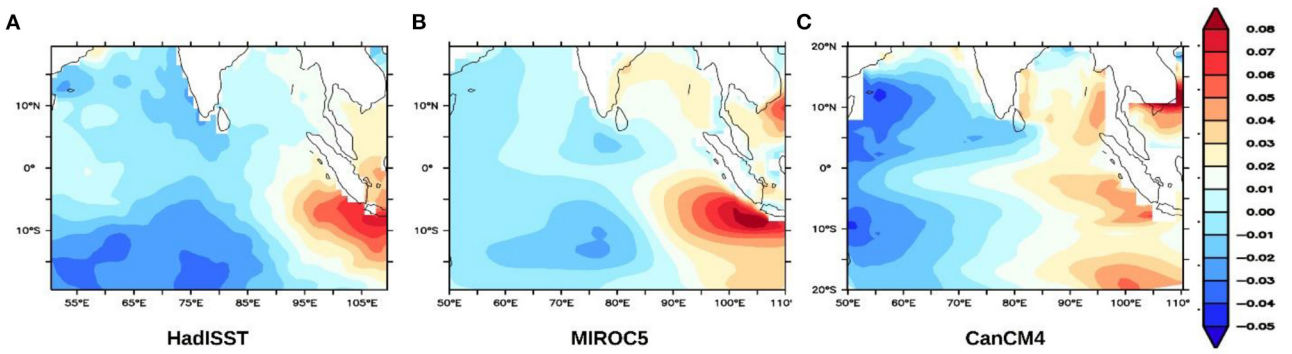



**FIGURE 3 |** The second mode from the EOF analysis of the tropical Indian Ocean SST, from **(A)** HadISST (1961–2011 period). **(B,C)** are the same as that of **(A)** but derived from MIROC5 and CanCM4 hindcasts for the same period (also see **Supplementary Figure 2**).

to a positive IOD event after 8–10 years, whereas a similar lead negative correlation coefficient in the Southern Ocean indicates a negative IOD event after such a time period. The signal leading the IODMI first appears in the sub-surface temperatures of the Southern Ocean just south-west of Africa between the depths of 300–800 m, 10 years before the occurrence of the IOD event. This signal is seen to propagate toward the east along the Antarctic Circumpolar Current (ACC, or West wind drift) to about 50–60°E about 6 years prior to the IOD event. From here, the signal in the heat content further propagates north, along the east of Madagascar, all the way to the Somali coast in a pathway that is coincident with the East Madagascar undercurrent (EMUC)—an intermediate undercurrent at the depth of thermocline and below. The period of the propagation is in agreement with the EMUC as described by Nauw et al. (2008). The EMUC transports about 2.8 ± 1.4 Sv of intermediate water equatorward. The signal upwells to the surface near Somalia (Somali upwelling) along with its propagation north (**Figure 5**).

The propagation of the signal from the deeper layers in the extra-tropics to the surface layers in the equatorial region can be explained by the Indian Ocean Meridional Overturning Circulation (IOMOC; Wang et al., 2014). The time period of the propagation is in agreement with the IOMOC. From the Somali coast, the lead heat content signal is seen to move east, reaching the central equatorial Indian Ocean 3 years before the IOD event. The signal remains in the equatorial Indian Ocean for the last 2–3 years before the occurrence of the IOD event (**Figure 6**). Furthermore, our correlation in **Figure 6** suggests an anomalous SST structure of opposite polarity in the equatorial Indian Ocean about a year before the occurrence of an IOD event. This is in conformation with Saji et al. (1999), who not only note a tight coupling between the intensity of the SST dipole anomaly and zonal wind anomaly but emphasize on a biennial tendency. The biennial tendency of the IOD is also noted by Meehl and Arblaster (2001), Ashok et al. (2003), etc. It is intriguing why the signal from the sub-surface at the equator does not manifest as an IOD event immediately. There are several studies that suggest potential drivers that can force an IOD such as the Indian summer monsoon, ENSO, and can account for the lead signals associated with the IOD with 1–2-year lead (Ashok et al., 2001; Ashok and Saji, 2007). As the current models do not show any lead skills of the Indian summer monsoon beyond a season, we can rule out that the long lead skills for the IODMI in the models come from the lead skills of monsoon or those related to ENSO (see **Figure 2A**). This might be explained by the fact that some IOD events are triggered by an atmospheric signal substantial enough to trigger the coupled evolution of the IOD (Shinoda et al., 2004; Saji, 2018) and hence this delay of 2–3 years.

