www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 17(2), 2026, 222-242. ISSN: 2007-1558. https://doi.org/10.61467/2007.1558.2026.v17i2.1291

Artificial Intelligence for Sustainable Ocean Management Using Satellite Data

Jorge A. Ruiz-Vanoye, Ocotlán Díaz-Parra, Francisco R. Trejo-Macotela, Julio C. Ramos-Fernández, Jaime Aguilar-Ortiz, José M. Liceaga-Ortiz-De-La-Peña, Marco Antonio Vera-Jiménez, Julio Cesar Salas López, Juvencio Sebastián Zarazúa Silva

[1] Polytechnic University of Pachuca, Mexico
E-mail: jorgeruiz@upp.edu.mx, ocotlan_diaz@upp.edu.mx, trejo_macotela@upp.edu.mx, jramos@upp.edu.mx, jao@upp.edu.mx, miguel.liop@gmail.com, marcovera@upp.edu.mx, salas@upp.edu.mx, zarazua@upp.edu.mx

Abstract. The integration of artificial intelligence and satellite remote sensing provides an innovative approach to sustainable ocean management. This study demonstrates how oceanographic sensors and AI-driven predictive models enhance the monitoring and governance of Marine Protected Areas (MPAs) and sustainable fishing zones. Multivariate datasets are used to map areas of high primary productivity in the Gulf of Mexico, employing QGIS and ArcGIS for spatial analysis. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) neural networks trained on historical time series forecast ecological risks, including hypoxic zones and harmful algal blooms. Unsupervised clustering and dimensionality reduction identify anomalies relative to natural oceanographic patterns, supporting more adaptive and precautionary ocean governance.
The integration of artificial intelligence and satellite remote sensing provides an innovative approach to sustainable ocean management. This study demonstrates how oceanographic sensors and AI-driven predictive models enhance the monitoring and governance of Marine Protected Areas (MPAs) and sustainable fishing zones. Multivariate datasets are used to map areas of high primary productivity in the Gulf of Mexico, employing QGIS and ArcGIS for spatial analysis. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) neural networks trained on historical time series forecast ecological risks, including hypoxic zones and harmful algal blooms. Unsupervised clustering and dimensionality reduction identify anomalies relative to natural oceanographic patterns, supporting more adaptive and precautionary ocean governance.
Keywords: Artificial Intelligence, Sustainable Ocean Management, Marine Protected Areas, Gated Recurrent Units, Long Short-Term Memory, Satellite Remote Sensing, Oceanographic Modelling.

Article Info
Received Dec 2, 2025
Accepted Jan 4, 2026

1 Introduction

Authors should submit their paper electronically, in either WORD (.doc file). Manuscripts should not exceed 20 pages, including figures, tables, references and appendices.

The current condition of the world's oceans is of significant concern owing to the multitude of global environmental challenges they face. Oceans constitute a critical component of the Earth's system, delivering essential ecosystem services that include climate regulation, oxygen generation, and the sequestration and redistribution of carbon. Nevertheless, they are increasingly subject to a range of anthropogenic and natural pressures that are detrimentally impacting their overall health and resilience:

1. Pollution: Marine pollution from plastics, pesticides and other chemical contaminants constitutes a major environmental threat. Plastic debris poses significant risks to marine fauna, as it is frequently ingested and bioaccumulates along trophic chains, leading to toxic effects and increased mortality rates.
2. Climate change: Global warming has accelerated sea-level rise, altered oceanographic conditions and amplifying extreme phenomena such as storm surges and typhoons. Disruption of the marine carbon cycle has further elevated dissolved carbon dioxide (CO_2) concentrations, driven acidification and reducing the capacity of seawater to support marine life.
3. Overfishing: The overexploitation of marine resources has resulted in critical population declines, attributable to indiscriminate harvesting and pressure on key species. These practices destabilise ecosystems and entail substantial socio-economic repercussions.
4. Coastal flooding: Sea-level rise associated with climate change is intensifying coastal flooding, degrading habitats and displacing human populations, thereby compounding ecological and humanitarian challenges.

To address these challenges, the use of smart tools to enable more efficient and effective ocean management is essential. Artificial Intelligence (AI), in conjunction with satellite data, plays a pivotal role in this endeavour:

- Continuous monitoring: AI-powered sensors and systems enable round-the-clock monitoring of marine health, providing detailed information on pollution levels, water temperature, oxygen concentrations and other critical parameters.
- Event prediction: By analysing historical trends and real-time data, AI supports the forecasting of extreme events such as tsunamis, storm surges and sea-ice collapses, facilitating faster and more effective responses.
- Optimising fishing strategies: AI-driven systems can process data on the distribution and movement of marine species, supporting both artisanal and commercial fishers in maximising sustainable yields while minimising the risk of overfishing.
- Protected area management: Satellite imagery enables the detection of areas vulnerable to pollution or human-induced degradation, supporting the designation of marine reserves where harmful practices can be regulated.

The oceans face significant challenges that require comprehensive management strategies and the innovative application of technologies such as Artificial Intelligence (AI) and satellite data to monitor, analyse and respond effectively to marine crises. These five indicators are standardised metrics established by the climate science community, particularly by the Intergovernmental Panel on Climate Change (IPCC) and agencies such as NOAA and NASA:

1. Atmospheric concentration of CO_2: Measured continuously since 1958 at the Mauna Loa Observatory by Charles David Keeling, this indicator reflects cumulative warming and the effectiveness of mitigation policies (Keeling et al., 1976).
2. Global temperature anomalies: Calculated using datasets such as GISS (NASA) and HadCRUT (Met Office Hadley Centre), this metric monitors variations in global land and ocean mean temperatures relative to a reference baseline period (Hansen et al., 2010).
3. Mean sea level: Derived from tide-gauge records and satellite altimetry since 1993, this indicator reveals the contribution of polar ice melt and thermal expansion resulting from global warming (Church & White, 2011).
4. Frequency and intensity of extreme events: As analysed in IPCC reports (2021), this metric evaluates changes in hurricanes, heatwaves and extreme precipitation, providing insight into the severity of short-term climate impacts.
5. Ocean acidification: As documented by Doney et al. (2009), this indicator measures reductions in marine pH due to CO_2 absorption and its consequences for coral reefs and broader marine ecosystems.

Satellite-Based Data Sources for Monitoring Ocean Health

The Copernicus, Sentinel, MODIS and SeaWiFS satellite programmes are vital components in the global effort to observe and monitor our planet from space. Below is a brief overview of each programme:

- Copernicus Programme: The Copernicus Programme is a European Earth observation initiative that provides global data on environmental, climatic and terrestrial conditions. It offers near real-time information on air pollution, forest fires, ocean monitoring, weather forecasting and more. The programme comprises a series of coordinated satellites delivering continuous and detailed observations. Within this framework, the Sentinel satellite series—particularly Sentinel-3—carries instruments such as the Ocean and Land Colour Instrument (OLCI) and the Sea and Land Surface Temperature Radiometer (SLSTR), providing near-daily global coverage of ocean colour and surface temperatures.

- Sentinel Programme: The Sentinel Programme is a collaborative initiative between the European Space Agency (ESA) and the United States National Oceanic and Atmospheric Administration (NOAA). It includes several satellite series, such as Sentinel-1, which monitors land-surface changes and sea-ice dynamics, and Sentinel-2, which captures high-resolution Earth imagery to support environmental management and agricultural monitoring. Sentinel data are freely available and provide high spatial resolution, with imagery reaching up to 10 metres.
- MODIS (Moderate Resolution Imaging Spectroradiometer): MODIS is a scientific instrument aboard several satellites, most notably NASA's Terra and Aqua. It captures spectral images across multiple bands, enabling the study of diverse terrestrial and atmospheric phenomena. MODIS data are widely used to monitor forest fires, air pollution, ice and snow cover, and for climate studies. It provides high temporal resolution (every one to two days) and spatial resolution (typically up to 500 metres). MODIS sensors on Aqua and Terra also capture multispectral images to derive sea surface temperature (SST) and chlorophyll concentrations in coastal and open-ocean waters.
- SeaWiFS (Sea-Viewing Wide Field-of-View Sensor): SeaWiFS was a NASA-operated satellite launched in October 1997 and placed in science mode in April 1998. It became renowned for providing data on oceanic primary productivity, delivering critical insights into biogeochemical cycles and marine ecosystem dynamics. Two major data versions were released: Version 2 (SeaWiFS-2) in August 2017 and Version 3, launched under the Copernicus Programme. Between 1997 and 2010, SeaWiFS generated essential ocean-colour imagery, mapping global chlorophyll-a concentrations and supporting long-term research in ocean biology and biogeochemistry.

