POLITECNICO DI TORINO
Repository ISTITUZIONALE

Space–time regression and interpolation of metocean measurements: A focus on satellite data for the offshore energy sector

Original
Space–time regression and interpolation of metocean measurements: A focus on satellite data for the offshore energy sector / Gambarelli, L., Pasta, E., Brandimarte, P., Giorgi, G.. - In: APPLIED OCEAN RESEARCH. - ISSN 0141-1187. - 170:(2026). [10.1016/j.apor.2026.104997]

Availability:
This version is available at: 11583/3008695 since: 2026-03-12T14:19:07Z

Publisher:
Elsevier

Published
DOI:10.1016/j.apor.2026.104997

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

01 November 2026

Applied Ocean Research 170 (2026) 104997



Contents lists available at ScienceDirect

Applied Ocean Research

journal homepage: www.elsevier.com/locate/apor



Review

Space–time regression and interpolation of metocean measurements: A focus on satellite data for the offshore energy sector

Leonardo Gambarelli [a,b,*], Edoardo Pasta [a,b,*], Paolo Brandimarte [c], Giuseppe Giorgi [a,b]

[a] Marine Offshore Renewable Energy Lab., Department of Mechanical and Aerospace Engineering (DIMEAS), Politecnico di Torino, 10129, Corso Duca degli Abruzzi 24, Turin (TO), Italy
[b] MESPAC S.r.l., Via Ottavio Revel 16, 10121, Turin (TO), Italy
[c] Department of Mathematical Sciences (DISMA), Politecnico di Torino, Corso Duca degli Abruzzi 24, Turin (TO), 10129, Italy

ARTICLE INFO

Keywords:
Interpolation
Regression
Machine learning
Satellite measurements
Metocean data
Offshore resource assessment
Environmental monitoring
Offshore energy assessment
Offshore wind energy
Wave energy

ABSTRACT

Ocean phenomena are characterised by a degree of complexity that prevents their effective modelling and simulation, making it necessary to rely on measuring their properties for monitoring and assessments. For these phenomena, obtaining a complete and exhaustive set of measurements is not trivial. Localised information on wind and wave conditions is of paramount importance for the proper design of offshore structures and offshore renewable energy plants. In such cases, a possible way to artificially extend the availability of information is by using regression and interpolation techniques. This review paper formalises the regression and interpolation problems with a focus on the offshore energy sector. It then presents the most common in-situ metocean measuring devices adopted in the offshore applications, followed by a description of the usage of satellites for performing measurements in this field. Subsequently, a brief description of the most adopted regression and interpolation techniques in the sector is provided. We review the state-of-the-art adoption of regression and interpolation techniques for extending measurements of such complex systems, focusing on the application of satellite measurements in the offshore renewable energy sector. The studies are categorised based on the family of regression and interpolation techniques applied, with each category further analysed, compared, and critically assessed. The main findings are reported, highlighting the adoption of machine learning techniques and the combinations of different regression and interpolation methods adopted to tackle the problems related to offshore energy sector resource assessment.

Contents

1. Introduction 2
2. Measuring technologies 4
   2.1. Offshore in-situ measuring devices 6
   2.2. Offshore satellite measurements 6
   2.3. Coastal radars measurements 7
3. Classification of the interpolation and regression techniques 8
   3.1. Traditional interpolation techniques 9
      3.1.1. Deterministic interpolation methods 9
      3.1.2. Statistical, geostatistical and data assimilation methods 10
      3.1.3. Machine-learning based methods 12
   3.2. Typical performance metrics 14
4. Description of the interpolation case studies 15
   4.1. Deterministic interpolation case studies 15
   4.2. Statistical and geostatistical interpolation case studies 15
   4.3. Interpolation through machine learning techniques case studies 17
   4.4. Hybrid techniques and comparison case studies 20
5. Inter-class analysis and critical comparison 21

* Corresponding authors.
E-mail addresses: leonardo.gambarelli@polito.it (L. Gambarelli), edoardo.pasta@polito.it (E. Pasta).

https://doi.org/10.1016/j.apor.2026.104997
Received 10 February 2025; Received in revised form 23 January 2026; Accepted 24 February 2026
Available online 11 March 2026
0141-1187/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

6. Conclusions 24
CRediT authorship contribution statement 25
Declaration of competing interest 25
Acknowledgements 25
References 25

1. Introduction

Nowadays the global energy demand continuously is increasing, as it is increasing the interest in renewable energy sources for supplying such demand. Among the many renewable sources available, offshore energy represent a largely untapped source with highly promising potential (González et al., 2021). In this context, one of the main challenges towards full commercialisation and optimal plant design is represented by the estimation of the potential of deployment sites in terms of energy source (Yavor et al., 2021). Properly scouting the plant deployment locations is fundamental to properly design and size the offshore conversion systems and their supporting structures, minimise the risk of faults, and maximise the energy extraction (Rusu, 2014). This is especially true for hybrid systems such as WEC-type floating breakwaters based on oscillating-buoy and oscillating-water-column principles (Cheng et al., 2022b,a), platform-mounted arrays of point-absorber WECs arranged around a free-floating platform (Cheng et al., 2024b) or single-point–moored cage–WEC hybrids (Cheng et al., 2024a), whose hydrodynamic loads and performance are tightly site-dependent. The analysis conducted during the scouting process necessitates tracking the evolution of metocean parameters of interest for each location within a specified time window. Parameters like the wave height and the wind speeds are fundamental for a correct estimate of the working conditions and the proper design of the plant. Moreover it is also important to assess the variability of the energy production (Widén et al., 2015) and to understand how this variability can change with respect to the different places where the devices are deployed. Such information is fundamental for the proper design of the offshore plant subsystems, such as the support structures (L. Wang et al., 2022), the control strategies (López-Queija et al., 2022; Edoardo Pasta et al., 2023), the mooring systems (Campanile et al., 2018), and the interactions with the electric grid system (Lumbreras and Ramos, 2013; Giglio et al., 2023). Similar issues, related to the availability of metocean data in specific positions, are also relevant in achieving proper environmental monitoring, since the continuous assessment of polluting indicators is fundamental in the detection of unhealthy environment scenarios (Shadrin et al., 2021; Mahmoodi et al., 2019).

To measure the evolution of parameters of interest for the design and operation of offshore installations, in-situ measuring instruments are typically adopted (Murthy and Rahi, 2017; Ma et al., 2019). However, the ocean environment is particularly harsh and challenging for measuring instruments (and installations in general). For this reason, computer-aided tools are typically used to conduct preliminary assessments of wind and wave resources and associated metocean conditions in these environments—for example, by running numerical simulations for offshore wind energy (Yang et al., 2022) and conducting wave-energy assessments (Guillou et al., 2020; Wan Nik et al., 2011). Although it is possible to study the spatio-temporal variability of offshore energy resources through these methods, they are limited by the size of the domain covered and analysed, by the required computational power, by the obtainable resolution (spatial and temporal) and by the accuracy provided (Ashton et al., 2014; Zhou et al., 2015). In this respect, the ocean environment is particularly critical. Indeed, the oceans are too vast and complex to be characterised in their entirety solely through in-situ measurements (the availability whose, therefore, is *sparse* in time and space). Moreover, given the harshness of the oceans, such kind of assessment is not economically viable, due to the high expenses associated to the deployment and constant maintenance of such instrumentation. At the same time, since their dynamics are highly nonlinear and complex, it is difficult to obtain accurate estimates numerically with limited computational effort and uncertainty. Indeed, numerical simulation often give accuracy lower than what is needed for specific engineering applications (Cervelli et al., 2022) and may require *ad hoc* post processing, like Bias-Correction through observation buoys (Penalba et al., 2023).

In this context, satellite observations provide a complementary source of information for metocean assessment rather than a replacement for numerical models. By performing measurements remotely, they can sample large ocean areas and locations where the deployment and maintenance of in-situ instruments would be impractical or too costly. Additionally, satellites are constantly moving their measurement focus, and are not restricted to a single position for recording parameters. Indeed, satellites are already being employed for energy assessments in the offshore energy sector (Alifdini et al., 2019; Goddijn-Murphy et al., 2015). Moreover, satellites have been recognised as a crucial tool for environmental assessments of the ocean in general (Brewin et al., 2023; Chunli Liu et al., 2023), as they facilitate the monitoring and evaluation of trends and variability in microorganisms, such as phytoplankton biomass (Zeng et al., 2022). However, the adoption of satellite measures is not free of drawbacks. Their primary limitations include a lower accuracy compared to traditional in-situ measurements and their dependence on fixed orbital trajectories (Medina-Lopez et al., 2021; Li et al., 2023).

Some applications in the environmental modelling and in the offshore engineering require highly accurate estimates of the metocean conditions, like the planning of the O&M activities and the site selection for the offshore energy systems (Konuk et al., 2023), but the spatial and temporal coverage provided by satellite measurements, even when combined with in-situ measurements, remains inadequate due to their insufficiency and sparseness in both space and time. These variables are needed for the analyses associated with both operational aspects and survivability. Accurate short- to medium-term estimates are needed to plan safe access, vessel limits, and weather downtime for O&M (Micallef and Rezaeiha, 2021; Gao et al., 2022), whereas long, homogeneous time series are required to characterise extremes and joint extremes that drive survivability design and maintenance risk reduction (Martinez-Iturricastillo et al., 2025; Görmüş et al., 2022; Mackay et al., 2021). Given these requirements, in-situ and satellite measurement networks, taken alone, seldom provide records that are simultaneously long, continuous, and spatially representative, and are therefore insufficient to fully characterise the parameters of interest. The only viable approach is to employ mathematical interpolation or regression techniques to extend the limited measurements available across both spatial and temporal dimensions (Gambarelli et al., 2023). Various interpolation techniques exist, each with differing levels of complexity and associated advantages and disadvantages. The optimal choice of an interpolation algorithm depends on numerous factors, which include the specific variable being interpolated and its typical behaviour, the source of the measurements, and the degree of sparsity and uncertainty inherent in these measurements, among others (Li, 2019). Furthermore, the selection of the performance metric used to evaluate the interpolation results can influence the choice of the most suitable interpolators and regressors.

In many practical workflows these interpolation and regression tools do not replace, but rather complement numerical hindcasts, reanalyses and observational products. State-of-the-art wave and atmospheric models already provide basin-scale reconstructions of metocean conditions and are widely used by industry and agencies, but their native

List of abbreviations

Abbreviations

3D-Var	3-Dimensional Variational Approach
4D-Var	4-Dimensional Variational Approach
ANN	Artificial Neural Network
BP ANN	Back-Propagation Artificial Neural Network
CHL-α	Chlorophyll- α
CNN	Convolutional Neural Network
COK	Cokriging
DA	Data Assimilation
DINCAE	Data INterpolating Convolutional Auto-Encoder
DINEOF	Data Interpolating Empirical Orthogonal Functions
DNN	Deep Neural Network
DT	Decision Tree
ELM	Extreme Learning Machine
EnOI	Ensemble Optimal Interpolation
EOF	Empirical Orthogonal Function
ERT	Extremely Randomised Tree
FFNN	Feedforward Neural Network
GA	Genetic Algorithm
GAM	Generalised Additive Model
GAMM	Generalised Additive Mixed Model
GEM	Gravest Empirical Modes
GLM	Generalised Linear Models
GLS	Generalised Least Squares
GPR	Gaussian process regression
GRNN	General Regression Neural Network
IDW	Inverse Distance Weighting
IEW	Inverse Exponential Weighting
KED	Kriging with External Drift
LETKF	Local Ensemble Transform Kalman Filter
LM	Linear Models
LPI	Local Polynomial Interpolation
LRTC-TNN	Low Rank Tensor Completion with Truncated Nuclear Norm
LS	Least Squares
LSTM NN	Long-Short Term Memory Neural Network
ML	Machine Learning
MLR	Multiple Linear Regression
MPGP	Marginalised Particle-filter Gaussian Process
MPNN	Multi Layer Perceptron Neural Network
OBIA	Object-Based Image Analysis
OI	Optimal Interpolation
OK	Ordinary Kriging
POD	Proper Orthogonal Decomposition
PWD	Peak Wave Direction
RBF	Radial Basis Functions
RF	Random Forest
RFRE	Random Forests-based Regression Ensemble
RK	Regression Kriging
RKHS	Reproducing Kernel Hilbert Spaces
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RT	Regression Tree
SAR	Synthetic Aperture Radar
SI	Scatter Index
SOM	Self-Organising Maps
SPGP	Sparse Pseudo-input Gaussian Process
SSGP	Sparse Spectrum Gaussian Process
SSH	Sea Surface Height
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
STK	Spatiotemporal Kriging
SVD	Singular Value Decomposition
SVR	Support Vector Regression
TI	Linear Temporal Interpolation
TIEOF	Tensor Interpolating Empirical Orthogonal Functions
TPS	Thin Plate Spline
UK	Universal Kriging
VE-DINEOF	Variable EOFs-DINEOF
WEC	Wave energy converter
WW3	WAVEWATCH III

spatial and temporal resolutions, coastal representation and residual biases might be insufficient for applications that require device-scale load time series, detailed nearshore gradients or consistent long-term statistics at specific sites. Likewise, fixed buoy networks and short-term deployment campaigns offer highly accurate point measurements but remain sparse in space and fragmented in time, so that many locations and periods of interest for design and planning are not directly sampled. These limitations, and the resulting need to combine numerical models, in-situ instruments and satellite observations to achieve the resolution and accuracy required by end users, are well documented in recent sea-state observing system assessments, such as the one presented in Ardhuin et al. (2019). In this context, interpolation and regression act as a bridge between the available numerical and observational information and the fine-scale, site-specific metocean inputs required by engineering design tools, energy-yield models and planning studies.

In this sense, the review does not propose interpolation and regression as a replacement for high-resolution numerical modelling, but as an alternative and increasingly common workflow: where long-term, device-scale simulations are too costly to run at full physics or across large ensembles, practitioners can combine basin-scale hindcasts, sparse observations and statistical or machine-learning interpolators to derive the metocean inputs needed by design and planning tools. We discuss the advantages of this observational–statistical route (e.g., computational efficiency, direct use of measurements) together with its limitations (e.g., dependence on data availability, challenges in representing rare extremes) relative to purely model-based approaches.

The novelty of this work lies in its comprehensive synthesis of interpolation techniques specifically tailored for offshore energy applications, integrating both satellite and in-situ data to address the spatial and temporal data gaps in metocean measurements with a critical assessment of recent advances in machine learning and data assimilation methods. With a particular focus on the offshore energy sector, this study aims to offer a comprehensive overview of interpolation techniques applied to address the challenges of metocean data surveying using in-situ and satellite measurements. The contributions of this work are as follows:

- A detailed description of in-situ and satellite measurement technologies within the offshore sector, highlighting the opportunities and pitfalls associated with each approach.

- An overview of the most commonly used interpolation techniques, accompanied by a critical classification of the various interpolation case studies tailored for the offshore siting application, and a discussion of the different features considered within the classification.
- An in-depth description and critical analysis of the studies within each class of interpolation algorithms, including a comprehensive comparative assessment of the techniques employed in each study and category.
- A concluding section summarising the main findings and discussing potential future developments in the field of metocean data interpolation and satellite measurement applications in the offshore energy sector.

Beyond a descriptive survey, we use the proposed classification and inter-class comparison to explicitly link each family of methods to the measurement geometry and error structure typical of offshore metocean datasets (sparse space–time sampling, heterogeneous sources, and non-negligible observation uncertainty). This synthesis highlights why purely deterministic interpolators are often too rigid for this setting, why statistical/geostatistical approaches remain a pragmatic baseline thanks to their explicit treatment of measurement errors and uncertainty, and why ML methods are rapidly gaining traction for their ability to capture nonlinear behaviour, leverage many covariates, and naturally exploit image-like satellite products.

From an end-user perspective, this review is intended to function as a methodological interface between metocean data providers and sector-specific tools. Technology developers can use the proposed classification and discussion of methods to select suitable algorithms for generating high-frequency metocean time series at device locations, to be fed directly into codes and processes for hydrodynamic, structural or control design. Resource assessment entities can combine the reviewed interpolation and regression schemes with farm-level energy production models to derive long-term yield estimates and associated uncertainty from heterogeneous data sources. Farm planners and policy-makers, in turn, can exploit the interpolated climate and variability maps produced with these techniques to rank candidate areas, inform zoning and tendering processes, and contextualise environmental monitoring. By summarising both the mathematical foundations and the practical behaviour of each family of methods, the paper aims to help these different communities deploy interpolation and regression tools in a transparent, informed way rather than as opaque black boxes.

Although offshore wind, wave, and tidal applications entail technology-specific engineering analyses, this review targets the technology-agnostic problem of reconstructing metocean variables from sparse in-situ and satellite data. Because the same variables and sampling geometries might recur across the different offshore energy technologies, we compare reconstruction methods and data sources dedicated to those variables rather than restricting the analysis to the ones associated with a single energy technology. The remainder of the paper is organised as follows. Section 2 introduces the most common in-situ (Section 2.1) and satellite sensors (Section 2.2) employed within the offshore sector to obtain information regarding the metocean parameters. Section 3 provides a classification of the most employed interpolation techniques in the metocean field, highlighting the performance metrics employed in the evaluation of their performances (Section 3.2). Section 4 provides a classification of the application studies considered, analysing each of them in terms of approach to the interpolation process, interpolated metocean variable, performance metrics and data sources. Section 5 provides a comparative analysis between the different classified applications, considering their similarities, advantages and pitfalls. Finally, Section 6 presents the paper's findings and discusses future research directions in this field. For clarity and navigation, Fig. 1 summarises the paper roadmap by outlining the main sections and how the classification of methods, the case-study mapping and the inter-class comparison are connected.

2. Measuring technologies

Traditional measuring instruments are typically stationary structures, designed to remain fixed in a single location. For example, specialised devices such as anemometers or windsocks are deployed to measure wind speed and direction (Dabiri et al., 2023). This principle applies similarly to other meteorological parameters, with barometers used for pressure measurements, rain gauges for quantifying precipitation, and pyranometers for assessing solar radiation, among others (Harrison, 2014). These instruments are generally installed at weather stations, where they consistently measure their respective parameters exclusively at that specific location.

In recent years, the advent of satellite technology has introduced the possibility of remote sensing for conducting environmental assessments (Burke et al., 2021; Tymków et al., 2019). Satellites provide a valuable complement to traditional in-situ measurements, as they are not confined to a single spatial point but instead traverse specific trajectories (Dicati, 2017). This allows them to measure parameters of interest at multiple locations, thereby offering data that is significantly less spatially sparse. On the other hand, when dealing with parameters that vary over time as well as space, both satellites (when measuring at a single point at a time) and in-situ devices capture data at only one point in time. Consequently, for dynamic parameters, satellite measurements are also sparse in the space–time domain, albeit with a different pattern of sparsity compared to in-situ devices (Gambarelli et al., 2023). However, this limitation does not apply when satellite measurements involve images, where each pixel can provide a value for the parameter of interest. Indeed, the most traditional use of satellites in remote sensing is through capturing pictures from above. These images can be analysed and refined using image analysis or machine learning techniques to extract the desired parameters (Rolf et al., 2021). In recent years, more advanced techniques have been employed for evaluating parameters through satellites, such as Synthetic Aperture Radar (SAR) systems (Moreira et al., 2013). However, it is important to note that all satellite-based measurements are indirect, and therefore it is unsurprising that their accuracy is typically lower than that of traditional in-situ measurements, which directly measure the parameter of interest. For the most commonly used metocean variables (significant wave height and 10-m wind speed), satellite–buoy collocations indicate satellite RMSE of $H_s \approx 0.2$–0.3 m and $U_{10} \approx 1$–1.5 m s^{-1}, while typical buoy accuracies are about ± 0.2 m for H_s and ± 1 m s^{-1} for U_{10}, with larger gaps in extreme sea states and near coasts (Yang and Zhang, 2019; Driesenaar et al., 2024; NOAA/National Data Buoy Center, 2025).

In the offshore energy sector, the most commonly used measuring instruments to take information on wave conditions are in-situ buoys (Uihlein and Magagna, 2016; Gambarelli et al., 2023; McLeod and Ringwood, 2022), equipped with sensors that record wave elevation over time, enabling the characterisation of the wave spectrum. During the design phase of these devices, the inclusion of a mooring system is typically required, underscoring their intended use as stationary measurement points. Although non-moored drifting buoys also exist, they are less common due to the practical challenges associated with their use (Barbariol et al., 2019). A similar approach applies to offshore wind measurements. It is increasingly common to install anemometers or costly lidars on measurement buoys, allowing for the simultaneous characterisation of both wave and wind resources at a single location.

Another challenge unique to the offshore sector is the harshness of the marine environment, which poses significant difficulties for measurement devices. There are two primary reasons for this challenge. First of all, the occurrence of extreme or highly energetic sea states can cause mechanical damage to the devices. The stress generated by waves can undermine the integrity of the hull or internal components (including measuring instruments) or damage the mooring system (Abdallah El-Reedy, 2020). Secondly, the corrosive nature of seawater can, over time, severely degrade the materials from which

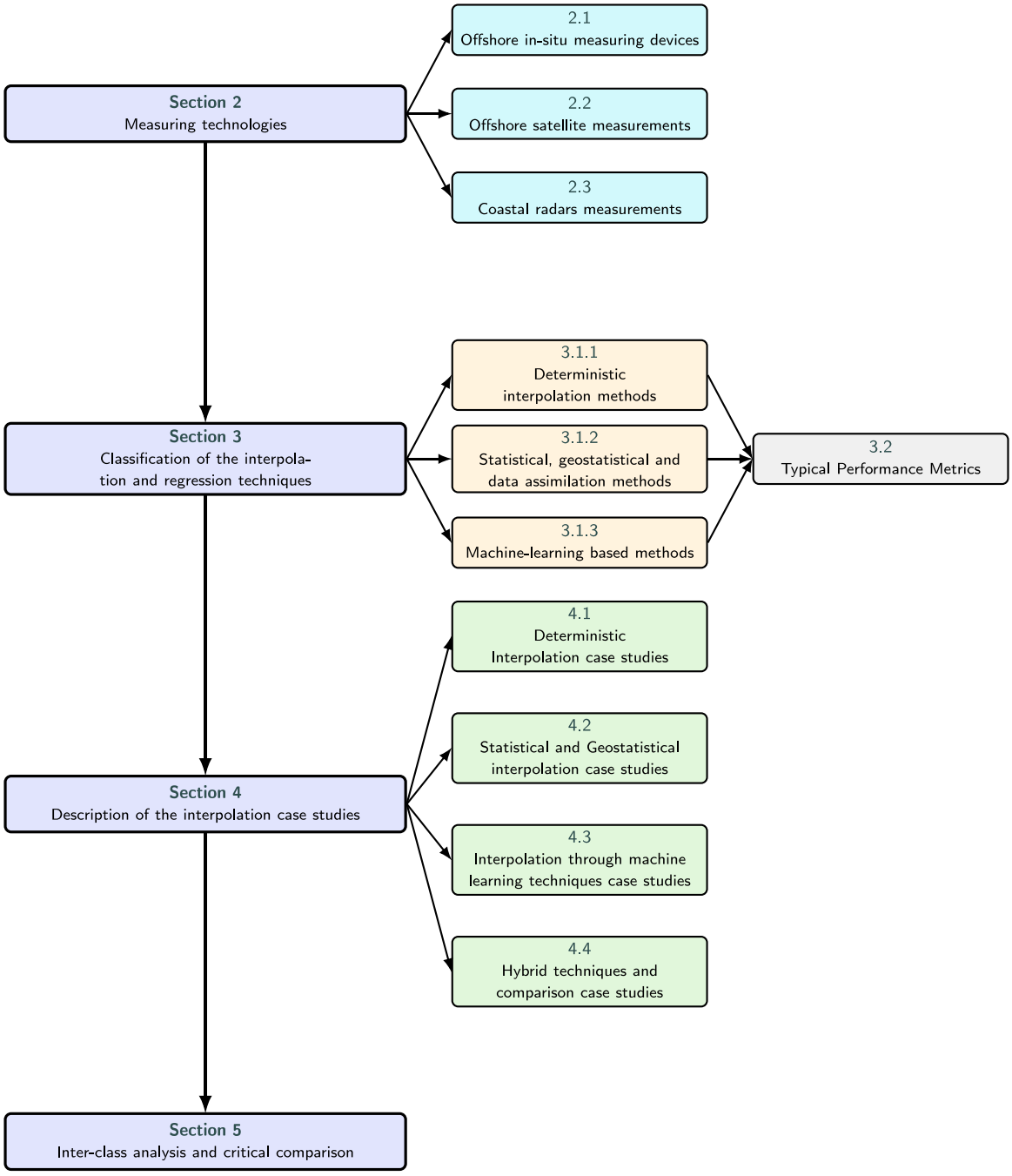


Fig. 1. Schematic roadmap of the review, outlining the main sections and the links between measurement technologies, method-family taxonomy, case-study mapping, and inter-class comparison.

these devices are constructed (Abdallah El-Reedy, 2020). As a result, offshore measuring devices are more prone to failure and require more frequent maintenance compared to onshore devices. Each maintenance operation necessitates reaching the device in the open sea, which can be exceedingly costly (Scheu et al., 2012).

The economic drawbacks associated with in-situ measurement devices in the offshore sector are a key factor contributing to the appeal of satellite measurements in this context. By employing a measuring device located above the ground level, satellite measurements effectively bypass the challenges posed by the harsh marine environment. This approach eliminates the need for devices to be deployed at sea, thereby avoiding the substantial maintenance costs typically incurred in offshore settings.

Alongside in-situ and satellite observations, a further and widely adopted category of metocean sensing in offshore and coastal contexts is represented by shore-based coastal radars (Huang et al., 2017). Installed on land or on coastal structures, these systems perform remote measurements of the sea surface over a fixed observation sector, thereby providing spatially distributed estimates of wave and wind properties with high temporal continuity. Compared to in-situ devices, coastal radars reduce the need for offshore deployment and recurrent open-sea maintenance operations, while still delivering area coverage rather than point measurements. Compared to satellites, they do not provide basin-wide sampling, but they can continuously observe the same region, partially mitigating revisit-time limitations and offering a complementary observing geometry. For these reasons, coastal radars

are often used as an additional data source for nearshore monitoring and, in some applications, as a bridge between sparse in-situ records and large-scale satellite products.

