Ocean Systems Engineering, *Vol. 14, No. 1 (2024) 85-99*
https://doi.org/10.12989/ose.2024.14.1.085

Prediction of ocean surface current: Research status, challenges, and opportunities. A review

Ittaka Aldini[1,2], Adhistya E. Permanasari*[1], Risanuri Hidayat[1] and Andri Ramdhani[2]

[1]*Department of Electrical and Information Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[2]*Indonesian Agency for Meteorology, Climatology, and Geophysics, Jakarta, Indonesia*

(Received December 23 2023, Revised February 21, 2024, Accepted March 5, 2024)

Abstract. Ocean surface currents have an essential role in the Earth's climate system and significantly impact the marine ecosystem, weather patterns, and human activities. However, predicting ocean surface currents remains challenging due to the complexity and variability of the oceanic processes involved. This review article provides an overview of the current research status, challenges, and opportunities in the prediction of ocean surface currents. We discuss the various observational and modelling approaches used to study ocean surface currents, including satellite remote sensing, in situ measurements, and numerical models. We also highlight the major challenges facing the prediction of ocean surface currents, such as data assimilation, model-observation integration, and the representation of sub-grid scale processes. In this article, we suggest that future research should focus on developing advanced modeling techniques, such as machine learning, and the integration of multiple observational platforms to improve the accuracy and skill of ocean surface current predictions. We also emphasize the need to address the limitations of observing instruments, such as delays in receiving data, versioning errors, missing data, and undocumented data processing techniques. Improving data availability and quality will be essential for enhancing the accuracy of predictions. The future research should focus on developing methods for effective bias correction, a series of data pre-processing procedures, and utilizing combined models and xAI models to incorporate data from various sources. Advancements in predicting ocean surface currents will benefit various applications such as maritime operations, climate studies, and ecosystem management.

Keywords: ocean current modelling; ocean current prediction challenge; ocean current prediction opportunities; ocean surface current prediction

1. Introduction

Ocean surface currents take a significant role in the global climate pattern, influencing weather, the cycling of gases, and the distribution of heat and critical nutrients to marine ecosystems (Siedler *et al.* 2013, WMO 2015). Understanding and predicting these currents is essential for a wide range of applications, including accurate weather forecasting, efficient and safe navigation, reliable search and rescue operations, prevention of pollution, as well as for marine resource management (Dohan 2017, Życzkowski *et al.* 2019).

*Corresponding author, Dr., E-mail: adhistya@ugm.ac.id

http://www.techno-press.org/?journal=ose&subpage=7 ISSN: 2093-6702 (Print), 2093-677X (Online)

Ocean surface currents are driven by complex forces, including wind, tides, and thermohaline circulation (Bolaños *et al.* 2014, Rahmstorf 2003). The ability to accurately predict ocean surface currents is critical for various industries, including shipping, oil and gas exploration, and fisheries management (Bao *et al.* 2015). Additionally, knowledge of ocean surface currents was essential in optimizing the efficiency and sustainability of a renewable energy systems (Rotor *et al.* 2023). However, it is challenging to predict ocean surface currents because of the many unknowns inherent in these natural processes. These unknowns include the ocean's unique physical characteristics, the difficulty of representing the ocean's complex flows, the scarcity of direct in-situ measurements, and the constraints imposed by the availability of observational data ("Unravelling ENSO complexity", 2023).

Recent advances in technology and modelling techniques have enabled significant progress in the prediction of ocean surface currents (Muhamed Ali *et al.* 2021, Sinha and Abernathey 2021). This article reviews the recent findings of technological advancement and techniques of modelling for ocean surface current prediction and highlights the key challenges and opportunities in this field. This article is composed of four parts and is organized as follows: First part is about introductions of ocean surface current predictions; Secondly elaborates on the method of ocean surface current predictions; The third part is about the challenges and opportunities of ocean surface current predictions. The last part is a summary of the article in discussion, and the conclusion as the core of the article.

2. Method for ocean current predictions

Ocean surface current predictions can be made using various method. Generally, the method can be divided into three types: traditional (physics-based), statistical, and soft computing method. Each method has its own advantages and disadvantages.