The following outlines how each monitored variable contributes to global understanding of the marine environment:

- Chlorophyll a: Ocean colour sensors—such as OLCI, SeaWiFS and MODIS—estimate chlorophyll concentrations, which indicate the abundance of phytoplankton. This is essential for global monitoring of primary production and algal blooms, offering insights into ecological dynamics and environmental impacts.
- Sea Surface Temperature (SST): Instruments such as SLSTR (on Sentinel 3) and MODIS provide SST measurements with an accuracy of up to 0.3 K. These data are critical for climate research and for advancing our understanding of marine ecological processes.
- Ocean acidification: Satellite data alone cannot directly measure ocean pH. However, initiatives such as OceanSODA MDB integrate satellite derived variables—such as SST and chlorophyll—with in situ carbonate chemistry measurements to assess long term acidification trends.
- Floating plastics: Emerging methods aim to detect macroscopic plastic debris by combining ocean colour data with high resolution satellite imagery. These remote observations are validated through in situ trawl surveys, improving the accuracy of plastic pollution assessments.

The Copernicus Programme complements its suite of satellite products with an extensive database of in situ observations collected from divers, research vessels, profiling floats (such as Argo), dredgers and aircraft. These observations are essential for the calibration and validation of satellite derived data.

The Copernicus Institutional Thematic Assembly Centre (INS TAC) compiles and quality controls essential ocean variable datasets that underpin operational oceanography. Its outputs are integrated with the Global Ocean Observing System (GOOS) and EMODnet networks, ensuring consistency between satellite and in situ measurements.

In real time forecasting systems, satellites provide broad spatial coverage, while in situ sensors deliver precise point based measurements. Both data sources are assimilated into ocean models to improve forecasting accuracy at global, regional and coastal scales.

Advanced approaches now fuse remote sensing and in situ data using machine learning frameworks—such as CLOINet—producing refined three dimensional reconstructions of key ocean variables, including temperature and salinity.

2 Applied AI Modelling Approaches

The utilisation of supervised learning algorithms for the analysis of satellite data represents a valuable avenue for advancing sustainable ocean management. By integrating sophisticated machine learning techniques with satellite-based observations, it is possible to develop predictive models that support informed decision-making in the management of marine ecosystems and resources.

Supervised algorithms enhance our understanding of marine connectivity, which is essential for ecological conservation efforts. For instance, integrating sea surface temperature data with machine learning techniques has proven instrumental in identifying critical marine areas and their dynamic properties, thereby facilitating the designation and management of Marine Protected Areas (MPAs) (Novi et al., 2021). The importance of machine learning is also exemplified by the Global Fishing Watch initiative, which transforms satellite data into actionable insights to combat illegal fishing activities, supporting sustainable fisheries management (Paolo et al., 2024). This constitutes a crucial mechanism for protecting biodiversity and strengthening marine governance frameworks.

Moreover, retrieving ocean subsurface temperatures from surface data using neural networks exemplifies how supervised learning can reconstruct thermal structures that are critical for understanding climate dynamics. Studies have demonstrated the application of these methods to predict subsurface conditions based on satellite-derived parameters such as sea surface temperature (SST) and salinity (Qi et al., 2022; Cheng et al., 2021). Integrating these predictive capabilities into oceanographic models enhances scientific understanding of ocean biogeochemistry and the carbon cycle, both vital for assessing the impacts of climate change (Li et al., 2019).

The potential for long-term predictions of ocean conditions and carbon uptake is another promising area for machine learning applications. Several research efforts have emphasised the use of decadal-scale forecasts to manage living marine resources effectively, including adapting to challenges such as ocean acidification (Wang et al., 2022).

In specific studies, machine learning has also been applied to technical tasks, such as correcting biases in satellite-derived salinity measurements and assimilating these corrected data into ocean models, improving predictive accuracy and operational efficiency (Zavala-Romero et al., 2024). This integration demonstrates the multifaceted approach required for sustainable ocean management, where algorithms function as decision-support systems in operational contexts.

The deployment of convolutional neural networks (CNNs) has demonstrated transformative potential in data assimilation processes for operational oceanography. The ability of CNNs to process and analyse extensive datasets from multiple satellite sources provides a nuanced understanding of sea surface anomalies, which are critical for informing marine resource management practices (Trossman & Bayler, 2022; Prochaska, 2021).

Monitoring the oceans is a crucial task in environmental science, climatology and marine resource management. In this context, supervised machine learning algorithms are emerging as effective tools for analysing and predicting complex oceanic phenomena. The following (Table 1) provides a summary of how these techniques are being applied to ocean monitoring, based on various recent publications.

Table 1. Comparative Table of Supervised Algorithms.

Algorithm	Type	Advantages	Disadvantages	Typical applications
Linear regression	Regression	Highly interpretable, fast	Only linear relationships, sensitive to outliers	Simple trends, continuous predictions
Logistic regression	Classification	Interpretable probabilities, well calibrated	Does not capture complex non-linearities	Binary diagnosis, medical classification
k-NN (k-Nearest Neighbours)	Classification/Regression	Intuitive, non-parametric	Slow with large data sets, sensitive to noise and scale	Recommendation systems, pattern detection
SVM (Support Vector Machines)	Classification/Regression	Effective in high dimensions, customisable kernels	Requires hyperparameter tuning, costly with large datasets	Bioinformatics, text, complex classification
Decision trees	Classification/Regression	Highly interpretable,	Overfitting, high variance	Segmentation, logical decisions

		no normalisation required		
Random Forest	Ensemble	Robust, reduces variance, does not require much tuning	Less interpretable, more computationally expensive	Tabular data, robust classification/regression
Boosting (XGBoost, LightGBM)	Ensemble	Very high accuracy, handles bias and variance	Intensive tuning, overfitting if not regulated	ML competitions, complex structured data
Naive Bayes	Classification	Fast, scalable, good baseline	Assumes independence of variables, low accuracy if not met	Text classification, spam
PLS-R	Regression	Good for data with collinearity, reduces dimension	Less interpretable, sensitive to noise	Chemometrics, bioinformatics
MLP (Multi-Layer Perceptron)	Deep Learning	Captures complex non-linear relationships	Black box, requires large amounts of data and tuning	Vision, time series, tabular classification
LSTM (Long Short-Term Memory)	Deep Learning / Time series	Captures long dependencies over time, handles sequential data	Expensive training, requires a lot of adjustment	Multivariate time series, weather forecasting, oceans, language
GRU (Gated Recurrent Unit)	Deep Learning / Time series	Similar to LSTM but more efficient, fewer parameters	Lower capacity than LSTM for long dependencies	Sequences, voice, IoT sensors
TCN (Temporal Convolutional Network)	Deep Learning / Time series	Processes sequences in parallel, faster training than	RNNs Requires greater architectural design	Time prediction, financial series, anomalies
Transformers (Time Series)	Deep Learning / Sequences	Handles global relationships in long sequences, total parallelisation	Very expensive, requires large datasets	Multivariate time prediction, text, weather, satellite

The use of deep learning algorithms, such as convolutional neural networks, has proven effective for ocean data inference and subgrid parametrisation (Bolton & Zanna, 2019). These methodologies not only capitalise on large oceanographic datasets but also enable robust inference under adverse conditions. For instance, the work of Fasnacht et al. (2022) demonstrates that it is possible to estimate ocean colour information from hyperspectral measurements, even in the presence of clouds and aerosols. This suggests that machine learning can overcome certain limitations inherent in traditional radiative transfer modelling methods.

Supervised algorithms applied to the estimation of ocean mixed-layer depth have shown how satellite and in situ observations can be integrated to provide more accurate estimates of these phenomena (Foster et al., 2021).
Another emerging field is the detection of plastics and other marine pollutants. Jamali and Mahdianpari (2021) present a cloud-based framework that employs multispectral satellite imagery and generative adversarial networks for marine plastic monitoring. This method represents a significant advancement in waste management and the protection of marine ecosystems.
However, it is crucial to approach the development of these models with caution. The review by Gray et al. (2024) warns of the risks of overfitting in models trained on limited datasets, which may result in poor performance when confronted with novel phenomena in a changing climate. Therefore, it is essential to use robust, representative data and to ensure that models are generalisable.

The future outlook for the use of machine learning techniques in marine analysis is encouraging, as noted by Zhang et al. (2024), who highlight the potential of remote sensing data and machine learning in the localisation of marine equipment, suggesting continued growth in its application and effectiveness in ocean science. Monitoring the oceans through unsupervised algorithms has gained importance in the fields of marine science and data analytics. These algorithms are used to detect patterns and anomalies in large volumes of data without the need for predefined labels (Table 2).

Table 2. Comparative Table of Unsupervised Algorithms.