A linear regression analysis carried out using HadISST and ORAS4 ocean temperature datasets (slope value = 0.46) indicates that the Southern Ocean signals at decadal lead explain about 18% of the inter-annual variability of IOD as suggested by a goodness-of-fit measure. To put it in perspective, ENSO, known as the most prominent driver of the Indian summer monsoon interannual variability, explains about 30% of the latter.

The source of decadal predictability for IOD in the MIROC5 and CanCM4 also apparently comes from the Southern Ocean (**Supplementary Figure 3**). We find statistically

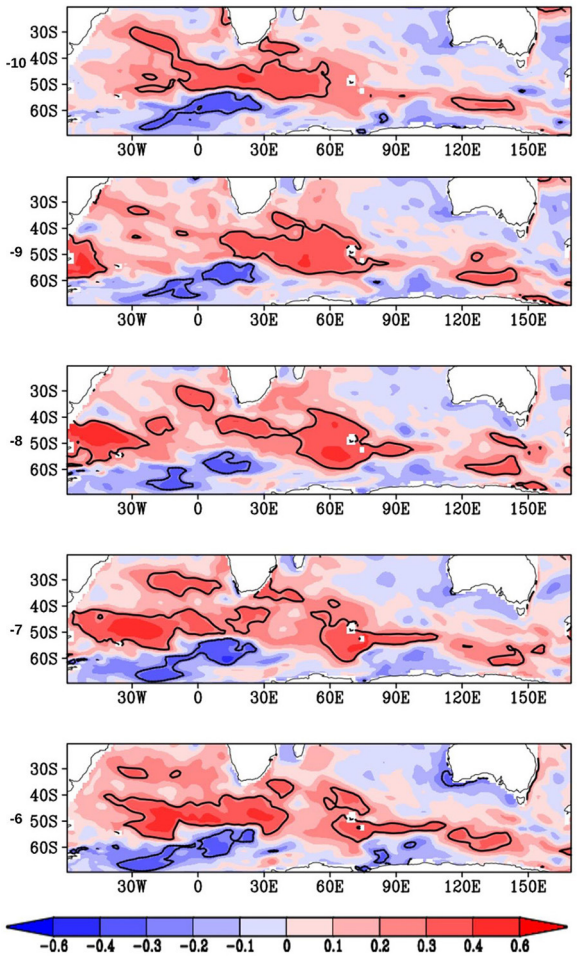



**FIGURE 4 |** Spatial distribution of the anomaly correlation of HadISST derived IODMI with vertically averaged subsurface temperatures from ORAS4 (over 300–800 m depth) from previous years averaged annually. The "lag number" is the years over which the IOD lags the Southern Ocean signal. In all panels, significant correlations at 95% confidence are shown as contour lines (value at 0.27).

significant anomaly correlations between the depth averaged from surface until ~800 m hindcast heat content with the IODMI with the heat content leading at 10–5 years, conforming to our results from the reanalysis. The source of signal is the Southern Ocean for both the models, but the simulated signal path from here to the equatorial Indian Ocean are, however, weaker and remain unclear specifically in the MIROC5 model. However, the signals re-emerge finally 2–3 years before the occurrence of the simulated IOD. The models also do not capture the biennial tendency of the IOD as in observations. Further understanding of other possible factors affecting IOD variability on decadal scales is needed.

