Check for updates

The Gulf of Mexico in trouble: Big data solutions to climate change science

Vishwamithra Sunkara[1]*, Jason McKenna[1], Soumyashree Kar[1], Iliyan Iliev[2] and Diana N. Bernstein[3]

[1]Ocean Enterprise, University of Southern Mississippi, Gulfport, MS, United States, [2]School of Social Science and Global Studies, University of Southern Mississippi, Hattiesburg, MS, United States, [3]School of Ocean Science and Engineering, University of Southern Mississippi, Hattiesburg, MS, United States

The latest technological advancements in the development and production of sensors have led to their increased usage in marine science, thus expanding data volume and rates within the field. The extensive data collection efforts to monitor and maintain the health of marine environments supports the efforts in data driven learning, which can help policy makers in making effective decisions. Machine learning techniques show a lot of promise for improving the quality and scope of marine research by detecting implicit patterns and hidden trends, especially in big datasets that are difficult to analyze with traditional methods. Machine learning is extensively used on marine science data collected in various regions, but it has not been applied in a significant way to data generated in the Gulf of Mexico (GOM). Machine learning methods using ocean science data are showing encouraging results and thus are drawing interest from data science researchers and marine scientists to further the research. The purpose of this paper is to review the existing approaches in studying GOM data, the state of the art in machine learning techniques as applied to the GOM, and propose solutions to GOM data problems. We review several issues faced by marine environments in GOM in addition to climate change and its effects. We also present machine learning techniques and methods used elsewhere to address similar problems and propose applications to problems in the GOM. We find that Harmful Algal Blooms (HABs), hypoxia, and sea-level rises have not received as much attention as other climate change problems and within the machine learning literature, the impacts on estuaries and coastal systems, as well as oyster mortality (also major problems for the GOM) have been understudied – we identify those as important areas for improvement. We anticipate this manuscript will act as a baseline for data science researchers and marine scientists to solve problems in the GOM collaboratively and/or independently.

1 Introduction

The Gulf of Mexico was formed by plate tectonics around 300 million years ago. Shallow continental shelf waters cover nearly half of the basin. The Gulf of Mexico's most significant biological and geological provinces are the coastal zone, continental shelf, continental slope, and abyssal plain. Tidal marshes, sandy beaches, mangrove-covered areas, and several bays, estuaries, and lagoons make up the coastal zone. Buried salt domes occur at varying depths on the shelf and on the slope that dips downward to the abyssal plain that are connected with economically valuable oil and natural gas resources. The Gulf of Mexico receives water from more than 150 rivers and runoff from 31 of the 50 states, making it a focus point for water quality research and improvement. The GOM has many natural resources and is home to many species of fish. Fishing is an important commercial activity, with red snapper, amberjack, tilefish, swordfish, and various grouper, as well as shrimp and crabs, being the most common catches. Many of the bays and sounds are also used to harvest oysters on a massive scale. Shipping, petrochemical processing and storage, military use, paper manufacturing, and tourism are all other key sectors along the Gulf coast (Worrall and Snelson, 1989).

The Gulf of Mexico has mixed semi-diurnal and diurnal tidal cycles due to the unusual form of its basin. The warm water temperature in the Gulf of Mexico can support major Atlantic hurricanes that bring widespread human death and destruction, as Hurricane Katrina did in 2005. The GOM approximately comprises of 30% continental shelf which is shallower than 200 m and the shallow waters in the northern GOM are warm temperate. Hurricanes do intensify over shallow waters despite the low tropical heat potential Potter et al. (2019). Although the water temperature drops when a hurricane passes through, it quickly recovers and becomes capable of sustaining another tropical cyclone (NASAEarthObservatory, 2005). Due to massive fresh water discharge from the Mississippi river and other rivers into the Gulf, the content of nitrogen and phosphates inflow into the Gulf has increased (Tian et al., 2020), and as a consequence led to hypoxic zones in the Gulf (Diaz and Rosenberg, 2008). In addition, the GOM is one of the World's major offshore petroleum-producing zones, accounting for one-sixth of all US output. The GOM has suffered from catastrophic oil spills like Deepwater Horizon and Ixotic I oil spills, and accounts for thousands of minor oil spills every year. The oil spills, agricultural run-offs, red-tide algae blooms, and hypoxic zones are some environmental threats in addition to the climate change impacting the GOM.

Oceans being the largest ecosystem on Earth are heavily impacted by climate change. On a local to global scale, oceans significantly affect the weather, and changes in climate can adversely impact many of the ocean's qualities. The primary indicators of climate change on oceans are the sea surface temperatures, sea-level rise, coastal flooding, and ocean acidity. The GOM has shown these primary indicators and hence is vulnerable to climate change (Epstein, 2005). Rising surface ocean temperatures, freshwater and nutrient imports, and CO_2 levels in the atmosphere will further aggravate these problems. Increase in ocean acidity and

freshwater run-off also exacerbate the hypoxic zones and algae blooms. The Deepwater Horizon oil spill was a catastrophicevent which led to expanding our knowledge of the GOM and restoration efforts. However, there is still much to be learned and we should not wait for any further disasters to happen.

Today, we are flooded with data originating from several sources in a wide range of application areas. It is now relevant to say that we are in an era of "big data (Data, 2008)". With the abundance of data, the decisions that were previously relying on mere speculation or simulated models can now be based on facts. Big data is revolutionizing scientific research and several sectors are benefiting by using historic and real-time data for better analysis. However, oceanography has yet to properly embrace the era of big data. Ocean data comes from models, satellites, buoys, research vessels, sensor networks, unmanned maritime vehicles, and others, but the majority of it comprises of remote sensing and model data. Most of the ocean data is hosted by regional and global ocean observation platforms like OceanCube (Stanic et al., 2020), the U.S. Integrated Ocean Observing System Program (IOOS), The Gulf of Mexico Coastal Ocean Observing System (GCOOS), etc. Several researchers, modelers, and industries use ocean observation platforms to get the data and make better use of it. However, our capacity for data analysis has not kept pace, and the rising gap is becoming a major bottleneck for making efficient use of the data available, as well as a roadblock to expanding data collecting even further.

Machine learning which itself is a subfield of artificial intelligence, is broadly used in various fields, such as epidemiology (Appice et al., 2020), physics (Carleo et al., 2019), ecology (Huang et al., 2018), data science (Huang et al., 2016), agriculture (Liakos et al., 2018), and language processing (Iliev et al., 2022), but is newly adapted by marine science. With the amplitude of ocean science data available, machine learning has great potential to answer several problems and expand marine science research. Machine learning techniques are used in addressing climate change, climate analysis, ecological environments, natural disasters, and global patterns Rasouli et al. (2012); Kim et al. (2014); Deo and Sahin (2015); Mosavi et al. (2018); Xiao et al. (2019); Mansfield et al. (2020); Rosso et al. (2020). However, climate change being a global problem and oceans covering more than 70% of land on Earth, machine learning is yet to be fully adapted by marine science to address any local and global problems using available ocean data. While machine learning is a hot topic engaged by numerous researchers from different fields it is not being utilized to its fullest with data available for the GOM region to address any climate change and regional problems.

This paper aims to provide an overview of critical problems in the GOM that need attention of machine learning experts to utilize the available ocean science data originating in the GOM. We discuss innovative engineering efforts and ocean observing platforms in the GOM that provide historical and real time data from several sources to marine researchers and data scientists to use the data to better understand the GOM. We highlight topics including but not limited to climate change and its impacts, CO_2 emissions, ocean acidification, algal blooms, oyster reefs, severe weather events, ports/coastal resilience, etc.

2 Climate change and other concerns for the Gulf of Mexico

The Gulf of Mexico is a vast ecological system with environmental and economic salience. It supports diverse ecosystems that provide significant benefits like wildlife habitat, erosion prevention, stable shorelines, hurricane buffer, fish nurseries, improved water quality, and tourism. The major sectors of the GOM's economy directly and indirectly rely on the coastal ecosystems. However, global warming and climate change are having a drastic impact on these ecosystems posing a serious threat and hence needs more attention from diverse researchers to understand, analyze, and mitigate any future impacts.

Many of the fundamental elements that regulate algal development, such as water temperature, nutrients, light, and grazers, are controlled by climate, and hence can be predicted to affect changes in the species composition, trophic structure, and function of aquatic ecosystems. Global temperatures are rising, a fact now widely acknowledged to be linked to human activity (Lee, 2007). Over the next century, average sea surface temperatures are anticipated to rise by up to 5°C, resulting in ice melt and altered precipitation in many maritime locations (Scavia et al., 2002; Levina et al., 2007; Doney, 2010; Hoegh-Guldberg and Bruno, 2010; Doney et al., 2012). The regional minimal and maximum temperatures have already increased by 0.8°C and 0.4°C since 1950 (Levina et al., 2007), with estimated warming rate of the GOM sea surface temperature of 0.19°C per decade (Wang et al., 2023). In the future climate, coupled model simulations show weakening of the Loop Current in the GOM during the 21st century (Liu et al., 2012). The study by 0 Misra et al. (2019) suggests that in the projected climate, the warming of the GOM will be enhanced as a result of the increase in heat flux from the atmosphere to the ocean and the increase in anomalous advective heat flux convergence due to the slowing of the heat transport across the Loop Current. Seasonal warming of ocean surfaces intensified by global warming and rising carbon dioxide levels in the atmosphere likely increase the water stratification and worldwide deoxygenation, aggravating red tides, hypoxia, aquatic species migration, and illnesses (Keeling et al., 2010; Cheng et al., 2017). At the same time, anthropogenic CO_2 intake is predicted to lower ocean pH and carbonate saturation levels (Sarmiento et al., 1998; Feely et al., 2009). This confluence of stressors could have serious consequences for ocean ecosystems and fisheries, particularly in coastal areas where eutrophication has already resulted in hypoxic and acidic conditions (Boyd and Doney, 2003; Bopp et al., 2013; Melzner et al., 2013; Altieri and Gedan, 2015; Breitburg et al., 2015; Flynn et al., 2015; Levin et al., 2015; Laurent et al., 2018).

Climate change is poised to exacerbate impacts of coastal eutrophication in the northern Gulf of Mexico (Laurent et al., 2018). Coastal eutrophication is often accompanied by hypoxia (Howarth, 1995; Nixon et al., 1995; Vitousek et al., 1997; Caraco and Cole, 1999; Bennett et al., 2001). The increase in frequency of hypoxia and anoxia formations in shallow, coastal and estuarine areas is highly likely a result of human activities (Diaz et al., 1995). One of the key stresses affecting estuarine and coastal ecosystems is nutrient over-enrichment from anthropogenic sources (Bricker, 1999; Howarth et al., 2000; Cloern, 2001). In many parts of the world, there is growing worry that an abundance of nutrients from various sources is having widespread biological repercussions on shallow coastal waters. Reduced light penetration, increased abundance of nuisance macroalgae, loss of aquatic habitat like seagrass or macroalgal beds, noxious and toxic algal blooms, hypoxia and anoxia, disruptions in trophic interactions and food webs, and impacts on living resources are just a few of the consequences (Vitousek et al., 1997; Schramm, 1999; Anderson et al., 2002; Rabalais, 2002; Rabalais et al., 2002). Excessive nitrogen inputs from rivers drive algae development and subsequent decomposition in bottom waters in many coastal locations. In combination with stratified waters, the ensuing oxygen consumption and dissolved inorganic carbon generation frequently result in hypoxic ($[O_2] < 62.5$ mmol O_2 m^3) and acidified situations (Rabouille et al., 2008; Cai et al., 2011). Hypoxic conditions caused by eutrophication are a common and growing problem that will most likely be aggravated by global warming (Diaz and Rosenberg, 2008; Rabalais et al., 2010).

Anthropogenic nutrient overabundance, combined with rising temperatures and a growing frequency of extreme hydrologic events (storms and droughts), is hastening eutrophication and encouraging the spread of harmful algal blooms (HABs) across the freshwater-to-marine continuum (Paerl et al., 2018). Water resources, fisheries, recreational usage, tourism, and property values are all threatened by cyanobacterial blooms. Climate change makes predicting changes in HABs frequency, intensity, and proliferation much more difficult (Paerl and Huisman, 2009; Havens and Paerl, 2015; Wells et al., 2015; Paerl et al., 2018). Changes in ocean and lake circulation, stratification, and upwelling, wind speed, and cyclone frequency and intensity, as well as global warming, altered precipitation patterns, and sea level rise, all play prominent roles in modulating HABs dynamics (Paerl and Paul, 2012). When combined with excessive fertilizer loading, hydrologic changes and climate change allow HABs to grow in size and remain longer (Paerl et al., 2016; Paerl, 2017). (Hu et al., 2011) suspects the rise of chl-a content after oil spills and provided results stating that the northeastern Gulf of Mexico became greener after the Deepwater Horizon oil spill. The GOM experiencing 1000s of oil spills every year may also be contributing to growth of chl-a. Within days after Hurricane Isabel in 2003, a large phytoplankton bloom developed in Chesapeake Bay, linked to increased nutrient loads (Miller et al., 2005). Algal blooms and extensive hypoxia/anoxia occurred shortly after several hurricanes affected the lagoon of the Neuse River Estuary in the 1990s (Burkholder and Glibert, 2006; Burkholder et al., 2006). A bloom of the picocyanobacterium *Synechococcus* in eastern Florida Bay, lasting more than 18 months, followed an input of nutrients from the highfreshwater discharge caused by Hurricanes Katrina, Rita, and Wilma in 2005 (Glibert et al., 2009).