Algorithm	Type	Advantages	Disadvantages	Typical Applications
K-means	Partitional clustering	Simple; efficient on large datasets	Sensitive to initialisation; requires prior choice of k	Customer segmentation; data compression
Hierarchical clustering	Agglomerative/divisive clustering	Does not require pre-set number of clusters; interpretable dendrogram	Computationally expensive; difficult to choose optimal cut	Taxonomies; phylogenetic analysis
DBSCAN	Density-based clustering	Detects arbitrarily shaped clusters; separates noise	Sensitive to ε and minPts; scales poorly in high dimensions	Anomaly detection; geospatial clustering
Gaussian mixture models	Probabilistic clustering	Models cluster overlap; provides membership probabilities	Requires parameter estimation; may converge to local optima	Density estimation; soft clustering
PCA (Principal Component Analysis)	Dimensionality reduction	Linear projection; components are interpretable	Captures only linear relationships; assumes major variance	Visualisation; preprocessing; compression
t-SNE	Non-linear dimensionality reduction	Excellent for 2D/3D visualisation	Computationally expensive; does not preserve global distances; parameter-sensitive	Data exploration; high-dimensional visualisation
Autoencoders	Dimensionality reduction/self-learning	Captures complex non-linearities; tunable with deep networks	Requires large datasets and hyperparameter tuning; black box	Anomaly detection; dimensionality reduction
Self-Organising	Self-organising maps	Mesh-based	Slow in high	Pattern analysis;

Map (SOM)	visualisation; preserves topological relationships	dimensions; complex grid-tuning	cluster visualisation

Unsupervised algorithms are essential tools for interpreting satellite data in ways that support sustainable ocean management. By revealing hidden patterns in complex marine systems, these methods improve the monitoring of ecological changes and inform decisions that promote resilience and long-term environmental sustainability.

The Probabilistic Clustering Model (PCM) has been employed to cluster Argo data in the Northwest Pacific Ocean, revealing how variations in ocean currents, such as the Kuroshio Extension, influence the ocean's vertical structure (Sambe & Suga, 2022).

In the detection and monitoring of harmful algal blooms (HABs), unsupervised classification techniques have shown considerable potential in processing satellite imagery to identify significant ecological changes. Optical approaches evaluating backscatter and absorption anomalies have been integrated with machine learning methods to monitor these blooms (Wolny et al., 2020). Such integration enables timely interventions to mitigate the threats posed by HABs to marine ecosystems and fisheries, thereby supporting sustainability in ocean management.

The application of unsupervised learning also extends to the detection and assessment of oceanic features. For example, researchers have developed algorithms for the automatic detection of ocean fronts using deep learning edge-detection models applied to satellite imagery (Wan et al., 2025).

For biological and biogeochemical monitoring, unsupervised techniques have been used to derive consistent ocean-biological products from various satellite ocean-colour sensors. This approach ensures improved integration and interpretation of satellite-derived data, supporting oceanographic research and marine resource management (Wang et al., 2020). Combining multiple satellite data sources enhances the reliability of assessments of ocean health and productivity.

Furthermore, exploring sediment and seabed characteristics through unsupervised classification of seabed colour from satellite imagery represents another innovative application. By analysing the optical properties of shallow coastal waters, this technique provides baseline data on sediment transport and habitat conditions, informing coastal zone management (Parsons, Amani, Moghimi 2021).

A significant application of unsupervised algorithms, such as k-means, is the classification of patterns in spatial and temporal data. For example, Espinoza-Guillen and Malpartida (2021) employed the k-means clustering algorithm to identify homogeneous regions of carbon monoxide in Metropolitan Lima, demonstrating the ability of these methods to group data according to similarities in the behaviour of measured variables. This approach could be equally adapted to monitor water quality in marine bodies, helping to identify areas with varying levels of pollution or differences in chemical composition, as suggested by Dharmarathne, et al. (2025).

In addition, the research of Rozenstein & Karnieli (2011) highlights the effectiveness of both supervised and unsupervised classification methods in assessing land cover, which can be closely related to the monitoring of marine habitats and the classification of different types of seabed substrates. Using satellite imagery and clustering techniques, it is possible to obtain accurate maps of marine seafloor cover, thereby facilitating the study of temporal changes in coastal ecosystems.

Another relevant example is provided by the study of Martínez-Mora et al. (2022), which recognises the value of unsupervised machine learning algorithms, noting that the ability to detect hidden patterns in complex data can lead to significant discoveries in various applications, including environmental health monitoring in oceans. This type of exploratory analysis is fundamental for revealing time series patterns and anomalous behaviours in marine data.

In practice, the implementation of clustering techniques, such as DBSCAN, as noted in the work of Kavanaugh, Oliver, Chavez & Letelier (2014), offers interesting perspectives for classifying risk levels associated with climatic phenomena and their impact on ecosystems, even though their focus is primarily on migration. By segmenting areas based on specific criteria, vulnerable regions can be identified, enabling researchers to direct conservation and restoration efforts more effectively.

The application of artificial intelligence techniques, specifically deep learning and machine learning methods, has become an essential tool for estimating and analysing critical parameters in environmental studies, such as chlorophyll-a concentration (Chl-a) and sea surface temperature (SST). These methodologies enable the efficient processing of large volumes of data, thereby improving the accuracy of predictive models compared with traditional approaches (Yang et al., 2023; Quang et al., 2021).

Studies employing machine-learning models to estimate Chl-a concentrations have yielded promising outcomes. For instance, Yang et al. (2023) implemented a residual network for Chl-a retrieval from HY-1C satellite data, demonstrating the effectiveness of deep-learning techniques for satellite imagery. Similarly, Ye et al. (2021) used a convolutional neural network to determine Chl-a concentrations in the Pearl River Estuary, illustrating the flexibility and adaptability of AI-based methods across diverse marine environments. Such advances are particularly significant given that chlorophyll is a principal indicator of aquatic-ecosystem health and productivity, often correlating with total suspended solids and water-quality conditions (Keller et al., 2018).

For SST, machine-learning techniques have been devised to extract meaningful patterns from satellite data, enhancing our understanding of ocean dynamics that affect marine ecosystems. These models have applications not only in predicting SST but also in assessing its influence on phytoplankton biomass (Cho et al., 2018).

With regard to TTS, machine-learning applications have facilitated the investigation of sedimentation in water bodies by correlating water-quality data with TTS and Chl-a measurements. By employing techniques such as random-forest regression and support-vector machines (SVMs), researchers have modelled TTS with greater fidelity than previous methods (Silveira Kupssinskü et al., 2020). Data-fusion approaches, which integrate satellite imagery observations and in situ measurements, have proven particularly effective (Jimeno-Sáez et al., 2020).

The effectiveness of these AI techniques resides in their capacity to process large volumes of multifaceted data and to uncover non-linear patterns discernible in the field.

3 Application in Marine Protected Areas and Sustainable Fishing Zones

The relationship between chlorophyll-a (Chl-a), sea surface temperature (SST) and marine pH is essential for understanding marine ecosystems and their response to climatic variations. Chl-a is a key indicator of phytoplankton biomass, which is strongly influenced by SST and ocean pH. Increased total suspended solids (TSS) lead to more pronounced stratification of the water column, reducing vertical mixing and, consequently, nutrient availability in the surface layers, thereby negatively affecting phytoplankton production and Chl-a concentration in the ocean (Raitsos et al., 2011; Wang et al., 2022).

In addition, ocean pH acts as a chemical thermometer of acidification, information that is critical in the context of increasing atmospheric CO2 and its effect on marine ecosystems. Acidification is associated with anthropogenic changes and can enhance eutrophication, affecting the health and sustainability of marine communities, thus influencing the availability of nutrients needed by phytoplankton (Cai et al., 2015). The interrelationship between TSS, pH and CHL implies that a change in one of these parameters can have knock-on effects on the others, underlining the importance of continuous monitoring of these variables in the context of climate change (Wang et al., 2022).

Evidence shows that, as global warming continues, an increase in the frequency and intensity of marine heatwave events (MHWs) is expected, which not only affect SST, but also impact ocean biodiversity and resources that depend on phytoplankton health (Oliver et al., 2018). Projections indicate that these changes may lead to dramatic alterations in planktonic communities, which serve as the basis of the oceanic food chain, representing a significant challenge for the sustainable management of marine resources (Costa & Rodrigues, 2021; Oliver et al., 2018).

The interaction between TSS, pH and chlorophyll-a is therefore fundamental to understanding how marine ecosystems respond to climate variability. The ability of phytoplankton communities to adapt to these changes is crucial not only for the health of marine ecosystems, but also for the food security and livelihoods of populations directly dependent on the ocean (Hartmann, 2015). Integrated monitoring that assesses TSS, CHL and pH is essential for future research and management of marine ecosystems in the context of climate change.
We tackled a major challenge: building a comprehensive, multivariate dataset for the Gulf of Mexico. Each record needed precise coordinates and reliable measurements. Accordingly, we focused on essential oceanographic parameter: Chlorophyll-a

(Chl-a), a cornerstone of climate studies. Chlorophyll-a (Chl-a) is a fundamental pigment for photosynthesis in phytoplankton and is therefore a key indicator of the health of marine ecosystems. Its concentration provides valuable information on primary production and biogeochemical dynamics in the ocean.