2.1 Traditional methods: Numerical modelling and data assimilation

Traditional Methods of ocean current prediction are typically relying on physics-based computational models of the earth system to predict ocean currents (Ness *et al.* 2022, Willard *et al.* 2020). These methods are based on well-established principles, referred as general circulation models (GCMs) and are often computationally expensive (Song *et al.* 2022, Storto *et al.* 2020). In this section, the traditional methods include Numerical Modelling and Data Assimilation.

2.1.1 Numerical modelling

Numerical modelling is a widely used method for simulating ocean circulation and predicting surface currents (Hays 2017, Miller 2007). The numerical models describe the fields of velocity and density in the ocean that based on a set of hydrodynamical-thermodynamical equations, incorporate the laws of momentum, mass, and energy conservation (Kowalik and Murty 1993). A variety of numerical models are available, including the Princeton Ocean Model (POM) (Xu *et al.* 2015), Regional Ocean Modelling System (ROMS) (Shchepetkin and McWilliams 2005), the Finite Volume Coastal Ocean Model (FVCOM) (Chen *et al.* 2012), the HYbrid Coordinate Ocean Model (HYCOM) (Bleck, 2002), the Nucleus for European Modelling of the Ocean (NEMO) (Kärnä *et al.* 2021), Modular Ocean Model (MOM) (Griffies 2012) and the Navy Coastal Ocean Model (NCOM) for ocean models; WRF (Skamarock *et al.* 2019) for atmosphere model; WAM (Günther *et al.* 1992), Wavewatch III (Tolman 2009) and SWAN (Booij *et al.* 1996) for wave model. These models use mathematical equations to simulate the physics of ocean circulation, taking into account factors such as wind, tides, and density gradients. The timely and accurate monitoring of oceanic networks, encompassing both in situ and satellite observations, is crucial for the effective implementation of data-assimilation techniques to refine models and validate forecast products. The accuracy of predictions from these models can be improved by incorporating data from a variety of sources, including in situ measurements and satellite observations (Tonani *et al.* 2015).

2.1.2 Data assimilation

Data assimilation is another commonly used method for ocean current predictions. Ocean surface current prediction using Data Assimilation (DA) applies the estimation theory/control theory involving statistical and numerical methods. DA combines observational data with numerical simulations for efficient, accurate, and realistic ocean estimations. Data assimilation (DA) is a method of analysis that involves integrating observed data into a model state by utilizing consistency constraints with respect to the laws of time evolution and physical properties. DA has the capability to furnish a time series comprising four dimensions of dynamically adapted fields, thereby constituting comprehensive and high-resolution data sets (Bouttier and Courtier 2002, Stanev and Schulz-Stellenfleth 2014).

Prediction using DA is generally classified into two types: a) Operational Ocean Prediction (Numerical Weather Prediction), which involves progressively updating the model state to get the best feasible forecast; and b) State / Parameter Estimation, which focuses on hindcast/reanalysis. This method compares the models to the observed temporal development in order to recreate the past. (Villas Bôas *et al.* 2019).

There are two categories of DA approaches: sequential and non-sequential. Sequential methods only consider past observations until the time of analysis (Bouttier and Courtier 2002). Direct Insertion (DI), Newtonian Relaxation (NR), Optimal Interpolation (Oke *et al.* 2010), Kalman Filter (KF) (Chen *et al.* 2009), 3D-VAR (Li *et al.* 2015), Singular Evolutive extended Kalman (SEEK), and Singular Evolutive Interpolated Kalman (SEIK) are included in sequential methods. Meanwhile, in the non-sequential method, consider future data observation such as reanalysis (which usually requires continual model restarting and data ingestion). A non-sequential method such as Smoother Approaches includes the four-dimensional analysis (4D-VAR) (Sugiura *et al.* 2008), Ensemble Kalman smoother (EnKS) (Nerger *et al.* 2014), and Asynchronous Ensemble Kalman Filter (AEnKF) (Stanev and Schulz-Stellenfleth 2014).