Human actions have altered the atmosphere and ocean environment in ways that affect storms and extreme climate occurrences. Changes in extremes, which are outside the bounds of prior weather, are the most common way that climate change is perceived. Trenberth (2012) and Solomon et al. (2007) state severe

weather events are also a result of climate change. Intensity of hurricanes and storm occurrences increased in the past decade. Increasing hurricanes and storm frequency can increase the loss of wetlands and coastal erosion. According to published assessments of tide gauge data, the average worldwide sea-level rise rate over the 20th century was 1.7 mm/yr (Bindoff et al., 2007). Climate scientists project a 0.25 to 0.5 m global sea level rise by 2100, which is more than double the pace of rise during the 20th century under some carbon emission scenarios (Meehl et al., 2007). Sea-level rise projections and impacts in coming years, global and regional sea-level rise scenarios, and the climate change impacts are very well documented in the reports (Reidmiller et al., 2017; Fox-Kemper et al., 2021; Sweet et al., 2022) generated by the National Oceanic and Atmospheric Administration (NOAA) and the Intergovernmental Panel on Climate Change. Assessments of coastal vulnerability indexes can reveal the relative risk of coastal change as a result of future sea-level rise. Barrier islands are dynamic habitats that experience both gradual and quick change as a result of waves, tides, and currents, as well as rapid change as a result of catastrophic storms. Due to increased sea-level rise and changes in the frequency and intensity of storm occurrences, these islands are likely to alter dramatically during the next century. Because of the dynamic nature of barrier islands and the importance of these areas, natural resource managers must understand how habitats on barrier islands are changing or may change over time in order to identify when and where management activities are necessary (Enwright et al., 2021). The findings of nearly three decades of marine geological research in the northwest Gulf of Mexico have been compiled in an effort to better understand the factors (e.g., sea-level rise, sediment supply, subsidence, and antecedent topography) that influenced coastal evolution during the last eustatic cycle (120 ka to present). The geological record shows that the northwestern Gulf of Mexico's low-gradient coastal ecosystems will continue to be adversely impacted by continuous sea-level rise, decreasing sediment supply, and human interference (Anderson et al., 2014).

The aquatic and wild life are also suffering the impacts of climate change. The GOM has around 22 percent of the world's nonarctic tidal marsh, accounting for about 62 percent of North America's tidal marsh habitat. These tidal marshes, which cover around 9,880 square kilometers of Gulf coast (Greenberg and Maldonado, 2006), feature both riverine and marine-dominated habitats with varied tidal effect, freshwater and sediment input, and plant species diversity. The biology and ecology of tidal marsh birds along the GOM are little understood, especially in comparison to Atlantic coast populations (Rush et al., 2009, 2010a; 2010b), and we have limited information on current population status and trends for most marsh birds (Conway, 2011). Despite the potential fragility of tidal marshes along the northern Gulf Coast, little is known about the life histories, behavior, distribution, and ecological affinities of the many marsh bird species. Predicting the future of tidal habitats and indigenous marsh bird species in the context of global climate change is difficult given current information gaps. Regional environmental changes that affect marine ecosystems will be brought on by a warming climate. These changes will include variations in ocean circulations, rising ocean acidity, changes in

riverine discharge, variations in precipitation and evaporation rates, and loss of coastal habitat due to flooding (Scavia et al., 2002; Roessig et al., 2004). Because of physiological and behavioral reactions to environmental gradients, fish and fisheries are likely to be impacted by changes to species ranges, loss or degradation of nearshore fish habitat, modification of larval dispersal pathways, and loss of nearshore fish habitat (Scavia et al., 2002; Perry et al., 2005; Hare et al., 2010). The Atlantic bluefin tuna (BFT) is a highly migratory species that feeds in frigid North Atlantic waters before migrating to tropical areas to reproduce. Projected future warmer climate will cause the upper ocean temperatures in the western Atlantic spawning habitat in GOM to rise dramatically, which may have an impact on the temporal and spatial range of BFT spawning activities. An ensemble of 20 climate models simulations were studied in (Muhling et al., 2011) to predict the mean temperature variations within GOM, in combination with larval BFT data to understand the effects of warming on the suitability of GOM as a spawning ground. The effects of climate, fishing, and other anthropogenic disturbances on three species (Atlantic Goliath grouper, red grouper, and tilefish) that are economically important to GOM and called habitat engineers are presented in (Coleman and Koenig, 2010). The intricate vertical (benthic-pelagic coupling) and horizontal (inshore to offshore) links between habitats, species, and marine strata are shown by the habitat engineers at various geographical and temporal scales. Oceanic currents connect them physically, and ontogenetic migrations and trophic interactions connect them biologically (Coleman and Koenig, 2010). Rising temperatures and decreasing freeze frequency/intensity may characterize climate change at temperate latitudes, potentially shifting the latitudinal limits of vegetation and animals (Bakkenes et al., 2002; Walther et al., 2002; Loarie et al., 2008). (Comeaux et al., 2012) examined the implications of mangrove expansion into coastal wetlands. Heavy rain and snowmelt in the Midwest prompted catastrophic flooding of the Mississippi River in the spring and summer of 2019, necessitating two openings of the Bonnet Carre Spillway (BCS) to relieve pressure on New Orleans' levees. That allowed a massive amount ´ of freshwater to flow into Lake Pontchartrain and then into the Mississippi Sound. If the frequency or duration of BCS openings increases as a result of increased precipitation, oyster populations in Mississippi may become unsustainable for harvesting unless future freshwater intrusions are factored into management plans (Gledhill et al., 2020). Although oyster reefs around the world are essential for maintaining healthy ecosystems, most of them have disappeared over the past 200 years (Beck et al., 2011; Grabowski et al., 2012). About 69 percent of the US commercial wild eastern oyster harvest is produced by the GOM, which is home to the greatest remaining wild oyster fisheries (Beck et al., 2011; NOAA, 2022).

Changing coastlines and rising sea levels may necessitate the relocation of highways, rail lines, or airport runways in the long run, with significant implications for port facilities and coastal infrastructure. Increased hurricanes and other extreme weather events could have ramifications for emergency evacuation planning, facility maintenance, and safety management for surface transportation, maritime vessels, and aviation. Rain and snowfall patterns, as well as periodic flooding patterns, may have an

impact on safety and maintenance. Storms, sea level rise, sedimentation and erosion rates and mechanisms, and changes in critical variables such as prevailing winds, waves, currents, and precipitation rates must all be addressed in order to gain a better understanding of the possible effects on ports and maritime shipping. Enhancing climate models to replicate storm events, determining the impact of severe events on ports and shipping, and simulating local sea level rise near ports and shipping channels are all research priorities. A summary of the literature cited in this section is presented in Figure 1 showing the proportion of the literature under specific categories.

3 Ocean of data

Data about America's seas and coasts comes from a variety of sources across several sectors and is used for a wide range of purposes. They are utilized to manage ocean ecosystems and fisheries, promote the blue economy's sustainable and equitable growth, conserve endangered marine species, and assist the global community in mitigating and preparing for climate change. Ocean science data, climate, and hydrodynamic models aid in the protection of endangered species by identifying acceptable areas for offshore wind energy production facilities, improving storm prediction models, understanding marine species migration/habitat change, warming waters, and planning for sea level rise and other climate change impacts on coastal populations to name a few. With newtechnology, new uses, and renewed national pledges to understanding and managing our oceans, ocean science data collection and access is undergoing a revolution. We have an unrivaled ability to gather and evaluate data on our environment and human usage of marine natural resources, as well as to provide enormous potential for scientific and decision-making improvement. The ability of government organizations to interpret and incorporate ocean data from new sources, including new technology, into decision-making is currently limited. Specifically, the data management infrastructure has not kept up with the virtually exponential increase in data being collected by the

public and private sectors. In this section, we discuss the ocean science data originating from the Gulf of Mexico, where this data is available, and present some characteristics of ocean science data.

3.1 Ocean data categories

Ocean science data is collected by several entities for various purposes. Academia, government agencies both state and federal, and private industries collect most ocean science data to fulfill their project requirements. Ocean data can be broadly categorized into five types: physical data, biological data, chemical data, geological data, and socioeconomic data (Trice et al., 2021).

3.1.1 Physical ocean data

Physical ocean data represents the physical attributes and dynamic processes of the oceans including ocean currents, coastal dynamics, internal waves and tides, temperature and salinity structure, and more. This data is captured by variables such as sea-surface temperature, sea surface salinity, ocean surface heat flux, wave conditions, surface and subsurface currents, bathymetry, and seabed forms. Physical ocean data is important in helping many industries and government agencies relying on oceans. Sea surface temperature and wave conditions are extensively used in weather prediction models. Physical data are also used to study climate change and also include atmospheric data such as heat and hydrological fluxes. Physical ocean data are also used by ports, research and commercial vessels in navigation as they face challenges due to shifting currents and tides and variable water levels. Physical ocean data collection is made by *in-situ* ocean instruments such as sensors and unmanned marine vehicles, and some datasets can also be collected remotely by satellites.

3.1.2 Biological ocean data

Biological data is more related to marine organisms, their ecology, and how marine organisms interact with the ocean environment. Biological ocean data is used to track and protect endangered species, maintain and boost ecosystem health,
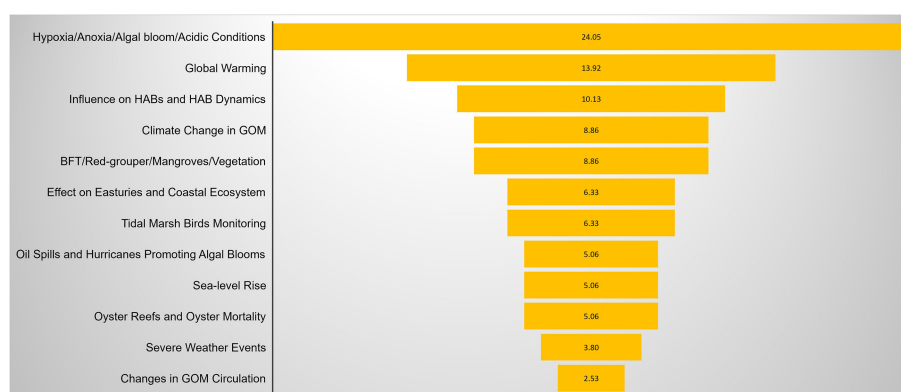

FIGURE 1
Illustration of the major categories of climate change based studies, represented per the proportion of the literature covered under each category in this article.

minimize and mitigate the impacts of human activities on marine wildlife, and study climate change effects. Fisheries data collected by commercial and recreational fishing industries, and fishery-independent data collected by researchers and government-led organizations are also good sources of biological data. Ocean sounds and acoustics data, for example, larval recruitment, biodiversity, marine mammals, and biotic sounds, and underwater videos and photos for example reef mosaicking, fish communities, species identification, etc. also contribute to biological ocean data. Biological data has been collected traditionally by taking the samples in situ, requiring human input. At sea, sample collection, data post-processing, and physical sample archiving are expensive and tedious. New emerging methods like image recognition and environmental DNA (eDNA) can help in reducing post-processing times, maintaining data storage, and reuse of data.

3.1.3 Chemical ocean data

Chemical ocean data represents the properties of ocean water, processes and cycles of ocean waters, and how ocean water interacts with the atmosphere and seafloor. These data help in understanding the ocean's role in climate change as the oceans are a major carbon sink. Ocean acidification is also a central topic investigated using chemical ocean data. Some of the chemical data variables are inorganic carbon, oxygen, pH, nutrients, ocean color, and dissolved organic carbon. Chemical ocean data collection is made using *in-situ* ocean sensors like pH sensors, pCO_2 sensors, etc., and also through remote sensing satellites.

3.1.4 Geological ocean data

Geological ocean data relates to the seafloor and sub-seafloor features. These data give insights into plate tectonics, volcanic processes, and other phenomena. Geological data may overlap with biological, chemical, and physical data to study ocean circulation, sedimentation patterns, and biological productivity, sediment coring, paleoceanography, stratigraphy, estuarine system, beach erosion, and sea level rise.