First, it is crucial to consider that Chl-a is the principal photosynthetic pigment in phytoplankton, which form the foundation of the marine food web. Graham et al. (2015) discuss how Chl-a concentrations derived from satellite data are essential for inferring iron supply regions in the Southern Ocean. The availability of this nutrient is a key determinant of phytoplankton growth, which in turn affects ocean primary production, as highlighted by Feng et al. (2015).

Furthermore, Rinaldi et al. (2013) emphasise that Chl-a is one of the most validated and widely available ocean colour parameters, preferred over other biological indicators that may be subject to greater uncertainty. This underscores the importance of Chl-a not only as a biological marker but also due to its ease of measurement through remote sensing techniques, enabling large-scale monitoring.

Another relevant aspect of Chl-a is its relationship with climate change and its impact on marine productivity. Roxy et al. (2016) discuss how satellite-measured chlorophyll concentrations can serve as an indicator of primary productivity, which is itself influenced by warming in tropical oceans. Such warming can increase nutrient limitation, potentially affecting phytoplankton negatively and, consequently, the marine food web.

Chl-a patterns also vary significantly in response to environmental drivers such as El Niño, as observed by Park et al. (2011). Their study uses satellite data to analyse the relationship between Chl-a variability and climate events, adding another layer of complexity to ocean health monitoring.

In specific regimes such as the Coral Sea, research by Welch et al. (2015) shows that Chl-a concentrations can define ecological regimes and are essential for comparing marine protected areas. This indicates that Chl-a dynamics not only have biological implications but also important applications in marine resource management. The use of satellite imagery and novel data-retrieval algorithms has improved the accuracy of Chl-a estimates. Clay et al. (2019) evaluate satellite-based tools used to retrieve Chl-a concentrations in different oceanic regions, highlighting their importance for studies in biogeochemical oceanography and climate change.

Satellite services provide complete databases—geo-referenced matrices filled with empty cells or null values—that increase file size without adding useful information. While comprehensive, these datasets are inefficient until we preserve the spatial and temporal integrity of valid data. To remedy this, we devised a step-by-step filter: first, discarding all observations with null values or outside plausible ranges; then, aligning the dates and spans of each variable at every geographic point.

The result is a single, streamlined and powerful CSV file in which each row represents a location in the Gulf by latitude, longitude and date, accompanied by its three cleaned measurements—ready for GIS analysis, statistical modelling or interactive visualisations. Quick to load, this dataset is perfect for extracting trends and patterns. Now, the resulting information flows with consistency and precision, ready to illuminate the Gulf's oceanographic dynamics.

We wrote a Python script that is as elegant as it is robust. In just a few lines, the official copernicusmarine library handles the API calls. With xarray, we open and manipulate the multidimensional variables (Chl-a, SST and pH) as though they were simple DataArrays; pandas then help us convert that data into a tidy, easy-to-merge DataFrame.

The workflow was structured as follows:

1. Initial configuration. The temporal range was defined from 1 June 2023 to 1 June 2025, and a geographical rectangle was established covering latitudes 18°N–30°N and longitudes –98° to –81°.
2. Data download and loading. On a daily basis, the function copernicusmarine.dataset.retrieve() was used to download one NetCDF file per variable. Each file was opened with xarray.open_dataset(), from which the desired layer was extracted, while dates or cells without valid data were discarded.
3. Intelligent filtering. Values that were either null or outside plausible ranges (for example, pH < 7 or > 9) were removed. Temporal and spatial dimensions of the three variables were then harmonised to ensure exact alignment.
4. Final unification. Each cleaned DataArray was converted into a DataFrame using to_dataframe().reset_index(). These were then merged using the pandas.merge() function, aligning CHL, SST and pH by date, latitude and longitude. The final result was stored in a CSV file, ready for GIS analysis or statistical modelling.

The dataset employed for chlorophyll-a was sourced from the Copernicus Marine Environment Monitoring Service under the ID cmems_obs-oc_glo_bgc-plankton_my_l4-gapfree-multi-4km_P1D and is internally designated as CHL; this Level 4, gap-free product provides daily global fields at a 4 km spatial resolution, blending satellite observations with in situ measurements to deliver a consistent time series of surface chlorophyll-a concentrations, which are widely used to monitor phytoplankton biomass, assess primary productivity and evaluate ocean health across the world's seas.

For each variable, an individual pandas DataFrame was created and cleaned using the dropna() method to remove any rows lacking measurements. These cleaned DataFrames were then consolidated into a single, well-structured CSV file.

The resulting file, named golfo_mexico_copernicus.csv, contains only records with valid values and is organised into the following fields:

- **fecha:** the date of the observation (format YYYY-MM-DD)
- **hora:** the exact time of the observation (UTC)
- **lat:** the latitude of the point
- **lon:** the longitude of the point
- **variable:** the name of the variable (one of chlor_a, sst or ph)
- **valor:** the measured or estimated value of the corresponding variable

These fields support direct geospatial analysis in GIS software such as QGIS or ArcGIS, as well as statistical analysis and visualisation in Python, R or Power BI.

The final CSV file occupies approximately 10 GB on disk and encompasses a total of 183,793,215 records, each of which has been verified to contain a valid measurement. Figure 1.

The temporal span of the dataset extends from 1 June 2023 to 1 June 2025, with observations recorded on a daily basis, subject to the availability of each variable's data on any given day.

```
fecha,hora,lat,lon,variable,valor
2023-06-01,00:09:02.605836,18.020832,-96.354164,chlor_a,61.387386
2023-06-01,00:09:02.605836,18.020832,-96.3125,chlor_a,62.8623
2023-06-01,00:09:02.605836,18.020832,-88.104164,chlor_a,5.549336
2023-06-01,00:09:02.605836,18.020832,-88.0625,chlor_a,4.2711163
2023-06-01,00:09:02.605836,18.020832,-88.02083,chlor_a,4.0192857
2023-06-01,00:09:02.605836,18.020832,-87.979164,chlor_a,4.005808
2023-06-01,00:09:02.605836,18.020832,-87.9375,chlor_a,4.300986
2023-06-01,00:09:02.605836,18.020832,-87.89583,chlor_a,0.77184194
2023-06-01,00:09:02.605836,18.020832,-87.854164,chlor_a,0.4857764
2023-06-01,00:09:02.605836,18.020832,-87.8125,chlor_a,0.31505963
2023-06-01,00:09:02.605836,18.020832,-87.77083,chlor_a,0.20763451
2023-06-01,00:09:02.605836,18.020832,-87.729164,chlor_a,0.1557293
2023-06-01,00:09:02.605836,18.020832,-87.6875,chlor_a,0.13009313
2023-06-01,00:09:02.605836,18.020832,-87.64583,chlor_a,0.11150729
2023-06-01,00:09:02.605836,18.020832,-87.604164,chlor_a,0.097730294
2023-06-01,00:09:02.605836,18.020832,-87.5625,chlor_a,0.09096241
2023-06-01,00:09:02.605836,18.020832,-87.52083,chlor_a,0.08321654
2023-06-01,00:09:02.605836,18.020832,-87.479164,chlor_a,0.07808659
2023-06-01,00:09:02.605836,18.020832,-87.4375,chlor_a,0.07432724
2023-06-01,00:09:02.605836,18.020832,-87.39583,chlor_a,0.07243411
```

Fig. 1. The final CSV file.

The application of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures in the context of Marine Protected Areas (MPAs) and sustainable fishing zones represents an emerging area of study within marine science. These recurrent neural network (RNN) architectures are particularly well suited to handling time-series data, which is critical for understanding ecological dynamics, species movements and environmental changes in these regions. LSTMs and GRUs are specialised forms of RNNs designed to capture long-term dependencies in sequential data. LSTMs, introduced by Hochreiter and Schmidhuber, address the vanishing gradient problem associated with traditional RNNs by employing gates that regulate the flow of information through time. GRUs, a simplified variant proposed by Cho et al. (2018), utilise fewer parameters by combining the forget and input gates into a single update gate, often resulting in faster training times and comparable performance (Yuan & Mahmoud, 2020).

Research has shown that LSTMs can effectively model complex relationships in marine environments. Recent studies demonstrate their potential in predicting critical factors such as marine currents, temperature variations and chlorophyll-a concentrations—parameters essential for assessing marine ecosystem health and MPA effectiveness (Ali et al., 2021). GRUs have been similarly employed for predictive tasks, providing efficient processing of the extensive datasets generated by marine monitoring systems.