## DISCUSSION

Our analysis of the CMIP5 decadal retrospective prediction products for the 1960–2011 period shows that two CMIP5 decadal prediction systems exhibit multi-year skill in predicting

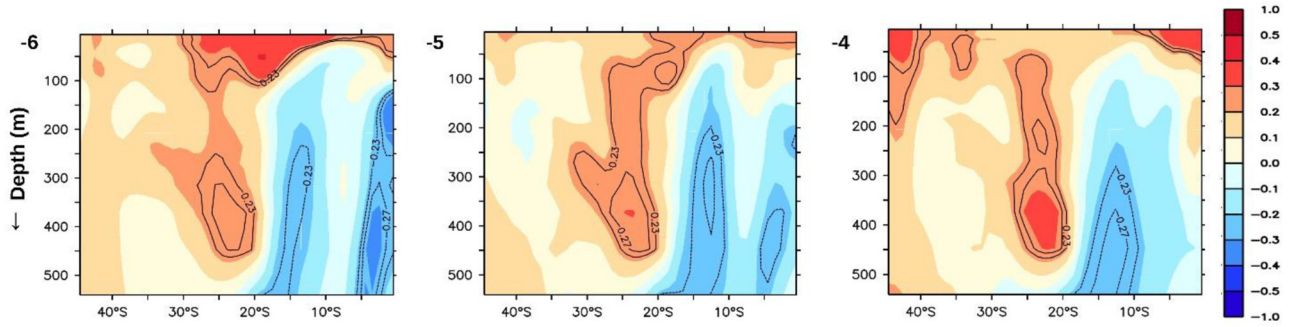



**FIGURE 5 |** Vertical distribution of anomaly correlation of IODMI with previous years' subsurface temperatures, averaged annually over longitudes 50–100°E. Significant correlations at 90% confidence (0.23) and at 95% confidence (0.27) are contoured.