3.1.5 Socioeconomic ocean data

Socioeconomic ocean data relates to ocean-based industries like shipping, fishing, tourism, and offshore renewable energy. Community impacts on coastal areas, oceans, and industries related to the oceans can also be considered as socioeconomic data. Analyzing and understanding the socioeconomic data can help find the climate change indicators as they affect employment, exports, demographics, and unemployment rates.

3.2 Ocean data platforms

The aforementioned ocean data types can be collected using different ocean technologies and methods *in-situ* and/or by remote sensing. Insitu data collection in the water uses portable ocean sensors, sensors on ships, submersibles, and unmanned marine vehicles. Sensing ocean data remotely without direct contact with the medium is made possible using satellites. Satellites are equipped with sensors that make measurements of the ocean like temperature, wave height, and watercolor. Sampling is also another data collection method where samples of specimens like water, sediments, etc are collected in situ. A mathematical model can also be developed to generate ocean data in a simulated environment. Below we discuss several data platforms that collect different sets of ocean science data. Ocean data platforms themselves are broad topics and discussing them in-depth is out of the scope of this paper.

3.2.1 Unmanned marine vehicles

UMVs are essentially vehicles that can be controlled remotely, either wired or wireless. Operators can control the vehicles either on a ship or onshore. UMVs can also be preprogrammed to go to a specific location and time for data collection. UMVs can be equipped with a variety of sensors for data collection for example temperature, pressure, dissolved oxygen, but also active and passive acoustic sensors. UMVs can store data on board which can be retrieved upon recovering the UMV while some UMVs can also transmit the data wirelessly *via* satellites, wifi, and/or cellular networks. UMVs are powered by batteries and can be recharged *via* solar energy or wind energy, while others can be recharged aboard the ship. Depending on the size, range, and endurance of UMVs, they can be deployed for longer time periods and collect data at desired sampling rates.

Gliders are also a type of preprogrammed unmanned marine vehicle that can essentially go underwater to follow a particular path. Gliders have sensors installed on them and record the data throughout their track until they reach the surface. This data can be used to generate a 3D visualization of the ocean data for example temperature and salinity, but also many more like acoustic measurements. Gliders transmit the recorded data to a satellite, and users can download it. Gliders can take new instructions from operators *via* the satellite link and can also change the data sampling rates.

Remotely operated vehicles (ROVs) are surface and underwater vehicles that are controlled by operators on the surface. ROVs are tethered to a main control board that can stay on the ship or can be connected to any other human-operated vehicle (HOV). ROVs and HOVs have an onboard cameras and manipulator arms that can be controlled by the operator. ROVs and HOVs send the real-time data feed to the mother ship *via* the tether cable. Thus they provide telepresence to the scientific community and public from Exploration Vessel (E/V) Nautilus, NOAA Ship Okeanos Explorer, and Schmidt Ocean Institute's (SOI) Research Vessel (R/V) Falkor during underwater exploration Raineault (2019). ROVs and HOVs are also used to collect specimen samples, and sediments, and capture marine organisms.

3.2.2 Buoys and floats

Buoys are platforms that comprise several instruments and sensors. While static buoys are anchored to the seafloor, some buoys like Argo buoys are free-floating and provide vertical profile of the water column by going up and down and drifting with the

current at a prescribed depth/pressure until they die. Buoys equipped with mission-specific sensors and instruments can be deployed at a desired location. Buoys have central control board that gathers the data from all the sensors on the buoy and transmits either averaged data or the whole collected data to a satellite. Nearshore buoys may use a cellular network for data transmission instead of expensive satellite data transmission rates. Some buoys can also be equipped with profilers that move through the water column vertically carrying instruments/sensors which sample the water column.

3.2.3 Research vessels

Research vessels which are floating labs also act as a data platforms. Vessels can carry instruments and sensors and collect the data or take specimen samples along their route. This data can be processed and analyzed onboard the vessel for any quick analysis.

3.2.4 Satellites and aircrafts

Satellites with the right instruments and sensors can be used to sense ocean data remotely and conduct non-invasive observations of the oceans. Satellites can be used to measure the ocean data variables like sea surface height, watercolor, sea surface temperature, etc. Satellites are also used to collect the data from buoys, and unmanned marine vehicles and transmit that data to land stations. Aircrafts are also used to collect remote observations of the ocean to obtain images or other data related to the ocean landscape or its properties. Earth observing system data and information system (EOSDIS) and National Environmental Satellite Data and Information Service (NESDIDS) maintain environmental data for land, ocean, and atmosphere applications collected by several satellites, aircraft, and field measurements. Remote sensing data is also used to study the spatial and temporal characteristics of earth's landscape which can be used to understand the effects of human activities, population growth, and natural and biological factors (Chen et al 2020a; 2021). Studies like Chen et al. (2022) have developed methods for extracting coastline information using remote sensing images to study the evolution of the coastlines.

3.2.5 Models

A mathematical model can simulate the ocean state and describe the ocean conditions and trends in a parameterized and quantitative way generating ocean data like temperature, salinity, waves, etc. This model data is also extensively used in understanding the oceans. With the rapid improvements in computing power, numerical simulations of the ocean have developed greatly and have become a big source of ocean science data. Some of the models that generate simulated ocean science data are Hybrid Coordinate Ocean Model (HYCOM), Amseas (Former Navy Coastal Ocean Model), Regional Ocean Modeling System (ROMS), Finite Volume Community Ocean Model (FVCOM), M.I.T. General Circulation Model (MITgcm), GFDL Modular Ocean Model (MOM), The Parallel Ocean Program (POP),and NGOFS2 (Northern Gulf of Mexico Operational Forecast System)

for ocean circulation simulations, Simulating WAves Nearshore (SWAN), and WAVEWATCH for wave simulations, ecosystem models (ECOSIM-ECOPATH) to understand dynamics of natural ecosystems, biophysical models such as Daigle et al. (2016) to study larval dispersal, and biogeochemical models such as Fennel ROMS to examine physical and biogeochemical mechanisms for the formation and destruction of seasonal hypoxia on the TX-LA shelf.

3.3 Data sources

Like in any other areas, Gulf of Mexico's ocean science data is collected by several ocean instruments – moored surface and underwater ocean buoys, floating ocean buoys, surface and sub-surface sensor networks, coastal ocean dynamics applications radar (CODAR) stations, remote sensing satellites, unmanned maritime vehicles, research vessels, and numerical ocean models. Regional, state, and national ocean observing platforms host most of the ocean science data. The data and the ocean observing platforms are managed by government, educational, and private organizations. The ocean observing platforms common goal is to maximize the full potential of ocean data. NOAA is responsible for one of the government's largest data inventories, collecting, organizing, and releasing data on everything from the deep ocean to the atmosphere and space. The agency founded in 1807 as the U.S. Coast and Geodetic Survey, now oversees a complex system of data on America's seas, Great Lakes, and coastlines in collaboration with other federal agencies, regional organizations, business, academic partners, and data producers and users. In addition, commercial firms, citizen science, non-governmental organizations, and others are using new, low-cost technology like drones and cellphones to contribute to the ever-growing pool of ocean data. For a number of legal, regulatory, private, or technological reasons, certain data may not be shared with government agencies or the general public, restricting its applicability. However, unrestricted ocean science data is publicly available for independent research purposes. The unrestricted ocean science data can be accessed *via* ERDDAP (https://coastwatch.pfeg.noaa.gov/erddap/index.html) servers and Open-source Project for a Network Data Access Protocol (OPeNDAP) depending on what protocols the data sources support. Accessing and downloading the ocean science data of interest is a tedious task for data scientists and marine researchers as the data of interest for a specific region may be collected by several ocean instruments and organizations but for different purposes. Thus, the similar data from a specific region but recorded by different or same instruments by different organizations may be available in the data platforms of specific regional, state, and federal organizations. Below we present some of the data sources commonly used to download and/or analyze GOM-related ocean science data that data scientists and marine researchers can resort to.

3.3.1 United States geological survey

USGS is one of the significant data contributors providing especially the water, energy, minerals, and other natural resources people rely on, natural hazards that can threaten lives and property,

the health of our ecosystems and environment, and the impacts of climate and land-use change. USGS is nation's largest water, earth, and biological science and civilian mapping agency. In addition to providing quality scientific data, USGS also provides web tools like interactive maps, alerts and notifications, data analysis and visualizations, and data repositories. USGS produces accurate geologic maps, topographic maps, and 3-D geologic frameworks that provide critical data for sustaining and improving the quality of life and economic vitality of the Nation. They also organize, maintain, and publish the geospatial baseline of the Nation's topography, natural landscape, built environment and more. USGS also has multimedia data like videos, images, audio, before-after imagery, webcams, and more. The USGS tracks and researches a wide range of water conditions and resources, including streamflow, groundwater, river discharge, water quality, and water usage and availability. USCG operates and maintains 500 laboratories nationwide, 60 science centers, and 5 volcano observatories. USGC supports research and also supply data to external users. USGS serve multiple types of users, including scientists (USGS, other federal agencies, and academic), regulators and resource managers (federal, state, and local), and private companies.

3.3.2 National centers for environmental information

One of the world's largest archives for atmospheric, coastal, geophysical, and oceanic research is managed by NCEI. NCEI houses a large portion of the data that NOAA scientists, observational systems, and research projects collect. It is in charge of a huge repository of environmental data that spans many different time periods, monitoring systems, scientific fields, and geographical locations. NCEI is holding 44 petabytes of data as of May 2022 and is expecting the data will grow to 250 petabytes by 2030 (NOAA, 2012). Figure 2 shows the current and forecasted NCEI data archival volume. NCEI's data access functionality

provides free access to NCEI's archive of global coastal, oceanographic, geophysical, climate, and historical weather data. These data include quality controlled hourly, sub-hourly, daily, monthly, seasonal, and yearly measurements of NOAA's archived environmental data. Data are available through direct download or sub-setting services. Customers can also order most of these data as certified hard copies for legal use. NCEI provides several data services to monitor, access, archive, and download multi domain ocean science data. NCEI generates a comprehensive annual summary report of the global climate system. This report summarizes the global and regional climate of the preceding calendar year with the input from hundreds of authors giving several directions for further research topics.

3.3.3 National data buoy center

National Data Buoy Center maintains a network of data collecting ocean buoys and coastal stations. NDBC provides meteorologic and oceanographic data collected by buoys and coastal stations. NDBC transmits hourly observations from its network of buoys and coastal stations to a ground facility operated by NOAA. This transmitted data is recorded and processed by the national oceanographic data center. The NDBC website displays all buoy details it is managing on a map along with some useful resources to download the observations data. The NDBC Distributed Oceanographic Data System (DODS) makes netCDF files available to the science community and general public *via* their website. It uses the Open Source Project for a Network Data Access Protocol software for data providers to share data with each other and the end users.

3.3.4 Integrated ocean observing systems

IOOS maintains near-real time observations data collected from 11 regional associations spanning the national coasts. It is a multidisciplinary system that provides data in the forms and at the speeds that decision makers require in order to accomplish

FIGURE 2
NCEI Archival Volume History and Forecast NCEI (2022).
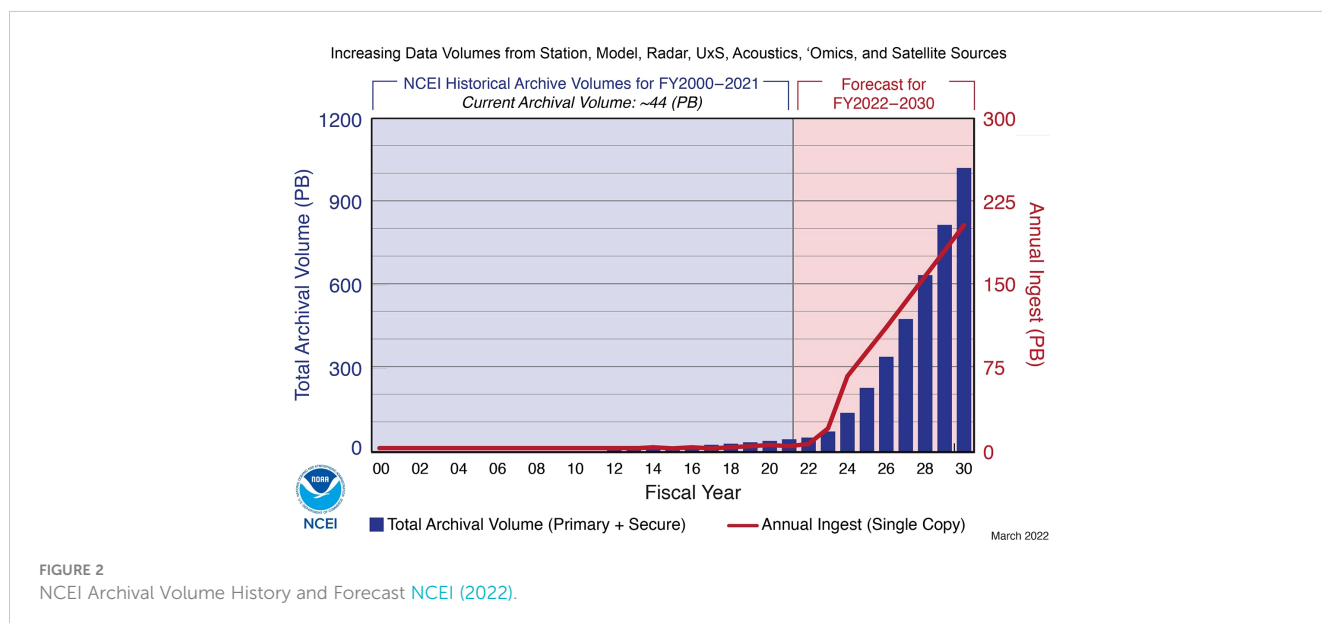
several societal goals. IOOS is a cooperative effort of federal and non-federal entities to provide new data, tools, and forecasts to improve marine safety, enhance the economy, and protect the U.S coastal and ocean environment. Several tools and resources to access and visualize multi-domain ocean science data are developed and readily available on IOOS web interface. IOOS also contributes to and shares the resources to global ocean observing system (GOOS). IOOS manages data assembly centers for gliders, animal telemetry network, and High Frequency (HF) Radar network. IOOS created the Coastal and Ocean Modeling Testbed that allows sharing of numerical ocean models data along with observations data, and utilizes the software tools for integration, scientific analysis, comparison, and data archival.