2.1. Offshore in-situ measuring devices

Offshore in-situ measuring devices are highly specialised instruments deployed directly in marine environments to monitor a range of parameters, including wave height, wind speed, and sea temperature. The majority of these devices are deployed as part of national and international initiatives (Centurioni et al., 2019). Numerous other programmes exist globally, many of which focus on specific regions and applications.

From an application perspective, offshore in-situ measuring devices can be classified into four main categories:

- Moored instruments: These are sensors and instruments that are fixed in position, typically anchored to the seafloor or attached to a buoy. Moored instruments are capable of collecting continuous data over extended periods and are widely used for monitoring deep-sea conditions such as temperature, salinity, and ocean currents (Doherty et al., 1999). They are particularly valuable for understanding long-term oceanographic trends and climate-related changes. An example of a programme deploying moored instruments is the Global Tropical Moored Buoy Array (GTMBA) (McPhaden et al., 2023), a network of moored buoys situated in tropical oceans, particularly in the Pacific, Atlantic, and Indian Oceans, to monitor climate-relevant parameters.
- Coastal structure instruments: These devices are typically installed near the shoreline, often attached to piers, jetties, or other coastal structures. They are employed to monitor nearshore and coastal processes, including wave heights, tides, water quality, and beach erosion (Sumer and Whitehouse, 2001). Such instruments are crucial for coastal management and hazard prediction. OceanSITES (Send and Lankhorst, 2011) is a global network of open-ocean and coastal reference stations dedicated to the sustained collection of high-quality oceanographic and atmospheric data, equipped with instruments to monitor a broad spectrum of physical and biogeochemical parameters.
- Drifting instruments: This category includes drifting buoys, floats, platforms, and drifting profilers (Mo et al., 2023). These instruments move with ocean currents, making them ideal for tracking water mass movements and studying ocean circulation patterns. They are often equipped with sensors to measure parameters such as temperature and salinity. Two programmes deploying those devices are the NOAA's Global Drifter Program (GDP) (Elipot et al., 2022), which operates over 1000 drifting buoys globally, designed to follow ocean currents and gather critical data and the Voluntary Observing Ship (VOS) Scheme (Kent, 2010), a programme that engages ships and vessels worldwide to voluntarily collect and report meteorological and oceanographic data to national services.
- Instruments on autonomous vehicles: Examples include autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs), which are mobile platforms equipped with various sensors and instruments for data collection (Sánchez et al., 2020). These vehicles can be programmed to follow specific routes or to explore targeted areas of interest and are commonly used for detailed studies of seafloor geology, marine life, and underwater habitats. The Argo programme (Roemmich et al., 2022) is an international ocean observing system consisting of a network of autonomous robotic floats that collect temperature and salinity data across the world's oceans.

These four categories encompass a broad range of oceanographic and environmental monitoring devices, each offering distinct advantages and applications. Scientists and researchers select the most appropriate category and specific instruments based on their research objectives, the particular marine or coastal environment under investigation, and the required duration of data collection. Fig. 2) summarises the different in-situ instrument categories, providing examples of the typical applications of each class.

Although these in-situ measuring devices are often regarded as the "ground truth" when compared to indirect satellite measurements or numerical simulation results, it is crucial to acknowledge that they, like all measuring instruments, are susceptible to errors. The magnitude of these errors can vary depending on several factors, such as the distance from the coast, the sea state, and the specific parameters being measured (Jensen et al., 2021; Magnusson et al., 2021).

2.2. Offshore satellite measurements

Recent technological advancements have enabled the widespread use of satellites for remote sensing applications. Satellites have been utilised for weather prediction since the late 1960s, and their role has progressively expanded to the point where their data is now indispensable for assessment and monitoring across numerous sectors. A comprehensive historical review of satellite usage, along with the technological innovations from the 1960s up to the present, is detailed in Eyre et al. (2020, 2022). These devices can remotely and indirectly measure various parameters using different physical principles and techniques, each with its own advantages and limitations. The resolution provided by different techniques can vary, as can their ability to penetrate the atmosphere. Below is a list of the most commonly used measurement techniques in the offshore sector, along with a brief description of the underlying physical principles:

- Microwave remote sensing: This technique involves the transmission and reception of microwave signals to acquire information about the Earth surface. The interaction of microwaves with targets, such as ocean surfaces or sea ice, causes variations in signal properties (e.g., intensity, polarisation), which can be analysed to infer characteristics of the observed targets. Due to their ability to penetrate cloud cover, microwaves are particularly effective for data acquisition under diverse weather conditions. Microwave sensors are widely employed to measure sea surface temperature and monitor sea ice extent, providing critical data for both meteorological and environmental studies (Ni et al., 2022).
- Infrared remote sensing: Infrared sensors detect thermal radiation emitted by objects, including the Earth surface. The emission of infrared radiation is dependent on the temperature of the material, with different surfaces emitting distinct radiation patterns. This property allows infrared sensors to accurately determine surface temperatures. In the offshore sector, infrared sensors are predominantly used to measure sea surface temperature by capturing the infrared radiation emitted by the ocean, which is vital for climate studies and weather forecasting (Xing et al., 2023).
- Radar altimeters: Radar altimeters operate by transmitting radar pulses from a satellite, which are reflected back from the Earth surface. By measuring the time delay between the transmission and reception of the radar signal, the distance between the satellite and the surface is calculated, enabling precise measurements of sea surface height. Radar altimetry is a critical technique for monitoring sea level changes, ocean circulation, and other oceanographic phenomena (Nencioli and Quartly, 2019).
- Scatterometers: Scatterometers emit microwave signals towards the Earth surface and measure the backscattered signals to assess surface roughness and wind speed. The scattering characteristics change in response to variations in wind speed, making scatterometry a valuable tool for oceanographic and meteorological

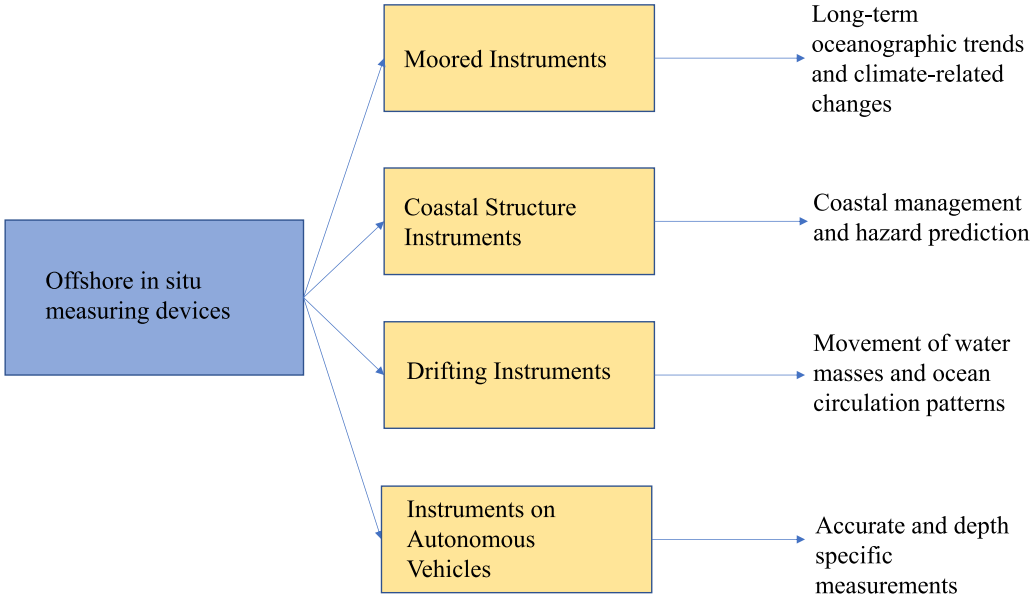


Fig. 2. Classification of the in-situ measuring devices and their main applications.

analysis. Scatterometers are extensively used to measure sea surface winds, which are essential for weather prediction, climate studies, and offshore operations (Surisetty et al., 2020).

- Synthetic Aperture Radar (SAR): SAR is an active radar system that generates high-resolution images of the Earth surface by emitting microwave pulses and recording the backscattered signals. The processing of these signals allows for detailed imaging of surface features, enabling the detection of wind speed, sea ice extent, and other environmental changes. SAR is extensively utilised for applications such as monitoring sea ice, mapping coastal areas, and assessing environmental impacts in marine environments (Moreira et al., 2013; Zen et al., 2021).

As for the in-situ instruments, Fig. 3 summarises the different satellite-based instrument categories, providing examples of the typical applications of each class.

While satellite measurements offer comprehensive spatial coverage and the ability to observe remote and inaccessible areas, they are not without limitations. The accuracy of satellite-derived data can be influenced by factors such as the resolution of the sensor, atmospheric interference, and the algorithms used to correct and interpret the signals (Wang et al., 2024). Additionally, certain parameters might be challenging to measure from space, necessitating the use of models or indirect methods to infer the desired information. Therefore, satellite data often require validation and calibration against in-situ measurements to ensure their reliability and accuracy. Despite these challenges, satellite remote sensing remains a cornerstone of modern oceanographic research, offering valuable insights into large-scale ocean and climate processes that would be difficult or impossible to obtain through in-situ methods alone.

2.3. Coastal radars measurements

In addition to offshore in-situ instruments and satellite remote sensing, coastal radar systems represent a third relevant class of observing technologies for metocean monitoring. These systems are typically land-based (or installed on fixed coastal infrastructure) and therefore avoid many of the deployment and maintenance constraints of offshore platforms, while still providing spatially distributed information over a coastal/ocean domain. Compared with satellites, coastal radars can deliver higher temporal sampling over a fixed region of interest; compared with point-like in-situ sensors, they provide area coverage, albeit limited to coastal ranges and dependent on local geometry and retrieval performance.

The two most common radar-based approaches employed for nearshore/offshore metocean sensing are:

- X-band marine radar: X-band systems (including nautical radars operated from coastal sites or fixed platforms) infer wave properties by processing the radar backscatter modulated by the sea surface. Early studies demonstrated the feasibility of using nautical radar as a wave monitoring instrument (Nieto Borge et al., 1999), and subsequent work has consolidated methodologies and applications for estimating wave and wind-related parameters (e.g., H_s, characteristic periods, wave direction and, in some configurations, wind information) (Huang et al., 2017).
- High-frequency (HF) surface wave radar: HF radars operate at longer wavelengths and exploit surface-wave propagation to sense the ocean over larger ranges, making them well-suited for continuous coastal monitoring. They can retrieve wave information, including the directional wave spectrum via inversion techniques (Wyatt, 1990), and can also provide wind-related estimates, such as sea-surface wind direction in bistatic configurations (Weimin Huang et al., 2012).

Overall, coastal radars provide an intermediate trade-off between satellites and offshore in-situ sensors: they deliver continuous, high-temporal-resolution observations over a fixed coastal domain, with spatial coverage that exceeds point measurements, while remaining constrained by coastal ranges and by the fact that they are indirect sensing systems requiring dedicated processing and validation. Fig. 4 presents these main types of coastal radars, along with the most commonly measured parameters.

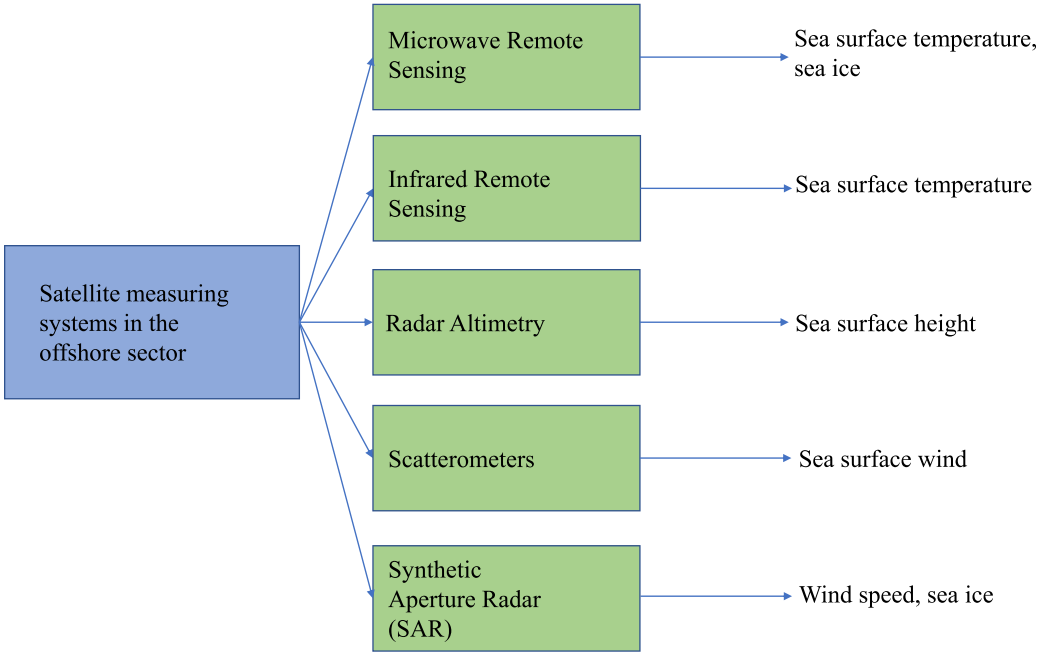


Fig. 3. Classification of typical satellite sensor systems in the offshore sector, along with the most commonly measured parameters.

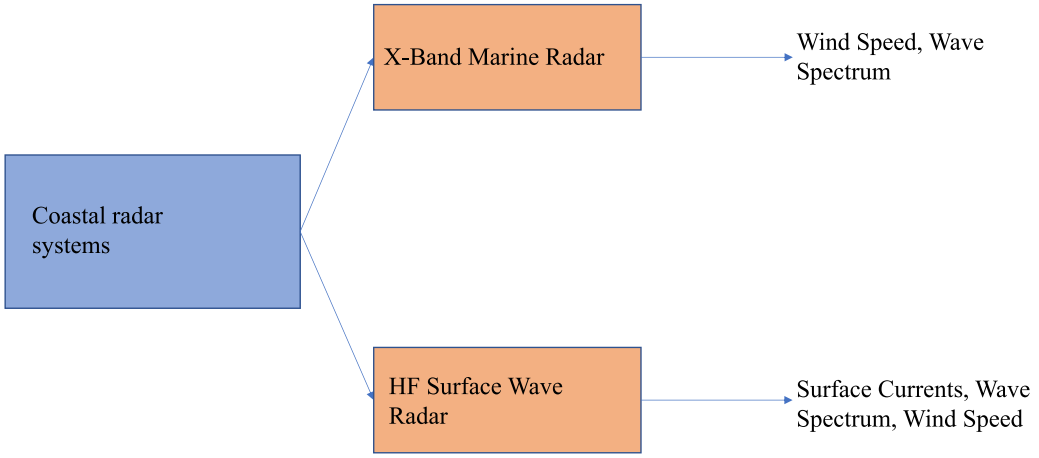


Fig. 4. Classification of typical coastal radar systems, along with the most commonly measured parameters.

3. Classification of the interpolation and regression[1] techniques

Interpolation and regression techniques play a critical role in the offshore energy sector, particularly in the analysis and extension of metocean data. Given the inherent challenges of collecting continuous and comprehensive measurements in marine environments — whether through in-situ devices or satellite observations — reliable interpolation methods are essential for filling gaps in both spatial and temporal data.

The need for interpolation and regression arises from the practical limitations of measuring systems. While in-situ devices offer high accuracy, their data is sparse and localised due to the high costs and maintenance demands of offshore deployment. On the other hand, satellite data, while offering broad spatial coverage, often lack the resolution and precision needed for site-specific applications. Thus, interpolation techniques are used to extend the available datasets and produce estimates for locations and times where direct measurements are unavailable.

Each interpolation and regression technique has its own strengths and limitations, making some methods more suitable for certain applications than others. To identify patterns and insights, key characteristics of the interpolation process have been highlighted for each study that applies these techniques.

[1] In this work, we use the terms interpolation and regression interchangeably, though they are mathematically distinct. Interpolation is generally an exact method that passes through all observed data points, aiming to estimate values within the range of the known data. In contrast, regression allows for errors in measurements by approximating the data, creating a model that may not pass through each point but minimises overall error to capture underlying trends.

First, the domain of interpolation was examined, whether it was applied purely in space, solely in time (such as for forecasting or hindcasting), or simultaneously across both space and time. Additionally, a distinction was made between interpolation based solely on observational data and techniques that integrate a system model, the latter often referred to as Data Assimilation (DA).

Interpolation and regression techniques were then categorised into three primary families. The first includes traditional deterministic mathematical methods, such as linear and spline interpolators, which assume that the parameter is known only at specific points and use predefined functions to estimate values for unknown points. The second category encompasses statistical and geostatistical techniques, such as Kriging methods and various DA approaches like Optimal Interpolation (OI) and three- or four-dimensional variational methods (3D-Var, 4D-Var). These techniques are designed to handle noisy data but often rely on simplifying assumptions about the statistical properties of both the noise and the system being modelled. For example, Kriging methods involve evaluating a variogram to capture spatial autocorrelation, while DA techniques focus on updating numerical models with available measurements. The third category comprises machine learning techniques, including methods like Random Forest (RF) and Neural Networks (NN). These techniques are characterised by the high dimensionality of their input space and the use of numerical optimisation algorithms, such as gradient descent, for tuning parameters. While machine learning methods are theoretically capable of modelling any phenomenon, they often require large amounts of training data and can lack interpretability, making them difficult to debug and troubleshoot. Finally, a fourth category includes hybrid techniques and studies that compare various interpolation methods, combining approaches from different families to develop new solutions.

The studies were further analysed based on the specific interpolation technique used, the parameter being interpolated, and the performance metric employed to assess the results. Different studies used various metrics, such as root mean square error (RMSE) or bias, depending on the nature of the data and the intended application. Finally, attention was given to the type of input data used for interpolation, which generally fell into one of four categories: in-situ measurements, which consist of spatially sparse point data; satellite image-based measurements, where the parameter is measured for each pixel except for areas that require reconstruction due to missing data (e.g., cloud cover); satellite along-track measurements, where data is collected along certain lines with gaps in the remaining areas; and numerical simulations, which provide data for specific points as if measured by instruments at those locations.

Those attributes here presented will be used for constructing the summary tables for all the case studies considered, divided by the family of the used interpolation technique.

3.1. Traditional interpolation techniques

In this section, the most commonly employed interpolation techniques are introduced, accompanied by a concise mathematical overview of their underlying principles. Hereafter, the points where the parameter is known and utilised for interpolation will be referred to as training points, while the points where the value is interpolated will be termed testing points, following a nomenclature typical of machine learning (ML) field. The term output will denote the value of the parameter of interest at a given point, whereas the specific geographical coordinates of that point will be referred to as the input. The estimation of the parameter at an unmeasured location is the output derived by using the spatial coordinates of that location as input.

In line with the classification adopted in this review, the interpolation and regression techniques are grouped into three main families: (i) deterministic methods, which construct the interpolated field from distance-based weights or smoothness constraints without explicitly modelling uncertainty; (ii) statistical and geostatistical methods (including data assimilation techniques), which treat the field as a stochastic process and explicitly incorporate error statistics and spatial/temporal correlations; and (iii) machine-learning-based regressors, which learn flexible (often highly nonlinear) input–output mappings from data and can exploit multiple covariates.

A schematic overview of the methods discussed in this subsection, organised according to these three families and their main characteristics, is shown in Fig. 5.

3.1.1. Deterministic interpolation methods

Deterministic techniques reconstruct the field by enforcing local similarity and smoothness, typically through distance-based weights or analytic basis functions. They do not explicitly describe the parameter as a stochastic process and, in their classical form, provide no formal uncertainty quantification. Nevertheless, their conceptual simplicity, limited number of tunable parameters, and low computational cost make them attractive baseline methods for metocean interpolation, especially when data and computational resources are scarce.

3.1.1.1. Inverse distance weighting. Inverse distance weighting (IDW) is a deterministic interpolation technique where the output at each testing point is determined as a linear combination of the outputs at the training points. The weight assigned to each training point is inversely proportional to the distance between the training and testing points. The underlying concept of this interpolation method is that points in close proximity tend to have similar values, while distant points exert minimal influence on one another (Luo et al., 2008). The primary equation governing this interpolator is:

$$\hat{f}(x_0) = \frac{\sum_{i=1}^{n} w_i \cdot f(x_i)}{\sum_{i=1}^{n} w_i} = \frac{\sum_{i=1}^{n} \frac{1}{d_i^p} \cdot f(x_i)}{\sum_{i=1}^{n} \frac{1}{d_i^p}}, \tag{1}$$

where the function f represents the input–output mapping to be reconstructed, x_i is the ith input training point, x_0 is the input testing point, and $\hat{f}(x_0)$ is the output estimate at the testing point provided by the algorithm. The weight w_i assigned to the ith training point is inversely proportional to $\frac{1}{d_i^p}$, where d_i is the Euclidean distance between x_i and x_0. The parameter p is a tunable aspect of the interpolator, determining the rate at which the correlation between distant points diminishes.

A common variant of this technique is the Inverse Exponential Weighting (IEW) method (Chutsagulprom et al., 2022):

$$\hat{f}(x_0) = \frac{\sum_{i=1}^{n} e^{-\lambda \cdot d_i} \cdot f(x_i)}{\sum_{i=1}^{n} e^{-\lambda \cdot d_i}}, \tag{2}$$

where the weights w_i are determined by the distances in a negative exponential form. The parameter λ serves as the tunable factor that establishes the correlation length. This technique, while easy to implement and computationally efficient, has limitations when applied to metocean parameters. These parameters often exhibit highly complex spatial and temporal variability, influenced by a combination of physical processes such as ocean currents, wind patterns, and wave dynamics. As such, the simplistic distance-based weighting of IDW may struggle to capture these intricate relationships, leading to less accurate interpolations in scenarios involving complex offshore environments.

3.1.1.2. Radial basis functions. Radial Basis Functions (RBF) represent another deterministic interpolation technique, wherein the surface of interest is reconstructed by sequentially placing radial functions over the training points. This process continues until a specified performance metric for the training set is achieved or the maximum number of placeable functions, also known as the Number of Neurons, is reached (this number serves as a tunable parameter).

$$\hat{f}(x_0) = \sum_{i=1}^{n} w_i \cdot \phi\left(\frac{|x_0 - x_i|}{l}\right), \tag{3}$$

where w_i are the weights determined by the algorithm to best match the training points, n is the aforementioned Number of Neurons, ϕ

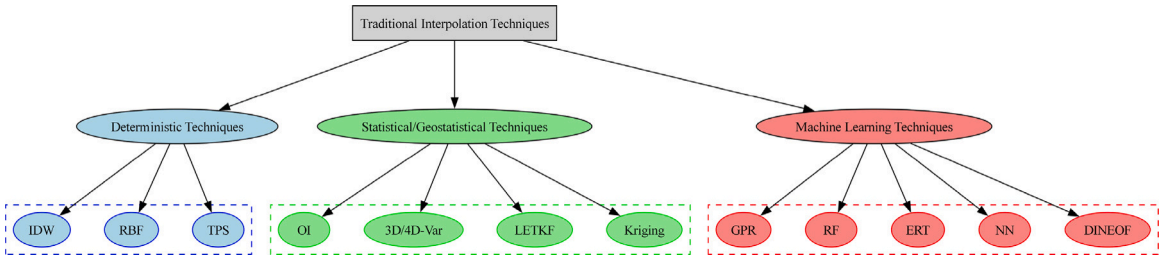


Fig. 5. Scheme of the traditional interpolation and regression techniques considered in this review, grouped into three main families: deterministic, statistical/geostatistical (including data assimilation), and machine-learning-based methods.

represents the selected radial function (termed 'radial' because its input depends solely on the distance between the two points), and l is another tunable parameter known as the spread, which determines the characteristic length of the reconstruction. Some common choices for the ϕ function are the Gaussian function $\phi(r) = e^{-\left(\frac{r}{l}\right)^2}$, the multiquadric function $\phi(r) = \sqrt{r^2 + l^2}$ and the inverse multiquadric function $\phi(r) = \frac{1}{\sqrt{r^2+l^2}}$. RBFs are particularly well-suited for integration with Machine Learning techniques, enabling adaptability to new scenarios (Cavoretto, 2021). However, special precautions must be taken to ensure scalability in large datasets (Esmaeilbeigi et al., 2020). The flexibility provided by the different possible choices for the basis function can be particularly valuable in the offshore energy resource assessment, as it allows the interpolation to adapt to the significantly different behaviours of various metocean parameters.

3.1.1.3. Thin plate spline. The Thin Plate Spline (TPS) is another deterministic interpolation technique that formulates interpolation problems as an optimisation task. The interpolated function is derived by minimising the following functional (Tahir et al., 2023):

$$E_{\text{tps,smooth}}(f) = \sum_{i=1}^{K} |y_i - f(x_i)|^2 + \lambda \iint \left[\left(\frac{\partial^2 f}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2} \right)^2 \right] dx_1 dx_2. \tag{4}$$

In Eq. (4), the first term (the summation) quantifies the *exactness* of the interpolation, essentially representing the sum of the squared residuals at the training points. Conversely, the second term (the integral) reflects the *bending energy* of the interpolated surface. The parameter λ serves as a trade-off factor, determining the relative importance of these two components. TPS offers notable advantages in offshore resource assessment due to its ability to produce smooth, continuous surfaces, making it well-suited for interpolating complex metocean variables like H_s. The balance between interpolation accuracy and surface smoothness ensures that TPS can handle noisy data effectively, which are common in offshore measurements.

3.1.2. Statistical, geostatistical and data assimilation methods

Statistical and geostatistical techniques explicitly represent the parameter of interest as a realisation of an underlying stochastic process. This allows them to incorporate measurement and model errors, spatial and temporal correlations, and, in many cases, to provide uncertainty estimates alongside the interpolated fields. Classical geostatistics (e.g., Kriging) directly model covariance structures, while data assimilation (DA) methods such as Optimal Interpolation, variational schemes and ensemble Kalman filters combine numerical model predictions with observations. These methods are generally more flexible and statistically grounded than deterministic interpolators, at the cost of higher computational complexity and stronger modelling assumptions.