In marine conservation, understanding species distribution and movement patterns is essential for effective management. The use of LSTM and GRU models facilitates the analysis of spatiotemporal patterns, enabling the identification of critical habitats within MPAs and informing decisions regarding the sustainable use of marine resources. For example, MacKeracher et al. (2018) integrate species movement data with economic factors to improve conservation planning for sharks and rays, an approach that could be enhanced by predictive models based on LSTM or GRU architectures.

The integration of LSTM and GRU architectures can also advance sustainable fisheries management within MPAs. Dharmarathne et al. (2025) highlight the socio-economic benefits of MPAs for small-scale fisheries. Predictive models forecasting fish populations and environmental conditions allow fisheries to adapt their practices to ecosystem health, supporting both conservation and economic viability. Furthermore, data derived from these models can inform stakeholders about best practices for sustainable fishing, optimising efforts within designated zones while protecting vulnerable species (Davies et al., 2021).

Dwyer et al. (2020) emphasise the effectiveness of MPAs in mitigating fishing pressure while enhancing fish populations. Leveraging models such as LSTM and GRU to analyse MPA datasets could provide deeper insights into how these areas act as refuges for diverse species, thereby supporting management strategies aimed at sustaining fisheries and promoting ecological recovery.

The following pseudocode outlines the end-to-end procedure for detecting anomalies in environmental time-series data using both LSTM and GRU architectures. It details the steps for loading and preprocessing multivariate observations (chlorophyll-a,), constructing sliding-window sequences, normalising the data, initialising and training recurrent models, and finally computing error thresholds to flag anomalous events. This structured approach ensures reproducibility and facilitates direct comparison of LSTM versus GRU performance.

```
#### PSEUDOCODE of LSTM y GRU ANOMALIES

BEGIN

1. LOAD DATASET

    - Read environmental dataset containing CHL, SST, and pH.

    - Convert 'date' and 'time' into a single 'datetime' object.

    - Pivot data into rows with columns: [datetime, lat, lon, CHL,
SST, pH].

2. SELECT COORDINATE

    - Identify the most frequently observed (lat, lon).

    - Filter the dataset to keep only records for that location.
```

```
    - Sort the time series in ascending order of datetime.

3. BUILD SEQUENCES

    FOR each timestep t from 0 to (length - window_size - 1) DO

        - Extract the previous 'window_size' values of all variables
→ input_sequence

        - Extract the next value at time t+1 → output_sequence

        - Append input_sequence to X

        - Append output_sequence to Y

    END FOR

4. NORMALISE DATA

    - Apply Min-Max scaling to all input and output values.

    - Split X and Y into training and testing sets.

5. INITIALISE MODELS

    FOR model_type IN [LSTM, GRU] DO

        - Create a recurrent neural network using model_type.

        - Add a dense output layer matching the number of predicted
variables.

        - Compile the model with a suitable loss function (e.g.,
MSE).

        - Train the model using the training data.

        - Predict outputs using the testing data.

    END FOR

6. COMPUTE PREDICTION ERROR

    - Calculate the error for each test sample (difference between
predicted and actual values).

    - Store error values for each model.

7. DETECT ANOMALIES

    - Determine a threshold based on a high quantile (e.g., 95th
percentile) of the errors.
```

```
        FOR each test sample DO

            IF error > threshold THEN

                Label sample as 'anomalous'

            ELSE

                Label sample as 'normal'

            END IF

        END FOR

    8. VISUALISE AND REPORT

        - Plot error values and mark anomaly threshold.

        - Display the number and position of detected anomalies.

        - Compare anomaly detection performance between LSTM and GRU.

    END
```

Both models (Table 3 and Table 4) are recurrent neural networks employed to forecast the evolution of the target variable at selected points across the Gulf of Mexico. The GRU (Gated Recurrent Unit) yields more dynamic forecasts that closely mirror the trends observed in the historical series. For example, the peak value of 1.9437 on 2 June 2025 at the first location suggests a considerable uptick following previously elevated measurements. Its mechanism for selectively retaining and discarding information enables it to capture abrupt changes and seasonal patterns with greater agility.

In contrast, the LSTM (Long Short-Term Memory), despite being architected for extended memory horizons, produces a notably conservative and almost uniform forecast in most instances (0.0777), aside from the first coordinate. This uniformity may result from a training regime that emphasised minimisation of overall error at the expense of oscillatory behaviour, or from hyperparameter settings—such as specific learning rates or cell-state sizes—that induce saturation in its forget gates, thereby dampening sensitivity to moderate fluctuations.

When anticipating sharp variations and pronounced peaks in the variable (for instance, during phytoplankton blooms), the GRU appears to deliver more adaptive and responsive forecasts. Conversely, the LSTM may be preferable if a steadier, more noise-resilient prediction is required, provided that its architectural and training parameters are meticulously tuned.

Table 3. Results of GRU.

Latitude	Longitude	Date	Forecasted value
18.1	-97.5	2025-06-02	1.9437
19.3	−94.7	2025-06-10	0.5817
21.8	−90.4	2025-07-05	1.5452
24.5	−87.2	2025-07-18	0.7884
28.0	−83.5	2025-07-22	0.8173

Table 4. Results of LSTM.

Latitude	Longitude	Date	Forecasted value
18.1	-97.5	2025-06-02	0.8960
19.3	−94.7	2025-06-10	0.0777
21.8	−90.4	2025-07-05	0.0777
24.5	−87.2	2025-07-18	0.0777
28.0	−83.5	2025-07-22	0.0777

The approach begins with the unsupervised identification of both natural patterns and anomalies by combining clustering with dimensionality-reduction techniques. In this workflow, DBSCAN is employed to delineate coherent ecological groupings, while data points that fall outside these clusters are flagged as potential disturbance hotspots. This dual strategy enables the detection of emerging irregularities without the need for prior labelling.

Once the clusters and outliers have been identified, their geographic coordinates and corresponding cluster labels are exported in a format compatible with Google My Maps or QGIS (together with the QuickMapServices plugin). These results are then overlaid upon a high-resolution satellite or bathymetric base map of the Gulf of Mexico. By mapping the clusters in QGIS or Google Maps, it becomes straightforward to verify whether they align with real-world features—such as coastal fringes, upwelling zones or areas at elevated risk—and to guide further ecological or management investigations.

This pseudocode outlines an unsupervised approach for uncovering intrinsic patterns and flagging anomalies within multivariate environmental data. By first harmonising chlorophyll-a (Chl-a), —alongside their spatial and temporal metadata—we then employ dimensionality-reduction techniques to render complex relationships visible. Subsequent clustering (via methods such as DBSCAN) reveals natural groupings, whilst outliers and noise points are identified as potential ecological disturbances. Finally, we annotate each record with its cluster assignment and anomaly status before exporting the enriched dataset for further geospatial analysis or visualisation. This workflow ensures a robust, reproducible pipeline for exploratory data analysis and anomaly detection in marine environments.

```
### Pseudocode: Unsupervised Pattern Discovery and Anomaly Detection

BEGIN

1. LOAD MULTIVARIATE DATA

   - Import dataset containing environmental variables: chlorophyll-
a (CHL).

   - Ensure data includes spatial coordinates (latitude, longitude)
and timestamps.

2. CLEAN AND STANDARDISE DATA

   - Remove missing or invalid values.

   - Standardise variable ranges for comparability.

3. APPLY CLUSTERING ALGORITHM

   - Choose a clustering method:
```

```
        • Use DBSCAN for irregular or noisy ecological patterns.

      - Fit the clustering algorithm to the reduced dataset.

   4. IDENTIFY ANOMALOUS POINTS

      - Mark points that do not belong to any cluster (outliers in
   DBSCAN) or lie far from their assigned group.

      - Consider these as potential ecological disturbances or
   anomalies.

   5. EXPORT RESULTS

      - Attach original latitude and longitude to each sample.

      - Assign cluster labels and anomaly flags to each record.

      - Export the enriched dataset as:

        • CSV or KML file for use in Google My Maps.

        • GeoJSON or shapefile for use in QGIS with QuickMapServices.

   6. OPTIONAL: VISUALISE

      - Generate a map showing cluster groupings and highlight detected
   anomalies.

      - Use colour coding for clusters and a distinct marker for
   anomalies.

   END
```

4 Algorithm Output and Visualisation

Cluster Mapping

The following table summarises the clusters, their mean values, new identifiers and descriptive labels:

Table 5. Cluster Mapping.