3.3.5 Gulf of Mexico coastal ocean observing system

GCOOS is at the center of data gathering for the Gulf of Mexico's ocean and coastal waters, collecting thousands of data points from sensors and ensuring that the data is trustworthy, timely, and correct before being disseminated to the ocean sectors that rely on it. GCOOS represents the academic, industrial, government, and non-government sectors, with organizations streaming data, information, and products on marine and estuarine systems to the GCOOS online platform, where thousands of users from ocean modelers to ship captains have instant access. GCOOS integrates physical, biological, meteorological, biogeochemical, bathymetric, and other ocean science data from diverse providers. GCOOS delivers timely and reliable data, products, and services to IOOS, decision-makers, and the general public to benefit human communities, the economy, and natural ecosystems. GCOOS is one of the 11 regional associations of the U.S. IOOS.

3.3.6 Coastal CUBEnet

The Coastal CUBEnet developed by the University of Southern Mississippi is a high-resolution, coastal ocean sensors, modelling, and data sharing network that provides the integrated, multi-dimensional, open infrastructure needed for collaborative ocean research products. The CUBEnet's centralized environmental intelligence resources are critical for sharing high-resolution data, model forecasts, and other related research products with end users and the community. The CUBEnet and its environmental intelligence tools serve as a platform that offers the opportunity to bring in expertise, insights, methods, and tools from multiple disciplines including oceanography, climate science, biology, natural resource management, computer/data science, public policy, and economics. The CUBEnet provides data collected by several ocean instruments like moored buoys, underwater sensor networks, HF Radar stations, unmanned maritime systems, satellites along with high resolution hydrodynamic models integrated for several GOM coastal areas. The web interface of CUBEnet provides tools and resources for visualization, analysis, and download the data. The CUBEnet is currently hosting data for Louisiana, Mississippi, Alabama, and West Florida coastal regions.

The CUBEnet's data is passed on to GCOOS for QA and QC, and is then archived at the Global Telecommunication System.

3.3.7 Earth observing system data and information system

The EOSDIS is the primary component of NASA's earth science data collected from satellites, aircrafts, field measurements, and other programs. EOSDIS provides an open source web application, "worldview," allowing users to browse satellite imagery interactively. Solar irradiance, oceanic, atmospheric, land surface, and subsurface data are only a few examples of the Earth observation data that EOSDIS supports. Worldview uses the Global Imagery Browse Services (GIBS) to rapidly retrieve its imagery for an interactive browsing experience. The GIBS system is a core EOSDIS component which provides a scalable, responsive, highly available, and community standards based set of imagery services. The GIBS imagery archive includes over 100 imagery products representing visualized science parameters.

3.3.8 National environmental satellite, data, and information service

The NOAA's mission is to understand and predict changes in climate, weather, oceans and coasts, to share that knowledge and information with others, and to conserve and manage coastal and marine ecosystems and resources as the Nation's authoritative environmental intelligence agency. NESDIS supports NOAA's mission of Science, Service and Stewardship through our satellite missions, data centers, data and information products and services as well as use-inspired science. The United States depends on NOAA to provide satellite data and imagery for meteorological forecasts and emergency services to support continuity of government. NESDIS' responsibility is to collect and provide the critical satellite Earth observations and other essential environmental information needed for disaster preparedness, all hazards response and recovery and the protection of the Nation's critical infrastructure and natural resources. The 24/7 global coverage provided by NESDIS generates an uninterrupted stream of information and products. These products and information enable services used across the country in preparation for events that impact our climate, weather, oceans, daily lives and national safety and provide essential information for national, regional and local planners and officials. Massive amounts of satellite data is processed to create products, tools, and services that help decision-makers better inform the public and safeguard our environment. This data is archived in NCEI and can be accessed through NOAA's data search platform, "One Stop."

3.4 Data issues

The ocean science data is as complex as the oceans are. As data density and resolutions improve with technological advancements, the volume of ocean data is expected to expand dramatically, surpassing 250 petabytes by 2030 (NOAA, 2012). Huge volumes of data from multiple disciples of ocean science create significant

challenges in handling and managing the data. The data output from any sensor instrument collecting ocean science data should describe the data as well as the file structure. Since different datasets may use different terms to describe ocean variables, descriptions of metadata in the output files guarantee correct interpretation of data. Network Common Data Form (netCDF) is one such data format that is widely used by several data platforms. Processing raw data and making ocean data useful for analysis that can guide management choices, clarify long-term patterns, and solve key scientific issues takes a significant amount of time and skill. The requirements vary depending on the kind of data and are frequently a result of how the data was obtained. The processes required to handle the data must also be accurately recorded in order for scientific reproducibility across research utilizing the same data to be possible. Data processing and analysis will be a severe bottleneck given the volume of data that is currently available and expected in the future with the traditional manual methods. Thus, data science and data driven skills and expertise is needed in the field of ocean related sciences and collaborations among data scientists and marine scientists will solve many data related problems.

Despite the significant amounts of ocean data currently available and expected in the future, because of the vastness of oceans and dependence of several economic sectors on the ocean, there are data needs and gaps identified for different ocean data categories. In addition to acquiring acceptable ocean data, data management and access systems must be capable of distributing data to decision makers and other users in ways that are helpful to them. Planning for data collection should be done in tandem with planning for data usage, taking into account a variety of purposes and users, including individuals whose desire to provide data may be influenced in large part by whether their requirements are satisfied and their interests are respected. Moreover, the data collected by some academic and private sector communities, the data collected by private industries that is withheld due to privacy and confidentiality purposes, and by scientific groups working on specific projects with short-term funding is not openly available to the public. As a result, large amounts of ocean science data is inaccessible, and thus shows the need for data sharing.

4 Machine learning and its application in Gulf of Mexico

Machine learning (ML) uses dynamic models to make data-driven choices, and ML approaches may be used to high-dimensional, complicated, non-linear, and big-data problems. ML can tackle issues that are unfeasible or too difficult for traditional methodologies, which need a large number of people, resources, time, and effort to achieve the requisite precision. ML not only delivers effective solutions, resilience, and accuracy, but also efficiency due to its ability to swiftly handle massive volumes of data. Furthermore, the ML technique works well even when the data is noisy. ML is adapted by several scientific fields and has proven to be effective with the fields that generate large volumes of data. As the ocean science data is real-time, near real-time, and historical, ML techniques and approaches have been effective in

addressing the ocean related problems. In this section, we discuss how ML is being used to solve some of the problems faced by GOM and other ocean waters.

Wind and wave conditions are critical for many marine industries and hence the accurate forecasting and prediction of wind and wave parameters is a valuable resource. With the knowledge of wind and wave characteristics, shipping routes can be optimized by avoiding rough seas, aquaculture harvesting can be improved, testing and performance evaluation of unmanned marine vehicles can be conducted for navy and military operations. While traditional physics based models are used in wave prediction (Hsu and Holland, 2007; Kalourazi et al., 2021), physics based wind and wave modelling is computationally expensive and thus machine learning approaches with better accuracy are preferred. The outputs of thousands of wave model runs is used to create the training dataset to develop a surrogate model as a data-driven technique to empirically approximate the response of a physics based model in (James et al., 2018). Artificial neural network is fed with reanalysis wind data to extend the observed time series data of significant wave heights in (Peres et al., 2015). Data recorded by the buoys carrying inertial sensors is used by a convolutional neural network to predict wave heights and period in (Liu et al., 2019). Support vector machines, random forests, and artificial neural networks are used to forecast hurricane wave height over the GOM utilizing the data collected from six different buoys at different locations (Mafi and Amirinia, 2017). While some data driven machine learning approaches use the real-time and historical observations data from buoys in forecasting and predictions of wind and wave characteristics, (Ellenson et al., 2020) presents a hybrid approach to improve predictions by correcting a physics based model using a machine learning method.

While satellite remote sensing is an important tool for spatial and temporal estimation of some of the ocean parameters, in many models satellite mapping is done for a specific oceanic region often dominated by a single major oceanic process. Thus, any numerical or data driven models developed to estimate ocean parameters for a specific region may have poor applicability in other regions. Sea surface temperature, pCO_2 and pH are critical parameters which play an important role in understanding climate change and global warming. Yet, the estimation of these variables mostly based on satellite remote sensing is difficult due to the complex relationship between these variables and other environmental variables. Statistical techniques such as multiple linear regression (MLR) (Lefevre et al., 2002; Olsen et al., 2004; Jamet et al., 2007), multiple polynomial regression (MPR) (Stephens et al., 1995; Ono et al., 2004; Gloege et al., 2022), and principal component regression (PCR) (Lohrenz and Cai, 2006; Lohrenz et al., 2010) are used in most research that estimates pCO_2, whether based on remote sensing pictures or observation data sets. Self-organizing maps (SOMs) and feedforward neural networks (Moussa et al., 2016) are examples of machine learning algorithms that have done well in estimating pCO_2. Statistical research, on the other hand, still lacks knowledge on the mechanical mechanisms that explain why coastal regions act as carbon sinks or sources (Dai et al., 2013). At the same time, neither physical, chemical, nor biological processes can explain the projected findings (Chen et al., 2017). Fu et al. (2020)

combines cubist and semi-mechanistic methods to predict ocean surface pCO_2 using sea surface temperature, sea surface salinity, chlorophyll, and diffuse attenuation coefficient. The GOM is chosen as the study area and has obtained a satisfied performance providing a solid foundation for extending its application to other areas with similar environmental and geographic conditions. Most common ML approaches used to study water quality and key parameters such as pH, temperature, chlorophyll-a (Chl-a), salinity, dissolved oxygen, are Artificial Neural Networks (ANN), Support Vector Machines (SVM), Radio Frequency (RF), Decision Trees (DT), Multilayer Perceptron (MLP), cubist, and Gaussian processes (GP) (Hassan and Woo, 2021). These ML approaches are also used to monitor the water quality at regional and global scales. However, a variety of indicators like turbidity, Chromophoric Dissolved Organic Matter (CDOM), Chl-a, Dissolved Oxygen (DO), suspended solids concentrations, harmful algae, e.t.c are used to assess the water quality of a water body while water temperature, pH, salinity, and DO are common water quality indicators. For most of the water quality prediction, forecasting, and monitoring ML models satellite remote sensing data is a common source for training datasets (Kim et al., 2014a; Chang et al., 2017; Shehhi and Kaya, 2020; Gómez et al., 2021; Zhu et al., 2022). However, real-time, near real-time, and historical sensor measurements are also used to predict water quality and the contributing parameters (Lu and Huang, 2009; Xiang and Jiang, 2009; Solanki et al., 2015; Khan and See, 2016; Lorenzo et al., 2019; Manimegalai et al., 2020).

Data-driven models (Minsker et al., 2006; Coopersmith et al., 2011; Li et al., 2020; Yu et al., 2020) are also developed to understand, predict and forecast hypoxic and anoxic waters along the coastal regions using the dissolved oxygen content data along with other environmental variables. The GOM being prone to hypoxic conditions along the coastlines affects the marine industries, aquacultures, and tourism. The freshwater discharges are related to hypoxic conditions and thus affect water quality along the coastline which is a major concern (Alizadeh et al., 2018; Dzwonkowski et al., 2018). In addition, the GOM is also prone to harmful algal blooms along the coastal waters resulting in fish kills and marine wildlife kills (Blondeau-Patissier et al., 2014; Le et al., 2019; Yñiguez and Ottong, 2020). Predicting the chlorophyll and ˜ dissolved oxygen contents is handled by several machine learning approaches, developed to investigate algal blooms (Kumar and Bhandarkar, 2017; Hill et al., 2020; Chen et al., 2020b; Yerrapothu, 2021; Yu et al., 2021). Computer vision techniques are also studied in detection and classification of algal blooms (Samantaray et al., 2018; Pant et al., 2020).