3.1.2.1. Optimal interpolation. Optimal Interpolation (OI) is one of the most widely employed geostatistical techniques for Data Assimilation (DA), particularly in the field of meteorology. DA is the process through which the outputs of a simulated numerical model are updated and corrected using actual measurements of the system. It is well established in this domain that geostatistical techniques demonstrate superior performance when compared to traditional deterministic methods (Li and Heap, 2011), due to their ability to account for spatial correlations and incorporate uncertainty in a more advanced manner. The fundamental concept (Barth et al., 2008; Janjić et al., 2018) is that the system of interest is represented by a state vector x, which encapsulates all the information about the system. However, direct measurement of x is not feasible; instead, access is limited to the observation vector y. Typically, the measurement vector is assumed to result from a linear mapping H applied to the state vector, expressed as $y = Hx$ (where H maps from the state space to the observation space). In a conventional framework, the vector x contains the values of the parameter of interest at the points of a regular grid, while y consists of measurements of this parameter at locations that do not align with the points of the grid. The challenge arises in real-world scenarios, where the actual measurements y^o are affected by noise ϵ. This can be mathematically expressed as follows: starting with an initial estimate for the state vector, denoted as x^b (also referred to as the background state and typically obtained through numerical simulations), one can utilise the information provided by the measurements y^o to update the state vector to x^a (known as the analysis state), ensuring it is as close as possible to the true but unknown state x^t. The relationship can be described by the equations:

$$y^o = Hx^t + \epsilon, \tag{5}$$

$$x^b = x^t + \eta^b, \tag{6}$$

where η^b represents the error associated with the background state. By assuming that the errors ϵ and η^b have zero mean and are uncorrelated, and by knowing their respective covariance matrices P^b and R, one can estimate the true state using the analysis state, as expressed by the following equation:

$$x^a = x^b + K(y^o - Hx^b), \tag{7}$$

where K is the Kalman gain matrix, defined as:

$$K = P^b H^\intercal (H P^b H^\intercal + R)^{-1}. \tag{8}$$

Moreover, with OI, it is possible to derive an uncertainty interval for x^a, which is characterised by the analysis error covariance P^a (the covariance matrix of the analysis error $\eta^a = x^a - x^t$), given by:

$$P^a = P^b - KHP^b. \tag{9}$$

Typically, to ensure that the problem remains tractable, particularly when handling large datasets, all aforementioned covariance matrices are approximated as an identity matrix scaled by a positive scalar. This assumption is equivalent to consider that all realisations of the errors are independent and identically distributed (i.i.d.). OI is particularly

valuable in offshore resource assessment, where it can be used to generate high-resolution spatial maps of key metocean variables, such as wave heights, wind speeds, or sea surface temperatures, by assimilating sparse in-situ measurements and satellite data. This is crucial for characterising resource availability and variability over large offshore areas. However, the assumption of i.i.d. errors can limit the accuracy of OI in regions with complex dynamics, such as coastal zones or areas with strong currents, where error structures may exhibit significant spatial correlation. In such cases, advanced covariance modelling techniques or hybrid methods that combine OI with machine learning approaches could offer improved performance.

3.1.2.2. Three- and four-dimensional variational data assimilation. Three- and Four-Dimensional Variational Data Assimilation (3D-Var and 4D-Var) are more advanced techniques for data assimilation (Hunt et al., 2007). These methods recast the interpolation problem as an optimisation task, where a cost function J is minimised. The cost function is expressed as:

$$J(x^a) = J^b(x^a) + J^o(x^a), \tag{10}$$

with

$$J^b(x^a) = (x^a - x^b)^\intercal (P^b)^{-1}(x^a - x^b), \tag{11}$$

$$J^o(x^a) = (H(x^a) - y^o)^\intercal R^{-1}(H(x^a) - y^o). \tag{12}$$

Here, J^b represents the background term, penalising the deviation between the analysis state x^a and the background state x^b, while J^o is the observation term, penalising the difference between the observation vector y^o and the predicted state $H(x^a)$. Unlike OI, which assumes a linear mapping for H, 3D-Var and 4D-Var allow H to be a nonlinear function. This flexibility enables the assimilation of a wider range of observational data which can often have complex, nonlinear relationships with the model state. However, this comes with the cost of increased computational complexity, as nonlinear mappings require iterative optimisation methods.

The key distinction between 3D-Var and 4D-Var lies in their treatment of time. In 3D-Var, only the three spatial dimensions are considered, treating each time instant independently. This means that an optimisation problem is solved at each time step. In contrast, 4D-Var includes the temporal evolution of the system, solving a single, comprehensive optimisation problem that propagates the uncertainty over time. Although 4D-Var can propagate uncertainty through time more effectively, it requires significantly greater computational resources compared to 3D-Var. These variational techniques are particularly useful when a numerical model provides a reasonably accurate initial guess x^b, which is then refined using available observations y^o. In offshore resource assessment, where metocean parameters exhibit complex spatiotemporal dynamics, 3D-Var and 4D-Var are particularly advantageous when integrating numerical model outputs with observational data. One of their strengths lies in their ability to provide a consistent estimate of the system state by assimilating observations over time (particularly in the case of 4D-Var), effectively reducing bias in the initial model state. However, the effectiveness of these techniques in resolving sudden, extreme events (e.g., storm surges or rogue waves) is more limited, as it depends on both the density of observational networks and the resolution of the underlying numerical model.

3.1.2.3. Local ensemble transform kalman filter. The core concept of the Local Ensemble Transform Kalman Filter (LETKF) is based on the assumption that measurements taken at points in close proximity are likely to be influenced by similar errors (Sluka et al., 2016). In contrast to the previously discussed weakly coupled DA techniques, which assume uncorrelated errors, LETKF is a strongly coupled DA method. Here, the measurement error at one location can be related to the error at another location, meaning that the error covariance matrices are no longer assumed to be diagonal. Unlike a simple scheme such as IDW — which uses distance-based weighting of the field and does not model observation errors — LETKF can explicitly incorporate a non-diagonal observation-error covariance (R), i.e., spatially correlated measurement errors, differing also from weakly coupled DA approaches, which typically assume a diagonal observation-error covariance (i.e., independent errors). Given the high computational complexity associated with considering interrelated errors across different points, a local ensemble is introduced. This ensemble defines a range within which the errors are treated as correlated, helping to manage the computational demands of this approach.

In the LETKF framework, for a given ensemble, the first step involves calculating the mean of the background state, denoted as $\bar{x}^b$, and the mean of the observations, $\bar{y}^o$. Next, the background perturbation matrix X^b is constructed by subtracting the mean from each ensemble member and combining them into a matrix. Defining the measurement perturbation background matrix Y^b as $H(X^b)$, and using $\bar{y}^b = H\bar{x}^b$ for the average background observations, the updated error covariance matrix can be expressed as:

$$\tilde{P}^a = [(k-1)I + (Y^b)^\intercal R^{-1} Y^b]^{-1}. \tag{13}$$

With this matrix, the weights $\bar{w}^a$, used to compute the analysis mean $\bar{x}^a$, can be determined by:

$$\bar{w}^a = \tilde{P}^a (Y^b)^\intercal R^{-1}(y^o - \bar{y}^b), \tag{14}$$

$$\bar{x}^a = \bar{x}^b + X^b \bar{w}^a. \tag{15}$$

The final step is to calculate the weights W^a for deriving the perturbations in the state space of the analysis ensemble X^a:

$$W^a = [(k-1)\tilde{P}^a]^{1/2}, \tag{16}$$

$$X^a = X^b W^a. \tag{17}$$

LETKF is expected to yield more accurate and realistic results compared to other DA methods, as it is a strongly coupled DA approach that directly accounts for the spatial autocorrelation of errors. However, since the inversion of non-diagonal matrices is required, this imposes limitations on the amount of data that can be processed and on the selection of the localisation range. This makes it particularly well-suited for capturing complex interactions in coastal or nearshore environments, where the influence of multiple interrelated factors — such as bathymetry, wind patterns, and tidal currents — plays a significant role. In these regions, where observational errors are often spatially correlated, LETKF provides a more realistic assimilation framework than techniques like OI, which assume uncorrelated errors. However, since the inversion of non-diagonal matrices is required, this imposes limitations on the amount of data that can be processed and on the selection of the localisation range. In practice, the accuracy improvement of LETKF relative to weakly coupled DA is problem-dependent. Gains are larger when observation/background errors are strongly correlated and well represented by the ensemble, and they vary with the measured parameter, the instruments and their sampling, and the operating regime (with noticeable sensitivity during extreme events).

3.1.2.4. Kriging techniques. The term Kriging refers to a family of geostatistical techniques developed in the mid-20th century as empirical methods for analysing the spatial distribution of underground minerals (Degré et al., 2015). In these techniques, the parameter of interest is treated as a stochastic process in space, which is characterised by a mean function and a variogram, both of which must be defined. The form of mean function is typically assumed a priori, with different assumptions leading to different Kriging methods (e.g., Ordinary Kriging (OK) assumes an unknown constant mean, while Universal Kriging (UK) assumes that the mean is a polynomial function of the spatial coordinates).

The variogram $C(h)$ is an empirically derived function of the distance h, such that for each value of h, it represents the average

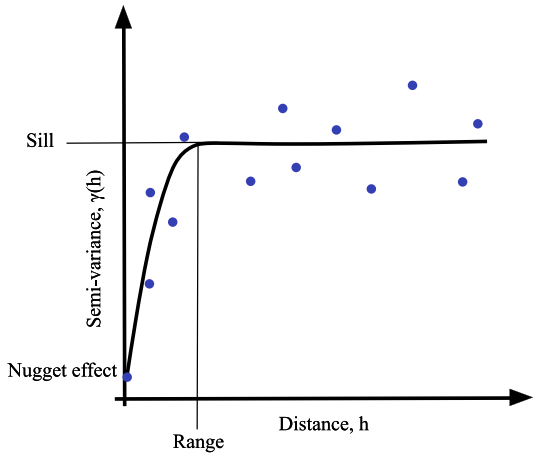


Fig. 6. Example of an empirical variogram with a Gaussian covariance, obtained by fitting a standard variogram model.
Source: Adapted from Degré et al. (2015).

difference between the values of the parameter of interest at two points separated by that distance. In practice, it is nearly impossible to calculate this variogram accurately and comprehensively, as it would require an infinite number of measurements. Instead, an empirical variogram is constructed as a cloud of points, which is then fitted to standard variogram models, as illustrated in Fig. 6.

Even though determining the specific type of variogram model (e.g., Gaussian, Linear, Exponential, Spherical) from a limited set of points can be challenging (Menezes et al., 2008; Setiyoko et al., 2019), three key parameters are typically derived from it:

- The Nugget Effect, representing the intrinsic measurement error (the average difference between measurements taken at two points that are very close together, effectively approximated as co-located).
- The Sill, which corresponds to the overall variance of the process (the average difference between measurements at two points separated by a large distance, theoretically infinite).
- The Range, which defines the maximum distance within which two points are considered correlated (the variogram values are lower than the sill for distances shorter than the range).

The choice of the specific variogram model is typically made through a trial-and-error process, using a held-out set of points for validation. Once the stochastic process is defined with a mean function and a variogram, the value of the parameter of interest at an unmeasured location can be estimated through statistical inference based on the locations where the parameter has been measured. It is also possible to provide an uncertainty interval for the estimate, though this interval should be regarded as a relative indicator of uncertainty, given the simplifying assumptions regarding the statistical distribution (Li and Heap, 2014).

More advanced variants of these methods exist, such as Kriging with External Drift (KED) and Co-Kriging (COK), which enable the use of additional parameters as covariables for the estimations. These techniques are particularly useful when information regarding a parameter related to the original one is available (Yang et al., 2019), although this can present challenges in traditional applications where measurements are typically location-destructive (García-Pérez, 2021). This is highly relevant in offshore resource assessment, where parameters such as wave height, wind speed, and sea surface temperature often exhibit correlated behaviours. For example, Co-Kriging can leverage sea surface temperature as a covariable to improve wave height estimation. However, Co-Kriging typically involves only one covariable, since additional variograms must be calculated for each pair of variable-covariable and covariable-covariable, adding computational complexity. Additionally, Kriging techniques can model anisotropy effects by calculating multiple variograms in different directions (Kim and Deutsch, 2022). This feature is particularly valuable in offshore environments where spatial dependencies may vary over different directions. However, the computational cost of Kriging increases significantly with large datasets, particularly when modelling anisotropy. Hybrid approaches, such as combining Kriging with machine learning techniques, could be explored to enhance scalability without sacrificing accuracy.

3.1.3. Machine-learning based methods

Machine-learning-based methods view interpolation and regression as supervised learning problems, where a flexible model is trained to approximate the mapping from inputs (e.g., spatial coordinates, time, and possibly additional metocean covariates) to outputs (the parameter of interest). In contrast to classical geostatistical techniques, these approaches do not rely on a prescribed variogram or covariance model: instead, they infer spatial and temporal structure directly from the data, often through optimisation of a loss function. These methods can naturally incorporate high-dimensional, heterogeneous inputs and capture strong nonlinearities and complex interactions, at the price of larger data requirements, higher computational costs, and, in many cases, reduced interpretability compared to simpler deterministic or parametric statistical models.

3.1.3.1. Gaussian process regression. Gaussian Process Regressions (GPRs) represent an abstraction and generalisation of Kriging techniques beyond the geostatistical domain, incorporating methods typical of machine learning for parameter tuning (Ebden, 2015; Christianson et al., 2022). The theoretical definition of a Gaussian Process (GP) is a collection of stochastic variables, any finite subset of which follows a multivariate normal distribution (Williams and Rasmussen, 2006). As with Kriging, the parameter of interest is modelled as realisations of a stochastic process, characterised by a mean function, $m(x)$, and a covariance or kernel function, $K(x, x')$ (equivalent to the variogram in Kriging techniques). This approach builds on foundational work in spatial statistics, as detailed in Cressie (1993).

Assuming a constant zero mean function and denoting X as the matrix of all training input points (analogous to the locations of measurements in Kriging), f as the training output values (the actual measurements at the training points), X_* as the testing input points (the points in the input space where we wish to estimate the output), and f_* as the estimated output at the testing points, this estimate follows a normal distribution with mean $K(X_*, X)K(X, X)^{-1}f$ and variance $K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*)$:

$$f_*|f, X, X_* \sim \mathcal{N}\left(K(X_*, X)K(X, X)^{-1}f, K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*)\right). \quad (18)$$

In the equation above, the $\mathcal{N}$ indicates the Normal distribution, while the | indicates a conditioning on the distribution. Not all functions can serve as kernels. However, new kernels can be constructed by taking linear combinations (with non-negative coefficients) of existing kernels. By combining multiple types of kernels, it is possible to capture distinct features of the dataset, with each kernel addressing a specific characteristic (Hamzi et al., 2021). The final kernel may contain a large number of parameters, particularly when the input dimension is high. These parameters are typically determined through iterative machine learning techniques, such as gradient descent, with respect to a chosen performance metric, such as the maximum likelihood estimator (MLE). However, more sophisticated methods for parameter tuning are also available (Manzhos and Ihara, 2023; Abdessalem et al., 2017).

A key advantage of Gaussian Process Regression (GPR) is its versatility. Despite being abstract enough to serve as reliable surrogate models for a wide range of applications (Forrester and Keane, 2009), GPR remains interpretable due to the simplicity of their underlying

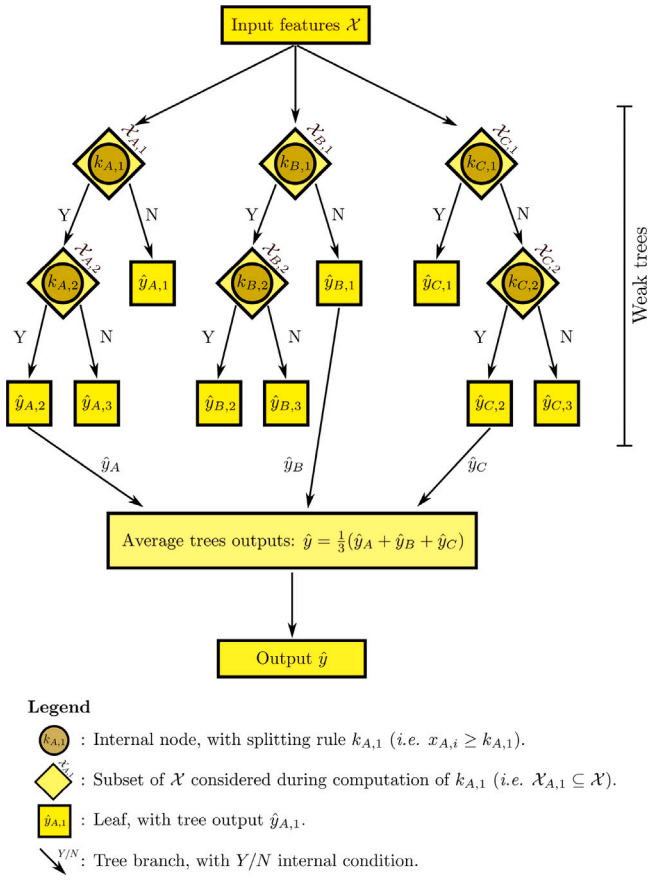


Fig. 7. Schematic representation of a Random Forest (RF), created as an ensemble of three Decision Trees (DTs).
Source: Adapted from E. Pasta et al. (2023).

concept. This interpretability distinguishes them from more complex machine learning techniques, such as neural networks, making it easier to incorporate physical laws and constraints into the models (Setiyoko et al., 2019). GPR's ability to provide both predictions and uncertainty estimates makes it highly suitable for offshore resource assessment, where understanding variability and risk is critical. This capability is particularly valuable in offshore environments, where direct measurements are sparse, and uncertainty quantification is crucial for decision-making. Moreover, the flexibility of kernel selection allows GPR to model different spatial and temporal scales of metocean parameters.

3.1.3.2. Decision tree and Random Forests. A Decision Tree (DT) is a supervised learning technique commonly used for regression tasks. It is represented as a directed graph with a tree-like structure, beginning with a single root node where the input is introduced. Each node processes the input it receives and produces an output based on the input values.

Each node has precisely one incoming edge but can have two or more outgoing edges, resembling the branches of a tree. When the output of a node is computed, it can be sent along different branches if multiple outgoing edges are present. The specific branch is chosen depending on the output itself. Typically, this process involves binary decision-making (e.g., a Yes/No bifurcation), as illustrated in Fig. 7. Additionally, the function governing the input–output relationship at each node usually contains tunable parameters that are calibrated during the algorithm training phase.

To minimise some performance metrics in the results, it is common practice to use an ensemble of uncorrelated decision trees rather than relying on a single tree. This ensemble approach is known as a Random Forest (RF) (Hastie and Tibshirani, 2008). Each individual DT produces an estimate, and the overall RF estimate is obtained by averaging the outputs of the individual trees. RFs are a specific example of bootstrap aggregation techniques, and, as with all bootstrap methods, they provide more robust predictions on unseen data compared to those produced by a single model. DTs and RFs have also been successfully applied to spatio-temporal interpolations (Akter et al., 2021; Requia et al., 2019), often yielding results comparable to, if not better than, traditional and geostatistical methods (Xiao et al., 2021). In offshore resource assessment, DTs and RFs offer significant advantages due to their ability to handle large, complex datasets with high dimensionality. In particular, RFs excel in situations where multiple covariates — such as wind speed, wave height, and sea surface temperature — interact in nonlinear ways. This makes them particularly well-suited for predicting metocean parameters across spatial and temporal scales, where traditional methods may struggle to capture intricate patterns. Another strength of RFs lies in their ability to rank the importance of different predictors, offering valuable insights into the relative influence of various metocean parameters.

3.1.3.3. Extremely randomised trees. Extremely Randomised Trees (ERT), also known as ExtraTrees, is a variant of the Random Forest algorithm, designed to improve efficiency and randomness during the training process. While Random Forest constructs each decision tree by selecting the best split from a random subset of features at each node, ExtraTrees takes this randomness a step further. In ExtraTrees, the splitting thresholds are selected randomly, and not based on the optimisation of some performance metrics (Geurts et al., 2006).

This increased randomness introduces several notable effects. On the positive side, ExtraTrees offers faster training by eliminating the need for computationally expensive split optimisation, making it particularly efficient for large datasets. It also enhances robustness to overfitting, as the additional randomness helps prevent the model from fitting too closely to noisy data. Furthermore, ExtraTrees maintains good generalisation performance by leveraging ensemble averaging, similar to Random Forests, which ensures reliable predictions across various tasks. However, this added randomness can sometimes result in less precise splits, potentially reducing accuracy on highly structured datasets. Additionally, like other ensemble methods, ExtraTrees suffers from limited interpretability, making it more challenging to understand the underlying decision-making process compared to single decision trees. Despite these drawbacks, Extremely Randomised Trees have proven effective in both regression and classification tasks, often matching or surpassing the performance of Random Forests, especially in scenarios that demand faster computation or involve high-dimensional data. In offshore resource assessment, Extremely Randomised Trees are particularly useful for tasks involving large-scale spatial and temporal interpolation of metocean parameters. The faster training time of ExtraTrees allows for rapid model updates as new data becomes available, which is critical in dynamic offshore environments where conditions can change rapidly. Another advantage is their robustness to noise, which is often present in offshore measurements due to sensor inaccuracies, environmental disturbances, or data gaps. ExtraTrees can effectively smooth out these inconsistencies, providing reliable estimates of key parameters such as wave energy potential or wind farm productivity. However, their reduced interpretability can be a drawback in scenarios where understanding the influence of individual predictors is crucial for decision-making, such as optimising the layout of offshore energy infrastructure or assessing the impact of environmental factors on energy production.

3.1.3.4. Neural networks. Similar to tree-based structures, Artificial Neural Networks (ANNs) are computational models represented by a directed graph. However, they differ from DTs in the sense that ANNs are inspired by the structure and functioning of biological brains. The most common type of neural network is the Deep Neural Network (DNN), where the nodes (often called neurons) are organised into layers. In

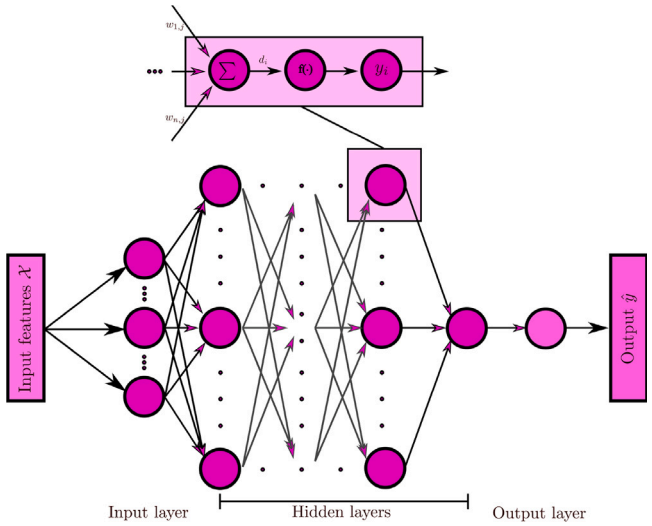


Fig. 8. Schematic representation of a DNN.
Source: Adapted from E. Pasta et al. (2023).

a fully connected network, each node in a given intermediate layer receives inputs from all the nodes in the previous layer and transmits its outputs to all the nodes in the subsequent layer, as depicted in Fig. 8. Each neuron performs a weighted sum of its inputs and then applies an activation function to produce its output. The weights and parameters within the activation function are tunable and must be optimised during the training phase. It can be shown that sufficiently complex ANNs are universal function approximators, meaning they can approximate any arbitrarily complex function (Hastie and Tibshirani, 2008). Beyond DNNs, various other architectures are employed, such as Recurrent Neural Networks (RNNs), where outputs can be fed back to previous neurons. A notable example of this is Long Short-Term Memory Networks (LSTM-NNs), which are particularly useful in forecasting tasks where the system exhibits memory properties (Ma et al., 2019). Other neural network architectures commonly used for spatial predictions and image processing include Convolutional Neural Networks (CNNs) and Encoder–Decoder Networks.

Neural networks offer several advantages and disadvantages when applied to interpolation and regression tasks. On the positive side, they excel at modelling highly complex, non-linear relationships between inputs and outputs, which is particularly valuable in domains like metocean data where interactions can be intricate and multi-dimensional. Additionally, they can generalise well when provided with sufficient training data, making them versatile for both interpolation and forecasting tasks. However, neural networks come with notable challenges. One major drawback is their lack of interpretability (it is often difficult to extract insights on how the model arrives at specific predictions, especially in deep architectures). This can be a limiting factor in fields where understanding the underlying data relationships is crucial. Moreover, they are computationally intensive, requiring significant computational resources and time, particularly for training large models or running hyperparameter optimisation. Neural networks are also data-hungry, often requiring large datasets to prevent overfitting and achieve robust performance. Neural networks have shown great promise in offshore resource assessment, where they can capture the complex, nonlinear interactions between various metocean parameters, such as wave height, wind speed, and ocean currents. However, the computational demands and lack of interpretability remain significant challenges. To address these challenges, hybrid approaches that combine neural networks with more interpretable models or post-hoc explainability tools can be explored.

3.1.3.5. Data interpolating empirical orthogonal functions. Data Interpolating Empirical Orthogonal Functions (DINEOF) is a technique primarily employed for reconstructing images with missing data, such as satellite images partially obscured by clouds. Typically, a collection of images is reshaped into a series of vectors, which are then combined into a single matrix. The mean is computed and subtracted, and the missing pixels are initially assigned a value of zero.

The Empirical Orthogonal Functions (EOFs) of the resulting matrix $X \in R^{NxM}$ are then calculated using Singular Value Decomposition (SVD). Here, N denotes the number of pixels per image, and M represents the number of images. Only the EOFs associated with singular values exceeding a predefined threshold are retained, and these are used to reconstruct an approximation $\tilde{X}$ of the original matrix X, including the previously missing pixels. This process is iteratively repeated on the updated matrix until a specified convergence criterion is satisfied. In matrix notation, the decomposition is expressed as:

$$X = U\Sigma V^{\intercal} \tag{19}$$

$$\tilde{X} = \sum_{i=1}^{k} u_i \sigma_i v_i^{\intercal}, \quad k < \min(N, M), \tag{20}$$

where $\Sigma \in R^{N\times M}$ is a rectangular diagonal matrix containing the singular values of the original matrix X (the eigenvalues of $XX^{\intercal}$), ordered in descending magnitude, with $\Sigma_{ii} = \sigma_i$. The matrix $U \in R^{N\times N}$ consists of the left singular vectors, whose columns $u_i \in R^N$ are the eigenvectors of $XX^{\intercal}$, while $V \in R^{M\times M}$ is the matrix of right singular vectors, with columns $v_i \in R^M$, the eigenvectors of $X^{\intercal}X$.