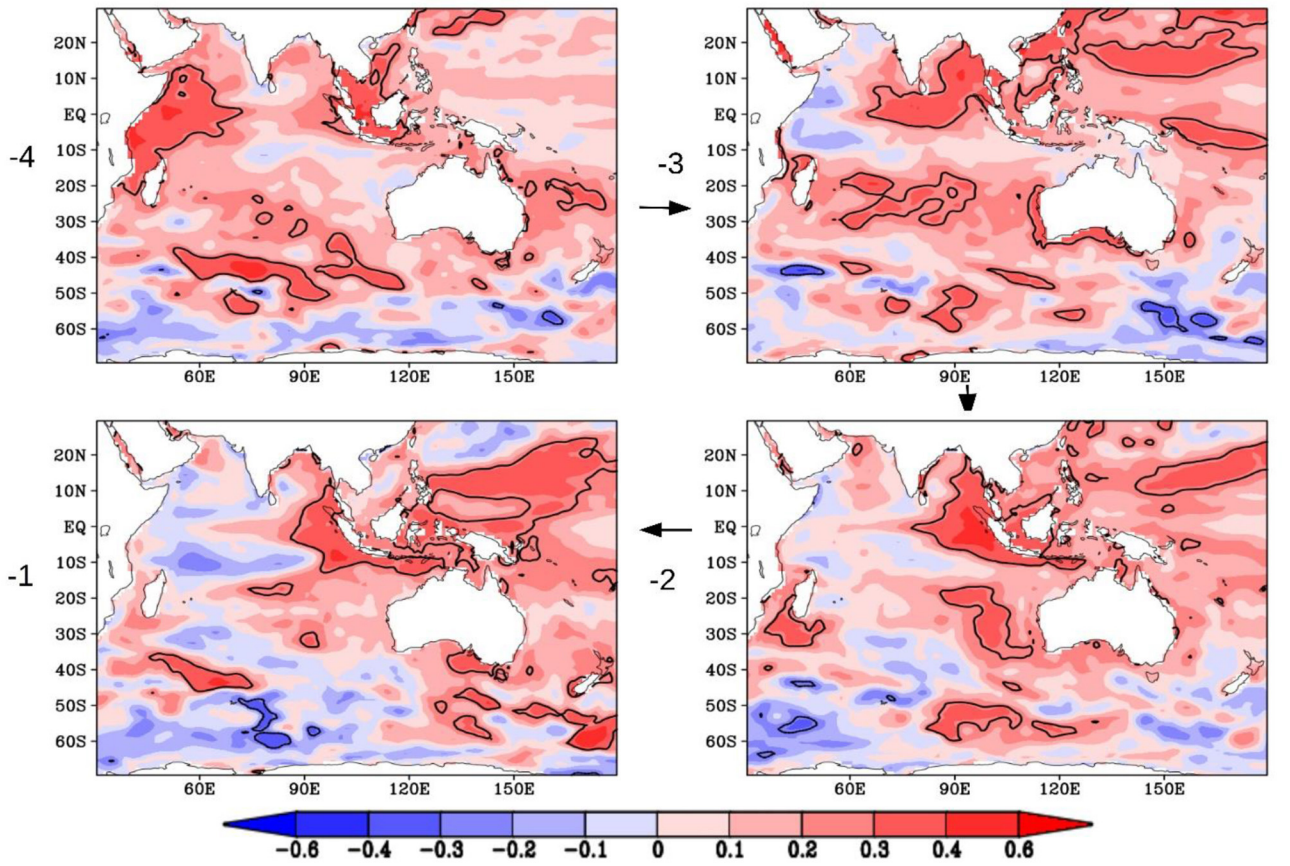



**FIGURE 6 |** Same as **Figure 4** but vertically averaged annually over surface to 100 m depth for the last 4 years leading up to an IOD event. Significant correlations at 95% confidence are shown as contour lines (value at 0.27).

the IODMI, higher than the 17-month lead skills for the ENSO, a leading climate driver. The physical processes responsible for the skill levels seems to originate in the Southern Ocean up to a decade before emerging in the Indian Ocean, suggesting that this multi-year skill could be extended even further. To our knowledge, this is the first study to show decadal skill in predicting the IOD.

Skill levels for surface fields and, in particular, rainfall, are not statistically indistinguishable from zero. This is in common with long-term predictions of other climate phenomena (Scaife and Smith, 2018) and indicates the relative immaturity of this form of forecasting. It remains a challenge to both improve models and to fully exploit decadal prediction skill. Nevertheless, the potential for long-term planning in both public and private sector

organizations that could be facilitated by such skillful predictions is great.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

FF, KA, and MC formulated the hypothesis and co-wrote the manuscript. FF conducted the analyses. All authors reviewed the final version of the manuscript.

## FUNDING

FF thanks MC and NERC/MoES SAPRISE project (NE/I022841/1) for the travel grant to visit MetOffice, UK. FF acknowledges CSIR for the JRF & SRF fellowship. FF also thanks C T Tejavath for assistance in data processing.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fclim.2021.736759/full#supplementary-material

## REFERENCES

Ashok, K., Chan, W.-L., Motoi, T., and Yamagata, T. (2004). Decadal variability of the Indian Ocean dipole. *Geophys. Res. Lett.* 31:L24207. doi: 10.1029/2004GL021345

Ashok, K., Guan, Z., and Yamagata, T. (2003). Influence of the Indian Ocean Dipole on the Australian winter rainfall. *Geophys. Res. Lett.* 30:1821. doi: 10.1029/2003GL017926

Ashok, K., Guan, Z. Y., and Yamagata, T. (2001). Impact of the Indian Ocean Dipole on the relationship between the Indian monsoon rainfall and ENSO. *Geophys. Res. Lett.* 28, 4499–4450. doi: 10.1029/2001GL013294

Ashok, K., and Saji, N. H. (2007). On the impacts of ENSO and Indian Ocean dipole events on sub-regional Indian summer monsoon rainfall. *Nat. Hazards.* 42, 273–285. doi: 10.1007/s11069-006-9091-0

Balmaseda, M. A., Mogensen, K., and Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *QJR Meteorol. Soc.* 139, 1132–1161. doi: 10.1002/qj.2063

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Shuhua, L., and DeWitt, D. G. (2012). Skill of real-time seasonal ENSO model predictions during 2002-11: is our capability increasing? *Bull. Amer. Meteor. Soc.* 93, 631–651. doi: 10.1175/BAMS-D-11-00111.1

Behera, S., Luo, J., and Masson, S. (2005). Paramount impact of the Indian Ocean dipole on the East African short rains: a CGCM study. *J. Clim.* 18, 4514–4530. doi: 10.1175/JCLI3541.1