Cluster	**Mean value**	**New ID**	**Label**
0	0.511	0	Normal

1	13.211	1	Intermediate
2	23.701	2	Atypical

The DBSCAN algorithm was applied to a large dataset of over seventy million records. A representative subset of 100,254 points was selected to train the model with ε (the neighbourhood radius) set to 0.5 and a minimum of 10 samples per cluster. The training completed in approximately fourteen seconds, revealing three distinct clusters alongside noise points (labelled –1). Cluster 0, with a mean value of 0.511, represents the "Normal" regime where the measured variable remains at background levels. Cluster 1 exhibits a higher mean of 13.211 and is labelled "Intermediate", indicating moderate deviations from the norm. Finally, Cluster 2's mean of 23.701 designates it as "Atypical", capturing extreme outliers or hotspots of disturbance.

This labelling and mapping process facilitates rapid identification of regions requiring further investigation, whether they be stable background conditions, areas of moderate change, or critical anomalies. By overlaying these clusters on a geographical base (for example, a satellite or bathymetric map of the Gulf of Mexico), one can visually inspect whether the "Atypical" points correspond to known ecological events, coastal upwellings, or other phenomena of interest.

DBSCAN Output

Total number of records: 70,178,384
Loading sample for DBSCAN…
→ Sample size: 100,254 points
Training DBSCAN (ε = 0.5, min_samples = 10)…
→ Completed in 14.3 s
→ Clusters found (excluding noise): 3
→ Labels present in sample: [–1 (noise), 0, 1, 2]

The figure 2 presents a comprehensive view of the Gulf of Mexico and its neighbouring shores, overlaid with the red-dot anomalies detected by DBSCAN. It illustrates the full spatial extent of sampling—from the western Yucatán Peninsula around to the Florida Keys and up to the Texas–Louisiana coastline—providing a clear sense of where the most extreme values cluster along the basin's perimeter.

Fig. 2. Overview of the Gulf of Mexico Study Area.

This (figure 3) close-up focuses on the Cuban archipelago, revealing concentrated clusters off the northern coast of Cuba and around Isla de la Juventud. The spatial pattern suggests persistent hotspots of elevated measurements, perhaps linked to regional upwelling zones or riverine outflows influencing chlorophyll or temperature anomalies in these waters.

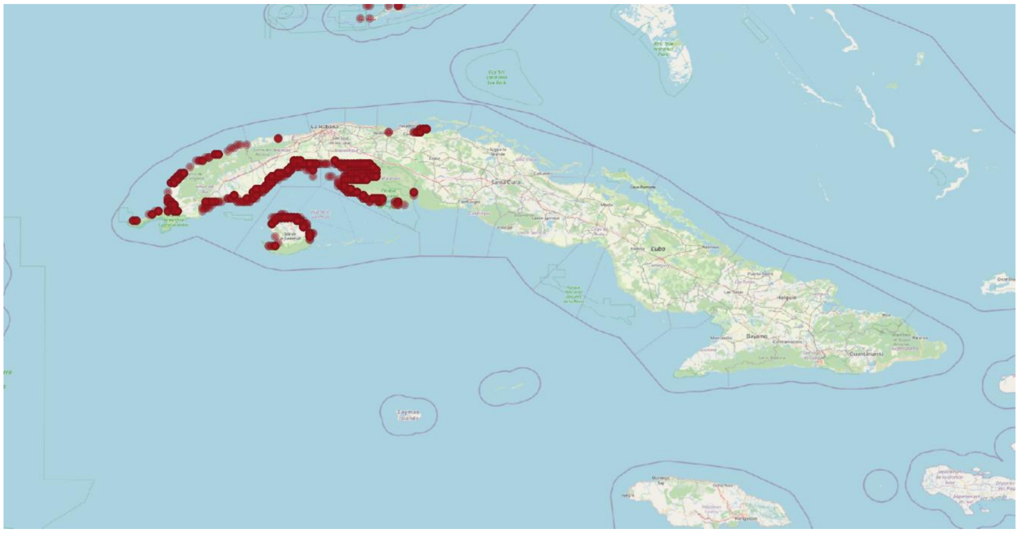

Fig. 3. Detailed View of Cuban Coastal Anomalies.

Figure 4, attention shifts to the Mexican littoral, where red-dot markers trace a near-continuous band of anomalous readings from the Laguna Madre near Tamaulipas down to the Campeche Sound. The density of points reflects both extensive data coverage and recurring perturbations likely associated with coastal shelf dynamics and terrestrial inputs along this stretch.

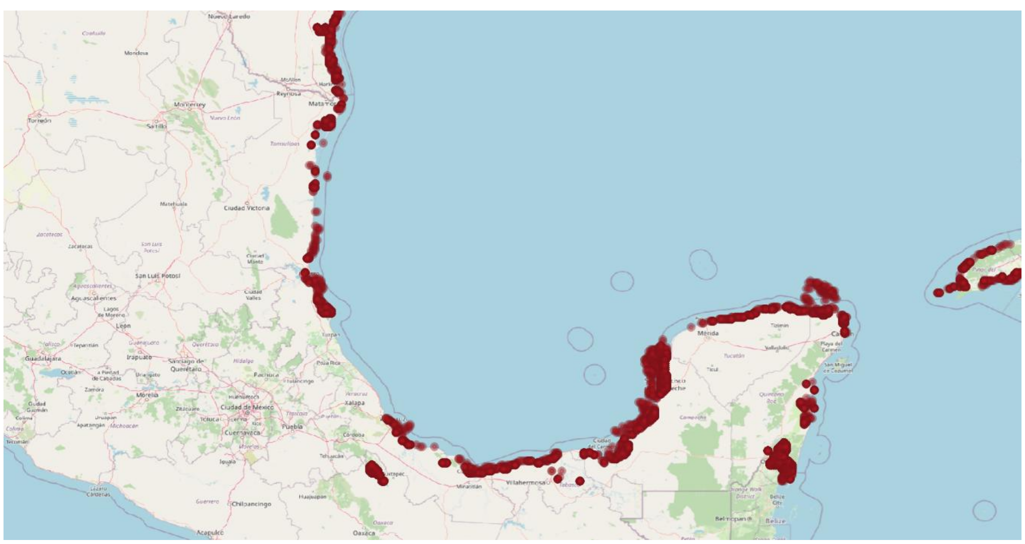

Fig. 4. Mexican Coastline Anomaly Distribution.

The final map zooms in on the US Gulf Coast (figure 5), from Brownsville, Texas, through the Mississippi Delta, and across to Florida’s panhandle and Keys. It highlights regions of particularly intense clustering—such as around Galveston Bay and the Louisiana coast—underscoring zones where environmental drivers (e.g., river discharges, nutrient plumes) may be generating pronounced anomalies.

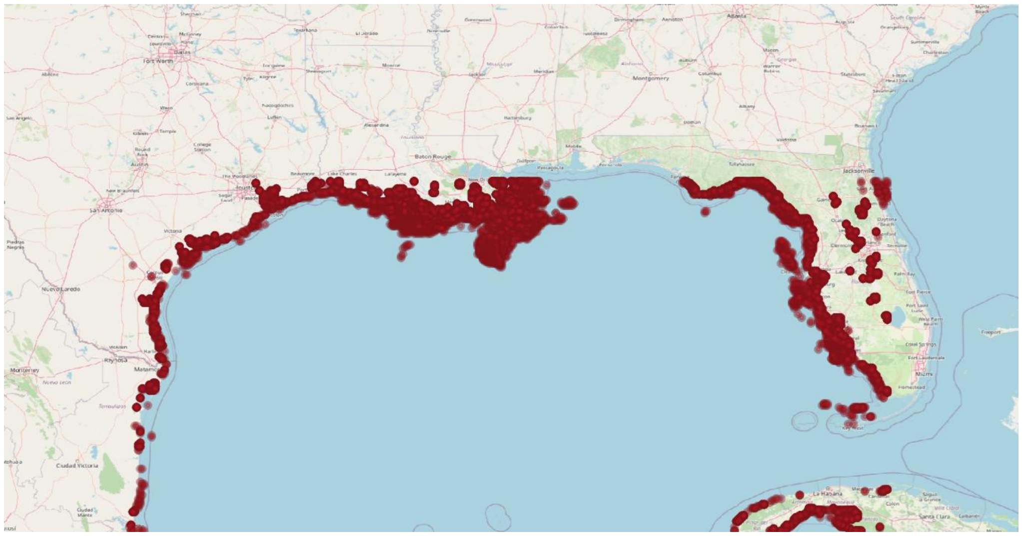

Fig. 5. Gulf Coast of the United States.

5 Conclusions and Future Directions

The successful application of Artificial Intelligence in ocean management represents a significant step towards more informed and sustainable decision-making. In this study, we have demonstrated how satellite data and AI technologies can substantially enhance the monitoring and management of marine resources.

The integration of multivariate satellite-derived data has facilitated the development of precise thematic maps that identify zones of high biological productivity and areas at ecological risk. These tools are essential for informed and responsible decision-making in the context of marine spatial planning and conservation efforts.