Typically, the matrix X is arranged such that each column represents all the data points at a specific time instant, and each row contains all the data for a particular point across different time instants. Under this configuration, the matrix U represents the most common spatial patterns within the dataset, arranged from the most dominant to the least, whereas V captures the temporal evolution of these spatial patterns (for instance, v_1 describes how much of the first spatial pattern is present at each time instant) (Kutz and Brunton, 2008). Therefore, the reconstruction using the first k principal singular values is the best rank-k approximation of the matrix X. The ability to separate spatial patterns from their temporal evolution makes this technique particularly well-suited for datasets consisting of multiple images taken over time from the same region. However, since this approach requires flattening the two-dimensional images into one-dimensional vectors, some spatial information is inevitably lost. To account for the inherent two-dimensional structure of images, a more advanced variant called Tensor Interpolating Empirical Orthogonal Functions (TIEOF) has been developed (Kupilik, 2016; Chen et al., 2020). In this method, the aforementioned steps are performed using tensors instead of conventional two-dimensional matrices, preserving more of the spatial structure. DINEOF is particularly well-suited for offshore resource assessment, where data gaps frequently arise in satellite observations due to cloud cover or technical limitations in sensor coverage. The separation of spatial and temporal patterns provided by EOFs enables DINEOF to effectively capture both the spatial heterogeneity and temporal variability of these parameters, offering insights into seasonal trends or regional anomalies in offshore environments. However, its reliance on matrix flattening introduces a trade-off between computational efficiency and the preservation of spatial relationships. For offshore applications requiring finer spatial resolution, TIEOF offers a compelling alternative by retaining more spatial structure, which can enhance the accuracy of resource assessments in complex regions like coastal zones or areas with significant bathymetric variability.

3.2. Typical performance metrics

The quality of interpolation is evaluated using one or more performance metrics, similar to traditional predictive algorithms. These

metrics are functions of the errors obtained during validation, each providing insight into different aspects of the errors. Below are the most commonly employed performance metrics, along with their mathematical formulations and practical explanations. Here, y^o denotes the observed (true) value of the parameter, while y^p represents the predicted value generated by the algorithm, such that the prediction error is given by $y^o - y^p$. The total number of errors is denoted by N.

- Bias or Mean Error (ME): As the name implies, this is the mean of all the errors, considering their sign:

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(y^o{}_i - y^p{}_i). \tag{21}$$

  A bias different from zero indicates a systematic error in the predictions, implying a consistent overestimation or underestimation of the parameter in question.
- Mean Absolute Error (MAE): This is the mean of the absolute values of the errors, without considering their sign:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y^o{}_i - y^p{}_i|. \tag{22}$$

  The MAE can be interpreted as the equivalent average error, where each individual error is weighted equally.
- Root Mean Square Error (RMSE): This is the square root of the sum of squared errors divided by the total number of errors:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y^o{}_i - y^p{}_i)^2}. \tag{23}$$

  Like the MAE, the RMSE represents an average error, but it tends to penalise larger errors more heavily due to the squaring of the error terms. Both the MAE and RMSE have normalised versions, which can be obtained by considering the relative errors, $\frac{y^o{}_i - y^p{}_i}{y^o{}_i}$, instead of the absolute errors $y^o{}_i - y^p{}_i$.
- Mean Absolute Percentage Error (MAPE): This is the mean of the absolute values of the percentage errors, without considering their signs (like the MAE but with percentage errors instead):

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y^o{}_i - y^p{}_i}{y^o{}_i}\right| \times 100 \tag{24}$$

  MAPE provides a percentage measure of prediction accuracy, making it intuitive for comparing error magnitudes across different scales.
- Coefficient of Determination (R^2) or Goodness of Fit: This performance metric is calculated as one minus the ratio of the sum of squared residuals to the total sum of squares:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i^o - y_i^p)^2}{\sum_{i=1}^{N}(y_i^o - \bar{y^o})^2}, \tag{25}$$

  where $\bar{y^o}$ is the average of the observed values, defined as $\bar{y^o} = \frac{1}{N}\sum_{i=1}^{N} y^o{}_i$. The total sum of squares, $\sum_{i=1}^{N}(y_i^o - \bar{y^o})^2$, represents the total variance of the observed data (ignoring a multiplicative constant). The R^2 value indicates the proportion of the variance in the observed data that is explained by the predictions, as the numerator $\sum_{i=1}^{N}(y_i^o - y_i^p)^2$ represents the residual variance after the predictions—essentially the variance of the prediction errors. Hence, R^2 is also referred to as the explained variance, since it quantifies the fraction of variance captured by the predictions.

4. Description of the interpolation case studies

As previously discussed, three main families of interpolation techniques have been considered in this work, along with a fourth category encompassing hybrid techniques and studies focusing on comparisons. Here with the term "case study" it is indicated an article that uses an interpolation or regression technique for some parameter in the offshore sector.

4.1. Deterministic interpolation case studies

Deterministic methods implicitly assume that metocean fields vary smoothly with distance and that local proximity is a sufficient proxy for similarity, without explicitly modelling measurement errors or uncertainty. They require relatively few inputs and are most commonly used in offshore workflows as quick baselines, for local gap filling, or for preprocessing/simplifying datasets when data are moderately dense and physics-informed uncertainty estimates are not essential.

Deterministic interpolators represent the simplest category of interpolation methods addressed here. Among the reviewed studies, only five articles focus specifically on deterministic techniques, though many hybrid or comparison-based studies also include these methods, such as Wei et al. (2019). All of these deterministic techniques involve fitting a predefined function to a set of training points across the entire domain. Due to their simplicity, these methods typically cannot incorporate any physical information from a model, and thus the interpolation is generally performed using only the available data.

Jones et al. (2015) and Liu et al. (2014) provide examples of traditional interpolation techniques. Jones et al. (2015) utilises the publicly available satellite database SOCAT to perform spatiotemporal interpolation of ocean surface CO2 concentrations using standard splines and harmonic functions. Validation of the reconstruction is conducted by ensuring that the statistical distribution of the process does not undergo any significant changes. In a similar vein, Jahanmard et al. (2022) applies linear interpolation, Inverse Distance Weighting (IDW), and Thin Plate Spline (TPS) methods to fit a set of points, although the focus of this study is on merging data from various sources (tide-gauge, hydrodynamic models, and satellite altimetry) and interpolating the errors relative to numerical simulations, rather than on sea level itself. The results indicate that IDW provides the most accurate and realistic outcomes.

Another comparative study on deterministic interpolation methods is found in Knysh et al. (2022), which compares three approaches for considering the vertical evolution of current velocity: averaging, linear interpolation, and Radial Basis Function (RBF) interpolation. Since the current velocity data is used as input for numerical simulations, the study evaluates how these different interpolation methods impact the simulation results. It demonstrates that the simplifications of linear interpolation, and even more so the simple mean, lead to less realistic outcomes compared to RBF interpolation.

More advanced deterministic interpolation techniques are presented by Liu et al. (2014) and Støle-Hentschel et al. (2021). Liu et al. (2014) introduces two methods — the wavelet-refined cubic method and the fractal method — for filling time gaps in in-situ measurements, comparing their performance against standard splines using data from two different sites. The proposed techniques outperform the spline method, with the wavelet method excelling at one site and the fractal method at the other. In a similar vein, Støle-Hentschel et al. (2021) explores the use of a deconvolution operator to reconstruct temporal gaps in in-situ wave measurements. Wave elevation is reconstructed using both simulated numerical data (JONSWAP) and real in-situ measurements. The robustness of the method is further tested by introducing artificial noise into the data, demonstrating that the technique can tolerate noise with an amplitude up to 10% of the signal amplitude.

Table 2 provides a summary of the reviewed study employing deterministic interpolation applications (for the meaning of the different employed symbols, the reader should refer to Table 1).

4.2. Statistical and geostatistical interpolation case studies

Statistical and geostatistical approaches assume an underlying stochastic structure (e.g., covariance/variogram or error statistics) and therefore treat both observations and, when available, model outputs as noisy information sources that can be optimally combined. They typically require enough data to estimate correlation/error parameters

Table 1
Reference guide to read Tables presented in this study.

Column	Brief description
Ref.	List of reviewed studies.
Domain	Type of relation that the model attempts to reconstruct (Space, Time).
Model	Information regarding the adoption of data generated from a numerical model: • ♦: Only data and no model information used. • ♦: Also the information from a numerical model is used.
Specific technique	Type of technique employed to solve the interpolation and regression problems.
Metocean variable	Metocean variable that is interpolated within the referenced study.
Performance metric	Metric employed to evaluate the performance of the interpolator/regressor during the validation process.
Input type	Source of data considered within the synthesis of the interpolator/regressor: • ■: Data coming from a simulated dataset. • ■: Data coming from in-situ measurements. • ■: Data coming from satellite image-like measurements. • ■: Data coming from satellite along-track measurements.

Table 2
Summary table of the reviewed study employing deterministic interpolation.

Ref.	Domain	Model	Specific technique	Metocean variable	Performance metric	Input type
Jones et al. (2015)	Space, Time	♦	Curve fitting	fCO_2	$RMSE, R^2$	■
Liu et al. (2014)	Time	♦	Wavelet refined method and fractal method	H_s	$RMSE, R^2$	■
Støle-Hentschel et al. (2021)	Time	♦	Deconvolution	Wave elevation	$RMSE$	■
Jahanmard et al. (2022)	Space	♦	Linear interpolator, IDW, TPS	Sea level	$Bias, RMSE$	■■
Knysh et al. (2022)	Space	♦	Mean, Linear interpolator, RBF	Current velocity	Effects on the simulation	■

(or to justify prescribed ones) and are particularly suited to offshore energy use cases where sparse in-situ/satellite observations must be merged with hindcasts or reanalyses while retaining uncertainty awareness (e.g., resource mapping, bias reduction, and operational decision support).

Statistical and geostatistical techniques represent more advanced interpolation methods, typically capable of directly addressing uncertainty arising from noise in the measurement system and/or incorporating information about the physical process through a model, as is common in Data Assimilation (DA) techniques. These methods can generally be divided into two main categories: Kriging methods (Urquhart et al., 2013; Saulquin et al., 2019; Jia and Taflanidis, 2013; Yin et al., 2022) and DA techniques (Barth et al., 2008; Rusu, 2015; Modi et al., 2022; Houghton et al., 2023; Smit et al., 2021; Sluka et al., 2016; Emmanouil et al., 2012; Houghton et al., 2022; Ni et al., 2022; Jincan Liu et al., 2023; Wu et al., 2023).

The only case study considered here that does not fall into one of these two categories is (Modi et al., 2022), which addresses gaps in satellite images of chlorophyll using a probability distribution-based approach and the Monte Carlo Multiple Imputation (MCMI) method. This statistical technique handles missing data by generating multiple imputed datasets, each containing plausible values for the missing data. These multiple imputations are then combined to provide a more accurate estimate of the variable of interest while also accounting for the uncertainty associated with the missing data.

The Kriging case studies discussed here primarily utilise Ordinary Kriging (OK), with some incorporating additional post-processing, except for Urquhart et al. (2013), which compares OK and Universal Kriging (UK) for gap-filling in satellite data on sea surface temperature (SST) and sea surface salinity (SSS) in the Chesapeake Bay. The findings indicate that UK outperformed OK, particularly in the upper bay for SSS and during summer months for both SSS and SST. This improvement is likely due to UK incorporating latitude as a predictor in its mean function, which is absent in traditional OK. With specific modifications, Kriging can also account for anisotropy in the process and extend the domain from spatial to spatiotemporal dimensions, as demonstrated in Saulquin et al. (2019), where both anisotropy and temporal differences are transformed into equivalent spatial distances. Kriging requires the inversion of a matrix with dimensions equal to the number of training points, which can lead to computational challenges when the dataset is too large. A potential solution to this issue is presented by Jia and Taflanidis (2013), where OK interpolation for hurricane risk is combined with Principal Component Analysis (PCA) to reduce the problem's dimensionality. A final example of OK is provided by Yin et al. (2022), where fish population density measurements from satellites and echosounders are spatially extrapolated using OK, and the results are analysed using a Generalised Additive Model (GAM), finding a significant relationship between fish population density and biomass distribution.

DA techniques involve updating an initial estimate of a system state (known as the background state) by incorporating measurements of the system, which are often subject to noise. Among the techniques discussed in the analysed studies, the most fundamental is the "vanilla" Optimal Interpolation (OI). OI is particularly effective for merging datasets from different measurement systems, as demonstrated in Barth et al. (2008), where OI is used to combine sea surface temperature (SST) and sea surface salinity (SSS) data from various satellite sources. Similarly, in Emmanouil et al. (2012), H_s measurements from satellites and in-situ observations are first processed with a statistical Kalman filter and then interpolated in regions lacking measurements using OI. In Ni et al. (2022), OI ability to merge data from different sources is leveraged to enhance the resolution of SST images. By combining low-resolution, cloud-penetrating microwave satellite data with high-resolution but cloud-sensitive infrared satellite images, OI effectively fuses the datasets, resulting in improved coverage and resolution while capitalising on the strengths and mitigating the weaknesses of each dataset. OI is also well-suited for long-term data assimilation. For instance, Rusu (2015) conducted a 15-year hindcast of H_s in the Black Sea, identifying it as a region of interest for wave energy exploitation. However, Smit et al. (2021) points out a limitation of OI when dealing with extended time periods: in cases where H_s is heavily influenced by wind speed, the improvements made to the background state by OI are less permanent, as the problem becomes more dynamic, akin to one driven by an external force. In contrast, when wind speed is less

Table 3
Summary table of the reviewed studies employing Kriging-based interpolation techniques.

Ref.	Domain	Model	Specific technique	Metocean variable	Performance metric	Input type
Urquhart et al. (2013)	Space	♦	OK, UK	SSS, SST	$Bias, MAE, RMSE, R^2$	■
Saulquin et al. (2019)	Space, Time	♦	Kriging with anisotropic covariance	$chl-\alpha$	$Bias, R, SI$	■
Jia and Taflanidis (2013)	Space, Time	♦	Kriging with PCA	Hurricane risk	MAE, R^2	■
Yin et al. (2022)	Space	♦	OK	Fish populations	R^2, R	■

Table 4
Summary table of the reviewed studies employing Data Assimilation and other statistical interpolation techniques.

Ref.	Domain	Model	Specific technique	Metocean variable	Performance metric	Input type
Barth et al. (2008)	Space	♦	OI	SSS, SST	$RMSE$	■
Rusu (2015)	Space	♦	OI	H_s	$MAE, RMSE, R, SI$	■
Modi et al. (2022)	Space	♦	MCMI	$chl-\alpha$	$RMSE$	■
Houghton et al. (2023)	Space, Time	♦	LETKF	H_s	$Bias, RMSE$	■
Smit et al. (2021)	Space, Time	♦	OI	H_s	$RMSE$	■
Sluka et al. (2016)	Space, Time	♦	LETKF	SSS, SST	$RMSE, R^2$	■
Emmanouil et al. (2012)	Space, Time	♦	Statistical Kalman filter with OI	H_s	$Bias, MAE, RMSE$	■
Houghton et al. (2022)	Space, Time	♦	OI	Wave directional spectra	$RMSE$	■
Mao et al. (2023)	Space, Time	♦	4d-Var	SSH, SST, SSS	$Bias, RMSE, MAE, R$	■
Ni et al. (2022)	Space, Time	♦	OI	SST	$Bias, RMSE$	■
Jincan Liu et al. (2023)	Space, Time	♦	Ensemble based OI	H_s	$Bias, RMSE, R$	■
Wu et al. (2023)	Space, Time	♦	Ensemble based OI	H_s	$Bias, RMSE, R$	■

significant, the system behaves more like an initial condition problem, and the updates remain more stable. A final example of standard OI usage is found in Houghton et al. (2022), where wave spectra data from over 600 Sofar Spotter buoys are assimilated into model spectra produced by the WW3 wave model. The assimilation process, performed using OI, significantly improves forecasts, with recorded enhancements of approximately 38% for H_s and around 45% for other wave parameters.

A variation of standard Optimal Interpolation (OI) is the Ensemble OI (EnOI), as employed in Wu et al. (2023) and Jincan Liu et al. (2023). EnOI combines ensemble-based methods with the OI approach. While traditional OI uses a single numerical simulation for the background state and observational data to estimate the system's state, EnOI involves running multiple simulations or model forecasts (ensemble members) to represent the uncertainty in the system. These ensemble members are generated by perturbing the initial conditions, boundary conditions, or physical parameters of the numerical model, thereby producing a range of plausible scenarios. This approach captures the variability in the system and enhances the robustness of the analysis. The resulting background states are then utilised during the assimilation process. In Jincan Liu et al. (2023), the use of EnOI demonstrates quantifiable improvements in H_s forecasts after assimilating satellite and in-situ measurements. The data assimilation reduces systematic prediction errors and reveals that the numerical simulations are generally more accurate offshore compared to nearshore. Similarly, Wu et al. (2023) applies EnOI for H_s data assimilation, with results validated against two in-situ devices. Improvements ranging from 3% to 11% were observed for one device, and from 7% to 44% for the other. In both studies, nearshore regions presented more challenges, suggesting that DA has a greater effect offshore. This could be attributed to the fact that both satellite measurements and numerical simulations are less reliable in shallow waters, where nonlinear processes are more prevalent. An example of the use of variational DA techniques in the offshore sector is provided by Mao et al. (2023), where satellite images and in-situ measurements are assimilated into a numerical model using 4D-Var to update predictions of sea surface height (SSH), SST, subsurface temperature, and subsurface salinity. After assimilation, all parameters showed reduced RMSE in validation, and the resulting distributions appeared more realistic than the original estimates. Sluka et al. (2016) and Houghton et al. (2023) are the only case studies in this review that perform strongly coupled DA using the Local Ensemble Transform Kalman Filter (LETKF). In Sluka et al. (2016), SST and SSS are updated using satellite data, while Houghton et al. (2023) integrates in-situ measurements from Sofar Spotter wave buoys and satellite altimetry data to perform DA on H_s from the WAVEWATCH III (WW3) model. Both studies show that strongly coupled DA using LETKF yields more accurate and realistic results than weakly coupled DA (e.g., standard OI), although it requires significantly more computational resources.

Tables 3 and 4 provide a summary of the reviewed studies employing statistical and geostatistical interpolation and regression techniques, with Table 3 reporting the Kriging-based case studies and Table 4 reporting the case studies DA or other statistical interpolation techniques.

4.3. Interpolation through machine learning techniques case studies

Machine-learning methods assume that the input–output relationship can be learned from data with limited a priori structure, enabling flexible nonlinear mappings and the use of many covariates (e.g., winds, bathymetry, currents, remote-sensing features). They generally require larger and more representative training datasets and are most attractive in offshore energy applications involving complex nonlinear dynamics, heterogeneous predictors, or image-like satellite products (e.g., gap filling of gridded fields, high-resolution downscaling, and surrogate modelling of numerical simulations).

The majority of the studies reviewed in this article use machine learning (ML) techniques for interpolation, which are particularly well-suited to the offshore sector due to the nonlinearities prevalent in marine environments. While the primary drawback of ML methods has traditionally been their need for vast amounts of data, this limitation is increasingly being alleviated by the expanding availability of satellite data. Of the 45 articles reviewed, nearly all can be classified into three main categories.:

- Tree-based algorithms: primarily Random Forest (RF) and Extremely Randomised Trees (ERT).
- Neural network (NN)-based algorithms, where the interpolation is performed by a NN-like model.
- Empirical Orthogonal Function (EOF)-based techniques, such as DINEOF, Principal Component Analysis (PCA), or Proper Orthogonal Decomposition (POD).

Across ML case studies reviewed, 23 employ neural-network models, 6 adopt tree-based regressors and 14 rely on EOF-family approaches. Only two papers deviate from these classifications: Wang and Chaib-draa (2017) and Albuquerque et al. (2018). Wang and Chaib-draa (2017) employs a Marginalised Particle Filter with a Gaussian Process

Regression (MPGP) to interpolate global surface temperatures, allowing the GPR to scale more effectively. The proposed method is then compared to two other sparse GPR variants — the Sparse Pseudo-Input Gaussian Process (SPGP) and the Sparse Spectrum Gaussian Process (SSGP) — which are alternative approaches for reducing the complexity of GPR, albeit at the expense of some accuracy. The results indicate that MPGP outperforms the other methods in both accuracy and computational efficiency. Instead, Albuquerque et al. (2018) utilises Multiple Linear Regression (MLR) to hindcast satellite altimeter data into the results of a numerical simulation. The proposed method enhances the accuracy of the numerical model and highlights the advantages of separating wave contributions from wind-sea and swell partitions.

In examining tree-based techniques, it appears that they are primarily applied to spatial prediction tasks, with only one exception (i.e. Chen et al. (2021)) which also addresses temporal interpolation. Both Li et al. (2013, 2016) explore the use of Random Forest (RF) for spatial prediction of seabed hardness based on in-situ measurements. Li et al. (2013) implements RF in R and demonstrates the model ability to differentiate between hard and soft substrates across four geomorphologically complex areas in the Joseph Bonaparte Gulf. In contrast, Li et al. (2016) reframes the problem as a classification task rather than a regression one, with the output indicating two possible hardness levels. Additionally, Li et al. (2016) highlights how RF is particularly advantageous when numerous covariates are available for predicting the target variable (in this case, 41 covariates were used, such as bathymetry and planar curvature). Both Chen et al. (2021) and Mounet et al. (2023) employ an RF-based surrogate model, trained and validated with in-situ data, to replicate the results obtained from SWAN simulations. These studies show that the RF model retains the accuracy of the numerical simulation while significantly reducing computational time. Mounet et al. (2023) further investigates the integration of the ship-as-wave-buoy concept with standard buoy networks, revealing that while the RF model delivers low prediction errors for scalar parameters (e.g., T_p and T_e), it produces substantial errors when estimating the principal directions of the wave spectrum. A different application of tree-based surrogate models is presented in Chen et al. (2019) and Park et al. (2019), where these models are used to reconstruct satellite images of chlorophyll-α ($chl-\alpha$) in ocean regions. Chen et al. (2019) trains a Random Forest Regression Ensemble (RFRE) to map $chl-\alpha$ in the Yellow Sea and East China Sea, showing that the RFRE can spatially extend measurements without compromising their quality. Similarly, Park et al. (2019) reconstructs ocean colour in polar regions, comparing RF with Extremely Randomised Trees (ERT). After an initial assessment, other methods such as linear and logistic regression were discarded. The results, validated with in-situ measurements, show that ERT slightly outperforms RF.

Unlike tree-based techniques, neural network (NN)-based methods are widely employed for interpolation tasks across space, time, and spacetime domains. Certain architectures are particularly suited to specific tasks: Long Short-Term Memory networks (LSTM-NNs) are commonly favoured for time and spacetime interpolation, whereas Convolutional Neural Networks (CNNs) are typically preferred for spatial predictions. Given that measuring buoys often experience downtime for maintenance, leading to temporal data gaps, both Vieira et al. (2020) and Cuadra et al. (2016) propose neural networks to address this issue. The former employs a Feedforward Neural Network (FFNN), while the latter uses an Extreme Learning Machine (ELM), a type of NN, to fill gaps in wave parameter recordings. Both studies demonstrate that these techniques can accurately fill temporal gaps with performance comparable to numerical models, without requiring specific a priori knowledge of the system—something typically needed by numerical models. Similarly, Oo and Zhang (2022) and Peres et al. (2015) use FFNNs for spatial interpolation of wave parameters from in-situ measurements. Oo and Zhang (2022) develops a spatial FFNN to assimilate in-situ data into the results of a nearshore numerical model (WW3), showing that the FFNN improves the simulation results, although these improvements diminish with increasing distance from the assimilation devices. Peres et al. (2015) applies an FFNN to extend H_s from in-situ measurements, incorporating 6-hour averaged 10 m wind speed from a reanalysis dataset as an additional input. The results, compared with those from a numerical model, show that the FFNN delivers superior accuracy with reduced computational time. Tapoglou et al. (2021) also utilises an ensemble of Artificial Neural Networks (ANNs) to analyse satellite images and generate spatial maps of H_s at an offshore wind farm. Validation with in-situ buoys shows that the errors are comparable to those of numerical models. Krasnopolsky et al. (2016) similarly employs an ensemble of NNs to fill gaps in satellite-derived ocean colour data, demonstrating that the ensemble approach outperforms a single NN in this task. Further advancements are seen in Medina-Lopez and Ureña-Fuentes (2019), where a Deep Neural Network (DNN) is developed to interpolate and extrapolate high-resolution Sea Surface Salinity (SSS) and Sea Surface Temperature (SST) using satellite and in-situ data. Despite being interpolated independently, the results affirm the known correlation between SSS and SST, with the algorithm producing high-resolution outputs with acceptable error margins. Chen and Hu (2017) also proposes an approach to enhance the spatial resolution of satellite-derived SSS measurements by using a Multi-Layer Perceptron (MLP) NN, which takes SST and satellite-measured reflectance as covariates. This model successfully produces near-real-time 1 km-resolution maps of SSS, though its performance declines in areas with algal blooms or upwellings. The MLP architecture is also applied in temporal reconstructions, as seen in James et al. (2018) and Feng et al. (2020), both of which integrate numerical simulations from the SWAN model. In James et al. (2018), the MLP is used to forecast H_s, while a Support Vector Machine (SVM) classifies T_p.[2] The results indicate that these surrogate models provide slightly less accuracy than the original numerical model but with only 10^{-3} times its computational time. Similarly, Feng et al. (2020) develops an MLP to forecast wave parameters on Lake Michigan, achieving comparable results to simulations using just from 5×10^{-5} to 10^{-4} times the original computational time. However, higher errors are observed in shallow water, likely due to increased nonlinearity in such regions. NNs can also be applied to global reconstructions, although with reduced resolution. For instance, Zeng et al. (2014) uses a Feedforward NN (FFNN) to perform spatio-temporal interpolation of sea surface CO2, achieving a $1^\circ \times 1^\circ$ grid resolution. The results indicate that the network effectively captures the system nonlinearities and that a single network suffices to model sea surface CO2 on a global scale.