Behera, S. K., Luo, J. J., Masson, S., Rao, S. A., Sakuma, H., and Yamagata, T. (2006). A CGCM study on the interaction between IOD and ENSO. *J. Clim.* 19, 1608–1705. doi: 10.1175/JCLI3797.1

Boer, G. J., Smith, D., Cassou, C., Doblas-Reyes, F., Danbasoglu, G., Kirtman, B., et al. (2016). The decadal climate prediction project (DCPP) contribution to CMIP6. *Geosci. Model. Dev.* 9, 3751–3777. doi: 10.5194/gmd-9-3751-2016

Cai, W. J., van Rensch, P., Cowan, T., and Hendon, H. H. (2011). Teleconnection pathways of ENSO and the IOD and the mechanisms for impacts on Australian rainfall. *J. Climate.* 24, 3910–3923. doi: 10.1175/2011JCLI4129.1

Chan, S., Behera, S., and Yamagata, T. (2008). Indian Ocean dipole influence on South American rainfall. *Geophys. Res. Lett.* 35:L14SL12 doi: 10.1029/2008GL034204

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., et al. (2006). GFDL's CM2 global coupled climate models. Part I: formulation and simulation characteristics. *J. Climate.* 19, 643–74. doi: 10.1175/JCLI3629.1

Doi, T., Behera, S. K., and Yamagata, T. (2016). Improved seasonal prediction using the SINTEX-F2 coupled model. *J. Adv. Model. Earth Syst*, 8, 1847–1867. doi: 10.1002/2016MS000744

Dutta, R., and Maity, R. (2018). Temporal evolution of hydroclimatic teleconnection and long-lead prediction of Indian summer monsoon rainfall using time varying model. *Sci. Rep.* 8:10778. doi: 10.1038/s41598-018-28972-z

Farneti, R. (2017). Modelling interdecadal climate variability and the role of the ocean. *Wiley Interdiscip. Rev. Climate Change.* 8:e441. doi: 10.1002/wcc.441

Gualdi, S., Navarra, A., Guilyardi, E., and Delecluse, P. (2003). Assessment of the tropical Indo-Pacific climate in the SINTEX CGCM. *Ann. Geophys.* 46, 1–26 doi: 10.4401/ag-3385

Ha, K.-J., Chu, J.-E., Lee, J.-Y., and Yun, K.-S. (2017). Interbasin coupling between the tropical Indian and Pacific Ocean on inter annual timescale: observation and CMIP5 reproduction. *Clim. Dyn.* 48, 459–475. doi: 10.1007/s00382-016-3087-6

Hardiman, S. S., Dunstone, N. J., Scaife, A., Smith, D. M., Knight, J. R., Davies, P., et al. (2020). Predictability of European winter 2019/20: Indian Ocean dipole impacts on the NAO. *Atmos. Sci. Lett.* 21:e1005. doi: 10.1002/asl.1005

Iizuka, S., Matsuura, T., and Yamagata, T. (2000). The Indian Ocean SST dipole simulated in a coupled general circulation model. *Geophys. Res. Lett.* 27, 3369–3372. doi: 10.1029/2000GL011484

Izumo, T., Vialard, J., Lengaigne, M., de Boyer Montégut, C., Behera, S., Luo, J.J., et al. (2010). Influence of the state of the Indian Ocean dipole on following year's El Niño. *Nat. Geosci.* 3, 168–172. doi: 10.1038/ngeo760

Jourdain, N. C., Sen Gupta, A., Taschetto, A. S., Ummenhofer, C. C., Moise, A. F., and Ashok, K. (2013). The Indo-Australian monsoon and its relationship to ENSO and IOD in reanalysis data and the CMIP3/CMIP5 simulations. *Clim. Dynam.* 41, 3073–3102. doi: 10.1007/s00382-013-1676-1

Luo, J.-J., Masson, S., Behera, S., and Yamagata, T. (2007). Experimental forecasts of the Indian Ocean dipole using a coupled OAGCM. *J. Clim.* 20, 2178–2190, doi: 10.1175/JCLI4132.1