With the extent of observational data available along the GOM, machine learning is used in predicting the severity of weather events (Ramachandra, 2019). Data-driving machine learning approach is used to predict the Loop Current evolution and the Loop Current ring formation in the Gulf of Mexico (Wang et al., 2019), including a forecast of the sea surface height of the Loop Current System (Zeng et al., 2015; Wang et al., 2021), and a forecast of velocity structures of the Loop Current and its eddies (Huang et al., 2021; Muhamed Ali et al., 2021; Huang et al., 2022b). All ocean basins are experiencing sea level rise and warming due to climate change and global warming, so predicting and understanding the sea-level rise is also done using the machine learning approaches (Roshni et al., 2019; Morovati et al., 2021; Nieves et al., 2021; Tur et al., 2021). Remote sensing data is quite helpful in detecting the oil spills by analyzing satellite images manually, however machine learning models can also help automate the detection and tracking of oil spills (Estes and Senger, 1971; Kubat et al., 1998; Shamsudeen, 2020). While remote sensing data is used to detect oil spills for a larger area, oil spill detection in confined areas like ports is also carried out with the help of aerial vehicles, thermal infrared images and a trained convolutional neural networks (De Kerf et al., 2020). To assist faster emergency oil spill responses a novel faster region-based convolutional neural network model is developed which uses satellite synthetic aperture radar imagery (Huang et al., 2022a). Flood warning systems and forecasting techniques play an important role in mitigating the hazards in flood prone areas and have a severe economic impact. While the physics based models have been long used to predict storms, rainfalls and other hydrologic events (Costabile et al., 2013; Fernández-Pato et al., 2016), data driven machine learning models show promising results faster than physics based models in forecasting floods (Xu and Li, 2002; Mekanik et al., 2013; Kim et al., 2016; Mosavi et al., 2017).

Within machine learning, Deep Learning (DL) models with a focus on ocean data analysis have also been developed in the recent years (Lou et al., 2021). A number of such DL models focus on object detection, sound, image or video processing of marine and ocean objects. Deep Learning models have been used to identify fish species in underwater drone footage (Meng et al., 2018) and coral reefs have been classified using such models with image processing applications (Mary and Dharma, 2017). With the advancement in underwater drones, ocean data can also take the form of videos and DL models have been applied for object detection in such videos when the quality is too low for manual detection (Sun et al., 2018). Computer vision and other DL models can successfully be used to detect numerous types of ocean objects, such as seagrass meadows in a plethora of ocean datasets (Moniruzzaman et al., 2017), some of the specific application include deep water sound processing to identify sources of sound pollution or detect specific marine species (Mishachandar and Vairamuthu, 2021) and to detect and classify fish call types in the northern Gulf of Mexico (Waddell et al., 2021). Satellite ocean data, such as what we discussed above, is also amenable to Deep Learning analysis (Ducournau and Fablet, 2016) including for ocean data forecasting (Choi et al., 2022). Deep Learning approaches also show significant promise for climate change modeling of ocean data such as wave energy forecasting (Bento et al., 2021), as well as sea surface temperature patterns to identify ocean extremes (Prochaska et al., 2021). They are also effective tools in ocean data quality control, such as when used for ocean temperature data with potential gross errors (Mieruch et al., 2021). To overcome sampling rates and low-resolution ocean data (Bolton and Zanna, 2019) use Deep Learning approaches to predict unresolved turbulent processes and subsurface flow fields. Figure 3 shows the summary of this section by illustrating the major areas of ML applications in the GoM with the proportion of the literature divided into specific categories of interest.
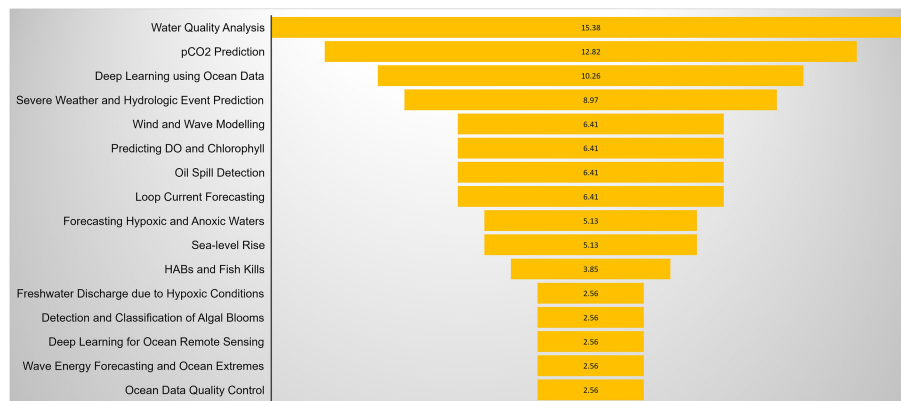
FIGURE 3
Illustration of the major areas of ML applications in the Gulf of Mexico, represented per the proportion of the literature covered under each area in this article.

5 Discussion and conclusions

The reviewed literature identifies several approaches that can be applied to GOM problems, and Figures 1, 3 depict the breadth of studies on a wide variety of applicable topics. The figures also illustrate areas that need further research both within the ML-based literature and the broader climate change literature. Algal blooms, HABs/fish kills, and sea-level rises (major problems with the GOM) have not received as much attention as other climate change problems. Within the ML-based literature, the effects on estuaries and coastal systems, and oyster mortality (both being major problem for the GOM) have been understudied and we identify those as important areas for improvement. Our overview, combined with the abundance of data and constantly developing methodologies, should serve as a template to identify areas of research most in need of further work. We have attempted to provide that template, suggestions for possible data sources, as well as methods that can be used to analyze the available data. As we pointed out, the GOM is of major interest economically, ecologically, and within public policy so such future research holds tremendous promise for a broad impact. We anticipate this paper will help data scientists looking to delve into ocean science to find a research problem and ocean science researchers trying to find ways how ML skills can be applied to their specific research problem.

Author contributions

VS wrote the manuscript with the support of JM, SK, II and DNB. All authors contributed to the article and approved the submitted version.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alizadeh, M. J., Kavianpour, M. R., Danesh, M., Adolf, J., Shamshirband, S., and Chau, K. W. (2018). Effect of river flow on the quality of estuarine and coastal waters using machine learning models. *Eng. Appl. Comput. Fluid Mech.* 12, 810–823. doi: 10.1080/19942060.2018.1528480

Altieri, A. H., and Gedan, K. B. (2015). Climate change and dead zones. *Global Change Biol.* 21, 1395–1406. doi: 10.1111/gcb.12754

Anderson, D. M., Glibert, P. M., and Burkholder, J. M. (2002). Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. *Estuaries* 25, 704–726. doi: 10.1007/BF02804901

Anderson, J. B., Wallace, D. J., Simms, A. R., Rodriguez, A. B., and Milliken, K. T. (2014). Variable response of coastal environments of the northwestern gulf of mexico to sea-level rise and climate change: Implications for future change. *Mar. Geol.* 352, 348–366. doi: 10.1016/j.margeo.2013.12.008

Appice, A., Gel, Y. R., Iliev, I., Lyubchich, V., and Malerba, D. (2020). A multi-stage machine learning approach to predict dengue incidence: a case study in mexico. *IEEE Access* 8, 52713–52725. doi: 10.1109/ACCESS.2020.2980634

Bakkenes, M., Alkemade, J. R. M., Ihle, F., Leemans, R., and Latour, J. B. (2002). Assessing effects of forecasted climate change on the diversity and distribution of european higher plants for 2050. *Global Change Biol.* 8, 390–407. doi: 10.1046/j.1354-1013.2001.00467.x

Beck, M. W., Brumbaugh, R. D., Airoldi, L., Carranza, A., Coen, L. D., Crawford, C., et al. (2011). Oyster reefs at risk and recommendations for conservation, restoration, and management. *Bioscience* 61, 107–116. doi: 10.1525/bio.2011.61.2.5

Bennett, E. M., Carpenter, S. R., and Caraco, N. F. (2001). Human impact on erodable phosphorus and eutrophication: a global perspective: increasing accumulation of phosphorus in soil threatens rivers, lakes, and coastal oceans with eutrophication. *BioScience* 51, 227–234. doi: 10.1641/0006-3568(2001)051[0227:HIOEPA]2.0.CO;2

Bento, P., Pombo, J., Mendes, R., Calado, M., and Mariano, S. (2021). Ocean wave energy forecasting using optimised deep learning neural networks. *Ocean Eng.* 219, 108372. doi: 10.1016/j.oceaneng.2020.108372

Bindoff, N. L., Willebrand, J., Artale, V., Cazenave, A., Gregory, J. M., Gulev, S., et al. (2007). *Observations: oceanic climate change and sea level* (Cambridge: Cambridge University Press).

Blondeau-Patissier, D., Gower, J. F., Dekker, A. G., Phinn, S. R., and Brando, V. E. (2014). A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Prog. oceanogr.* 123, 123–144. doi: 10.1016/j.pocean.2013.12.008

Bolton, T., and Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *J. Adv. Modeling Earth Syst.* 11, 376–399. doi: 10.1029/2018MS001472

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., et al. (2013). Multiple stressors of ocean ecosystems in the 21st century: projections with cmip5 models. *Biogeosciences* 10, 6225–6245. doi: 10.5194/bg-10-6225-2013

Boyd, P. W., and Doney, S. C. (2003). "The impact of climate change and feedback processes on the ocean carbon cycle," in *Ocean biogeochemistry* (Heidelberg Berlin: Springer), 157–193.

Breitburg, D. L., Salisbury, J., Bernhard, J. M., Cai, W.-J., Dupont, S., Doney, S. C., et al. (2015). And on top of all that… coping with ocean acidification in the midst of many stressors. *Oceanography* 28, 48–61. doi: 10.5670/oceanog.2015.31

Bricker, S. B. (1999). National estuarine eutrophication assessment: effects of nutrient enrichment in the nation's estuaries. *NOAA National Ocean Service Special Projects Office and the National Centers for Coastal Ocean Science*. Silver Spring, MD, 77p.

Burkholder, J. M., Dickey, D. A., Kinder, C. A., Reed, R. E., Mallin, M. A., McIver, M. R., et al. (2006). Comprehensive trend analysis of nutrients and related variables in a large eutrophic estuary: a decadal study of anthropogenic and climatic influences. *Limnol. Oceanogr.* 51, 463–487. doi: 10.4319/lo.2006.51.1_part_2.0463

Burkholder, J. M., and Glibert, P. M. (2006). Intraspecific variability: an important consideration in forming generalisations about toxigenic algal species. *Afr. J. Mar. Sci.* 28, 177–180. doi: 10.2989/18142320609504143

Cai, W.-J., Hu, X., Huang, W.-J., Murrell, M. C., Lehrter, J. C., Lohrenz, S. E., et al. (2011). Acidification of subsurface coastal waters enhanced by eutrophication. *Nat. Geosci.* 4, 766–770. doi: 10.1038/ngeo1297

Caraco, N. F., and Cole, J. J. (1999). Human impact on nitrate export: an analysis using major world rivers. *Ambio* 28, 167–170.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., et al. (2019). Machine learning and the physical sciences. *Rev. Modern Phys.* 91, 045002. doi: 10.1103/RevModPhys.91.045002

Chang, N. B., Bai, K., and Chen, C. F. (2017). Integrating multisensor satellite data merging and image reconstruction in support of machine learning for better water quality management. *J. Environ. Manage.* 201, 227–240. doi: 10.1016/j.jenvman.2017.06.045

Chen, C., Chen, H., Liao, W., Sui, X., Wang, L., Chen, J., et al. (2020a). Dynamic monitoring and analysis of land-use and land-cover change using landsat multitemporal data in the zhoushan archipelago, china. *IEEE Access* 8, 210360–210369. doi: 10.1109/ACCESS.2020.3036128

Chen, H., Chen, C., Zhang, Z., Lu, C., Wang, L., He, X., et al. (2021). Changes of the spatial and temporal characteristics of land-use landscape patterns using multi-temporal landsat satellite data: A case study of zhoushan island, china. *Ocean Coast. Manage.* 213, 105842. doi: 10.1016/j.ocecoaman.2021.105842

Chen, S., Hu, C., Cai, W. J., and Yang, B. (2017). Estimating surface pco2 in the northern gulf of mexico: Which remote sensing model to use? *Continental Shelf Res.* 151, 94–110. doi: 10.1126/sciadv.160154

Chen, C., Liang, J., Xie, F., Hu, Z., Sun, W., Yang, G., et al. (2022). Temporal and spatial variation of coastline using remote sensing images for zhoushan archipelago, china. *Int. J. Appl. Earth Observ. Geoinf.* 107, 102711. doi: 10.1016/j.jag.2022.102711