As previously mentioned, Convolutional Neural Networks (CNNs) are among the most commonly used NN architectures for spatial interpolation tasks, as demonstrated in Jörges et al. (2023). Additionally, they are frequently employed for "space forecasting", where the input consists of current 2D maps, and the output is a predicted future map, as seen in Rajabi-Kiasari et al. (2023) and Bai et al. (2022). In Bolton and Zanna (2019), a CNN is trained using satellite altimetry images to reconstruct subfilter eddy momentum forcings, S_x and S_y, across both space and time. The study shows that the algorithm can accurately predict both subgrid and large-scale processes, and it is capable of incorporating physical constraints, such as momentum conservation, through careful hyperparameter tuning. However, the computational cost is relatively high, and the paper concludes by suggesting various approaches to reduce this, albeit at the expense of some accuracy. In Jörges et al. (2023), a CNN is proposed to spatially interpolate H_s measurements from in-situ data, incorporating bathymetry as a covariate. Since bathymetry is dynamic, its evolution is simulated using random field simulations, which are effectively Gaussian Process Regression (GPR) reconstructions. The results, validated against held-out in-situ devices and compared with SWAN simulations, show that

[2] Although T_p is a continuous variable, it was discretised into distinct categories, allowing it to be treated as a classification problem for the SVM model.

the CNN achieves near-simulation-level performance (with an RMSE of 0.23 m, compared to 0.218 m for SWAN) while reducing computational costs by a factor of 300,000. A different application of CNNs is presented in Wei and Davison (2022), where the algorithm is used to forecast nearshore wave conditions. The training and testing data are generated by the non-hydrostatic numerical model SWASH, and the results show that the CNN is able to accurately predict the dynamics of the simulated system. The remaining two papers, Rajabi-Kiasari et al. (2023) and Bai et al. (2022), focus on bidimensional forecasting. In Rajabi-Kiasari et al. (2023), a 2D-CNN is developed to take as input an accurate map of the current topography, alongside information on wind speed and direction, surface pressure, and temperature from various datasets and satellite measurements, with the output being predicted future topography maps. Similarly, Bai et al. (2022) develops a 2D-CNN for forecasting wave (H_s) and wind (U10 and V10) parameters, training the algorithm on a numerical dataset and validating the results against in-situ measurements.

Recurrent Neural Networks (RNNs) are a class of neural networks that incorporate feedback loops into their architecture, allowing the model to retain a memory of previous states. This key feature makes RNNs particularly well-suited to handling time-domain problems, especially in the case of Long Short-Term Memory Networks (LSTM-NNs), a specific RNN architecture. In Pirhooshyaran and Snyder (2020), the potential of RNNs for temporal gap filling in buoy-based wave measurements is explored. The study combines RNNs with Bayesian hyperparameter optimisation and regularisation techniques, specifically Elastic Net, testing various architectures, including LSTM-NNs and sequence-to-sequence NNs (s-t-s NNs). The results demonstrate that these algorithms can effectively fill temporal gaps in H_s recordings from buoys using only data from neighbouring buoys. However, the best-performing algorithm varies depending on the chosen validation metric.

Similarly, Ahn et al. (2022) uses an LSTM-NN to forecast H_s measurements from three in-situ devices, incorporating wind speed as an additional covariate. While the inclusion of wind speed yields marginal improvements, these are mainly seen in short-term forecasts. The study also highlights that extreme H_s values (below 0.5 m and above 4 m) are not accurately reconstructed, although the performance is particularly strong for predicting the most frequent sea states. LSTM-NNs are also capable of handling spatio-temporal interpolation, as demonstrated in Chen et al. (2022b). In this work, an LSTM-NN is trained to learn the spatio-temporal patterns of ocean waves from simulations and in-situ measurements and is applied to fill gaps in the measurement systems. The achieved accuracy is slightly lower than that of numerical simulations, but the neural network operates with significantly reduced computational resource requirements. Some specialised applications of NN-based techniques are presented in Silva et al. (2018) and Özger (2009). Silva et al. (2018) performs temporal gap filling of sea surface wind speed using a Nonlinear Autoregressive Artificial Neural Network with Exogenous Inputs (NARX ANN). The study compares the results with those obtained from a standard Feedforward Neural Network (FFNN) and numerical simulations using the JONSWAP spectrum, demonstrating the superiority of the proposed method. Similarly, Özger (2009) applies a neuro-fuzzy approach to fill spatial gaps in H_s measurements from in-situ devices, showing that the proposed algorithm outperforms a Multiple Linear Regression (MLR) model.

EOF-based techniques are predominantly employed for reconstructing missing pixels in satellite imagery, as exemplified by the DINEOF method. While they are mostly applied to satellite images, they are occasionally used with satellite along-track data and in-situ measurements as well. Typically, the interpolated parameters are $chl-\alpha$, SST, or SSS, as these are more readily captured in accurate enough image distributions from satellites, unlike more hydrodynamic parameters such as H_s. Several studies, including (Liu and Wang, 2022, 2018; Li and He, 2014; Alvera-Azcárate et al., 2016; Sirjacobs et al., 2008; Beckers and Rixen, 2003; Andrawina et al., 2021), provide standard examples of the DINEOF technique being used to generate gap-free datasets for these parameters. In Liu and Wang (2022, 2018), and Li and He (2014), it is noted that DINEOF is capable of capturing seasonal variations, which are easily discernible from the EOF decomposition due to the interpretability of the method. Beckers and Rixen (2003) introduces an interesting approach for quantifying the uncertainty of the reconstruction through cross-validation. However, in most of these case studies, the initial data is relatively dense, with typically only 60%–70% of the dataset missing (though in Alvera-Azcárate et al. (2016), this figure reaches 90%). Additionally, Alvera-Azcárate et al. (2016) highlights that the DINEOF interpolation process automatically reduces noise in the original dataset. DINEOF also facilitates the merging of image data from different satellites, each with its own set of missing pixels, as demonstrated in Liu and Wang (2019), where $chl-\alpha$ is interpolated in both space (at a 9 km resolution) and time. The analysis of the EOF decomposition revealed large-scale and mesoscale features that were not visible prior to the merging of satellite data. A variation of the traditional DINEOF method is proposed in Ping et al. (2016) for reconstructing SST satellite images. The study employs a modified DINEOF technique, known as VE-DINEOF (Variable EOFs DINEOF), where the number of EOFs used for reconstruction can adapt to each specific image. The results show that this variation provides better accuracy and requires less computational time compared to the traditional DINEOF method. DINEOF can also be modified to include covariates. For instance, in Alvera-Azcárate et al. (2007), a multivariate DINEOF technique is used to reconstruct SST surfaces using satellite data. This modified version incorporates additional variables, with $chl-\alpha$ and wind speed serving as covariates. The results indicate improved performance when these additional variables are included. Furthermore, the study emphasises that this approach is more suitable than traditional methods like Optimal Interpolation (OI) for handling multiple predictive parameters, as it does not require any assumptions about the statistical distributions of these variables. Sijia Wang et al. (2022) investigates the variability of dimethyl sulphide concentration in the sea by first performing a hindcast using a Generalised Additive Mixed Model (GAMM) and subsequently applying DINEOF to interpolate and fill spatial gaps. The GAMM incorporates satellite-derived SST and $chl-\alpha$ as covariates. The findings reveal that short-term variations (2–3 years) are driven solely by SST, whereas long-term fluctuations (6–8 years) are influenced by both SST and $chl-\alpha$.

For more precise interpolation, the TIEOF algorithm can be employed. Unlike DINEOF, TIEOF directly accounts for the two-dimensional spatial structure of the images. In Kulikov et al. (2021), TIEOF is used to reconstruct satellite images of $chl-\alpha$ in Lake Baikal, resulting in a 69% improvement in accuracy compared to the standard DINEOF method. By incorporating an additional dimension, TIEOF better captures the spatial correlations within the system, albeit at a significantly higher computational cost. Further examples of EOF-based techniques for reconstructing offshore parameters are provided in Meunier et al. (2022) and Hamzi et al. (2021). Meunier et al. (2022) uses satellite altimetry data and in-situ measurements to examine the three-dimensional structure of Loop Current Rings. The study reconstructs vertical profiles of sea salinity and temperature from sea surface elevation data using the Gravest Empirical Mode (GEM), an EOF-based technique. In contrast, Hamzi et al. (2021) combines Proper Orthogonal Decomposition (POD) with Reproducing Kernel Hilbert Spaces (RKHS) to forecast weekly-averaged SST, utilising an open-source NOAA dataset. This method essentially applies Gaussian Process Regression (GPR) after reducing the dimensionality of the input through a POD-based change of basis and projection. The study compares this approach with a POD-LSTM-NN, demonstrating that while the POD-RKHS technique performs better and is computationally more efficient, the LSTM-NN could potentially be enhanced through more refined hyperparameter tuning.

Tables 5–8 provide a summary of the reviewed studies employing machine learning regression and interpolation techniques, with Table 5 reporting the NN-based techniques case studies, Table 6 the RF-based ones, Table 7 the EOF-based ones and Table 8 the ones using other ML techniques.

Table 5
Summary table of the reviewed studies employing NN-based ML interpolation techniques (Contractor and Roughan, 2021; Zhou et al., 2021).

Ref.	Domain	Model	Specific technique	Metocean variable	Performance metric	Input type
Tapoglou et al. (2021)	Space, Time	♦	EnANN	H_s	$RMSE, R^2$	
Cuadra et al. (2016)	Time	♦	ELM	H_s, T_z, T_s	$RMSE, R$	
Vieira et al. (2020)	Time	♦	FFNN	H_s, T_p, Power spectral density	$RMSE, Bias, SI$	
Contractor and Roughan (2021)	Space, Time	♦	LSTM-NN	SST, Oxygen and Nutrients	$MSE, MAPE$	
Krasnopolsky et al. (2016)	Space	♦	EnNN	$chl-\alpha$	$RMSE, Bias, R$	
Silva et al. (2018)	Space, Time	♦	NARX ANN, FFNN	Sea surface wind speed	$RMSE, SI$	
Bolton and Zanna (2019)	Space, Time	♦	CNN	Eddy momentum forcing	$RMSE, R$	
Medina-Lopez and Ureña-Fuentes (2019)	Space	♦	DNN	SST, SSS	$RMSE, MAE, R$	
Chen and Hu (2017)	Space, Time	♦	MLP	SSS	$RMSE, MAE, R^2$	
Zeng et al. (2014)	Space	♦	FFNN	fCO_2	$SI, RMSE, Bias$	
Peres et al. (2015)	Space	♦	FFNN	H_s	$R^2, RMSE, Bias$ accuracy-score	
Chen et al. (2022b)	Space, Time	♦	LSTM-NN	H_s, T_p, T_m Wave direction	$R^2, RMSE, MAE, MAPE$	
Pirhooshyaran and Snyder (2020)	Time	♦	RNN, s-t-s NN, LSTM	H_s, Wave power	$RMSE, MAE, R^2, HuberLoss$	
Jörges et al. (2023)	Space	♦	CNN	H_s	$R^2, RMSE, MAE, MAPE, Bias$	
Wei and Davison (2022)	Space, Time	♦	CNN	Water elevation and velocity	$RMSE, NMAE, R^2$	
James et al. (2018)	Time	♦	MLP and SVM	H_s, T_p	$R^2, RMSE, LassoLoss$	
Feng et al. (2020)	Time	♦	MLP	H_s, T_p	$RMSE, Bias, R^2$	
Zhou et al. (2021)	Space, Time	♦	ConvLSTM-NN	H_s	$Bias, RMSE, MAE$	
Bai et al. (2022)	Space, Time	♦	2D-CNN	H_s	$R, RMSE, MAE$	
Oo and Zhang (2022)	Space	♦	FFNN	H_s, T_p	$RMSE, Bias, R, SI$	
Ahn et al. (2022)	Time	♦	LSTM-NN	H_s	$R, RMSE$	
Rajabi-Kiasari et al. (2023)	Space, Time	♦	2D-CNN	Dynamic topography	$R^2, RMSE$	

Table 6
Summary table of the reviewed studies employing RF-based ML interpolation techniques.

Ref.	Domain	Model	Specific technique	Metocean variable	Performance metric	Input type
Chen et al. (2021)	Space, Time	♦	RF	H_s, T_z, T_p, Wave direction	$RMSE, R^2$,	
Chen et al. (2019)	Space	♦	RFRE	$chl-\alpha$, Rayleigh reflectance	$R^2, RMSE, SI$	
Li et al. (2013)	Space	♦	RF	Seabed hardness	MSE	
Park et al. (2019)	Space	♦	RF and ERT	$chl-\alpha$	$R^2, RMSE, MAE$ accuracy-score	
Li et al. (2016)	Space	♦	RF	Seabed hardness	Accuracy and confusion matrix	
Mounet et al. (2023)	Space	♦	RF	Wave spectrum	$R^2, RMSE, MAE, SI$	

Table 7
Summary table of the reviewed studies employing EOF-based ML interpolation techniques.

Ref.	Domain	Model	Specific technique	Metocean variable	Performance metric	Input type
Liu and Wang (2022, 2018)	Space, Time	♦	DINEOF	$chl-\alpha$	Mean and median of the errors	
Sirjacobs et al. (2008)	Space, Time	♦	DINEOF	$chl-\alpha, SST$	R^2	
Beckers and Rixen (2003)	Space	♦	DINEOF	Image reconstruction	$RMSE$	
Andrawina et al. (2021)	Space, Time	♦	DINEOF	SST	Expected error	
Ping et al. (2016)	Space, Time	♦	VE-DINEOF	SST	$RMSE, MAE, R$	
Alvera-Azcárate et al. (2007)	Space, Time	♦	DINEOF	SST	$RMSE, Bias, R$	
Liu and Wang (2019)	Space, Time	♦	DINEOF	$chl-\alpha$,	$RMSE, SI$	
Li and He (2014)	Space, Time	♦	DINEOF	SST	$RMSE, R^2$	
Alvera-Azcárate et al. (2016)	Space, Time	♦	DINEOF	SSS	$RMSE, CRMSE, Bias, R, SI$	
Kulikov et al. (2021)	Space, Time	♦	TIEOF	$chl-\alpha$	$RMSE$	
Sijia Wang et al. (2022)	Space, Time	♦	DINEOF	Dimethyl sulphide concentration	R, R^2, SI	

Table 8
Summary table of the reviewed studies employing other ML interpolation techniques.

Ref.	Domain	Model	Specific technique	Metocean variable	Performance metric	Input type
Meunier et al. (2022)	Space	♦	GEM	Loop current rings	$RMSE, R^2$	
Wang and Chaib-draa (2017)	Space	♦	MPGP	SST	MSE	
Hamzi et al. (2021)	Space, Time	♦	POD-RKHS	SST	$RMSE, MAE, R^2$	
Özger (2009)	Space	♦	Fuzzy inference system	H_s	$R, RMSE, MAE$	
Albuquerque et al. (2018)	Space, Time	♦	MLR	H_s	SI	

4.4. Hybrid techniques and comparison case studies

Hybrid and comparison studies typically assume that no single family is uniformly optimal across variables, sampling geometries, and regimes, and therefore either benchmark multiple methods under a shared validation protocol or explicitly combine complementary components (e.g., statistical error modelling with ML feature learning). Their data requirements are driven by the most demanding component, and they are particularly relevant to offshore energy workflows that need both scalability and robustness—such as long-term site screening with quantified uncertainty, or operational products that blend model consistency with observation-driven corrections.

This section encompasses case studies that do not fall precisely into the previously discussed categories, ranging from papers that compare different techniques across various families, such as Wei et al. (2019) and Stock et al., to those that propose hybrid approaches by integrating multiple methods, such as Zhisong Huang et al. (2012) and Barth et al. (2020). The primary objective here is to explore the potential of different classes of techniques in a variety of application scenarios and to highlight possible synergies and hybrid approaches for addressing the metocean gap-filling problem.

Both Alvera-Azcárate et al. (2005) and Elken et al. (2019) compare DINEOF reconstructions with Optimal Interpolation (OI). Alvera-Azcárate et al. (2005) investigates the reconstruction of satellite images of SST by artificially introducing noise to simulate cloud coverage, while Elken et al. (2019) utilises a dataset of daily averages, reconstructing SST and SSS from selected parts. Alvera-Azcárate et al. (2005) finds comparable performance between DINEOF and OI but a 30-fold reduction in runtime for DINEOF. In contrast, Elken et al. (2019) reports that OI significantly outperforms DINEOF for SST, and achieves only a slight advantage for SSS. The paper suggests that OI is more suitable when data is denser, making the Gaussian distribution assumption more applicable. The differing results may stem from DINEOF being particularly effective for image reconstruction, as in Alvera-Azcárate et al. (2005), while Elken et al. (2019) works with a complete dataset. In Wei et al. (2019), the assessment of wind-generated power potential in Jiangsu Province required wind speed data interpolation. The study compared several methods, including Spline interpolation, Nearest Neighbour, Inverse Distance Weighting (IDW), and Kriging. Cross-validation indicated that Kriging achieved the lowest error, so it was for the final assessment. A more focused analysis of interpolation techniques is presented in Stock et al., where ten algorithms (Temporal Interpolation, IDW, Ordinary Kriging (OK), Spatiotemporal Kriging, DINEOF, Ridge Regression with and without nearby $chl-\alpha$ as a covariate, Random Forest (RF) with and without nearby $chl-\alpha$, and Self-Organising Maps) are compared for gap-filling satellite images of $chl-\alpha$ distributions. Performance varies by spatial and temporal domain, yet the two leading methods (OK and DINEOF) are closely matched. Rufino et al. (2021) proposes a methodology for selecting the most suitable spatial interpolation technique. It compares a range of methods, from basic approaches such as IDW and Thin Plate Splines (TPS) to more advanced ones like Kriging, RF, and Generalised Linear Models (GLM), in interpolating marine species distributions. The key conclusion of the study is that optimality varies markedly with the chosen metric, the temporal frame, and the species considered. Similarly, Li et al. (2011) compares multiple techniques for mapping seabed sediments, using data from the MARS database. The study evaluates methods such as IDW, Generalised Least Squares Trend Estimation (GLS), Kriging with External Drift (KED), Ordinary Cokriging (COK), OK, Universal Kriging (UK), Decision Trees (DT), TPS, General Regression Neural Networks, Support Vector Machines (SVM), and combinations of OK with linear models, GLM, GLS, and RF. Among the tested methods, OK integrated with RF performs the best and exhibits the greatest robustness.

Several case studies focus specifically on comparing machine learning (ML) interpolation methods, as seen in Penalba et al. (2022), Zeng et al. (2017), Alexandre et al. (2015), and Diesing et al. (2014). In Penalba et al. (2022), long-term time series of significant wave height (H_s) and peak wave period (T_p) are analysed from in-situ measurements. The study develops a Random Forest (RF), a Support Vector Regression (SVR),[3] and a Neural Network (NN) to forecast the evolution of H_s. After comparing the three algorithms, it is found that SVR performs best on the test set, although RF yields superior results during training. Zeng et al. (2017), a continuation of Zeng et al. (2014), compares three ML techniques: a Feedforward Neural Network (FNN), a Support Vector Machine (SVM), and a Self-Organising Map (SOM) to map sea surface CO2 levels from a database of satellite along-track measurements. The algorithms incorporate various covariates, such as sea surface temperature (SST), sea surface salinity (SSS), and $chl-\alpha$. The results indicate that the SVM performs best, while the SOM performs worst. However, the SVM has the highest computational cost, whereas the SOM is the most computationally efficient. Therefore, the author recommends SOM when handling large datasets. Alexandre et al. (2015) proposes a method to fill temporal gaps in H_s measurements from in-situ buoys, particularly gaps caused by extreme events. Initially, a Genetic Algorithm (GA) combined with an Extreme Learning Machine (ELM) selects a subset of wave parameters that minimises the reconstruction error. These selected variables are then passed to the regression algorithm. Three algorithms — Gaussian Process Regression (GPR), SVR, and another ELM — are tested and compared, with GPR outperforming SVR and ELM on the reconstruction task. Diesing et al. (2014) compares three techniques for reconstructing seabed sediments from in-situ and acoustic measurements. The study examines manual interpretation, Object-Based Image Analysis (OBIA), RF, and cokriging algorithms. The results suggest that RF is the best-performing interpolation technique, although the optimal method may vary depending on the performance metric used.

The final case studies present the development and testing of novel interpolation techniques, including Zhisong Huang et al. (2012), Barth et al. (2020, 2022), and Chen et al. (2022a). In Zhisong Huang et al. (2012), a Least Squares Support Vector Machine (LS-SVM) Kriging method is developed for spatially interpolating sea surface salinity (SSS) and sea surface height (SSH) using data from the GODAS system. The technique eliminates the need to manually select a variogram, as the LS-SVM automatically determines the appropriate variogram from the data. The study demonstrates that the proposed LS-SVM Kriging algorithm outperforms traditional Kriging methods that rely on standard variograms. Barth et al. (2020) introduces a new approach called DINCAE (Data Interpolating Convolutional Auto-Encoder), which reconstructs sea surface temperature (SST) fields from satellite observations where data are missing (e.g., due to cloud cover). This method essentially combines the DINEOF technique with an autoencoder neural network. Various variants of the neural network are tested, differing in the connections within the network, and are compared to traditional DINEOF. Results indicate that DINCAE captures more variability and provides better performance than the conventional DINEOF method. Building on this work, Barth et al. (2022) updates the DINCAE method to handle multivariate data, using wind speed and $chl-\alpha$ as additional covariates. The updated method, DINCAE 2.0, incorporates a new interpolation layer that processes data on a grid before reconstruction, enabling the method to work with unstructured data, such as along-track satellite observations, rather than satellite images alone. DINCAE 2.0 is compared to the DIVAnd method (a variational technique similar to Optimal Interpolation for spatial field interpolation). The findings suggest that DINCAE 2.0 slightly outperforms DIVAnd in accuracy, and its uncertainty estimates are more reliable. In Chen et al. (2022a), a spatiotemporal gap-filling method for in-situ measurements is developed, using a nonconvex low-rank tensor completion (LRTC) model. This method frames the interpolation as an optimisation problem, similar to the TIEOF algorithm, but instead of minimising the tensor rank, it minimises the truncated nuclear norm (TNN) as an approximation of the rank. This approach follows a similar methodology to Chen et al. (2020). The proposed LRTC-TNN method is compared against linear and cubic spline interpolation techniques, showing superior performance and demonstrating its potential for integration with additional covariates.

Table 9 provides a summary of the reviewed hybrid and comparative studies.

5. Inter-class analysis and critical comparison

By examining Table 2 and Fig. 9, it is clear that deterministic interpolation techniques are not particularly prominent in the offshore sector, as indicated by the limited number of studies focusing on

[3] Support Vector Regression (SVR) and Support Vector Machine (SVM) are similar but different techniques both based on the support vector framework. SVM is used for classification tasks, where the model finds an optimal hyperplane that maximises the margin between data points from different classes. In contrast, SVR is used for regression tasks, fitting a regression function that minimises errors while keeping the predictions within an epsilon-insensitive tube. This tube defines a margin of tolerance around the true values, allowing for robust predictions with controlled complexity.

Table 9
Summary table of the reviewed study employing hybrid interpolation or presenting comparison analyses.

Ref.	Domain	Model	Specific technique	Metocean variable	Performance metric	Input type
Penalba et al. (2022)	Time	♦	RF, SVR, ANN	H_s, T_p	$RMSE, R, MAE$	
Wei et al. (2019)	Space	♦	Spline, Natural Neighbour, IDW, OK	Wind speed	$RMSE$	
Stock et al.	Space, Time	♦	DINEOF, TI, IDW, OK, STK, SOM	$chl-\alpha$	$RMSE$	
Rufino et al. (2021)	Space	♦	LM, GAM, IDW, TPS, OK, GLM, RF	Marine species	$RMSE, MAE$	
Zhisong Huang et al. (2012)	Space	♦	LS SVM-Kriging	SSS, SSH	$Bias, MAE, RMSE$	
Li et al. (2011)	Space	♦	IDW, GLS, RT, TPS, GRNN, SVM, Various Krigings	Seabed sediments	$RMAE$	
Zeng et al. (2017)	Space	♦	FNN, SVM, SOM	fCO_2	R^2	
Alexandre et al. (2015)	Time	♦	GA-ELM, GPR, SVM, ELM	H_s	$RMSE, SI$	
Diesing et al. (2014)	Space	♦	OBIA, RF, COK	Seabed sediments	Accuracy, Purity	
Alvera-Azcárate et al. (2005)	Space	♦	DINEOF, OI	SST	$RMSE$	
Elken et al. (2019)	Space, Time	♦	DINEOF, OI	SST, SSS	$RMSE$	
Barth et al. (2020)	Space	♦	DINCAE	SST	$RMSE, SI, Bias$	
Barth et al. (2022)	Space, Time	♦	DINCAE, DIVAnd	SST	$RMSE, SI$	
Chen et al. (2022a)	Space, Time	♦	LRTC-TNN, Linear and Spline interpolator	H_s, T_z	$MAE, RMSE, SI, R^2$	

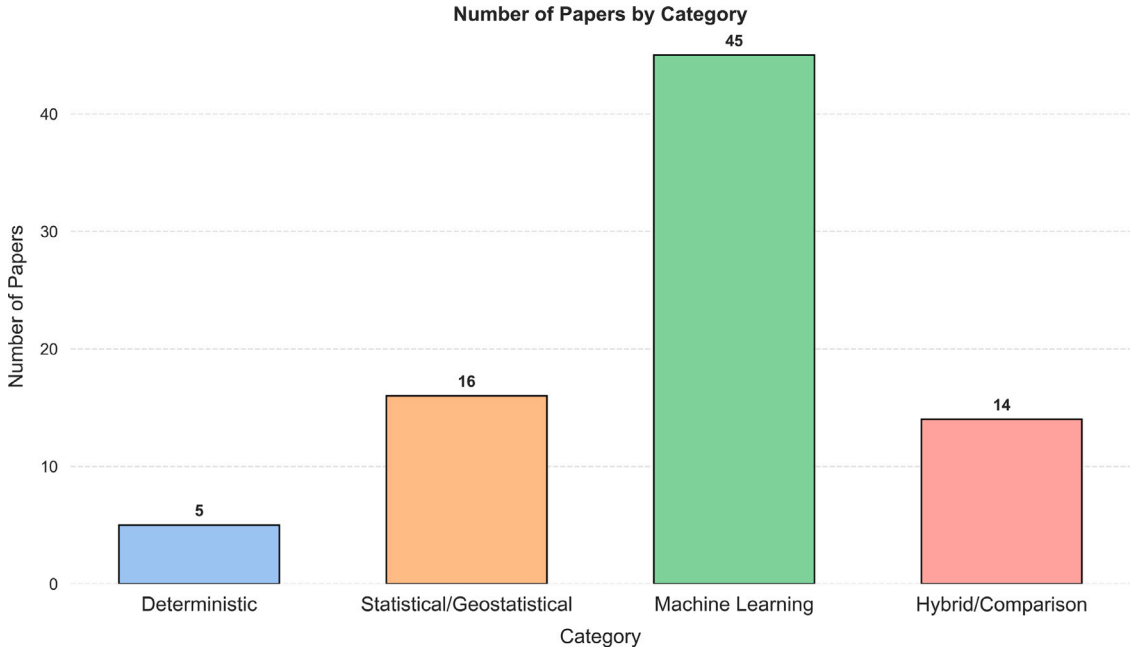


Fig. 9. Number of reviewed paper belonging to the different classes of techniques and studies.

them. This impression is further corroborated by the temporal analysis in Fig. 10, where deterministic studies remain marginal and exhibit limited growth compared with statistical and ML approaches. Additionally, these techniques are generally unable to incorporate information from numerical models used alongside the data. Although deterministic methods are applied across all the considered domains, the input data they handle are typically restricted to along-track satellite measurements or in-situ observations. These techniques are less suitable for filling gaps in satellite images due to their inherent simplicity.