The global sharing of marine data and scientific knowledge can significantly enhance our collective understanding of oceanic systems. Strengthening international collaboration in the field of marine artificial intelligence presents an opportunity to advance more coherent and effective policies for the protection and sustainable use of marine resources.

The establishment of ethical standards and regulatory frameworks for the application of AI in marine environments is increasingly urgent. Ensuring that emerging technologies are deployed responsibly requires a coordinated effort among researchers, industry stakeholders, policymakers, and civil society organisations. A shared governance model is necessary to guide the ethical development and deployment of AI in support of ocean sustainability.

As the integration of artificial intelligence and satellite data continues to reshape the landscape of ocean monitoring and management, it becomes increasingly important to chart a course for future action. While recent advances have demonstrated the potential of AI to enhance ecological understanding and support more informed decision-making, further progress depends on sustained investment, institutional coordination, and ethical foresight. The following priority areas outline key directions for advancing the field and ensuring its long-term contribution to ocean sustainability:

- **Augmentation of Capacity Building and Training.** Expanding the training and education of professionals in the use of advanced ocean technologies is vital to fully realise the potential of artificial intelligence in marine contexts. Equipping scientists, technicians, and decision-makers with the necessary digital and analytical skills will ensure that AI-driven tools are deployed effectively and responsibly.
- **Development of World-Leading Technology.** Investment in the research and development of cutting-edge marine AI technologies should be encouraged, with an emphasis on adaptability to diverse oceanographic and socio-environmental conditions worldwide. Promoting innovation in sensor design, data fusion techniques, and predictive modelling will be key to addressing emerging challenges in ocean management.
- **Establishment of Interoperability Networks.** The creation of international and cross-institutional networks to promote the ongoing exchange of ocean data, tools, and expertise is essential. Interoperable systems and open

standards will facilitate collaboration and support the development of globally coordinated ocean governance frameworks.
- **Promoting Sustainable Environmental Policies.** Technological advancements must be implemented alongside robust environmental policies and inclusive management strategies. AI should serve as a tool to strengthen responses to major global challenges, including climate change, ocean acidification, and biodiversity loss, ensuring that innovation contributes to sustainability and long-term ocean resilience.

The application of Artificial Intelligence to ocean governance brings immense potential, but also raises a number of pressing ethical and technical concerns. As marine datasets become increasingly accessible and AI technologies continue to evolve, the need for responsible governance frameworks becomes paramount.

One key ethical dilemma lies in balancing open access to oceanographic data with the protection of strategically sensitive information. While open data promotes collaboration and scientific progress, unrestricted access could expose economically and geopolitically valuable data, posing risks to national security. Clear international agreements must be established to determine which data should remain confidential and how shared resources can be used responsibly for mutual benefit.

On a technical level, there remains a significant digital divide between nations. Countries with advanced infrastructure and expertise are more likely to lead in developing and applying AI for marine management, while others risk marginalisation. In response, it is vital to promote global inclusion through equitable access to technology, transparent validation of predictive models, and targeted capacity-building in AI ethics. Only through coordinated international efforts can marine AI be deployed in a sustainable, fair, and accountable manner.
In the field of ocean monitoring, several ethical and technical challenges arise, particularly concerning open access to maritime data, the technological gap in the use of environmental artificial intelligence (AI), and the validation and auditing of predictive models.

Open access to ocean monitoring data is essential for research and sustainable development. However, it raises concerns about the privacy and security of strategic information. Unrestricted sharing of data carries the risk that sensitive information, such as the locations and conditions of protected areas, could be misused for unethical purposes, including illegal fishing or maritime trafficking. Balancing open access with the need to safeguard sensitive data requires appropriate regulation and management strategies to ensure responsible data sharing, promote scientific collaboration and protect maritime resources.

The technological gap between countries in applying artificial intelligence to address environmental issues is a critical challenge. Many developing countries lack the necessary infrastructure to implement advanced AI solutions, limiting their ability to monitor and protect their ocean resources. This gap is both technical and educational, with limited access to high-quality technological education exacerbating existing disparities. Addressing this issue calls for international cooperation and capacity-building programmes that facilitate access to AI technologies, enabling these countries to engage more effectively in broader conservation efforts.

The validation and auditing of predictive models are essential in the context of ocean monitoring. While these models are valuable, they can produce misleading results if not properly validated, affecting environmental management. A robust framework for model auditing is required to ensure transparency and accuracy. Continuous auditing can help identify non-intuitive or erroneous variables that impact prediction accuracy. The emergence of machine learning techniques has introduced more systematic approaches to model auditing, yet these systems still require rigorous and ongoing evaluation to adapt to new conditions and data, particularly in such a dynamic environment as the ocean.

The ethical and technical challenges in ocean monitoring are multifaceted and interrelated. As technology advances, it is crucial to implement integrative measures that promote equitable access to data and technology, help reduce technological disparities, and ensure the integrity of predictive models. Only through a collaborative and responsible approach can progress be made in protecting our oceans.

References

Bu, X., Liu, K., Liu, J., & Ding, Y. (2023). A harmful algal bloom detection model combining moderate resolution imaging spectroradiometer multi-factor and meteorological heterogeneous data. Sustainability, 15(21), 15386. https://doi.org/10.3390/su152115386

Cai, W., Borlace, S., Lengaigne, M., van Rensch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., McPhaden, M. J., Wu, L., England, M. H., Wang, G., Guilyardi, E., & Jin, F.-F. (2014). Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature Climate Change, 4*, 111–116. https://doi.org/10.1038/nclimate2492

Cheng, H., Sun, L., & Li, J. (2021). *Neural network approach to retrieving ocean subsurface temperatures from surface parameters observed by satellites*. Water, 13(3), 388. https://doi.org/10.3390/w13030388

Cho, K. H., Kang, J. H., Park, Y., Cha, S. M., & Kim, J. H. (2018). *Impact of SST on phytoplankton biomass: A machine-learning approach.* Ecological Modelling, 369, 42–55.

Church, J. and White, N. (2011). sea-level rise from the late 19th to the early 21st century. Surveys in Geophysics, 32(4-5), 585-602. https://doi.org/10.1007/s10712-011-9119-1

Clay, S., Peña, M., DeTracey, B., & Devred, É. (2019). *Evaluation of satellite-based algorithms to retrieve chlorophyll-a concentration in the Canadian Atlantic and Pacific oceans*. Remote Sensing, 11(22), 2609. https://doi.org/10.3390/rs11222609

Costa, N., & Rodrigues, R. (2021). *Future summer marine heatwaves in the western South Atlantic*. Geophysical Research Letters, 48(22). https://doi.org/10.1029/2021GL094509

Davies, B. F. R., Holmes, L., Rees, A., Attrill, M. J., Cartwright, A., & Sheehan, E. V. (2021). *Ecosystem approach to fisheries management works—How switching from mobile to static fishing gear improves populations of fished and non-fished species inside a marine-protected area*. Journal of Applied Ecology, 58(11), 2463–2478. https://doi.org/10.1111/1365-2664.13986

Dharmarathne, G., Abekoon, A. M. S. R., Bogahawaththa, M., Alawatugoda, J., & Meddage, D. P. P. (2025). A review of machine learning and internet-of-things on water quality assessment: Methods, applications and future trends. Results in Engineering, 26, 105182. https://doi.org/10.1016/j.rineng.2025.105182

Doney, S., Fabry, V., Feely, R., & Kleypas, J. (2009). ocean acidification: the other co2problem. Annual Review of Marine Science, 1(1), 169-192. https://doi.org/10.1146/annurev.marine.010908.163834

Dwyer, R. G., Krueck, N. C., Udyawer, V., Heupel, M. R., Chapman, D. D., Pratt, H. L., & Simpfendorfer, C. A. (2020). *Individual and population benefits of marine reserves for reef sharks*. Current Biology, 30(3), 480–489.e5. https://doi.org/10.1016/j.cub.2019.12.005

Feng, J., Durant, J., Stige, L., Hessen, D., Hjermann, D., Zhu, L., … & Stenseth, N. (2015). *Contrasting correlation patterns between environmental factors and chlorophyll levels in the global ocean*. Global Biogeochemical Cycles, 29(12), 2095–2107. https://doi.org/10.1002/2015GB005216

Foster, D., Gagne, D., & Whitt, D. (2021). Probabilistic machine learning estimation of ocean mixed layer depth from dense satellite and sparse in situ observations. Journal of Advances in Modeling Earth Systems, 13(12). https://doi.org/10.1029/2021ms002474

Foster, D., Gagne, D., & Whitt, D. (2021). Probabilistic machine learning estimation of ocean mixed layer depth from dense satellite and sparse in situ observations. Journal of Advances in Modeling Earth Systems, 13(12). https://doi.org/10.1029/2021ms002474

García-Morales, R., García-Gómez, J., & López, P. (2017). *Chlorophyll-a patterns in tropical waters.* Oceanologia, 59(3), 345–358.