Chen, Y. J., Nicholson, E., and Cheng, S. T. (2020b). Using machine learning to understand the implications of meteorological conditions for fish kills. *Sci. Rep.* 10, 1–13. doi: 10.1038/s41598-020-73922-3

Cheng, L., Trenberth, K. E., Fasullo, J., Boyer, T., Abraham, J., and Zhu, J. (2017). Improved estimates of ocean heat content from 1960 to 2015. *Sci. Adv.* 3, 3. doi: 10.1126/sciadv.1601545

Choi, Y., Park, Y., Hwang, J., Jeong, K., and Kim, E. (2022). Improving ocean forecasting using deep learning and numerical model integration. *J. Mar. Sci. Eng.* 10, 450. doi: 10.3390/jmse10040450

Cloern, J. E. (2001). Our evolving conceptual model of the coastal eutrophication problem. *Mar. Ecol. Prog. Ser.* 210, 223–253. doi: 10.3354/meps210223

Coleman, F. C., and Koenig, C. C. (2010). The effects of fishing, climate change, and other anthropogenic disturbances on red grouper and other reef fishes in the gulf of mexico. *Integr. Comp. Biol.* 50, 201–212. doi: 10.1093/icb/icq072

Comeaux, R. S., Allison, M. A., and Bianchi, T. S. (2012). Mangrove expansion in the gulf of mexico with climate change: Implications for wetland health and resistance to rising sea levels. *Estuar. Coast. Shelf Sci.* 96, 81–95. doi: 10.1016/j.ecss.2011.10.003

Conway, C. J. (2011). Standardized north American marsh bird monitoring protocol. *Waterbirds* 34, 319–346. doi: 10.1675/063.034.0307

Coopersmith, E. J., Minsker, B., and Montagna, P. (2011). Understanding and forecasting hypoxia using machine learning algorithms. *J. Hydroinform.* 13, 64–80. doi: 10.2166/hydro.2010.015

Costabile, P., Costanzo, C., and Macchione, F. (2013). A storm event watershed model for surface runoff based on 2d fully dynamic wave equations. *Hydrol. processes* 27, 554–569. doi: 10.1002/hyp.9237

Dai, M., Cao, Z., Guo, X., Zhai, W., Liu, Z., Yin, Z., et al. (2013). Why are some marginal seas sources of atmospheric co2? *Geophys. Res. Lett.* 40, 2154–2158. doi: 10.1002/grl.50390

Daigle, R. M., Chassé, J., and Metaxas, A. (2016). The relative effect of behaviour in larval dispersal in a low energy embayment. *Prog. Oceanogr.* 144, 93–117. doi: 10.1016/j.pocean.2016.04.001

Data, B. (2008). Community cleverness required. *Nature* 455. doi: 10.1038/455001a

De Kerf, T., Gladines, J., Sels, S., and Vanlanduit, S. (2020). Oil spill detection using machine learning and infrared images. *Remote Sens.* 12, 4090. doi: 10.3390/rs12244090

Deo, R. C., and Sahin, M. (2015). Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern australia. *Atmos. Res.* 153, 512–525. doi: 10.1016/j.atmosres.2014.10.016

Diaz, R. J., and Rosenberg, R. (2008). Spreading dead zones and consequences for marine ecosystems. *science* 321, 926–929. doi: 10.1126/science.1156401

Diaz, R. J., and Rosenberg, R. (1995). Marine benthic hypoxia: a review of its ecological effects and the behavioural responses of benthic macrofauna. *Oceanogr. Mar. Biol. Annu. Rev.* 33, 245–203.

Doney, S. C. (2010). The growing human footprint on coastal and open-ocean biogeochemistry. *science* 328, 1512–1516. doi: 10.1126/science.1185198

Doney, S. C., Ruckelshaus, M., Duffy, E., J., B., P., J., Chan, F., et al. (2012). Climate change impacts on marine ecosystems. *Annu. Rev. Mar. Sci.* 4, 11–37. doi: 10.1146/annurev-marine-041911-111611

Ducournau, A., and Fablet, R. (2016). "Deep learning for ocean remote sensing: an application of convolutional neural networks for super-resolution on satellite-derived sst data," in *2016 9th IAPR Workshop on Pattern Recogniton in Remote Sensing (PRRS)*. (IEEE) 1–6.

Dzwonkowski, B., Fournier, S., Reager, J. T., Milroy, S., Park, K., Shiller, A. M., et al. (2018). Tracking sea surface salinity and dissolved oxygen on a river-influenced, seasonally stratified shelf, mississippi bight, northern gulf of mexico. *Continental Shelf Res.* 169, 25–33. doi: 10.1016/j.csr.2018.09.009

Ellenson, A., Pei, Y., Wilson, G., Özkan-Haller, H. T., and Fern, X. (2020). An application of a machine learning algorithm to determine and describe error patterns within wave model output. *Coast. Eng.* 157, 103595. doi: 10.1016/j.coastaleng.2019.103595

Enwright, N. M., Wang, L., Dalyander, P. S., Wang, H., Osland, M. J., Mickey, R. C., et al. (2021). Assessing habitat change and migration of barrier islands. *Estuar. Coasts* 44, 2073–2086. doi: 10.1007/s12237-021-00971-w

Epstein, P. R. (2005). Climate change and human health. *New Engl. J. Med.* 353, 1433–1436. doi: 10.1056/NEJMp058079

Estes, J. E., and Senger, L. W. (1971). The multispectral concept as applied to marine oil spills. *Remote Sens. Environ.* 2, 141–163. doi: 10.1016/0034-4257(71)90088-5

Feely, R. A., Doney, S. C., and Cooley, S. R. (2009). Ocean acidification: Present conditions and future changes in a high-co2 world. *Oceanography* 22, 36–47. doi: 10.5670/oceanog.2009.95

Fernández-Pato, J., Caviedes-Voullième, D., and García-Navarro, P. (2016). Rainfall/runoff simulation with 2d full shallow water equations: Sensitivity analysis and calibration of infiltration parameters. *J. hydrol.* 536, 496–513. doi: 10.1016/j.jhydrol.2016.03.021

Flynn, K. J., Clark, D. R., Mitra, A., Fabian, H., Hansen, P. J., Glibert, P. M., et al. (2015). Ocean acidification with (de) eutrophication will alter future phytoplankton growth and succession. *Proc. R. Soc. B.: Biol. Sci.* 282, 20142604. doi: 10.1098/rspb.2014.2604

Fox-Kemper, B., Hewitt, H., Xiao, C., Aðalgeirsdóttir, G., Drijfhout, S., Edwards, T., et al. (2021). *Ocean, cryosphere and Sea level change* (Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press), 1211–1362. doi: 10.1017/9781009157896.011

Fu, Z., Hu, L., Chen, Z., Zhang, F., Shi, Z., Hu, B., et al. (2020). Estimating spatial and temporal variation in ocean surface pco2 in the gulf of mexico using remote sensing and machine learning techniques. *Sci. Total Environ.* 745, 140965. doi: 10.1016/j.scitotenv.2020.140965

Gledhill, J. H., Barnett, A. F., Slattery, M., Willett, K. L., Easson, G. L., Otts, S. S., et al. (2020). Mass mortality of the eastern oyster crassostrea virginica in the western mississippi sound following unprecedented mississippi river flooding in 2019. *J. Shellfish Res.* 39, 235–244. doi: 10.2983/035.039.0205

Glibert, P. M., Heil, C. A., Rudnick, D. T., Madden, C. J., Boyer, J. N., and Kelly, S. P. (2009). Florida Bay: water quality status and trends, historic and emerging algal bloom problems. *Contrib. Mar. Sci.* 38, 5–17.

Gloege, L., Yan, M., Zheng, T., and McKinley, G. A. (2022). Improved quantification of ocean carbon uptake by using machine learning to merge global models and pco2 data. *J. Adv. Modeling Earth Syst.* 14, e2021MS002620. doi: 10.1029/2021MS002620

Gómez, D., Salvador, P., Sanz, J., and Casanova, J. L. (2021). A new approach to monitor water quality in the menor sea (spain) using satellite data and machine learning methods. *Environ. pollut.* 286, 117489. doi: 10.1016/j.envpol.2021.117489

Grabowski, J. H., Brumbaugh, R. D., Conrad, R. F., Keeler, A. G., Opaluch, J. J., Peterson, C. H., et al. (2012). Economic valuation of ecosystem services provided by oyster reefs. *Bioscience* 62, 900–909. doi: 10.1525/bio.2012.62.10.10

Greenberg, R. S., and Maldonado, J. E. (2006). Diversity and endemism in tidal marsh vertebrates. *Vertebr. Tidal Marshes* 32, 32–53.

Hare, J. A., Alexander, M. A., Fogarty, M. J., Williams, E. H., and Scott, J. D. (2010). Forecasting the dynamics of a coastal fishery species using a coupled climate–population model. *Ecol. Appl.* 20, 452–464. doi: 10.1890/08-1863.1

Hassan, N., and Woo, C. (2021). "Machine learning application in water quality using satellite data," in *IOP Conference Series: Earth and Environmental Science*, (IOP Publishing)Vol. 842. 012018.

Havens, K. E., and Paerl, H. W. (2015). Climate change at a crossroad for control of harmful algal blooms. *Environmental Science and Technology* (2015) 49, 12605–12606. doi: 10.1021/acs.est.5b03990

Hill, P. R., Kumar, A., Temimi, M., and Bull, D. R. (2020). Habnet: Machine learning, remote sensing-based detection of harmful algal blooms. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 13, 3229–3239. doi: 10.1109/JSTARS.2020.3001445

Hoegh-Guldberg, O., and Bruno, J. F. (2010). The impact of climate change on the world's marine ecosystems. *Science* 328, 1523–1528. doi: 10.1126/science.1189930

Howarth, R. (1995). "Transport to and processing of phosphorus in near-shore and oceanic waters," in *Phosphorus in the global environment, transfers, cycles and management* SCOPE.

Howarth, R. W., Anderson, D., Cloern, J. E., Elfring, C., Hopkinson, C. S., Lapointe, B., et al. (2000). Nutrient pollution of coastal rivers, bays, and seas. *Issues Ecol.* 7, 1–16.

Hsu, W. R. J. K. L., and Holland, R. J. D. J. K. (2007). Forecasting and hindcasting waves with the swan model in the southern california bight. *Coast. Eng.* 54, 1–15. doi: 10.1016/j.coastaleng.2006.06.011

Hu, C., Weisberg, R. H., Liu, Y., Zheng, L., Daly, K. L., English, D. C., et al. (2011). Did the northeastern gulf of mexico become greener after the deepwater horizon oil spill? *Geophys. Res. Lett.* 38.

Huang, X., Iliev, I. R., Brenning, A., and Gel, Y. R. (2016). Space-time clustering with?stability probe while riding downhill. *Proc. SIGKDD Min. Learn. Time Ser.* doi: 10.1145/1235

Huang, X., Iliev, I. R., Lyubchich, V., and Gel, Y. R. (2018). Riding down the bay: Space-time clustering of ecological trends. *Environmetrics* 29, e2455. doi: 10.1002/env.2455

Huang, Y., Tang, Y., Zhu, X., Zhuang, H., and Cherubin, L. (2022b). Physics-coupled spatio-temporal active learning for dynamical systems. *IEEE Access* 10, 112909–112920. doi: 10.1109/ACCESS.2022.3214544

Huang, Y., Tang, Y., Zhuang, H., VanZwieten, J., and Cherubin, L. (2021). Physics-informed tensor-train convlstm for volumetric velocity forecasting of the loop current. *Front. Artif. Intell.* 4, 780271. doi: 10.3389/frai.2021.780271

Huang, X., Zhang, B., Perrie, W., Lu, Y., and Wang, C. (2022a). A novel deep learning method for marine oil spill detection from satellite synthetic aperture radar imagery. *Mar. pollut. Bull.* 179, 113666. doi: 10.1016/j.marpolbul.2022.113666

Iliev, I., Rishel, T., and Li, B. (2022). Semantic analysis of arab leaders on social media. *Soc. Sci. J.*, 1–26. doi: 10.1080/03623319.2021.2001224

James, S. C., Zhang, Y., and O'Donncha, F. (2018). A machine learning framework to forecast wave conditions. *Coast. Eng.* 137, 1–10. doi: 10.1016/j.coastaleng.2018.03.004

Jamet, C., Moulin, C., and Lefèvre, N. (2007). "Estimation of the oceanic pco 2 in the north atlantic from vos lines in-situ measurements: parameters needed to generate seasonally mean maps," in *Annales geophysicae (Copernicus GmbH)*, vol. 25. , 2247–2257.