In contrast, statistical and geostatistical techniques, specially Kriging, appear to be more relevant, as evidenced by the higher number of studies investigating their application in the offshore domain. In particular, Kriging often emerges as the best-performing interpretable method because it explicitly models spatial dependence via the variogram/covariance (including anisotropy and, in universal Kriging, large-scale trends) while retaining a linear, transparent estimator that yields both predictions and an uncertainty (kriging-variance) maps—a practical balance between simplicity and expressive power. However, as shown in Table 3, kriging-based approaches are almost never employed in a purely temporal domain: the few spatio-temporal applications found in the literature (e.g. Saulquin et al. (2019), Jia and Taflanidis (2013)) treat time as an additional coordinate or rely on *ad hoc* covariance structures, and no case study explicitly targets forecasting or hindcasting. For such temporal interpolation tasks, alternative methods such as machine-learning regressors are generally preferred, as they are more naturally suited to long time series and prediction-oriented setups. When a numerical model is available, data assimilation (DA) techniques prove to be more effective in integrating the model information with the data. Furthermore, most techniques within this category are adaptable to the different input data types considered.

When high accuracy is required and/or the available data are sparse, techniques such as LETKF, 4D-Var, and EnOI are capable of extracting the maximum amount of information from the available data, although their computational cost can be prohibitive when dealing with large datasets. On the other hand, for cases where abundant data are available, Optimal Interpolation (OI) and Kriging methods tend to be more appropriate. As noted by Elken et al. (2019), both techniques are more effective when measurements are dense, as the assumption of a normal distribution is more likely to be valid under these conditions. Moreover, the computational complexity of Kriging with large datasets can be mitigated through ad hoc preprocessing techniques, as demonstrated in Jia and Taflanidis (2013).

As evidenced by the large number of articles on machine learning (ML) interpolation techniques, this appears to be the most promising family of methods. ML techniques are employed for spatiotemporal, temporal, and spatial interpolation. Some studies incorporate numerical models for interpolation, although this is less common than in the case

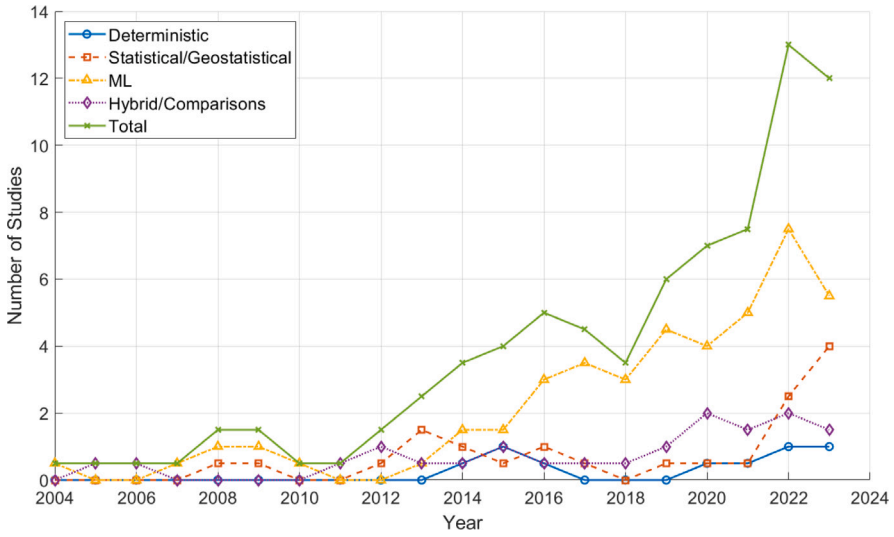


Fig. 10. Number of case studies published each year, per category.

of statistical and geostatistical approaches. This is primarily because, while ML algorithms can integrate information from numerical models, doing so is not as straightforward as with DA techniques. Many ML algorithms function as "black boxes" and are difficult to interpret.

Among these ML techniques, the DINEOF algorithm is the most widely used for reconstructing satellite images, having been specifically developed for that purpose. However, its main drawbacks include its limited performance when the training points are sparse or few in number, and its difficulty in incorporating additional covariables, though this can be achieved with certain modifications, as shown by Alvera-Azcárate et al. (2007). The reliance of DINEOF on image data also restricts its applicability to only a subset of parameters of interest in the offshore sector. Another noteworthy method is the TIEOF, which, despite its higher computational complexity and more advanced mathematics, is capable of addressing the bidimensional structure of the images.

Random Forest (RF) models are also particularly promising, especially for spatial interpolation tasks, primarily due to their capacity to incorporate information from a large number of covariates. Even with over 40 covariates, RF can automatically identify the most relevant ones. Neural Networks (NNs), on the other hand, are the most versatile of these methods, having been applied to spatial, temporal, and spatiotemporal interpolation, with different architectures favoured for specific applications. A key advantage of NNs in the offshore sector is their ability to handle nonlinear systems, such as ocean dynamics. Additionally, the typical overfitting problem associated with both NNs and RFs can be mitigated by employing ensemble methods, as demonstrated by Tapoglou et al. (2021) and Chen et al. (2019).

When considering studies that compare different techniques, it becomes evident that the DINEOF method consistently outperforms or at least matches the performance of other algorithms in satellite image reconstruction, while requiring significantly less computational time (Alvera-Azcárate et al., 2005; Stock et al.). However, no single algorithm is universally optimal for all scenarios. Studies that evaluate a wide range of algorithms across different regions and metrics frequently report that the best-performing method varies depending on the specific context (Stock et al.; Penalba et al., 2022; Diesing et al., 2014). Nevertheless, algorithms that frequently rank among the top performers or demonstrate robustness include Kriging, Random Forest (RF), and DINEOF. Hybrid methods, particularly those combining statistical and machine learning techniques, consistently outperform their standalone counterparts (Zhisong Huang et al., 2012; Barth et al., 2022), as they harness the strengths of multiple approaches. A noteworthy observation from several case studies is that many algorithms struggle to effectively capture extreme values of the parameter of interest, while showing lower error rates for more typical values (Ahn et al., 2022), indicating that extreme values pose additional challenges in interpolation.

Fig. 10 shows how many articles are published each year, divided by the category of the interpolation used. It is evident that there has been a significant increase in the number of studies published in recent years, especially in the categories of ML and statistical methods. This trend reflects a growing interest and reliance on more sophisticated techniques for addressing spatio-temporal challenges in the offshore energy sector, particularly for metocean data.

Machine learning, in particular, has seen a sharp rise in the number of studies since around 2015. This surge can be attributed to the increasing accessibility of computational resources and the development of more advanced ML algorithms capable of handling large datasets. In the context of offshore measurements and metocean data, the complexity of ocean phenomena has driven researchers towards ML methods, which offer the ability to model non-linear interactions, to process vast amounts of satellite and in-situ data, and to provide more accurate predictions for oceanic conditions. The flexibility and adaptability of ML models make them highly suitable for environments characterised by uncertainty and variability, like the ocean.

Statistical methods have also shown an uptick in publications, albeit at a slower pace compared to ML. This increase highlights the continued importance of traditional statistical techniques, which remain foundational in the field. These methods offer valuable insights into the spatio-temporal interpolation of offshore data, often serving as a benchmark for evaluating the performance of newer approaches like ML.

Deterministic methods, on the other hand, have seen a relatively smaller increase rate, suggesting a shift away from simpler, less flexible approaches towards more robust techniques that can handle the complexities of real-world data. Similarly, hybrid methods, which combine various approaches, have maintained steady growth as they offer a way to balance the strengths and weaknesses of different techniques.

Overall, the data clearly indicate a paradigm shift towards more complex and data-driven methodologies, with ML at the forefront of this evolution. This trend is likely to continue as the offshore energy sector faces increasing demands for accurate, efficient, and scalable solutions for metocean data analysis.

One of the primary reasons for this increasing interest in ML is the rapid growth in data availability. As technological advances, particularly in satellite-based measurements and in-situ sensors, generate ever-larger datasets, traditional statistical methods face growing limitations. Many statistical techniques, especially those rooted in geostatistics

or linear regression, rely heavily on matrix inversion—a process that becomes computationally burdensome or even intractable as data scales up. This limitation is particularly severe for covariance-based geostatistical models, where the dimension of the covariance matrix grows with the number of observations, leading to cubic scaling in both computational cost and memory. In contrast, simpler regression models whose main cost depends on the number of predictors rather than on the number of data points can remain computationally affordable, unless they are enriched with complex basis expansions or spatial correlation structures. For instance, Kriging methods, widely used for spatial interpolation, require inverting a covariance matrix that scales with the number of available data points. As the amount of data increases, this requirement places a significant strain on computational resources, making these methods less feasible for large-scale applications. In large-data regimes, this creates a clear trade-off between accuracy, complexity and computational cost: exact Kriging-type approaches typically provide high accuracy and rigorous uncertainty quantification but suffer from poor scalability, so practitioners often resort to approximate or localised variants that reduce computational burden at the price of some loss in accuracy. Conversely, many ML methods — such as tree-based ensembles or neural networks trained with mini-batch optimisation — are designed to exploit parallel hardware and scale more gracefully with sample size, although they may require higher implementation and tuning effort and often provide less transparent uncertainty estimates than classical geostatistical models.

Indeed, Machine learning is inherently more resilient in handling vast datasets. ML models, particularly deep learning architectures such as neural networks, are designed to scale efficiently with data. Instead of relying on matrix inversions, these models leverage optimisation algorithms like gradient descent, which allow them to process large datasets while managing complexity. Moreover, ML models can handle the non-linearity and high-dimensionality typical of metocean data, offering significant advantages over traditional methods in capturing complex spatial and temporal patterns. As a result, they are increasingly favoured for large-scale offshore energy assessments, where vast amounts of data from satellites and buoys must be processed to make accurate predictions.

Another factor driving the rise in ML usage is the ability of these methods to learn directly from data, often without requiring the detailed domain knowledge or assumptions about underlying distributions that statistical methods necessitate. This adaptability makes ML approaches particularly suitable for dynamic environments like oceans, where conditions can change rapidly and unpredictably. In addition, ML's capacity to integrate different data sources — such as combining satellite imagery with in-situ measurements — allows researchers to construct models that are more robust and comprehensive than those based solely on statistical assumptions.

While statistical techniques remain valuable for smaller datasets or when interpretability is paramount, the offshore energy sector is increasingly turning to ML to meet the demands of modern data-rich environments. The growing use of hybrid approaches, which combine the strengths of statistical models with the scalability of ML, further underscores this shift. As the offshore industry continues to generate more abundant and complex data, it is likely that the reliance on machine learning will continue to grow, driven by the need for methods that can handle this data effectively and deliver precise insights for resource assessment and environmental monitoring.

6. Conclusions

Based on the reviewed literature, several key conclusions can be drawn. Machine learning (ML) techniques are emerging as highly promising for interpolation and prediction tasks related to offshore parameters and satellite data. A substantial proportion of the reviewed studies focus on ML approaches, reflecting the increasing prevalence of these techniques in the field. In comparisons across different categories, ML methods or hybrid techniques that integrate ML components frequently outperform others, as evidenced in the works of Li et al. (2011), Diesing et al. (2014), and Barth et al. (2022).

DINEOF (Data Interpolating Empirical Orthogonal Functions) consistently demonstrates superior performance when applied to satellite imagery, owing to its general applicability and interpretability. However, its reliance on satellite data limits its use to parameters where such data is available. For more complex scenarios, where preserving spatial structure or achieving higher accuracy is critical, TIEOF (Tensor Interpolating Empirical Orthogonal Functions) provides a viable alternative, though it comes with a higher computational cost, as highlighted by Chen et al. (2021), Mounet et al. (2023), and Kulikov et al. (2021).

Neural Networks (NNs) and Random Forests (RFs) are particularly effective in handling scenarios with a large number of covariates, with RFs demonstrating exceptional performance in such cases, as shown by Li et al. (2016). These algorithms have been proven to deliver surrogate models that achieve accuracy comparable to, or only slightly below, that of numerical simulations, but with a significantly reduced computational cost, as evidenced by Jörges et al. (2023).

Overall, the inter-class comparison provides a practical reading key for method selection in offshore applications: deterministic techniques are attractive for their simplicity but are frequently ill-suited to sparse, noisy observations and complex metocean dynamics, whereas statistical/geostatistical methods offer a robust compromise by combining flexible space–time estimation with transparent uncertainty handling. Consistently with the publication trends observed across categories, ML approaches are increasingly preferred when nonlinearity, high-dimensional covariates, or gridded/image satellite data are central to the task, and the most promising direction is often the use of hybrid pipelines that retain statistical uncertainty quantification while benefiting from ML representational power.

From a practitioner standpoint, these differences translate into distinct downstream uses. For early-stage site screening and zoning, spatial reconstructions and climatological maps are typically sufficient and often favour fast statistical baselines or scalable ML surrogates, whereas long-term resource assessment requires methods that preserve distributions and enable uncertainty-aware aggregation into yield metrics. In contrast, device-scale design and O&M inputs (e.g., high-frequency load time series, joint extremes, access/weather windows) demand spatio-temporal products that remain robust under sparsity and observation noise, making statistical–ML hybrid pipelines particularly attractive.

In more complex situations, it may be necessary to develop bespoke hybrid techniques that combine the strengths of different methods to address specific challenges, as demonstrated by Zhisong Huang et al. (2012) and Barth et al. (2020). Furthermore, as highlighted by Diesing et al. (2014) and Pirhooshyaran and Snyder (2020), it is clear that identifying a universally "optimal" interpolation technique for a given task is difficult. The performance of various methods can differ considerably based on factors such as the geographic region, time frame, and the evaluation metrics employed.

It is also important to acknowledge that the spatial and temporal resolutions used in the reviewed studies vary widely, and in some instances, they are not clearly defined. This variability may explain some of the contradictory findings reported across different investigations. Although statistical/geostatistical methods and ML-based approaches generally appear to be the most robust, the most suitable technique for a given application should always be identified through validation using a dedicated dataset. Future research should continue to explore these methodologies, particularly focusing on how they can be effectively integrated to address complex and domain-specific challenges.

CRediT authorship contribution statement

Leonardo Gambarelli: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Edoardo Pasta:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Paolo Brandimarte:** Writing – review & editing, Supervision. **Giuseppe Giorgi:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Leonardo Gambarelli reports financial support was provided by Ministero dell'Università e Ricerca. Edoardo Pasta reports financial support was provided by European Union. Giuseppe Giorgi reports financial support was provided by European Union. Paolo Brandimarte reports financial support was provided by European Union. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is part of the project NODES which has received funding from the MUR - M4C2 1.5 of PNRR with grant agreement no. ECS00000036. The work also received funding as part of the project AIMS: Artificial Intelligence to Monitor our Seas – funded by the European Union - NextGeneration EU - Ministry of University and Research - under the PRIN 2022 PNRR Call for Proposals (D.D.1409 of 14/09/2022).

References

Abdallah El-Reedy, Mohamed, 2020. Offshore Structures. Elsevier, ISBN: 9780128161913, http://dx.doi.org/10.1016/C2017-0-03037-9.

Abdessalem, Anis Ben, Dervilis, Nikolaos, Wagg, David J., Worden, Keith, 2017. Automatic kernel selection for Gaussian processes regression with approximate Bayesian computation and sequential Monte Carlo. Front. Built Environ. (ISSN: 22973362) 3 (August 2017), 1–13. http://dx.doi.org/10.3389/fbuil.2017.00052.

Ahn, Seongho, Tran, Trung Duc, Kim, Jongho, 2022. Systematization of short-term forecasts of regional wave heights using a machine learning technique and long-term wave hindcast. Ocean Eng. (ISSN: 00298018) 264, 112593. http://dx.doi.org/10.1016/j.oceaneng.2022.112593, URL https://linkinghub.elsevier.com/retrieve/pii/S0029801822018765.

Akter, Farzina, Bishop, Thomas F.A., Vervoort, R. Willem, 2021. Space-time modelling of groundwater level and salinity. Sci. Total Environ. (ISSN: 18791026) 776, 145865. http://dx.doi.org/10.1016/j.scitotenv.2021.145865.

Albuquerque, João, Antolínez, Jose A.A., Rueda, Ana, Méndez, Fernando J., Coco, Giovanni, 2018. Directional correction of modeled sea and swell wave heights using satellite altimeter data. Ocean. Model. 131 (September), 103–114. http://dx.doi.org/10.1016/j.ocemod.2018.09.001.

Alexandre, E., Cuadra, L., Nieto-Borge, J.C., Candil-García, G., del Pino, M., Salcedo-Sanz, S., 2015. A hybrid genetic algorithm-extreme learning machine approach for accurate significant wave height reconstruction. Ocean. Model. (ISSN: 14635003) 92, 115–123. http://dx.doi.org/10.1016/j.ocemod.2015.06.010.

Alifdini, Inovasita, Kismawardhani, Ratu Almira, Wirasatriya, Anindya, Sugianto, Denny Nugroho, Bin Yaakob, Omar, Widodo, Adrian Bela, 2019. Identification of tidal energy resources using satellite altimetry data for Indonesian Seas. In: 2018 4th International Symposium on Geoinformatics, ISyG 2018. IEEE, ISBN: 9781538670859, pp. 8–13. http://dx.doi.org/10.1109/ISYG.2018.8611798.

Alvera-Azcárate, A., Barth, A., Beckers, J.M., Weisberg, R.H., 2007. Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. J. Geophys. Res.: Ocean. (ISSN: 21699291) 112 (3), 1–11. http://dx.doi.org/10.1029/2006JC003660.

Alvera-Azcárate, Aida, Barth, Alexander, Parard, Gaëlle, Beckers, Jean Marie, 2016. Analysis of SMOS sea surface salinity data using DINEOF. Remote Sens. Environ. (ISSN: 00344257) 180, 137–145. http://dx.doi.org/10.1016/j.rse.2016.02.044.

Alvera-Azcárate, A., Barth, A., Rixen, M., Beckers, J.M., 2005. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea surface temperature. Ocean. Model. (ISSN: 14635003) 9 (4), 325–346. http://dx.doi.org/10.1016/j.ocemod.2004.08.001.

Andrawina, Yochi Okta, Kismawardhani, Ratu Almira, Rejeki, Hasti Amrih, 2021. DINEOF reconstruction for creating long-term cloud-free sea surface temperature data records: A Case study in Lombok Strait, Indonesia. IOP Conf. Ser.: Earth Environ. Sci. (ISSN: 17551315) 893 (1), http://dx.doi.org/10.1088/1755-1315/893/1/012069.

Ardhuin, Fabrice, Stopa, Justin E., Chapron, Bertrand, Collard, Fabrice, Husson, Romain, Jensen, Robert E., Johannessen, Johnny, Mouche, Alexis, Passaro, Marcello, Quartly, Graham D., Swail, Val, Young, Ian, 2019. Observing sea states. Front. Mar. Sci. 6, 124. http://dx.doi.org/10.3389/fmars.2019.00124.

Ashton, I., Van-Nieuwkoop-McCall, J.C.C., Smith, H.C.M., Johanning, L., 2014. Spatial variability of waves within a marine energy site using in-situ measurements and a high resolution spectral wave model. Energy (ISSN: 03605442) 66, 699–710. http://dx.doi.org/10.1016/j.energy.2013.12.065, URL https://linkinghub.elsevier.com/retrieve/pii/S0360544213011353.

Bai, Gen, Wang, Zhifeng, Zhu, Xianye, Feng, Yanqing, 2022. Development of a 2-D deep learning regional wave field forecast model based on convolutional neural network and the application in South China Sea. Appl. Ocean Res. (ISSN: 01411187) 118 (October 2021), 103012. http://dx.doi.org/10.1016/j.apor.2021.103012.

Barbariol, F., Benetazzo, A., Bertotti, L., Cavaleri, L., Durrant, T., McComb, P., Sclavo, M., 2019. Large waves and drifting buoys in the Southern Ocean. Ocean Eng. (ISSN: 0029-8018) 172, 817–828. http://dx.doi.org/10.1016/j.oceaneng.2018.12.011, URL https://www.sciencedirect.com/science/article/pii/S0029801818313878.

Barth, Alexander, Alvera-Azcárate, Aida, Licer, Matjaz, Beckers, Jean Marie, 2020. DINCAE 1.0: A convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. Geosci. Model. Dev. (ISSN: 19919603) 13 (3), 1609–1622. http://dx.doi.org/10.5194/gmd-13-1609-2020.

Barth, Alexander, Alvera-Azcárate, Aida, Troupin, Charles, Beckers, Jean-Marie, 2022. DINCAE 2.0: multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations. Geosci. Model. Dev. (ISSN: 1991-9603) 15 (5), 2183–2196. http://dx.doi.org/10.5194/gmd-15-2183-2022, URL https://gmd.copernicus.org/articles/15/2183/2022/.

Barth, Alexander, Azc, Aida Alvera, Joassin, Pascal, Troupin, Charles, 2008. Introduction to optimal interpolation and variational analysis. In: ReVision. SESAME Summer School, September.

Beckers, J.M., Rixen, M., 2003. EOF calculations and data filling from incomplete oceanographic datasets. J. Atmos. Ocean. Technol. (ISSN: 07390572) 20 (12), 1839–1856. http://dx.doi.org/10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2.

Bolton, Thomas, Zanna, Laure, 2019. Applications of deep learning to ocean data inference and subgrid parameterization. J. Adv. Model. Earth Syst. (ISSN: 19422466) 11 (1), 376–399. http://dx.doi.org/10.1029/2018MS001472.

Brewin, Robert J.W., Sathyendranath, Shubha, Kulk, Gemma, Rio, Marie-Hélène, Concha, Javier A., Bell, Thomas G., Bracher, Astrid, Fichot, Cédric, Frölicher, Thomas L., Galí, Martí, Hansell, Dennis Arthur, Kostadinov, Tihomir S., Mitchell, Catherine, Neeley, Aimee Renee, Organelli, Emanuele, Richardson, Katherine, Rousseaux, Cécile, Shen, Fang, Stramski, Dariusz, Tzortziou, Maria, Watson, Andrew J., Addey, Charles Izuma, Bellacicco, Marco, Bouman, Heather, Carroll, Dustin, Cetinić, Ivona, Dall'Olmo, Giorgio, Frouin, Robert, Hauck, Judith, Hieronymi, Martin, Hu, Chuanmin, Ibello, Valeria, Jönsson, Bror, Kong, Christina Eunjin, Kovač, Žarko, Laine, Marko, Lauderdale, Jonathan, Lavender, Samantha, Livanou, Eleni, Llort, Joan, Lorinczi, Larisa, Nowicki, Michael, Pradisty, Novia Arinda, Psarra, Stella, Raitsos, Dionysios E., Ruescas, Ana Belén, Russell, Joellen L., Salisbury, Joe, Sanders, Richard, Shutler, Jamie D., Sun, Xuerong, Taboada, Fernando González, Tilstone, Gavin H., Wei, Xinyuan, Woolf, David K., 2023. Ocean carbon from space: Current status and priorities for the next decade. Earth-Sci. Rev. (ISSN: 00128252) 240, 104386. http://dx.doi.org/10.1016/j.earscirev.2023.104386, URL https://linkinghub.elsevier.com/retrieve/pii/S0012825223000752.

Burke, Marshall, Driscoll, Anne, Lobell, David B., Ermon, Stefano, 2021. Using satellite imagery to understand and promote sustainable development. Science (ISSN: 10959203) 371 (6535), http://dx.doi.org/10.1126/science.abe8628.

Campanile, A., Piscopo, V., Scamardella, A., 2018. Mooring design and selection for floating offshore wind turbines on intermediate and deep water depths. Ocean Eng. (ISSN: 0029-8018) 148, 349–360. http://dx.doi.org/10.1016/j.oceaneng.2017.11.043, URL https://www.sciencedirect.com/science/article/pii/S0029801817307126.

Cavoretto, Roberto, 2021. Adaptive radial basis function partition of unity interpolation: A bivariate algorithm for unstructured data. J. Sci. Comput. (ISSN: 15737691) 87 (2), 1–24. http://dx.doi.org/10.1007/s10915-021-01432-z.

Centurioni, Luca R., Turton, Jonathan D., Lumpkin, Rick, Braasch, Lancelot, Brassington, Gary, Chao, Yi, Charpentier, Etienne, Chen, Zhaohui, Corlett, Gary, Dohan, Kathleen, Donlon, Craig, Gallage, Champika, Hormann, Verena, Ignatov, Alexander, Ingleby, Bruce, Jensen, Robert, Kelly-Gerreyn, Boris A., Koszalka, Inga M., Lin, Xiaopei, Lindstrom, Eric, Maximenko, Nikolai, Merchant, Christopher J., Minnett, Peter, O'Carroll, Anne G., Paluszkiewicz, Theresa, Poli, Paul, Poulain, Pierre, Reverdin, Gilles, Sun, Xiujun, Swail, Val, Thurston, Sidney, Wu, Lixin, Yu, Lisan, Wang, Bin, Zhang, Dongxiao, 2019. Global in-situ observations of essential climate and ocean variables at the air-sea interface. Front. Mar. Sci. (ISSN: 22967745) 6 (JUL), 1–23. http://dx.doi.org/10.3389/fmars.2019.00419.