The sources of data for DA systems range from satellite remote sensing observations, remotely sensed observations from coastal HF Radar, in situ observation instruments (profiling Argo, mooring arrays, CTD, and XBT) also from the observations from autonomous vehicles and ocean gliders. DA methods are capable of aggregating or downscaling the remote sensing data. Data assimilation can also be used to interpolate and extrapolate the observational data. By assimilating observational data into numerical models, researchers can improve the accuracy of ocean current forecasts. The data assimilation techniques have been shown to improve the accuracy of predictions in various oceanic regions (Storto *et al.* 2020).

2.2 Statistical method

Statistical methods for ocean surface current prediction typically involve the use of statistical techniques to analyze historical data and identify patterns and trends that can be used to make predictions about future currents (Allen *et al.* 2018, Egger and Carpi 2008, Sinha and Abernathey 2021). Unlike traditional physical-based methods, which rely on mathematical models of the physical processes that drive ocean currents, and soft computing methods, which use artificial intelligence techniques to learn from data and make predictions, statistical methods do not rely on a priori knowledge of the underlying physical processes, but rather on the analysis of data and the identification of patterns and relationships within the data (Lammers and Badia 2005, Mishra *et al.* 2019).

One example of a statistical method for ocean surface current prediction is the use of an Empirical Orthogonal Functions (EOF) as empirical statistical model that developed to relate the non-tidal motion of the ocean surface currents off the Oregon coast to forecasts of the coastal winds by Zelenke (2005) The empirical statistical model based on bilinear regression is then used to produce predictions of the surface currents that are evaluated for their agreement with measured currents. Correlations between the predictions and the measurements were found in excess of 0.7 in both the V and U components, and the average error in this prediction remained less than 4 cm/s out to 48 hours into the future.

Another example of a statistical method for ocean surface current prediction is the use of linear Auto-Regression (AR) statistical model. Frolov *et al*. (2012) used empirical orthogonal functions (EOFs) to capture spatial correlations in the HF-radar data and used a linear autoregression model to predict the temporal dynamics of the EOF coefficients. The developed statistical model was tested using historical observations of surface currents in Monterey Bay, California. The developed model is found to be capable of learning tidal variability directly and found to be more accurate than the existing operational model in Monterey Bay (GPO2009 empirical model), also better in capturing of the spatial and temporal correlations.

The other application of statistical approach was conducted in grid-based spatial ARIMA (auto-regressive integrated moving average) (Pongto *et al.* 2020). It estimated the short-term forecast values of Ocean Current Patterns. The implemented method showed that the approach outperformed other methods in V component prediction and a historical dataset of 1 day and 7h prior compared with different existing approaches.

In summary, statistical methods for ocean surface current prediction involve the use of statistical techniques to analyze historical data and identify patterns and trends that can be used to make predictions about future currents. While these methods do not rely on a priori knowledge of the underlying physical processes, they have been shown to be effective in predicting surface currents over short time horizons.

2.3 Soft computing methods

Soft Computing Methods typically rely on machine learning algorithms to learn patterns in data and make predictions. These methods are based on data-driven models and are often computationally efficient (Houssein *et al.* 2023, Pugliese *et al.* 2021, Sarker 2021). The most frequently used method of soft computing is Artificial Intelligence method (Sarker 2022).

2.3.1 Artificial intelligence

Artificial intelligence (AI) has been increasingly used in oceanography and marine science to understand better and predict changes in marine ecosystems (Song *et al.* 2023). AI is described as the science and engineering of intelligent machines, with a special emphasis on intelligent computer programs (Xu *et al.* 2021). By providing a more accurate and efficient approach to predicting ocean