Luo, J.-J., Zhang, R., Behera, S. K., Masumoto, Y., Jin, F.-F., Lukas, R., et al. (2010). Interaction between El Niño and extreme Indian Ocean dipole. *J. Clim.* 23, 726–742. doi: 10.1175/2009JCLI3104.1

Marathe, S., Terray, P., and Karumuri, A. (2021). Tropical Indian Ocean and ENSO relationships in a changed climate. *Clim. Dynam.* 56, 3255–3276. doi: 10.1007/s00382-021-05641-y

Meehl, G., and Arblaster, J. (2001). The tropospheric biennial oscillation and Indian monsoon rainfall. *Geophys Res Lett.* 28, 1731–1734. doi: 10.1029/2000GL012283

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., et al. (2014). Decadal climate prediction an update from the trenches. *Bull. Am. Meteorol. Soc.* 95, 243–267. doi: 10.1175/BAMS-D-12-00241.1

Meehl, G. A., Goddard, L., Murphy, J., Boer, G., Danabasoglu, G., Dixon, K. W., et al. (2009). Can It Be Skillful? *Bull. Am. Meteorol. Soc.* 90, 1467–1486. doi: 10.1175/2009BAMS2778.1

Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., et al. (2013). The Canadian Seasonal to interannual prediction system. Part I: models and initialization. *Mon. Wea. Rev.* 141, 2910–2945. doi: 10.1175/MWR-D-12-00216.1

Murtugudde, R., McCreary, J. P., and Busalacchi, A. J. (2000). Oceanic processes associated with anomalous events in the Indian Ocean with relevance to 1997-98. *J. Geophys. Res.* 105, 3295–3306. doi: 10.1029/1999JC900294

Nauw, J. J., van Aken, H. M., Webb, A., Lutjeharms, J. R. E., and de Ruijter, W. P. M. (2008). Observations of the southern East Madagascar Current

and undercurrent and countercurrent system. *J. Geophys. Res.* 113:C08006. doi: 10.1029/2007JC004639

Pal, M., Maity, R., Ratnam, J. V., Nonaka, M., and Behera, S. K. (2020). Long-lead prediction of ENSO modoki index using machine learning algorithms. *Sci. Rep.* 10:365. doi: 10.1038/s41598-019-57183-3

Park, J. H., Kug, J. S., Li, T., and Behera, S. K. (2018). Predicting El Niño beyond 1-year lead: effect of western hemisphere warm pool. *Sci. Rep.* 8:14957. doi: 10.1038/s41598-018-33191-7

Rayner, N. A., Parker, D. E., Horton, E.B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmosp.* 108:4407. doi: 10.1029/2002JD002670

Saji, N. H. (2018). "The Indian Ocean Dipole," in *Oxford Research Encyclopedia of Climate Science*, ed H. von Storch (Oxford: Oxford University Press).

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., and Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature* 401, 360–363. doi: 10.1038/43854

Saji, N. H., and Yamagata, T. (2003). Possible impacts 209 of Indian Ocean dipole mode events on global climate. *Clim. Res.* 25:151–169. doi: 10.3354/cr025151

Scaife, A. A., and Smith, D. (2018). A signal-to-noise paradox in climate science. *NPJ Clim. Atmos. Sci.* 1:28. doi: 10.1038/s41612-018-0038-4

Shinoda, H., Hendon, H., and Alexander, M. A. (2004). Surface and subsurface dipole variability in the Indian Ocean and its relation with ENSO. *Deep-Sea Res.* 51, 619–635. doi: 10.1016/j.dsr.2004.01.005

Smith, D. M., Scaife, A. A., Boer, G. J., Caian, M., Guemas, V., Hawkins, E., et al. (2013). Real-time multi-model decadal climate predictions. *Clim. Dyn.* 41, 2875–2888. doi: 10.1007/s00382-012-1600-0

Sun, Q., Bo, W. U., Zhou, T. J., and Yan, Z. X. (2018). ENSO hindcast skill of the IAP-DecPreS near term climate prediction system: comparison of full field and anomaly initialization. *Atmos. Oceanic Sci. Lett.* 11, 54–62, doi: 10.1080/16742834.2018.1411753