Cervelli, G., Parinello, L., Moscoloni, C., Giorgi, G., 2022. Comparison of the ERA5 wave forecasting dataset against buoy record. Instrum. Mes. Métrol. 21, http://dx.doi.org/10.18280/i2m.210301.

Chen, Jiaxin, Ashton, Ian G.C., Pillai, Ajit C., 2022a. Wave record Gap-Filling using a Low-Rank tensor completion model. In: Volume 8: Ocean Renewable Energy. American Society of Mechanical Engineers, ISBN: 978-0-7918-8593-2, p. 10. http://dx.doi.org/10.1115/OMAE2022-79897, URL https://asmedigitalcollection.asme.org/OMAE/proceedings/OMAE2022/85932/V008T09A070/1148025.

Chen, Jiaxin, Ashton, Ian G.C., Steele, Edward C.C., Pillai, Ajit C., 2022b. A Real-Time spatiotemporal machine learning framework for the prediction of nearshore wave conditions. Artif. Intell. Earth Syst. 2 (1), http://dx.doi.org/10.1175/aies-d-22-0033.1.

Chen, Shuangling, Hu, Chuanmin, 2017. Estimating sea surface salinity in the northern Gulf of Mexico from satellite ocean color measurements. Remote Sens. Environ. (ISSN: 00344257) 201 (September), 115–132. http://dx.doi.org/10.1016/j.rse.2017.09.004.

Chen, Shuangling, Hu, Chuanmin, Barnes, Brian B., Xie, Yuyuan, Lin, Gong, Qiu, Zhongfeng, 2019. Improving ocean color data coverage through machine learning. Remote Sens. Environ. (ISSN: 00344257) 222 (December 2018), 286–302. http://dx.doi.org/10.1016/j.rse.2018.12.023.

Chen, Jiaxin, Pillai, Ajit C., Johanning, Lars, Ashton, Ian, 2021. Using machine learning to derive spatial wave data: A case study for a marine energy site. Environ. Model. Softw. 142 (April), 105066. http://dx.doi.org/10.1016/j.envsoft.2021.105066.

Chen, Xinyu, Yang, Jinming, Sun, Lijun, 2020. A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation. Transp. Res. C: Emerg. Technol. 117 (March), 102673. http://dx.doi.org/10.1016/j.trc.2020.102673.

Cheng, Y., Fu, L., Dai, S.S., Collu, M., Cui, L., Yuan, Z.M., Incecik, A., 2022a. Experimental and numerical analysis of a hybrid WEC-breakwater system combining an oscillating water column and an oscillating buoy. Renew. Sustain. Energy Rev. 169, 112909.

Cheng, Y., Fu, L., Dai, S.S., Collu, M., Ji, C.Y., Yuan, Z.M., Incecik, A., 2022b. Experimental and numerical investigation of WEC-type floating breakwaters: a single-pontoon oscillating buoy and a dual-pontoon oscillating water column. Coast. Eng. 177, 104188.

Cheng, Y., Gong, J., Zhang, J., 2024a. Hydrodynamic investigation on a single-point moored offshore cage–wave energy converter hybrid system. Ocean Eng. 299, 116848.

Cheng, Y., Liu, W., Dai, S., Yuan, Z., Incecik, A., 2024b. Wave energy conversion by multi-mode exciting wave energy converters arrayed around a floating platform. Energy 313, 133621.

Christianson, Ryan B., Pollyea, Ryan M., Gramacy, Robert B., 2022. Machine learning and geospatial methods for large-scale mining data. ArXiv 1–35, URL http://arxiv.org/abs/2207.10138.

Chutsagulprom, N., Chaisee, K., Wongsaijai, B., Inkeaw, P., Oonariya, C., 2022. Spatial interpolation methods for estimating monthly rainfall distribution in Thailand. Theor. Appl. Climatol. (ISSN: 0177-798X) 148 (1–2), 317–328. http://dx.doi.org/10.1007/s00704-022-03927-7, URL https://link.springer.com/10.1007/s00704-022-03927-7.

Contractor, Steefan, Roughan, Moninya, 2021. Efficacy of feedforward and LSTM neural networks at predicting and gap Filling Coastal Ocean timeseries: Oxygen, nutrients, and temperature. Front. Mar. Sci. (ISSN: 22967745) 8 (May), 1–17. http://dx.doi.org/10.3389/fmars.2021.637759.

Cressie, Noel, 1993. Statistics for Spatial Data, Revised ed. Wiley, New York, ISBN: 978-0-471-00255-0.

Cuadra, L., Salcedo-Sanz, S., Nieto-Borge, J.C., Alexandre, E., Rodríguez, G., 2016. Computational intelligence in wave energy: Comprehensive review and case study. Renew. Sustain. Energy Rev. (ISSN: 18790690) 58, 1223–1246. http://dx.doi.org/10.1016/j.rser.2015.12.253.

Dabiri, John O., Howland, Michael F., Fu, Matthew K., Goldshmid, Roni H., 2023. Visual anemometry for physics-informed inference of wind. Nat. Rev. Phys. (ISSN: 2522-5820) 5 (10), 597–611. http://dx.doi.org/10.1038/s42254-023-00626-8.

Degré, Aurore, Tech, Gembloux Agro-bio, Passage, Soil-water Systems, 2015. Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale : a review PoPuPS | Different methods for spatial interpolation of rainfall data for .... Base (ISSN: 1780-4507) 17 (2013), 1–10, URL lake.

Dicati, Renato, 2017. Meteorological satellites. In: Stamping the Earth from Space. Springer International Publishing, Cham, ISBN: 978-3-319-20756-8, pp. 207–263. http://dx.doi.org/10.1007/978-3-319-20756-8_6.

Diesing, Markus, Green, Sophie L., Stephens, David, Lark, R. Murray, Stewart, Heather A., Dove, Dayton, 2014. Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. Cont. Shelf Res. 84, 107–119. http://dx.doi.org/10.1016/j.csr.2014.05.004.

Doherty, K.W., Frye, D.E., Liberatore, S.P., Toole, J.M., 1999. A Moored profiling instrument. J. Atmos. Ocean. Technol. 16 (11), 1816–1829. http://dx.doi.org/10.1175/1520-0426(1999)016<1816:AMPI>2.0.CO;2.

Driesenaar, Tilly, de Kloe, Jos, Belmonte Rivas, Maria, Giesen, Rianne, et al., 2024. Quality Information Document: Global Ocean Wind Production Centre (WIND_GLO_PHY_L3_NRT_012_002 and WIND_GLO_PHY_L3_MY_012_005). EUMETSAT OSI SAF and Copernicus Marine Service, URL https://documentation.marine.copernicus.eu/QUID/CMEMS-WIND-QUID-012-002-005.pdf.

Ebden, Mark, 2015. Gaussian processes: A quick introduction. http://dx.doi.org/10.48550/arXiv.1505.02965, (August).

Elipot, Shane, Sykulski, Adam, Lumpkin, Rick, Centurioni, Luca, Pazos, Mayra, 2022. A dataset of hourly sea surface temperature from drifting buoys. Sci. Data 9, http://dx.doi.org/10.1038/s41597-022-01670-2.

Elken, Jüri, Zujev, Mihhail, She, Jun, Lagemaa, Priidik, 2019. Reconstruction of Large-Scale sea surface temperature and salinity fields using Sub-Regional EOF patterns from models. Front. Earth Sci. (ISSN: 22966463) 7 (September), 1–20. http://dx.doi.org/10.3389/feart.2019.00232.

Emmanouil, George, Galanis, George, Kallos, George, 2012. Combination of statistical Kalman filters and data assimilation for improving ocean waves analysis and forecasting. Ocean. Model. (ISSN: 14635003) 59–60, 11–23. http://dx.doi.org/10.1016/j.ocemod.2012.09.004,

Esmaeilbeigi, M., Chatrabgoun, O., Hosseinian-Far, A., Montasari, R., Daneshkhah, A., 2020. A low cost and highly accurate technique for big data spatial-temporal interpolation. Appl. Numer. Math. (ISSN: 01689274) 153, 492–502. http://dx.doi.org/10.1016/j.apnum.2020.03.009.

Eyre, J.R., Bell, W., Cotton, J., English, S.J., Forsythe, M., Healy, S.B., Pavelin, E.G., 2022. Assimilation of satellite data in numerical weather prediction. Part II: Recent years. Q. J. R. Meteorol. Soc. (ISSN: 1477870X) 148 (743), 521–556. http://dx.doi.org/10.1002/qj.4228.

Eyre, Jonathan Robert, English, Stephen J., Forsythe, Mary, 2020. Assimilation of satellite data in numerical weather prediction. Part I: The early years. Q. J. R. Meteorol. Soc. (ISSN: 1477870X) 146 (726), 49–68. http://dx.doi.org/10.1002/qj.3654.

Feng, Xi, Ma, Gangfeng, Su, Shih-feng, Huang, Chenfu, Boswell, Maura K, Xue, Pengfei, 2020. A multi-layer perceptron approach for accelerated wave forecasting in Lake Michigan. Ocean Eng. (ISSN: 0029-8018) 211 (April), 107526. http://dx.doi.org/10.1016/j.oceaneng.2020.107526.

Forrester, Alexander I.J., Keane, Andy J., 2009. Recent advances in surrogate-based optimization. Prog. Aerosp. Sci. (ISSN: 03760421) 45 (1–3), 50–79. http://dx.doi.org/10.1016/j.paerosci.2008.11.001.

Gambarelli, Leonardo, Pasta, Edoardo, Giorgi, Giuseppe, 2023. On spatial interpolation of ocean energy source variables: A comparative analysis. In: Proceedings of the European Wave and Tidal Energy Conference. Vol. 15, http://dx.doi.org/10.36688/ewtec-2023-619, URL https://submissions.ewtec.org/proc-ewtec/article/view/619.

Gao, Zhi-teng, Feng, Xing-ya, Zhang, Zi-tan, Liu, Zheng-liang, Gao, Xiao-xia, Zhang, Li-jun, Li, Shan, Li, Ye, 2022. A brief discussion on offshore wind turbine hydrodynamics problem. J. Hydrodyn. 34 (1), 15–30. http://dx.doi.org/10.1007/s42241-022-0002-y.

García-Pérez, Alfonso, 2021. New robust cross-variogram estimators and approximations of their distributions based on saddlepoint techniques. Mathematics (ISSN: 22277390) 9 (7), http://dx.doi.org/10.3390/math9070762.

Geurts, Pierre, Ernst, Damien, Wehenkel, Louis, 2006. Extremely randomized trees. Mach. Learn. 63 (1), 3–42. http://dx.doi.org/10.1007/s10994-006-6226-1, URL https://link.springer.com/article/10.1007/s10994-006-6226-1.

Giglio, Enrico, Novo, Riccardo, Mattiazzo, Giuliana, Fioriti, Davide, 2023. Reserve provision in the optimal planning of Off-Grid power systems: Impact of storage and renewable energy. IEEE Access 11, 100781–100797. http://dx.doi.org/10.1109/ACCESS.2023.3313979.

Goddijn-Murphy, Lonneke, Martín Míguez, Belén, McIlvenny, Jason, Gleizon, Philippe, 2015. Wave energy resource assessment with altika satellite altimetry: A case study at a wave energy site. Geophys. Res. Lett. (ISSN: 00948276) 42 (13), 5452–5459. http://dx.doi.org/10.1002/2015GL064490, URL http://doi.wiley.com/10.1002/2015GL064490.

González, A., Dunning, P., Howard, I., McKee, K., Wiercigroch, M., 2021. Is wave energy untapped potential? Int. J. Mech. Sci. 205, 106544. http://dx.doi.org/10.1016/j.ijmecsci.2021.106544.

Görmüş, Tahsin, Ayat, Berna, Aydoğan, Burak, 2022. Statistical models for extreme waves: Comparison of distributions and Monte Carlo simulation of uncertainty. Ocean Eng. 248, 110820. http://dx.doi.org/10.1016/j.oceaneng.2022.110820.

Guillou, Nicolas, Lavidas, George, Chapalain, Georges, 2020. Wave energy resource assessment for exploitation-A review. J. Mar. Sci. Eng. 8 (9), http://dx.doi.org/10.3390/JMSE8090705.

Hamzi, B., Maulik, R., Owhadi, H., 2021. Simple, low-cost and accurate data-driven geophysical forecasting with learned kernels. Proc. R. Soc. A: Math. Phys. Eng. Sci. (ISSN: 1364-5021) 477 (2252), http://dx.doi.org/10.1098/rspa.2021.0326, URL https://royalsocietypublishing.org/doi/10.1098/rspa.2021.0326.

Harrison, R. Giles, 2014. Meteorological Measurements and Instrumentation. Wiley, ISBN: 9781118745809, http://dx.doi.org/10.1002/9781118745793.

Hastie, Jerome Friedman Trevor, Tibshirani, Robert, 2008. The elements of statistical learning.

Houghton, Isabel A., Hegermiller, Christie, Teicheira, Camille, Smit, Pieter B., 2022. Operational assimilation of spectral wave data from the sofar spotter network. Geophys. Res. Lett. (ISSN: 19448007) 49 (15), http://dx.doi.org/10.1029/2022GL098973.

Houghton, Isabel, Penny, Stephen Gregory, Hegermiller, Christie, Cesaretti, Moriah, Teicheira, Camille, Smit, Pieter Bart, 2023. Ensemble-Based data assimilation of significant wave height from sofar spotters and satellite altimeters with a global operational wave model. Ocean. Model. (ISSN: 1463-5003) 183 (January), 102200. http://dx.doi.org/10.1016/j.ocemod.2023.102200, URL http://dx.doi.org/10.22541/essoar.167458064.45317697/v1.

Huang, Weimin, Gill, Edward W., Wu, X., Li, L., 2012. Measurement of sea surface wind direction using bistatic High-Frequency radar. IEEE Trans. Geosci. Remote Sens. 50 (10), 4117–4122. http://dx.doi.org/10.1109/TGRS.2012.2188298.

Huang, Weimin, Liu, Xin, Gill, Edward W., 2017. Ocean wind and wave measurements using X-Band marine radar: A comprehensive review. Remote. Sens. 9 (12), 1261. http://dx.doi.org/10.3390/rs9121261.

Huang, Zhisong, Wang, Huizan, Zhang, Ren, 2012. An improved kriging interpolation technique based on SVM and its recovery experiment in oceanic missing data. Am. J. Comput. Math. 2012 (41276036), 56–60.

Hunt, Brian R., Kostelich, Eric J., Szunyogh, Istvan, 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Phys. D: Nonlinear Phenom. (ISSN: 01672789) 230 (1–2), 112–126. http://dx.doi.org/10.1016/j.physd.2006.11.008.

Jahanmard, Vahidreza, Delpeche-Ellmann, Nicole, Ellmann, Artu, 2022. Towards realistic dynamic topography from coast to offshore by incorporating hydrodynamic and geoid models. Ocean. Model. (ISSN: 14635003) 180, 102124. http://dx.doi.org/10.1016/j.ocemod.2022.102124, URL https://linkinghub.elsevier.com/retrieve/pii/S146350032200138X.

James, Scott C., Zhang, Yushan, Donncha, Fearghal O., 2018. A machine learning framework to forecast wave conditions. Coast. Eng. 137, 1–10. http://dx.doi.org/10.1016/j.coastaleng.2018.03.004.

Janjić, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L., Losa, S.N., Nichols, N.K., Potthast, R., Waller, J.A., Weston, P., 2018. On the representation error in data assimilation. Q. J. R. Meteorol. Soc. (ISSN: 1477870X) 144 (713), 1257–1278. http://dx.doi.org/10.1002/qj.3130.

Jensen, Robert Edward, Swail, Val, Bouchard, Richard Harry, 2021. Quantifying wave measurement differences in historical and present wave buoy systems. Ocean. Dyn. (ISSN: 16167228) 71 (6–7), 731–755. http://dx.doi.org/10.1007/s10236-021-01461-0.

Jia, Gaofeng, Taflanidis, Alexandros A., 2013. Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment. Comput. Methods Appl. Mech. Engrg. (ISSN: 00457825) 261–262, 24–38. http://dx.doi.org/10.1016/j.cma.2013.03.012.

Jones, Steve D., Le Quéré, Corinne, Rödenbeck, Christian, Manning, Andrew C., Olsen, Are, 2015. A statistical gap-filling method to interpolate global monthly surface ocean carbon dioxide data. J. Adv. Model. Earth Syst. (ISSN: 19422466) 7 (4), 1554–1575. http://dx.doi.org/10.1002/2014MS000416, URL http://doi.wiley.com/10.1002/2014MS000416.

Jörges, Christoph, Berkenbrink, Cordula, Gottschalk, Hanno, Stumpe, Britta, 2023. Spatial ocean wave height prediction with CNN mixed-data deep neural networks using random field simulated bathymetry. Ocean Eng. (ISSN: 00298018) 271 (June 2022), http://dx.doi.org/10.1016/j.oceaneng.2023.113699.

Kent, Elizabeth, 2010. The voluntary observing ship (VOS) scheme. In: Proceedings of OceanObs09 Conference. ISBN: 978-3-86987-200-1, pp. 518–528. http://dx.doi.org/10.5270/OceanObs09.cwp.48.

Kim, Jinpyo, Deutsch, Clayton V., 2022. Calculation and modeling of variogram anisotropy. Geostat. Lessons 1–9, URL https://geostatisticslessons.com/lessons/variogramanisotropy.

Knysh, Alexander, Drach, Andrew, Fredriksson, David, Dewhurst, Tobias, Tsukrov, Igor, 2022. Methodology for multidimensional approximation of current velocity fields around offshore aquaculture installations. Aquac. Eng. (ISSN: 0144-8609) 99, 102284. http://dx.doi.org/10.1016/j.aquaeng.2022.102284, URL https://www.sciencedirect.com/science/article/pii/S0144860922000607.

Konuk, Erim-Bora, Centeno-Telleria, Manu, Zarketa-Astigarraga, Ander, Aizpurua, Jose-Ignacio, Giorgi, Giuseppe, Bracco, Giovanni, Penalba, Markel, 2023. On the definition of a comprehensive Technology-Informed accessibility metric for offshore renewable energy site selection. J. Mar. Sci. Eng. (ISSN: 2077-1312) 11 (9), http://dx.doi.org/10.3390/jmse11091702, URL https://www.mdpi.com/2077-1312/11/9/1702.

Krasnopolsky, Vladimir, Nadiga, Sudhir, Mehra, Avichal, Bayler, Eric, Behringer, David, 2016. Neural networks technique for filling gaps in satellite measurements: Application to ocean color observations. Comput. Intell. Neurosci. (ISSN: 16875273) 2016, http://dx.doi.org/10.1155/2016/6156513.

Kulikov, Leonid, Inkova, Natalia, Cherniuk, Daria, Teslyuk, Anton, Namsaraev, Zorigto, 2021. TIEOF: Algorithm for recovery of missing multidimensional satellite data on water bodies based on Higher-Order tensor decompositions. Water (ISSN: 2073-4441) 13 (18), 2578. http://dx.doi.org/10.3390/w13182578, URL https://www.mdpi.com/2073-4441/13/18/2578.

Kupilik, Matthew, 2016. An introduction to Gaussian processes for spatial data (predictions!).

Kutz, J. Nathan, Brunton, Steven L., 2008. Data driven science and engineering.

Li, Jin, 2019. A Critical Review of Spatial Predictive Modeling Process in Environmental Sciences with Reproducible Examples in R. MDPI, http://dx.doi.org/10.3390/app9102048.

Li, Yizhen, He, Ruoying, 2014. Spatial and temporal variability of SST and ocean color in the gulf of Maine based on cloud-free SST and chlorophyll reconstructions in 2003–2012. Remote Sens. Environ. (ISSN: 00344257) 144, 98–108. http://dx.doi.org/10.1016/j.rse.2014.01.019.

Li, Jin, Heap, Andrew D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. Ecol. Inform. 6 (3–4), 228–241. http://dx.doi.org/10.1016/j.ecoinf.2010.12.003.

Li, Jin, Heap, Andrew D., 2014. Spatial interpolation methods applied in the environmental sciences: A review. Environ. Model. Softw. (ISSN: 13648152) 53, 173–189. http://dx.doi.org/10.1016/j.envsoft.2013.12.008.

Li, Jin, Heap, Andrew D., Potter, Anna, Huang, Zhi, Daniell, James J., 2011. Can we improve the spatial predictions of seabed sediments? A case study of spatial interpolation of mud content across the southwest Australian margin. Cont. Shelf Res. (ISSN: 02784343) 31 (13), 1365–1376. http://dx.doi.org/10.1016/j.csr.2011.05.015.

Li, Jin, Justy, P., Siwabessy, W., Tran, Maggie, Huang, Zhi, Heap, Andrew D., 2013. Predicting seabed hardness using random forests in R. Data Mining Applications with R. (1984), pp. 299–329. http://dx.doi.org/10.1016/B978-0-12-411511-8.00011-6.

Li, Jin, Tran, Maggie, Siwabessy, Justy, 2016. Selecting optimal random forest predictive models: A case study on predicting the spatial distribution of seabed hardness. PLoS One (ISSN: 19326203) 11 (2), 1–29. http://dx.doi.org/10.1371/journal.pone.0149089.

Li, Zheng, Wan, Bingcheng, Duan, Zexia, He, Yuanhong, Yu, Yingxin, Chen, Huansang, 2023. Evaluation of HY-2C and CFOSAT satellite retrieval offshore wind energy using weather research and forecasting (WRF) simulations. Remote. Sens. (ISSN: 2072-4292) 15 (17), 4172. http://dx.doi.org/10.3390/rs15174172, URL https://www.mdpi.com/2072-4292/15/17/4172.

Liu, Chunli, Li, Jiepeng, Ma, Wenjuan, Tang, Xue, Zhang, Xutao, Wang, Sufen, Tang, Danling, 2023. Progress of research on satellite remote sensing application in oceanography: A case study in China. Reg. Stud. Mar. Sci. (ISSN: 23524855) 103055. http://dx.doi.org/10.1016/j.rsma.2023.103055, URL https://linkinghub.elsevier.com/retrieve/pii/S2352485523002451.

Liu, Xiaoming, Wang, Menghua, 2018. Gap filling of missing data for VIIRS global ocean color products using the DINEOF method. IEEE Trans. Geosci. Remote Sens. (ISSN: 01962892) 56 (8), 4464–4476. http://dx.doi.org/10.1109/TGRS.2018.2820423.

Liu, Xiaoming, Wang, Menghua, 2019. Filling the gaps of missing data in the merged VIIRS SNPP/NOAA-20 ocean color product using the DINEOF method. Remote. Sens. (ISSN: 20724292) 11 (2), http://dx.doi.org/10.3390/rs11020178.

Liu, Xiaoming, Wang, Menghua, 2022. Global daily gap-free ocean color products from multi-satellite measurements. Int. J. Appl. Earth Obs. Geoinf. (ISSN: 15698432) 108, 102714. http://dx.doi.org/10.1016/j.jag.2022.102714, URL https://linkinghub.elsevier.com/retrieve/pii/S030324342200040X.

Liu, Xin, Xia, Jianhong, Gunson, Jim, Wright, Graeme, Arnold, Lesley, 2014. Comparison of wave height interpolation with wavelet refined cubic spline and fractal methods. Ocean Eng. (ISSN: 00298018) 87, 136–150. http://dx.doi.org/10.1016/j.oceaneng.2014.05.013.

Liu, Jincan, Xu, Bin, Wang, Jichao, 2023. Ensemble-based assimilation of wave model predictions: Contrasting the impact of assimilation in nearshore and offshore forecasting at different distances from assimilated data. Appl. Ocean Res. (ISSN: 01411187) 140, 103726. http://dx.doi.org/10.1016/j.apor.2023.103726, URL https://linkinghub.elsevier.com/retrieve/pii/S0141118723002675.

López-Queija, Javier, Robles, Eider, Jugo, Josu, Alonso-Quesada, Santiago, 2022. Review of control technologies for floating offshore wind turbines. Renew. Sustain. Energy Rev. (ISSN: 1364-0321) 167, 112787. http://dx.doi.org/10.1016/j.rser.2022.112787, URL https://www.sciencedirect.com/science/article/pii/S1364032122006712.

Lumbreras, S., Ramos, A., 2013. Offshore wind farm electrical design: a review. Wind. Energy 16 (3), 459–473. http://dx.doi.org/10.1002/we.1498.

Luo, W., Taylor, M.C., Parker, S.R., 2008. A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. Int. J. Climatol. (ISSN: 08998418) 28 (7), 947–959. http://dx.doi.org/10.1002/joc.1583, URL https://onlinelibrary.wiley.com/doi/10.1002/joc.1583.

Ma, Jun, Ding, Yuexiong, Cheng, Jack C.P., Jiang, Feifeng, Wan, Zhiwei, 2019. A temporal-spatial interpolation and extrapolation method based on geographic long Short-Term Memory neural network for PM2.5. J. Clean. Prod. (ISSN: 09596526) 237, 117729. http://dx.doi.org/10.1016/j.jclepro.2019.117729.

Mackay, Ed, de Hauteclocque, Guillaume, Vanem, Erik, Jonathan, Philip, 2021. The effect of serial correlation in environmental conditions on estimates of extreme events. Ocean Eng. 242, 110092. http://dx.doi.org/10.1016/j.oceaneng.2021.110092.

Magnusson, Anne Karin, Jensen, Robert, Swail, Val, 2021. Spectral shapes and parameters from three different wave sensors. Ocean. Dyn. (ISSN: 16167228) 71 (9), 893–909. http://dx.doi.org/10.1007/s10236-021-01468-7.

Mahmoodi, Kumars, Ghassemi, Hassan, Razminia, Abolhassan, 2019. Temporal and spatial characteristics of wave energy in the Persian Gulf based on the ERA5 reanalysis dataset. Energy 187, http://dx.doi.org/10.1016/j.energy.2019.115991.

Manzhos, Sergei, Ihara, Manabu, 2023. On the optimization of hyperparameters in Gaussian process regression with the help of low-order high-dimensional model representation. J. Math. Chem. 14.