Table 1 Summary of types and models for Ocean Current Prediction Method

Method of Ocean Current Prediction	Types	Models
Traditional	Numerical Modelling	POM (Xu *et al.* 2015), ROMS (Shchepetkin and McWilliams 2005), FVCOM (Chen *et al.* 2012), HYCOM (Bleck 2002), NEMO (Kärnä *et al.* 2021), MOM (Griffies, 2012), NCOM, WRF (Skamarock *et al.* 2019), WAM (Günther *et al.* 1992), WAVEWATCH III (Tolman 2009), SWAN (Booij *et al.* 1996)
	Data Assimilation	SEQUENTIAL (Direct Insertion (DI), Newtonian Relaxation (NR), Optimal Interpolation (OI) (Oke *et al.* 2010), Kalman Filter (KF) (Chen *et al.* 2009), 3D-VAR (Li *et al.* 2015)) and NON-SEQUENTIAL (4DVAR (Sugiura *et al.* 2008), EnKS (Nerger *et al.* 2014), AEnKF (Stanev and Schulz-Stellenfleth 2014))
Statistical	Statistical Models	Empirical Statistical Model (Zelenke 2005), AR Model (Frolov *et al.* 2012), Grid-based ARIMA (Pongto *et al.* 2020)
Soft Computing	Artificial Intelligence	Machine Learning (Temporal kNN, ANN) (Dauji *et al.* 2015, Jirakittayakorn *et al.* 2017, Moreno *et al.* 2022, Schultz *et al.* 2021, Sinha and Abernathey 2021)
		Deep Learning (CNN-GRU (Thongniran *et al.* 2019), LSTM (Bayindir 2019), STCANet (Xie *et al.* 2023), SDPNet (Zhang *et al.* 2023))

currents, AI has the potential to revolutionize the field of oceanography (Wen *et al.* 2021). The application of AI models for prediction and forecast has demonstrated greater performance in various areas, including weather forecasting, traffic forecasting, video prediction, energy estimation, and so on, compared to traditional statistical approaches (Xu *et al.* 2021). Since it can deal with unclear, partial, and even conflicting data, an AI approach to forecasting in the ocean environment has potential advantages over other approaches (Durazo and Baumgartner 2002).

The major parts of AI are Machine Learning (ML) and Deep Learning (DL), frequently used interchangeably in recent years. Artificial intelligence, or AI, is a subfield of computer science that aims to mimic human intelligence by developing computational capabilities that mimic human cognitive and logical abilities in order to tackle difficult issues. In order to accomplish this goal, AI develops computational systems capable of performing activities normally associated with human intelligence.

On the other hand, ML approaches attempt to discover relationships (such as pattern recognition) between input data and output data using or without mathematical representations of problems (algorithms). Once the machine-learning models have been adequately trained using the training dataset, decision-makers can retrieve the forecasting output values by feeding the forecasted input data into the well-trained models to perform the aforementioned tasks. A third terminology often written as DL, is also in use. Deep learning (DL) is a machine learning subfield that uses flexibility and scalability of neural networks trained with several processing layers (Tapeh and Naser 2023).

Machine Learning (ML) has grown rapidly in a variety of study domains, including computer vision (Khan and Al-Habsi 2020), natural language processing (Otter *et al.* 2020), science (Mjolsness and DeCoste 2001) and engineering (Willard *et al.* 2022). The literature, in particular, reveals that the use of ML techniques, ranging from traditional methods to deep neural networks, is present in practically all aspects of spatio-temporal problems. Adopting ML techniques improves forecasting of complex high-dimensional dynamics significantly. In recent years, machine learning techniques also have been used to predict ocean currents such as the utilization of Temporal kNN (Jirakittayakorn *et al.* 2017), and Artificial Neural Network (ANN) (Dauji *et al.* 2015, Kwon *et al.* 2023, Moreno *et al.* 2022, Sinha and Abernathey 2021). These techniques can be used to uncover patterns in observational data that typical numerical models cannot capture (Schultz *et al.* 2021). These methods were be able to extract the spatial and temporal characteristic ocean current patterns that may be challenging for conventional models (Li *et al.* 2021). This can improve the accuracy of predictions in coastal regions and areas with complex topography (Zandi *et al.* 2022).

Deep learning is the most rapidly expanding AI sub-domain (Szczepanski 2019). DL is a collection of approaches that make use of deep neural networks, or networks with two or more hidden layers. The capacity to tackle jobs utilizing the end-to-end method is the primary advantage of deep architectures. Because a signal or picture vector is utilized as an input to the network, and the network independently finds the regularities tying the input vector to the target variable, this approach lowers the requirements for preliminary data processing. The network selects the significant features, which is a time-consuming and complex procedure. This network's operation substantially simplifies the task of researcher. However, these benefits are only seen with a large enough amount of training data and a properly constructed neural network architecture.