Kalourazi, M. Y., Siadatmousavi, S. M., Yeganeh-Bakhtiary, A., and Jose, F. (2021). Wavewatch-iii source terms evaluation for optimizing hurricane wave modeling: A case study of hurricane ivan. *Oceanologia* 63, 194–213. doi: 10.1016/j.oceano.2020.12.001

Keeling, R. F., Körtzinger, A., and Gruber, N. (2010). Ocean deoxygenation in a warming world. *Annu. Rev. Mar. Sci.* 2, 199–229. doi: 10.1146/annurev.marine.010908.163855

Khan, Y., and See, C. S. (2016). "Predicting and analyzing water quality using machine learning: a comprehensive model," in *2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*. (IEEE) 1–6.

Kim, Y. H., Im, J., Ha, H. K., Choi, J. K., and Ha, S. (2014). Machine learning approaches to coastal water quality monitoring using goci satellite data. *GISci. Remote Sens.* 51, 158–174. doi: 10.1080/15481603.2014.900983

Kim, S., Matsumi, Y., Pan, S., and Mase, H. (2016). A real-time forecast model using artificial neural network for after-runner storm surges on the tottori coast, japan. *Ocean Eng.* 122, 44–53. doi: 10.1016/j.oceaneng.2016.06.017

Kubat, M., Holte, R. C., and Matwin, S. (1998). Machine learning for the detection of oil spills in satellite radar images. *Mach. Learn.* 30, 195–215. doi: 10.1023/A:1007452223027

Kumar, A. C., and Bhandarkar, S. M. (2017). "A deep learning paradigm for detection of harmful algal blooms," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. (IEEE) 743–751.

Laurent, A., Ko, D.S., and Lehrter, J. (2018). Climate change projected to exacerbate impacts of coastal eutrophication in the northern gulf of mexico. *J. Geophys. Res.: Oceans* 123, 3408–3426. doi: 10.1002/2017JC013583

Le, D., Huang, W., and Johnson, E. (2019). Neural network modeling of monthly salinity variations in oyster reef in apalachicola bay in response to freshwater inflow and winds. *Neural Comput. Appl.* 31, 6249–6259. doi: 10.1007/s00521-018-3436-y

Lee, H. (2007). Intergovernmental panel on climate change. *AR4 Climate Change 2007: Synthesis Report*. doi: 10.1017/CBO9780511546013

Lefevre, N., Aiken, J., Rutllant, J., Daneri, G., Lavender, S., and Smyth, T. (2002). Observations of pco2 in the coastal upwelling off chile: Spatial and temporal extrapolation using satellite data. *J. Geophys. Res.: Oceans* 107, 8–1. doi: 10.1029/2000JC000395

Levin, L. A., Liu, K.-K., Emeis, K.-C., Breitburg, D. L., Cloern, J., Deutsch, C., et al. (2015). Comparative biogeochemistry–ecosystem–human interactions on dynamic continental margins. *J. Mar. Syst.* 141, 3–17. doi: 10.1016/j.jmarsys.2014.04.016

Levina, E., Jacob, J., Bustillos, L. E. R., and Ortiz, I. (2007). *Policy frameworks for adaptation to climate change in coastal zones: The case of the gulf of mexico* Vol. 7 (Paris, France: OECD Papers), 1–68.

Li, W., Fang, H., Qin, G., Tan, X., Huang, Z., Zeng, F., et al. (2020). Concentration estimation of dissolved oxygen in pearl river basin using input variable selection and machine learning techniques. *Sci. Total Environ.* 731, 139099. doi: 10.1016/j.scitotenv.2020.139099

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors* 18, 2674. doi: 10.3390/s18082674

Liu, Y., Lee, S.-K., Muhling, B. A., Lamkin, J. T., and Enfield, D. B. (2012). Significant reduction of the loop current in the 21st century and its impact on the gulf of mexico. *J. Geophys. Res.: Oceans* 117. doi: 10.1029/2011JC007555

Liu, T., Zhang, Y., Qi, L., Dong, J., Lv, M., and Wen, Q. (2019). "Wavenet: learning to predict wave height and period from accelerometer data using convolutional neural network," in *IOP Conference Series: Earth and Environmental Science*, (IOP Publishing) Vol. 369. 012001.

Loarie, S. R., Carter, B. E., Hayhoe, K., McMahon, S., Moe, R., Knight, C. A., et al. (2008). Climate change and the future of california's endemic flora. *PloS One* 3, e2502. doi: 10.1371/journal.pone.0002502

Lohrenz, S. E., and Cai, W.-J. (2006). Satellite ocean color assessment of air-sea fluxes of co2 in a river-dominated coastal margin. *Geophys. Res. Lett.* 33.

Lohrenz, S. E., Cai, W.-J., Chen, F., Chen, X., and Tuel, M. (2010). Seasonal variability in air-sea fluxes of co2 in a river-influenced coastal margin. *J. Geophys. Res.: Oceans* 115.

Lorenzo, A. R., Dula, A. Y., Valeroso, N. A. C., Munda, J. D. C., Supang, B. N. G., Padilla, M. V. C., et al. (2019). "Dissolved oxygen (do) meter hydrological modelling using predictive algorithms," in *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*. (IEEE) 1–6.

Lou, R., Lv, Z., Dang, S., Su, T., and Li, X. (2021). Application of machine learning in ocean data. *Multimed. Syst.*, 1–10. doi: 10.1007/s00530-020-00733-x

Lu, J., and Huang, T. (2009). "Data mining on forecast raw water quality from online monitoring station based on decision-making tree," in *2009 Fifth International Joint Conference on INC, IMS and IDC*. (IEEE) 706–709.

Mafi, S., and Amirinia, G. (2017). Forecasting hurricane wave height in gulf of mexico using soft computing methods. *Ocean Eng.* 146, 352–362. doi: 10.1016/j.oceaneng.2017.10.003

Manimegalai, R., and Jerom, A. B., (2020). "An iot based smart water quality monitoring system using cloud," in *2020 International conference on emerging trends in information technology and engineering (ic-ETITE)*. (IEEE) 1–7.

Mansfield, L. A., Nowack, P. J., Kasoar, M., Everitt, R. G., Collins, W. J., Voulgarakis, A., et al. (2020). Predicting global patterns of long-term climate change from short-term simulations using machine learning. *NPJ Climate Atmos. Sci.* 3, 1–9. doi: 10.1038/s41612-020-00148-5

Mary, N. A. B., and Dharma, D. (2017). Coral reef image classification employing improved ldp for feature extraction. *J. Visual Commun. Image Represent.* 49, 225–242. doi: 10.1016/j.jvcir.2017.09.008

Meehl, G., Stocker, T., Collins, W., Friedlingstein, P., Gaye, A., Gregory, J., et al. (2007). "Chapter 10: Global climate projections," in *Ipcc fourth assessment report*. Available at: https://www.ipcc.ch/site/assets/uploads/2018/03/ar4_wg2_full_report.pdf

Mekanik, F., Imteaz, M., Gato-Trinidad, S., and Elmahdi, A. (2013). Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. *J. Hydrol.* 503, 11–21. doi: 10.1016/j.jhydrol.2013.08.035

Melzner, F., Thomsen, J., Koeve, W., Oschlies, A., Gutowska, M. A., Bange, H. W., et al. (2013). Future ocean acidification will be amplified by hypoxia in coastal habitats. *Mar. Biol.* 160, 1875–1888. doi: 10.1007/s00227-012-1954-1

Meng, L., Hirayama, T., and Oyanagi, S. (2018). Underwater-drone with panoramic camera for automatic fish recognition based on deep learning. *IEEE Access* 6, 17880–17886. doi: 10.1109/ACCESS.2018.2820326

Mieruch, S., Demirel, S., Simoncelli, S., Schlitzer, R., and Seitz, S. (2021). Salaciaml: A deep learning approach for supporting ocean data quality control. *Front. Mar. Sci.* 8, 611742. doi: 10.3389/fmars.2021.611742

Miller, W., Harding, L., and Adolf, J. (2005). "The influence of hurricane isabel on chesapeake bay phytoplankton dynamics," in *Hurricane Isabel in perspective* (Edgwater, MD: Chesapeake Research Consortium Publication), 05–160.

Minsker, B., Maidment, D., Hodges, B., Montagna, P., and Bonner, J. (2006). "An environmental information system for hypoxia in corpus christi bay: A waters network testbed," in *AGU fall meeting abstracts* (San Francisco, CA; American Geophysical Union), vol. 2006, H21F–1432.

Mishachandar, B., and Vairamuthu, S. (2021). Diverse ocean noise classification using deep learning. *Appl. Acoust.* 181, 108141. doi: 10.1016/j.apacoust.2021.108141

Misra, V., Mishra, A., and Bhardwaj, A. (2019). A coupled ocean-atmosphere downscaled climate projection for the peninsular florida region. *J. Mar. Syst.* 194, 25–40. doi: 10.1016/j.jmarsys.2019.02.010

Moniruzzaman, M., Islam, S. M. S., Bennamoun, M., and Lavery, P. (2017). "Deep learning on underwater marine object detection: A survey," in *International Conference on Advanced Concepts for Intelligent Vision Systems*. (Springer) 150–160.

Morovati, K., Nakhaei, P., Tian, F., Tudaji, M., and Hou, S. (2021). A machine learning framework to predict reverse flow and water level: A case study of tonle sap lake. *J. Hydrol.* 603, 127168. doi: 10.1016/j.jhydrol.2021.127168

Mosavi, A., Ozturk, P., and wing Chau, K. (2018). Flood prediction using machine learning models: Literature review. *Water* 10, 11. doi: 10.3390/w10111536

Mosavi, A., Rabczuk, T., and Varkonyi-Koczy, A. R. (2017). "Reviewing the novel machine learning tools for materials design," in *International Conference on Global Research and Education*. (Springer) 50–58.

Moussa, H., Benallal, M. A., Goyet, C., and Lefèvre, N. (2016). Satellite-derived co2 fugacity in surface seawater of the tropical atlantic ocean using a feedforward neural network. *Int. J. Remote Sens.* 37, 580–598. doi: 10.1080/01431161.2015.1131872

Muhamed Ali, A., Zhuang, H., VanZwieten, J., Ibrahim, A. K., and Chérubin, L. (2021). A deep learning model for forecasting velocity structures of the loop current system in the gulf of mexico. *Forecasting* 3, 934–953. doi: 10.3390/forecast3040056

Muhling, B. A., Lee, S. K., Lamkin, J. T., and Liu, Y. (2011). Predicting the effects of climate change on bluefin tuna (thunnus thynnus) spawning habitat in the gulf of mexico. *ICES J. Mar. Sci.* 68, 1051–1062. doi: 10.1093/icesjms/fsr008

NASA Earth Observatory (2005) *Warm waters provide fuel for potential storms*. Available at: https://earthobservatory.nasa.gov/images/5862/warm-waters-provide-fuel-for-potential-storms (Accessed 20-Sep-2005).

NCEI (2022). *Ncei archive: Growth and change* (Asheville, North Carolina: NCEI). Available at: https://www.ncei.noaa.gov/news/ncei-archive-growth-and-change.

Nieves, V., Radin, C., and Camps-Valls, G. (2021). Predicting regional coastal sea level changes with machine learning. *Sci. Rep.* 11, 1–6. doi: 10.1038/s41598-021-87460-z

Nixon, S. W., Granger, S. L., and Nowicki, B. L. (1995). An assessment of the annual mass balance of carbon, nitrogen, and phosphorus in narragansett bay. *Biogeochemistry* 31, 15–61. doi: 10.1007/BF00000805

NOAA (2012). *2011 report to congress on data and information management* (Washington DC: National Oceanic and Atmospheric Administration NOAA).

NOAA (2022). *Noaa (national oceanic and atmospheric administration) fisheries* (Washington D.C.: NOAA). Available at: https://tinyurl.com/y4yhnre3.