Swapna, P., Roxy, M. K., Aparna, K., Kulkarni, K., Prajeesh, A. G., Ashok, K., et al. (2015). The IITM earth system model: transformation of a seasonal prediction model to a long term climate model. *Bull. Am. Meteorol. Soc.* 96, 1351–1368. doi: 10.1175/BAMS-D-13-00276.1

Tang, Y., Rong-Hua, Z., Ting, L., Wansuo, D., Dejian, Y., Fei, Z., et al. (2018). Progress in ENSO prediction and predictability study. *Natl. Sci. Rev.* 5, 826–839. doi: 10.1093/nsr/nwy105

Tanizaki, C., Tozuka, T., Doi, T., and Yamagata, T. (2017). Relative importance of the processes contributing to the development of SST anomalies in the eastern pole of the Indian Ocean Dipole and its implication for predictability. *Clim. Dyn.* 49, 1289–1304. doi: 10.1007/s00382-016-3382-2

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.* 93, 485–498. doi: 10.1175/BAMS-D-11-00094.1

Tozuka, T., Luo, J. J., Masson, S., and Yamagata, T. (2007). Decadal modulations of the Indian Ocean Dipole in the SINTEX-F1coupled GCM. *J. Clim.* 20, 2881–2894. doi: 10.1175/JCLI4168.1

Trenberth, K. E. (1997). The definition of El Niño. *Bull. Am. Met. Soc.* 78, 2771–2778. doi: 10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2

von Storch, H., and Zwiers, F. W. (1999). *Statistical Analysis in Climate Research* (Cambridge: Cambridge University Press), 484.

Wajsowicz, R. C. (2005). Potential predictability of tropical Indian Ocean SST anomalies. *Geophys. Res. Lett.* 32:L24702. doi: 10.1029/2005GL024169

Wang, W., Zhu, X., Wang, C., and Köhl, A. (2014). Deep meridional overturning circulation in the Indian Ocean and its relation to Indian Ocean Dipole. *J. Clim.* 27, 4508–4520. doi: 10.1175/JCLI-D-13-00472.1

Watanabe, M., Suzuki, T., O'ishi, R, Komuro, Y., Watanabe, S., Emori, S., et al. (2010). Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. *J. Clim.* 23, 6312–6335. doi: 10.1175/2010JCLI3679.1

Webster, P. J., Moore, A., Loschnigg, J., and Leban, M. (1999). Coupled dynamics in the Indian Ocean during 1997-1998. *Nature* 401, 356–360. doi: 10.1038/43848

Wu, T., Li, W., Ji, J., Xin, X., Li, L., Wang, Z., et al. (2013). Global carbon budgets simulated by the Beijing climate center climate system model for the last century. *J. Geophys. Res. Atmos.* 118, 4326–4347. doi: 10.1002/jgrd.50320

Yamagata, T., Behera, S., Rao, S., Guan, Z. Y., Ashok, K., and Hameed, S. (2003). Comments on Dipoles, temperature gradients, and tropical climate anomalies. *Bull. Am. Meteorol. Soc.* 84, 1418–1422. doi: 10.1175/BAMS-84-10-1418

Yamagata, T., Behera, S. K., Luo, J.-J., Masson, S., Jury, M., and Rao, S. A. (2004). "Coupled ocean-atmosphere variability in the tropical Indian Ocean," in: *Earth Climate: The Ocean-Atmosphere Interaction*, eds. C. Wang, S.-P. Xie, and J. A. Carton (Washington, DC: AGU), 189–212.

Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., and Teng, H. (2012). A decadal prediction case study: late twentieth century North Atlantic Ocean heat content. *J. Clim.* 25, 5173–5189. doi: 10.1175/JCLI-D-11-00595.1

Zhang, S., Harrison, M. J., Rosati, A., and Wittenberg, A. T. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Rev.* 135, 3541–3564. doi: 10.1175/MWR3466.1

**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.

# 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:** frontiersin.org/about/contact

**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