One example of a deep learning application for ocean surface current prediction is the use of CNN-GRU spatio-temporal deep learning model to capture both spatial and temporal effect of ocean surface current (Thongniran *et al.* 2019). When the model evaluated and compared to the other eight existing models on the current data in the Gulf of Thailand from 2014 to 2016, it is and found that the CNN-GRU combination method outperforms almost all baselines in terms of the Root Mean Square (RMSE) as the measures of the average difference between a statistical model's predicted values and the actual values. Another deep learning prediction approach namely Long Short-Term Memory (LSTM) was applied to the ocean current velocity timeseries to find its prediction performance (Bayindir 2019). The algorithm was tested to four months experimental data collected by NOAA in Massachusetts Bay and found that the RMSE predicted time series for the ocean current velocity is less than 0.07 for both u and v ocean current surface components.

A similar ocean current pattern forecast value was conducted using spatio-temporal coupled attention deep network model STCANet (Xie *et al.* 2023). It utilizes the Spatial Channel Attention Module (SCAM) to capture the spatial correlations and dependencies, the Gated-Recurrent-Unit (GRU) to extract temporal relationships, and finally the nearest neighbour time attention module to extract the interdependences of ocean currents between adjacent times to improve the accuracy of ocean current prediction. A series of comparative experiments to the Dataset showed that the prediction quality of the model can outperforms the other benchmark models.

Another application of deep learning method was done using namely skipped dual path network (SDPNet) (Zhang *et al.* 2023). Specifically, SDPNet has a 1D CNN) module a recurrent neural network (RNN) module with a dual-path structure (LSTM-GRU) to enable mining the intrinsic change pattern contained in ocean surface current timeseries data. Experiments are conducted on the South China Sea utilizing the data set from ocean reanalysis dataset of South China Sea (REDOS). SDPNet can achieves average accuracy of 74% in predicting the coming 7 days of ocean surface current values and directions and performs better than state-of-the-art machine learning methods.

3. Limitation and challenges

3.1 Multiple sources of errors from numerical model

Regional ocean models are extremely useful tools for operational oceanography (Wilkin and Hunter 2013). However, the use of numerical models in computing ocean currents is unavoidably hampered by errors from many sources which can have a negative impact on model's performance and accuracy particularly in highly dynamic coastal locations where conditions vary fast (Hernandez-Lasheras *et al.* 2021).

The numerical model errors can be caused by both formulation and discretization problems. This is because the mathematical model can only be an approximation of the genuine processes in nature. Physical processes are typically complex, poorly understood, or, in some cases, altogether unknown, and so can only be described in a simplified form. Parts of the real dynamics are consequently ignored in the final mathematical model, introducing the first layer of error, the formulation error. The mathematical model is then discretized and approximated on a computer, introducing the second component of model errors, known as discretization error. This word refers to all errors that have their origins in the discretization of the continuous equations and boundary conditions of the mathematical model (Jan Ackmann 2017).

Other example of errors includes problems in properly specifying a model's initial and boundary conditions (Moore *et al.* 2011), faults in setting atmospheric parameters (wind, air temperature, and so on), uncertainties in flow forcing (tide, winds, pressure) (Karri *et al.* 2013), and uncertainties in bathymetric data. These inaccuracies are frequently accompanied with other flaws in the core of modelling (Fringer *et al.* 2019). These errors may accumulate over time and often deviate the model from the actual ocean trajectory.

3.2 Requirement of high computation

The field of ocean prediction has been dominated by numerical models for decades. Nonetheless, there are boundaries to what can be modelled numerically. To begin, a high-powered computing infrastructure capable of producing reliable future state predictions is essential for running numerical models. Typically, supercomputers are required to do the calculations necessary for producing high-precision worldwide forecasts. Second, the atmospheric forcing field is a crucial input for numerical models. Since weather forecasts is limited within two weeks, accurate long-range ocean prediction is challenging to obtain (Zhao *et al.* 2020).