Olsen, A., Triñanes, J. A., and Wanninkhof, R. (2004). Sea–Air flux of co2 in the caribbean sea estimated using *in situ* and remote sensing data. *Remote Sens. Environ.* 89, 309–325. doi: 10.1016/j.rse.2003.10.011

Ono, T., Saino, T., Kurita, N., and Sasaki, K. (2004). Basin-scale extrapolation of shipboard pco2 data by using satellite sst and chl a. *Int. J. Remote Sens.* 25, 3803–3815. doi: 10.1080/01431160310001657515

Paerl, H. W. (2017). Controlling harmful cyanobacterial blooms in a climatically more extreme world: management options and research needs. *J. Plankton Res.* 39, 763–771. doi: 10.1093/plankt/fbx042

Paerl, H. W., Gardner, W. S., Havens, K. E., Joyner, A. R., McCarthy, M. J., Newell, S. E., et al. (2016). Mitigating cyanobacterial harmful algal blooms in aquatic ecosystems impacted by climate change and anthropogenic nutrients. *Harmful Algae* 54, 213–222. doi: 10.1016/j.hal.2015.09.009

Paerl, H. W., and Huisman, J. (2009). Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environ. Microbiol. Rep.* 1, 27–37. doi: 10.1111/j.1758-2229.2008.00004.x

Paerl, H. W., Otten, T. G., and Kudela, R. (2018). Mitigating the expansion of harmful algal blooms across the freshwater-to-marine continuum. *Environmental Science and Technology* (2018) 52(10), 5519–5529. doi: 10.1021/acs.est.7b05950

Paerl, H. W., and Paul, V. J. (2012). Climate change: links to global expansion of harmful cyanobacteria. *Water Res.* 46, 1349–1363. doi: 10.1016/j.watres.2011.08.002

Pant, G., Yadav, D., and Gaur, A. (2020). Resnext convolution neural network topology-based deep learning model for identification and classification of pediastrum. *Algal Res.* 48, 101932. doi: 10.1016/j.algal.2020.101932

Peres, D. J., Iuppa, C., Cavallaro, L., Cancelliere, A., and Foti, E. (2015). Significant wave height record extension by neural networks and reanalysis wind data. *Ocean Model.* 94, 128–140. doi: 10.1016/j.ocemod.2015.08.002

Perry, A. L., Low, P. J., Ellis, J. R., and Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. *Science* 308, 1912–1915. doi: 10.1126/science.1111322

Potter, H., DiMarco, S. F., and Knap, A. H. (2019). Tropical cyclone heat potential and the rapid intensification of hurricane harvey in the texas bight. *J. Geophys. Res.: Oceans* 124, 2440–2451. doi: 10.1029/2018JC014776

Prochaska, J. X., Cornillon, P. C., and Reiman, D. M. (2021). Deep learning of sea surface temperature patterns to identify ocean extremes. *Remote Sens.* 13, 744. doi: 10.3390/rs13040744

Rabalais, N. N. (2002). Nitrogen in aquatic ecosystems. *AMBIO: J. Hum. Environ.* 31, 102–112. doi: 10.1579/0044-7447-31.2.102

Rabalais, N. N., Diaz, R. J., Levin, L. A., Turner, R. E., Gilbert, D., and Zhang, J. (2010). Dynamics and distribution of natural and human-caused hypoxia. *Biogeosciences* 7, 585–619. doi: 10.5194/bg-7-585-2010

Rabalais, N. N., Turner, R. E., and Wiseman, J. W. J. (2002). Gulf of mexico hypoxia, aka "the dead zone". *Annu. Rev. Ecol. Syst.*, 235–263. doi: 10.1146/annurev.ecolsys.33.010802.150513

Rabouille, C., Conley, D., Dai, M., Cai, W.-J., Chen, C., Lansard, B., et al. (2008). Comparison of hypoxia among four river-dominated ocean margins: The changjiang (yangtze), mississippi, pearl, and rhône rivers. *Continental Shelf Res.* 28, 1527–1537. doi: 10.1016/j.csr.2008.01.020

Raineault, N. (2019). New frontiers in ocean exploration: The e/v nautilus, noaa ship okeanos explorer, and r/v falkor 2018 field season. *Oceanography* 33(1). doi: 10.5670/oceanog.2019.supplement.01

Ramachandra, V. (2019). Weather event severity prediction using buoy data and machine learning. arXiv:1911.09001.

Rasouli, K., Hsieh, W. W., and Cannon, A. J. (2012). Daily streamflow forecasting by machine learning methods with weather and climate inputs. *J. Hydrol.* 414, 284–293. doi: 10.1016/j.jhydrol.2011.10.039

Reidmiller, D. R., Avery, C. W., Easterling, D. R., Kunkel, K. E., Lewis, K. L., Maycock, T. K., et al. (2017). "Impacts, risks, and adaptation in the united states," in *Fourth national climate assessment* (Washington D.C.; U.S. Global Change Research Program), vol. ii.

Roessig, J. M., Woodley, C. M., Cech, J. J., and Hansen, L. J. (2004). Effects of global climate change on marine and estuarine fishes and fisheries. *Rev. fish Biol. fish.* 14, 251–275. doi: 10.1007/s11160-004-6749-0

Roshni, T., Samui, P., and Drisya, J. (2019). Operational use of machine learning models for sea-level modeling. *Indian Journal of Geo Marine Sciences* 48 (09), 1427–1434.

Rosso, I., Mazloff, M. R., Talley, L. D., Purkey, S. G., Freeman, N. M., Maze, G., et al. (2020). Water mass and biogeochemical variability in the kerguelen sector of the southern ocean: A machine learning approach for a mixing hot spot. *J. Geophys. Res.: Oceans* 125, 3. doi: 10.1029/2019JC015877

Rush, S. A., Mordecai, R., Woodrey, M. S., and Cooper, R. J. (2010a). Prey and habitat influences the movement of clapper rails in northern gulf coast estuaries. *Waterbirds* 33, 389–396. doi: 10.1675/063.033.0317

Rush, S. A., Olin, J. A., Fisk, A. T., Woodrey, M. S., and Cooper, R. J. (2010b). Trophic relationships of a marsh bird differ between gulf coast estuaries. *Estuar. Coasts* 33, 963–970. doi: 10.1007/s12237-010-9281-6

Rush, S. A., Soehren, E. C., Woodrey, M. S., Graydon, C. L., and Cooper, R. J. (2009). Occupancy of select marsh birds within northern gulf of mexico tidal marsh: current estimates and projected change. *Wetlands* 29, 798–808. doi: 10.1672/08-174.1

Samantaray, A., Yang, B., Dietz, J. E., and Min, B.-C. (2018). Algae detection using computer vision and deep learning. *arXiv*.

Sarmiento, J. L., Hughes, T.M.C, Stouffer, R. J., and Manabe, S. (1998). Simulated response of the ocean carbon cycle to anthropogenic climate warming. *Nature* 393, 245–249. doi: 10.1038/30455

Scavia, D., Field, J. C., Boesch, D. F., Buddemeier, R. W., Burkett, V., Cayan, D. R., et al. (2002). Climate change impacts on us coastal and marine ecosystems. *Estuaries* 25, 149–164. doi: 10.1007/BF02691304

Schramm, W. (1999). "Factors influencing seaweed responses to eutrophication: some results from eu-project eumac," in *Sixteenth International Seaweed Symposium*. (Springer) 583–592.

Shamsudeen, T. Y. (2020). Advances in remote sensing technology, machine learning and deep learning for marine oil spill detection, prediction and vulnerability assessment. *Remote Sens.* 12, 3416. doi: 10.3390/rs12203416

Shehhi, M. R. A., and Kaya, A. (2020). Time series and machine learning to forecast the water quality from satellite data. *arXiv*.

Solanki, A., Agrawal, H., and Khare, K. (2015). Predictive analysis of water quality parameters using deep learning. *Int. J. Comput. Appl.* 125, 975–8887. doi: 10.5120/ijca2015905874

Solomon, S., Manning, M., Marquis, M., and Qin, D. (2007). "Climate change 2007-the physical science basis," in *Working group I contribution to the fourth assessment report of the IPCC*, vol. Vol. 4. (Cambridge: university press).

Stanic, S., Bernard, L., Delgado, R., Braud, J., Jones, B., Fanguy, P., et al. (2020). The 4-dimension ocean cube training test and evaluation area. *Global Oceans 2020*. doi: 10.1109/IEEECONF38699.2020.9388980

Stephens, M. P., Samuels, G., Olson, D. B., Fine, R. A., and Takahashi, T. (1995). Sea-Air flux of co2 in the north pacific using shipboard and satellite data. *J. Geophys. Res.: Oceans* 100, 13571–13583. doi: 10.1029/95JC00901

Sun, X., Shi, J., Liu, L., Dong, J., Plant, C., Wang, X., et al. (2018). Transferring deep knowledge for object recognition in low-quality underwater videos. *Neurocomputing* 275, 897–908. doi: 10.1016/j.neucom.2017.09.044

Sweet, W. V., Hamlington, B. D., Koop, R. E., Weaver, C. P., Barnard, P. L., Bekaert, D., et al. (2022). *2022: Global and regional sea level rise scenarios for the united states: Updated mean projections and extreme water level probabilities along u.s. coastlines* (Washington D.C.: noaa technical report).

Tian, H., Xu, R., Pan, S., Yao, Y., Bian, Z., Cai, W.-J., et al. (2020). Long-term trajectory of nitrogen loading and delivery from mississippi river basin to the gulf of mexico. *Global Biogeochem. Cycl.* 34, e2019GB006475. doi: 10.1029/2019GB006475

Trenberth, K. E. (2012). Framing the way to relate climate extremes to climate change. *Clim. Change* 115, 283–290. doi: 10.1007/s10584-012-0441-5

Trice, A., Robbins, C., Philip, N., and Rumsey, M. (2021). *Challenges and opportunities for ocean data to advance conservation and management* (Washington DC: Ocean Conservancy).

Tur, R., Tas, E., Haghighi, A. T., and Mehr, A. D. (2021). Sea Level prediction using machine learning. *Water* 13, 3566. doi: 10.3390/w13243566

Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., et al. (1997). Human alteration of the global nitrogen cycle: sources and consequences. *Ecol. Appl.* 7, 737–750. doi: 10.1890/1051-0761(1997)007[0737:HAOTGN]2.0.CO;2

Waddell, E. E., Rasmussen, J. H., and Širović, A. (2021). Applying artificial intelligence methods to detect and classify fish calls from the northern gulf of mexico. *J. Mar. Sci. Eng.* 9, 1128. doi: 10.3390/jmse9101128

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J., et al. (2002). Ecological responses to recent climate change. *Nature* 416, 389–395. doi: 10.1038/416389a

Wang, Z., Boyer, T., Reagan, J., and Hogan, P. (2023). Upper oceanic warming in the gulf of mexico between 1950 and 2020. *J. Climate*, 1–32. doi: 10.1175/JCLI-D-22-0409.1

Wang, J. L., Zhuang, H., Chérubin, L. M., Ibrahim, A. K., and Muhamed Ali, A. (2019). Medium-term forecasting of loop current eddy cameron and eddy darwin formation in the gulf of mexico with a divide-and-conquer machine learning approach. *J. Geophys. Res.: Oceans* 124, 5586–5606. doi: 10.1029/2019JC015172

Wang, J. L., Zhuang, H., Chérubin, L., Muhamed Ali, A., and Ibrahim, A. (2021). Loop current ssh forecasting: a new domain partitioning approach for a machine learning model. *Forecasting* 3, 570–579. doi: 10.3390/forecast3030036

Wells, M. L., Trainer, V. L., Smayda, T. J., Karlson, B. S., Trick, C. G., Kudela, R. M., et al. (2015). Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful algae* 49, 68–93. doi: 10.1016/j.hal.2015.07.009

Worrall, D., and Snelson, S. (1989). "Evolution of the northern gulf of mexico," in *The geology of north america; an overview: Geological society of America* Boulder, (Colorado: The Geological Society of America), 97–138.

Xiang, Y., and Jiang, L. (2009). "Water quality prediction using ls-svm and particle swarm optimization," in *2009 Second International Workshop on Knowledge Discovery and Data Mining.* (IEEE) 900–904.

Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., et al. (2019). A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environ. Model. Softw.* 120. doi: 10.1016/j.envsoft.2019.104502

Xu, Z., and Li, J. (2002). Short-term inflow forecasting using an artificial neural network model. *Hydrol. Processes* 16, 2423–2439. doi: 10.1002/hyp.1013

Yerrapothu, B. T. S. (2021). *Application of machine learning techniques to forecast harmful algal blooms in gulf of mexico* (Hattiesburg MS: The University of Southern Mississippi).

Yñiguez, A. T., and Ottong, Z. J. (2020). Predicting fish kills and toxic blooms in an intensive mariculture site in the philippines using a machine learning model. *Sci. Total Environ.* 707, 136173. doi: 10.1016/j.scitotenv.2019.136173

Yu, P., Gao, R., Zhang, D., and Liu, Z.-P. (2021). Predicting coastal algal blooms with environmental factors by machine learning methods. *Ecol. Indic.* 123, 107334. doi: 10.1016/j.ecolind.2020.107334

Yu, X., Shen, J., and Du, J. (2020). A machine-learning-based model for water quality in coastal waters, taking dissolved oxygen and hypoxia in chesapeake bay as an example. *Water Resour. Res.* 56, e2020WR027227. doi: 10.1029/2020WR027227

Zeng, X., Li, Y., and He, R. (2015). Predictability of the loop current variation and eddy shedding process in the gulf of mexico using an artificial neural network approach. *J. Atmos. Oceanic Technol.* 32, 1098–1111. doi: 10.1175/JTECH-D-14-00176.1

Zhu, B., Bai, Y., Zhang, Z., He, X., Wang, Z., Zhang, S., et al. (2022). Satellite remote sensing of water quality variation in a semi-enclosed bay (yueqing bay) under strong anthropogenic impact. *Remote Sensing* (2022) 14 (3), 550. doi: 10.3390/rs14030550