Mao, Shun, He, Ruoying, Bane, John, Gawarkiewicz, Glen, Todd, Robert E., 2023. A data-assimilative modeling investigation of Gulf Stream variability. Deep. Sea Res. II: Top. Stud. Ocean. (ISSN: 09670645) 211, 105319. http://dx.doi.org/10.1016/j.dsr2.2023.105319, URL https://linkinghub.elsevier.com/retrieve/pii/S0967064523000693.

Martinez-Iturricastillo, Nahia, Nic Guidhir, Meabh, Ulazia, Alain, Ringwood, John V., 2025. Extreme wave analysis for marine renewable energies in Ireland. Energy Convers. Manag.: X 26, 100972. http://dx.doi.org/10.1016/j.ecmx.2025.100972.

McLeod, Iain, Ringwood, John V., 2022. Powering data buoys using wave energy: a review of possibilities. J. Ocean. Eng. Mar. Energy (ISSN: 2198-6444) 8 (3), 417–432. http://dx.doi.org/10.1007/s40722-022-00240-3.

McPhaden, Michael J., Connell, Kenneth J., Foltz, Gregory R., Perez, Renellys C., Grissom, Karen, 2023. Tropical ocean observations for weather and climate: A decadal overview of the global tropical Moored Buoy Array. Oceanography http://dx.doi.org/10.5670/oceanog.2023.211.

Medina-Lopez, E., McMillan, D., Lazic, J., Hart, E., Zen, S., Angeloudis, A., Bannon, E., Browell, J., Dorling, S., Dorrell, R.M., Forster, R., Old, C., Payne, G.S., Porter, G., Rabaneda, A.S., Sellar, B., Tapoglou, E., Trifonova, N., Woodhouse, I.H., Zampollo, A., 2021. Satellite data for the offshore renewable energy sector: Synergies and innovation opportunities. Remote Sens. Environ. (ISSN: 00344257) 264, 112588. http://dx.doi.org/10.1016/j.rse.2021.112588, URL https://linkinghub.elsevier.com/retrieve/pii/S0034425721003084.

Medina-Lopez, Encarni, Ureña-Fuentes, Leonardo, 2019. High-resolution sea surface temperature and salinity in coastal areas worldwide from raw satellite data. Remote. Sens. (ISSN: 20724292) 11 (19), http://dx.doi.org/10.3390/rs11192191.

Menezes, Raquel, Garcia-Soidán, Pilar, Febrero-Bande, Manuel, 2008. A kernel variogram estimator for clustered data. Scand. J. Stat. (ISSN: 03036898) 35 (1), 18–37. http://dx.doi.org/10.1111/j.1467-9469.2007.00566.x.

Meunier, Thomas, Pérez-Brunius, Paula, Bower, Amy, 2022. Reconstructing the Three-Dimensional structure of loop current rings from satellite altimetry and in situ data using the gravest empirical modes method. Remote. Sens. (ISSN: 20724292) 14 (17), http://dx.doi.org/10.3390/rs14174174.

Micallef, Daniel, Rezaeiha, Abdolrahim, 2021. Floating offshore wind turbine aerodynamics: Trends and future challenges. Renew. Sustain. Energy Rev. 152, 111696. http://dx.doi.org/10.1016/j.rser.2021.111696.

Mo, Zonglai, Li, Kexiang, Xie, Kefeng, Li, Jun, Li, Yanjun, 2023. Offshore, unmanned Auto-Leveling Sea-Surface drifting platform with compact size. Mar. Sci. Eng. 11 (3), 506. http://dx.doi.org/10.3390/jmse11030506.

Modi, Aditi, Roxy, M.K., Ghosh, Subimal, 2022. Gap-filling of ocean color over the tropical Indian Ocean using Monte-Carlo method. Sci. Rep. (ISSN: 20452322) 12 (1), 1–12. http://dx.doi.org/10.1038/s41598-022-22087-2.

Moreira, Alberto, Prats-Iraola, Pau, Younis, Marwan, Krieger, Gerhard, Hajnsek, Irena, Papathanassiou, Konstantinos P., 2013. A tutorial on synthetic aperture radar. IEEE Geosci. Remote. Sens. Mag. 1 (1), 6–43. http://dx.doi.org/10.1109/MGRS.2013.2248301.

Mounet, Raphaël E G, Chen, Jiaxin, Nielsen, Ulrik D, Brodtkorb, Astrid H, Pillai, Ajit C, Ashton, Ian G C, Steele, Edward C C, 2023. Deriving spatial wave data from a network of buoys and ships. Ocean Eng. (ISSN: 0029-8018) 281 (February), 114892. http://dx.doi.org/10.1016/j.oceaneng.2023.114892.

Murthy, K.S.R., Rahi, O.P., 2017. A comprehensive review of wind resource assessment. Renew. Sustain. Energy Rev. (ISSN: 18790690) 72 (July 2016), 1320–1342. http://dx.doi.org/10.1016/j.rser.2016.10.038.

Nencioli, Francesco, Quartly, Graham D., 2019. Evaluation of sentinel-3A wave height observations near the coast of southwest England. Remote. Sens. 11 (24), http://dx.doi.org/10.3390/rs11242998.

Ni, Jinyang, Feng, Jiajun, Sun, Runxia, Zhang, Yuanzhi, 2022. Assessing sea surface temperatures estimated from fused infrared and microwave data. Water (ISSN: 2073-4441) 14 (21), 3357. http://dx.doi.org/10.3390/w14213357, URL https://www.mdpi.com/2073-4441/14/21/3357.

Nieto Borge, Juan Carlos, Reichert, Klaus, Dittmer, Jörg, 1999. Use of nautical radar as a wave monitoring instrument. Coast. Eng. 37 (3), 331–342. http://dx.doi.org/10.1016/S0378-3839(99)00032-0.

NOAA/National Data Buoy Center, 2025. What are the sensor reporting, sampling, and accuracy specifications?. https://www.ndbc.noaa.gov/faq/rsa.shtml.

Oo, Ye Htet, Zhang, Hong, 2022. Spatial wave assimilation by integration of artificial neural network and numerical wave model. Ocean Eng. (ISSN: 00298018) 247, 110752. http://dx.doi.org/10.1016/j.oceaneng.2022.110752, URL https://linkinghub.elsevier.com/retrieve/pii/S0029801822002025.

Özger, Mehmet, 2009. Neuro-fuzzy approach for the spatial estimation of ocean wave characteristics. Adv. Eng. Softw. 40 (9), 759–765. http://dx.doi.org/10.1016/j.advengsoft.2009.02.004.

Park, Jinku, Kim, Jeong Hoon, Kim, Hyun Cheol, Kim, Bong Kuk, Bae, Dukwon, Jo, Young Heon, Jo, Naeun, Lee, Sang Heon, 2019. Reconstruction of ocean color data using machine learning techniques in polar regions: Focusing on off Cape Hallett, Ross Sea. Remote. Sens. (ISSN: 20724292) 11 (11), http://dx.doi.org/10.3390/rs11111366.

Pasta, E., Carapellese, F., Faedo, N., Brandimarte, P., 2023. Data-driven control of wave energy systems using random forests and deep neural networks. Appl. Ocean Res. (ISSN: 01411187) 140, 103749. http://dx.doi.org/10.1016/j.apor.2023.103749.

Pasta, Edoardo, Faedo, Nicolás, Mattiazzo, Giuliana, Ringwood, John V., 2023. Towards data-driven and data-based control of wave energy systems: Classification, overview, and critical assessment. Renew. Sustain. Energy Rev. (ISSN: 1364-0321) 188, 113877. http://dx.doi.org/10.1016/j.rser.2023.113877.

Penalba, Markel, Aizpurua, Jose Ignacio, Martinez-Perurena, Ander, Iglesias, Gregorio, 2022. A data-driven long-term metocean data forecasting approach for the design of marine renewable energy systems. Renew. Sustain. Energy Rev. (ISSN: 18790690) 167 (May), http://dx.doi.org/10.1016/j.rser.2022.112751.

Penalba, Markel, Guo, Chao, Zarketa-Astigarraga, Ander, Cervelli, Giulia, Giorgi, Giuseppe, Robertson, Bryson, 2023. Bias correction techniques for uncertainty reduction of long-term metocean data for ocean renewable energy systems. Renew. Energy (ISSN: 0960-1481) 219, 119404. http://dx.doi.org/10.1016/j.renene.2023.119404, URL https://www.sciencedirect.com/science/article/pii/S0960148123013198.

Peres, D.J., Iuppa, C., Cavallaro, L., Cancelliere, A., Foti, E., 2015. Significant wave height record extension by neural networks and reanalysis wind data. Ocean. Model. (ISSN: 14635003) 94, 128–140. http://dx.doi.org/10.1016/j.ocemod.2015.08.002, URL https://linkinghub.elsevier.com/retrieve/pii/S1463500315001432.

Ping, Bo, Su, Fenzhen, Meng, Yunshan, 2016. An improved DINEOF algorithm for filling missing values in spatio-temporal sea surface temperature data. PLoS One (ISSN: 19326203) 11 (5), 1–12. http://dx.doi.org/10.1371/journal.pone.0155928.

Pirhooshyaran, Mohammad, Snyder, Lawrence V., 2020. Forecasting, hindcasting and feature selection of ocean waves via recurrent and sequence-to-sequence networks. Ocean Eng. (ISSN: 00298018) 207 (October 2019), 107424. http://dx.doi.org/10.1016/j.oceaneng.2020.107424.

Rajabi-Kiasari, Saeed, Delpeche-Ellmann, Nicole, Ellmann, Artu, 2023. Forecasting of absolute dynamic topography using deep learning algorithm with application to the Baltic Sea. Comput. Geosci. (ISSN: 00983004) 178, 105406. http://dx.doi.org/10.1016/j.cageo.2023.105406, URL https://linkinghub.elsevier.com/retrieve/pii/S0098300423001103.

Requia, Weeberb J., Coull, Brent A., Koutrakis, Petros, 2019. Evaluation of predictive capabilities of ordinary geostatistical interpolation, hybrid interpolation, and machine learning methods for estimating PM2.5 constituents over space. Environ. Res. (ISSN: 10960953) 175 (April), 421–433. http://dx.doi.org/10.1016/j.envres.2019.05.025.

Roemmich, Dean, Wilson, W., Gould, William, Owens, W., Traon, Pierre-Yves, Freeland, Howard, King, Brian, Wijffels, Susan, Sutton, Philip, Zilberman, Nathalie, 2022. The argo program. In: OceanObs09 Conference Proceedings. ISBN: 9780323904278, pp. 53–69. http://dx.doi.org/10.1016/B978-0-323-90427-8.00004-6.

Rolf, Esther, Proctor, Jonathan, Carleton, Tamma, Bolliger, Ian, Shankar, Vaishaal, Ishihara, Miyabi, Recht, Benjamin, Hsiang, Solomon, 2021. A generalizable and accessible approach to machine learning with global satellite imagery. Nat. Commun. (ISSN: 20411723) 12 (1), 1–11. http://dx.doi.org/10.1038/s41467-021-24638-z.

Rufino, Marta M., Albouy, Camille, Brind'Amour, Anik, 2021. Which spatial interpolators I should use? A case study applying to marine species. Ecol. Model. (ISSN: 03043800) 449 (April), 109501. http://dx.doi.org/10.1016/j.ecolmodel.2021.109501.

Rusu, Eugen, 2014. Evaluation of the wave energy conversion efficiency in various coastal environments. Energies (ISSN: 19961073) 7 (6), 4002–4018. http://dx.doi.org/10.3390/en7064002.

Rusu, Liliana, 2015. Assessment of the wave energy in the Black Sea based on a 15-year hindcast with data assimilation. Energies (ISSN: 19961073) 8 (9), 10370–10388. http://dx.doi.org/10.3390/en80910370.

Sánchez, Pedro José Bernalte, Papaelias, Mayorkinos, Márquez, Fausto Pedro García, 2020. Autonomous underwater vehicles: Instrumentation and measurements. IEEE Instrum. Meas. Mag. 23 (2), 105–114. http://dx.doi.org/10.1109/MIM.2020.9062680.

Saulquin, Bertrand, Gohin, Francis, Fanton d'Andon, Odile, 2019. Interpolated fields of satellite-derived multi-algorithm chlorophyll-a estimates at global and European scales in the frame of the European Copernicus-Marine environment monitoring service. J. Oper. Ocean. (ISSN: 17558778) 12 (1), 47–57. http://dx.doi.org/10.1080/1755876X.2018.1552358.

Scheu, Matti, Matha, Denis, Hofmann, Matthias, Muskulus, Michael, 2012. Maintenance strategies for large offshore wind farms. Energy Procedia (ISSN: 1876-6102) 24, 281–288. http://dx.doi.org/10.1016/j.egypro.2012.06.110.

Send, Uwe, Lankhorst, Matthias, 2011. The global component of the US ocean observatories initiative and the global OceanSITES project. In: OCEANS'11 MTS - IEEE KONA. http://dx.doi.org/10.23919/OCEANS.2011.6106959.

Setiyoko, A., Arymurthy, A.M., Basaruddin, T., Arief, R., 2019. Semivariogram fitting based on SVM and GPR for DEM interpolation. IOP Conf. Ser.: Earth Environ. Sci. (ISSN: 17551315) 311 (1), http://dx.doi.org/10.1088/1755-1315/311/1/012076.

Shadrin, Dmitrii, Nikitin, Artyom, Tregubova, Polina, Terekhova, Vera, Jana, Raghavendra, Matveev, Sergey, Pukalchik, Maria, 2021. An automated approach to groundwater quality monitoring-geospatial mapping based on combined application of gaussian process regression and bayesian information criterion. Water (Switz.) (ISSN: 20734441) 13 (4), 1–19. http://dx.doi.org/10.3390/w13040400.

Silva, Murilo T., Gill, Eric W., Huang, Weimin, 2018. An improved estimation and Gap-Filling technique for sea surface wind speeds using NARX Neural Networks. J. Atmos. Ocean. Technol. 35, 1521–1532. http://dx.doi.org/10.1175/JTECH-D-18-0001.1.

Sirjacobs, Damien, Alvera-Azcárate, Aida, Barth, Alexander, Park, Young Je, Nechad, Bouchra, Ruddick, Kevin, Beckers, Jean Marie, 2008. Cloud Filling of Total Suspended Matter, Cholorphyll and Sea Surface Temperature Remote Sensing Products by the Data Interpolation with Empirical Orthogonal Functions Methodology, Application to the BELCOLOUR-1 Database. European Space Agency, (Special Publication) ESA SP, (666 SP), ISSN 03796566.

Sluka, Travis C., Penny, Stephen G., Kalnay, Eugenia, Miyoshi, Takemasa, 2016. Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. Geophys. Res. Lett. (ISSN: 19448007) 43 (2), 752–759. http://dx.doi.org/10.1002/2015GL067238.

Smit, P.B., Houghton, I.A., Jordanova, K., Portwood, T., Shapiro, E., Clark, D., Sosa, M., Janssen, T.T., 2021. Assimilation of significant wave height from distributed ocean wave sensors. Ocean. Model. (ISSN: 14635003) 159, 101738. http://dx.doi.org/10.1016/j.ocemod.2020.101738, URL https://linkinghub.elsevier.com/retrieve/pii/S1463500320302407.

Stock, Andy, Subramaniam, Ajit, Dijken, Gert L Van, Wedding, Lisa M, Comparison of Cloud-Filling algorithms for marine satellite data. Remote. Sens..

Støle-Hentschel, Susanne, Nieto Borge, José Carlos, Trulsen, Karsten, 2021. The deconvolution as a method to deal with gaps in ocean wave measurements. Ocean Eng. (ISSN: 00298018) 219, 108373. http://dx.doi.org/10.1016/j.oceaneng.2020.108373, URL https://linkinghub.elsevier.com/retrieve/pii/S0029801820312804.

Sumer, Alf Tørum B. Mutlu, Whitehouse, Richard J.S., 2001. Scour around coastal structures: a summary of recent research. Coast. Eng. 44 (2), 153–190. http://dx.doi.org/10.1016/S0378-3839(01)00024-2.

Surisetty, V.V. Arun Kumar, Garlapati, Nagababu, Sharma, Raj Kumar Rashmi, 2020. Synergetic use of multiple scatterometers for offshore wind energy potential assessment. Ocean Eng. 196, http://dx.doi.org/10.1016/j.oceaneng.2019.106745.

Tahir, Anam, Haghbayan, Hashem, Böling, Jari M., Plosila, Juha, 2023. Energy-Efficient Post-Failure reconfiguration of swarms of unmanned aerial vehicles. IEEE Access 11, 24768–24779. http://dx.doi.org/10.1109/ACCESS.2022.3181244.

Tapoglou, Evdokia, Forster, Rodney M., Dorrell, Robert M., Parsons, Daniel, 2021. Machine learning for satellite-based sea-state prediction in an offshore windfarm. Ocean Eng. (ISSN: 00298018) 235 (June), 109280. http://dx.doi.org/10.1016/j.oceaneng.2021.109280.

Tymków, Przemysław, Jóźków, Grzegorz, Walicka, Agata, Karpina, Mateusz, Borkowski, Andrzej, 2019. Identification of water body extent based on remote sensing data collected with unmanned aerial vehicle. Water (Switz.) (ISSN: 20734441) 11 (2), http://dx.doi.org/10.3390/w11020338.

Uihlein, Andreas, Magagna, Davide, 2016. Wave and tidal current energy - A review of the current state of research beyond technology. Renew. Sustain. Energy Rev. (ISSN: 18790690) 58, 1070–1081. http://dx.doi.org/10.1016/j.rser.2015.12.284.

Urquhart, Erin A., Hoffman, Matthew J., Murphy, Rebecca R., Zaitchik, Benjamin F., 2013. Geospatial interpolation of MODIS-derived salinity and temperature in the Chesapeake Bay. Remote Sens. Environ. (ISSN: 00344257) 135, 167–177. http://dx.doi.org/10.1016/j.rse.2013.03.034.

Vieira, Filipe, Cavalcante, Georgenes, Campos, Edmo, Taveira-Pinto, Francisco, 2020. A methodology for data gap filling in wave records using Artificial Neural Networks. Appl. Ocean Res. (ISSN: 01411187) 98 (March), http://dx.doi.org/10.1016/j.apor.2020.102109.

Wan Nik, W.B., Sulaiman, O.O., Rosliza, R., Prawoto, Y., Muzathik, AM, 2011. Wave energy resource assessment and review of the technologies. Int. J. Energy Environ. 2 (6), 1101–1112.

Wang, Yali, Chaib-draa, Brahim, 2017. An online Bayesian filtering framework for Gaussian process regression: Application to global surface temperature analysis. Expert Syst. Appl. (ISSN: 09574174) 67, 285–295. http://dx.doi.org/10.1016/j.eswa.2016.09.018.

Wang, L., Kolios, A., Liu, X., Venetsanos, D., Cai, R., 2022. Reliability of offshore wind turbine support structures: A state-of-the-art review. Renew. Sustain. Energy Rev. (ISSN: 1364-0321) 161, 112250. http://dx.doi.org/10.1016/j.rser.2022.112250, URL https://www.sciencedirect.com/science/article/pii/S136403212200171X.

Wang, Sijia, Sun, Qun, Li, Siyu, Shen, Jiawei, Liu, Qian, Zhao, Liang, 2022. Interannual variability of dimethylsulfide in the Yellow Sea. J. Ocean. Limnol. (ISSN: 2096-5508) 40 (2), 551–562. http://dx.doi.org/10.1007/s00343-020-0480-0, URL https://link.springer.com/10.1007/s00343-020-0480-0.

Wang, Dong, Tang, Bo-Hui, Li, Zhao-Liang, 2024. Evaluation of five atmospheric correction algorithms for multispectral remote sensing data over Plateau Lake. Ecol. Inform. (ISSN: 1574-9541) 82, 102666. http://dx.doi.org/10.1016/j.ecoinf.2024.102666, URL https://www.sciencedirect.com/science/article/pii/S1574954124002085.

Wei, Zhangping, Davison, Andrew, 2022. A convolutional neural network based model to predict nearshore waves and hydrodynamics. Coast. Eng. (ISSN: 0378-3839) 171 (November 2021), 104044. http://dx.doi.org/10.1016/j.coastaleng.2021.104044.

Wei, Xianglin, Duan, Yuewei, Liu, Yongxue, Jin, Song, Sun, Chao, 2019. Onshore-offshore wind energy resource evaluation based on synergetic use of multiple satellite data and meteorological stations in jiangsu province, China. Front. Earth Sci. (ISSN: 20950209) 13 (1), 132–150. http://dx.doi.org/10.1007/s11707-018-0699-7.

Widén, Joakim, Carpman, Nicole, Castellucci, Valeria, Lingfors, David, Olauson, Jon, Remouit, Flore, Bergkvist, Mikael, Grabbe, Mårten, Waters, Rafael, 2015. Variability assessment and forecasting of renewables: A review for solar, wind, wave and tidal resources. Renew. Sustain. Energy Rev. (ISSN: 13640321) 44, 356–375. http://dx.doi.org/10.1016/j.rser.2014.12.019.

Williams, C.K.I., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning. MIT Press.

Wu, Mengmeng, Wang, Hui, Wan, Liying, Wang, Juanjuan, Wang, Yi, Wang, Jiuke, 2023. The impacts of the application of the ensemble optimal interpolation method in global ocean wave data assimilation. Atmosphere (ISSN: 2073-4433) 14 (5), 818. http://dx.doi.org/10.3390/atmos14050818, URL https://www.mdpi.com/2073-4433/14/5/818.

Wyatt, L.R., 1990. A relaxation method for integral inversion applied to HF radar measurement of the ocean wave directional spectrum. Int. J. Remote Sens. 11 (8), 1481–1494. http://dx.doi.org/10.1080/01431169008955106.

Xiao, Qingyang, Geng, Guannan, Cheng, Jing, Liang, Fengchao, Li, Rui, Meng, Xia, Xue, Tao, Huang, Xiaomeng, Kan, Haidong, Zhang, Qiang, He, Kebin, 2021. Evaluation of gap-filling approaches in satellite-based daily PM2.5 prediction models. Atmos. Environ. (ISSN: 18732844) 244 (June 2020), 117921. http://dx.doi.org/10.1016/j.atmosenv.2020.117921.

Xing, Qinwang, Yu, Haiqing, Wang, Hui, Ito, Shin-ichi, 2023. An improved algorithm for detecting mesoscale ocean fronts from satellite observations: Detailed mapping of persistent fronts around the China Seas and their long-term trends. Remote Sens. Environ. (ISSN: 00344257) 294, 113627. http://dx.doi.org/10.1016/j.rse.2023.113627, URL https://linkinghub.elsevier.com/retrieve/pii/S0034425723001785.

Yang, Xiu, Barajas-Solano, David, Tartakovsky, Guzel, Tartakovsky, Alexandre M., 2019. Physics-informed CoKriging: A Gaussian-process-regression-based multifidelity method for data-model convergence. J. Comput. Phys. 395, 410–431. http://dx.doi.org/10.1016/j.jcp.2019.06.041.

Yang, Zihao, Lin, Yifan, Dong, Sheng, 2022. Weather window and efficiency assessment of offshore wind power construction in China adjacent seas using the calibrated SWAN model. Ocean Eng. (ISSN: 00298018) 259, 111933. http://dx.doi.org/10.1016/j.oceaneng.2022.111933, URL https://linkinghub.elsevier.com/retrieve/pii/S0029801822012690.

Yang, Jie, Zhang, Jian, 2019. Validation of Sentinel-3A/3B satellite altimetry wave heights with Buoy and Jason-3 data. Sensors 19 (13), 2914. http://dx.doi.org/10.3390/s19132914.

Yavor, Kim Maya, Bach, Vanessa, Finkbeiner, Matthias, 2021. Resource assessment of renewable energy systems—a review. Sustain. (Switz.) 13 (11), http://dx.doi.org/10.3390/su13116107.

Yin, Xiaoqing, Yang, Dingtian, Du, Ranran, 2022. Fishery resource evaluation in shantou seas based on remote sensing and hydroacoustics. Fishes (ISSN: 2410-3888) 7 (4), 163. http://dx.doi.org/10.3390/fishes7040163, URL https://www.mdpi.com/2410-3888/7/4/163.

Zen, Simone, Hart, Edward, Medina-Lopez, Encarni, 2021. The use of satellite products to assess spatial uncertainty and reduce life-time costs of offshore wind farms. Clean. Environ. Syst. (ISSN: 26667894) 2 (December 2020), 100008. http://dx.doi.org/10.1016/j.cesys.2020.100008.

Zeng, Jialing, Liu, Chunli, Li, Xue, Zhao, Hui, Lu, Xiaoling, 2022. Comparative study of the variability and trends of phytoplankton biomass between spring and winter upwelling systems in the South China Sea. J. Mar. Syst. (ISSN: 09247963) 231, 103738. http://dx.doi.org/10.1016/j.jmarsys.2022.103738, URL https://linkinghub.elsevier.com/retrieve/pii/S0924796322000392.

Zeng, Jiye, Matsunaga, Tsuneo, Saigusa, Nobuko, Shirai, Tomoko, Nakaoka, Shin Ichiro, Tan, Zheng Hong, 2017. Technical note: Evaluation of three machine learning models for surface ocean CO2 mapping. Ocean. Sci. (ISSN: 18120792) 13 (2), 303–313. http://dx.doi.org/10.5194/os-13-303-2017.

Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M., Nakaoka, S., 2014. A global surface ocean fCO2 climatology based on a feed-forward neural network. J. Atmos. Ocean. Technol. (ISSN: 15200426) 31 (8), 1838–1849. http://dx.doi.org/10.1175/JTECH-D-13-00137.1.

Zhou, Guoqing, Huang, Jingjin, Yue, Tao, Luo, Qingli, Zhang, Guangyun, 2015. Temporal-spatial distribution of wave energy: A case study of Beibu Gulf, China. Renew. Energy (ISSN: 09601481) 74, 344–356. http://dx.doi.org/10.1016/j.renene.2014.08.014.

Zhou, Shuyi, Xie, Wenhong, Lu, Yuxiang, Wang, Yuanlin, Zhou, Yulong, Hui, Nian, Dong, Changming, 2021. ConvLSTM-Based wave forecasts in the South and East China Seas. Front. Mar. Sci. (ISSN: 22967745) 8 (June), 1–10. http://dx.doi.org/10.3389/fmars.2021.680079.