García-Morales, R., López-Martínez, J., Váldez-Holguín, J., Herrera-Cervantes, H., & Espinosa-Chaurand, L. (2017). *Environmental variability and oceanographic dynamics of the central and southern coastal zone of Sonora in the Gulf of California*. Remote Sensing, 9(9), 925. https://doi.org/10.3390/rs9090925

Graham, R., Boer, A., Sebille, E., Kohfeld, K., & Schlösser, C. (2015). *Inferring source regions and supply mechanisms of iron in the southern ocean from satellite chlorophyll data*. Deep Sea Research Part I: Oceanographic Research Papers, 104, 9–25. https://doi.org/10.1016/j.dsr.2015.05.007

Hansen, J., Rüedy, R., Sato, M., & Lo, K. (2010). global surface temperature change. Reviews of Geophysics, 48(4). https://doi.org/10.1029/2010rg000345

Hartmann, D. (2015). *Pacific sea surface temperature and the winter of 2014*. Geophysical Research Letters, 42(6), 1894–1902. https://doi.org/10.1002/2015GL063083

Intergovernmental Panel on Climate Change. (2021). *Climate change 2021: The physical science basis. Summary for policymakers*. Cambridge University Press.

Jimeno-Sáez, P., Vicente, V., & Tomás, C. (2020). *Estimating total suspended solids (TSS) using multispectral satellite data.* Environmental Monitoring and Assessment, 192, 340.

Kavanaugh, M. T., Oliver, M. J., Chavez, F. P., & Letelier, R. M. (2014). Hierarchical and dynamic seascapes: A quantitative framework for scaling pelagic ecosystem monitoring. Progress in Oceanography, 120, 291–304.

Keeling, C., Bacastow, R., Bainbridge, A., Ekdahl, C., Guenther, P., Waterman, L., … & Chin, J. (1976). atmospheric carbon dioxide variations at mauna loa observatory, hawaii. Tellus a Dynamic Meteorology and Oceanography, 28(6), 538. https://doi.org/10.3402/tellusa.v28i6.11322

Keller, A., Taylor, D., Oviatt, C., & Kavanaugh, M. (2018). *Chlorophyll as an indicator of ecosystem productivity.* Estuaries and Coasts, 41, 162–176.

Li, H., Ilyina, T., Müller, W. A., & Landschützer, P. (2019). *Predicting the variable ocean carbon sink*. Science Advances, 5(4), eaav6471. https://doi.org/10.1126/sciadv.aav6471

Li, S., Yu, X., & Zhao, H. (2018). *Satellite-based evaluation of marine primary productivity.* Remote Sensing, 10(12), 1923.

Momenimovahed, A., Gagné, S., Martens, P., Jakobi, G., Czech, H., Wichmann, V., … & Thomson, K. (2021). Comparison of black carbon measurement techniques for marine engine emissions using three marine fuel types. Aerosol Science and Technology, 56(1), 46-62. https://doi.org/10.1080/02786826.2021.1967281

Novi, L., Bracco, A., & Falasca, F. (2021). *Uncovering marine connectivity through sea surface temperature*. Scientific Reports, 11(1), 8839. https://doi.org/10.1038/s41598-021-87711-z

Oliver, E. C. J., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., … Wernberg, T. (2018). *Longer and more frequent marine heatwaves over the past century.* Nature Communications, 9, 1324. https://doi.org/10.1038/s41467-018-03732-9

Paolo, F. S., Kroodsma, D., Raynor, J., Hochberg, T., Davis, P., Cleary, J., Marsaglia, L., Orofino, S., Thomas, C., & Halpin, P. (2024). Satellite mapping reveals extensive industrial activity at sea. Nature, 625, 85–91. https://doi.org/10.1038/s41586-023-06825-8

Park, J., Kug, J., Park, J., Yeh, S., & Jang, C. (2011). *Variability of chlorophyll associated with El Niño–Southern Oscillation and its possible biological feedback in the equatorial Pacific*. Journal of Geophysical Research: Oceans, 116(C10). https://doi.org/10.1029/2011JC007056

Parsons, S.; Amani, M.; Moghimi, A. (2021). *Ocean colour mapping using remote sensing technology and an unsupervised machine learning algorithm*. The Journal of Ocean Technology, 16(3), 103–115

Prochaska, X. (2021). Deep learning of sea surface temperature patterns to identify ocean extremes.. https://doi.org/10.5281/zenodo.5537791

Qi, J., Liu, C., Chi, J., Li, D., Gao, L., & Yin, B. (2022). *An ensemble-based machine learning model for estimation of subsurface thermal structure in the South China Sea*. Remote Sensing, 14(13), 3207. https://doi.org/10.3390/rs14133207

Quang, N. H., Nguyen, L. H. K., & Nguyen, T. T. (2021). *Deep learning models for chlorophyll-a estimation using satellite imagery.* Environmental Modelling & Software, 140, 105031.

Raitsos, D.E., Hoteit, I., Prihartato, P., Chronis, T., Triantafyllou, G., & Abualnaja, Y. (2011). Abrupt warming of the red sea. Geophysical Research Letters, 38(14), n/a-n/a. https://doi.org/10.1029/2011gl047984

Rinaldi, M., Fuzzi, S., Decesari, S., Marullo, S., Santoleri, R., Provenzale, A., … & Facchini, M. (2013). *Is chlorophyll-a the best surrogate for organic matter enrichment in submicron primary marine aerosol?* Journal of Geophysical Research: Atmospheres, 118(10), 4964–4973. https://doi.org/10.1002/jgrd.50417

Roxy, M., Modi, A., Murtugudde, R., Valsala, V., Swapna, P., Kumar, S., … & Lévy, M. (2016). *A reduction in marine primary productivity driven by rapid warming over the tropical Indian Ocean*. Geophysical Research Letters, 43(2), 826–833. https://doi.org/10.1002/2015GL066979

Rozenstein, O., & Karnieli, A. (2011). Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. Applied Geography, 31(2), 533–544. https://doi.org/10.1016/j.apgeog.2010.11.006

Silveira Kupssinskü, L., Thomassim Guimarães, T., Menezes de Souza, E., Zanotta, D. C., Veronez, M. R., Gonzaga, L., & Mauad, F. F. (2020). A method for chlorophyll-*a* and suspended solids prediction through remote sensing and machine learning. Sensors, 20(7), 2125. https://doi.org/10.3390/s20072125

Su, H., Xu, H., Liu, G., & Zeng, T. (2021). *Deep-learning-based SST prediction using OLCI data from Sentinel platforms.* Remote Sensing of Environment, 265, 112671.

Trossman, D. and Bayler, E. (2022). An algorithm to bias-correct and transform arctic smap-derived skin salinities into bulk surface salinities.. https://doi.org/10.5281/zenodo.6353521

Wan, X., Zhang, L., Ma, X., Xu, W., Chen, Q., Zhao, R., & Zeng, M. (2025). Dynamic gradient orientation and multi-scale fusion network for ocean front detection. Journal of Sea Research, 206, 102601. https://doi.org/10.1016/j.seares.2025.102601

Wang, D., Fang, G., Jiang, S., Xu, Q., Wang, G., Wei, Z., … & Xu, T. (2022). satellite-detected sea surface chlorophyll-a blooms in the japan/east sea: magnitude and timing.. https://doi.org/10.5194/egusphere-2022-547

Wang, S., Wang, W., Chen, J., Zhang, B., Li, Z., & Jiang, X. (2020). Characteristics of dissolved organic matter and its role in lake eutrophication at the early stage of algal blooms—a case study of lake taihu, china. water, 12(8), 2278. https://doi.org/10.3390/w12082278

Welch, H., Pressey, R., Heron, S., Ceccarelli, D., & Hobday, A. (2015). *Regimes of chlorophyll-a in the Coral Sea: Implications for evaluating adequacy of marine protected areas*. Ecography, 39(3), 289–304. https://doi.org/10.1111/ecog.01450

Yang, G., Ye, X., Xu, Q., Yin, X., & Xu, S. (2023). *Chl-a retrieval using a residual network with HY-1C satellite data.* Remote Sensing, 15(14), 3696. https://doi.org/10.3390/rs15143696

Ye, H., Tang, S., & Yang, C. (2021). *Deep learning for chlorophyll-a retrieval in the Pearl River Estuary.* Remote Sensing, 13(18), 3717. https://doi.org/10.3390/rs13183717

Zavala-Romero, O., Bozec, A., Chassignet, E., & Miranda, J. (2024). Convolutional neural networks for sea surface data assimilation in operational ocean models: test case in the gulf of mexico.. https://doi.org/10.5194/egusphere-2024-1293