Meanwhile, for Data Assimilation, although it obtains more reliable ocean state estimates by routinely constraining and adjust the model outputs with incoming ocean data through, however, DA also relies on supercomputer facilities to assimilate massive amounts of data gathered by monitoring systems in real-time to generate the advanced general circulation models (Hoteit *et al.* 2018).

3.3 Uncertainties factors for data assimilation

Although Data Assimilation can solve data availability problems by generating data at a specific spatial and temporal scale, DA has big issues that can affect the reliabilities of the prediction result. Data assimilation has uncertainties that can be inherited from the incomplete ocean models dynamics (Tokmakian and Challenor 2019), measurement sensors (Reichle 2008), and various parameters and inputs.

3.4 Limitation of AI model

There are promising potential for the widespread use of AI. However, there are a number of challenges, and resolving them would open up exciting new opportunities for the use of AI in predicting ocean surface currents. There are several limitations of AI Applications, including: 1) Requirement toward large amounts of data; 2) Difficulties in data collection and preliminary processing; 3) Insufficient labeled data and tedious data labeling; 4) Data biases and technical limitations; 5) The need of improve the model and accelerate the learning.

3.5 Limitation of instruments lead to limited data availability and data issues

In predictions, observational data from ocean observing instruments are used in several ways: (a) validating numerical model output, (b) employing observational data directly or via a short-term predictive model, or (c) assimilating observational data into ocean models that are then used for predictions.

Measurements of the ocean are collected and distributed with varying time lags from various observing platforms such as satellites, anchored surface and subsurface buoys, drifters, floats, dedicated manned and unmanned vehicles, research ships, and vessels of opportunity. These instruments contain a number of difficulties, including delayed and duplicate data receipts, versioning errors, missing data and metadata, and undocumented data processing techniques. Because of a lack of metadata, the accuracy of observational data is frequently unclear or underestimated within historical data records. This also prevents the implementation of appropriate bias-correction processes and may result in the incorrect definition of instrumental errors, which may compromise data quality.

On the other hand, each of the observing instruments has its own disadvantages. In-situ observing instruments, although they can generate higher data resolutions, these instruments, besides the high operational costs, are also lacking in providing measurements for larger areas. In addition, surface drifters, as drifting buoys, face some challenges in interpretations due to several factors, such as motions due to slip caused by windage, surface gravity wave rectification, and stokes drift (Lumpkin *et al.* 2017).

In the meantime, remote sensing observation instruments are at the forefront of many elements of global monitoring. They are less expensive, more uniform, cover more areas, and have the potential for synergistic data mining. Valid data acquisition is limited by a number of satellite constraints, and obtaining a continuous data series is seldom accomplished in practice. The altimeter observing system has two key drawbacks when it comes to tracking ocean surface currents: (1) it can only deduce the geostrophic component of the currents, and (2) in some regions, it can only do so for a limited range of spatial scales (Dotto *et al.* 2018).

In terms of high-frequency coastal radar (HF Radar), although generally having spatial resolution larger than satellite altimetry, due to long-range propagation circumstances, HF radars are susceptible to external Radio Frequency Interference (RFI) caused by diurnal fluctuations in the ionosphere, resulting in greater noise levels (Cook *et al.* 2008).

4. Opportunities for further research

4.1 Development of algorithm for ocean data pre-processing

The methods of ocean data pre-processing have significant influence on the creation of the models, and these approaches in turn depend heavily on the requirements of various applications and the cognitive background of the users. In order to improve the overall efficiency of model creation and reduce the large amount of time spent on ocean data preparation, future algorithm development (particularly employing AI) should be directed toward numerous data formats, especially from the different ocean observing equipment. To illustrate, the ocean current data can be obtained from HF Radar and Acoustic Doppler Current Profiler (ADCP), while the data format of that two devices are different. The data obtained from the HF Radar instrument has a format of either *.RUV, * (radial velocity) or *.TUV (total velocity). Besides, other observing instrument such as satellite, wave buoy, ocean glider, moored buoy, and Argo float, mostly stored the data in *.NC (NetCDF) format. In contrast, ADCP instrument has a format in binary form (*.ENR, *.ENS, *.ENX, *.N1R, *.STA, *.LOG). Uniformize all of those formats to ASCII format such as Comma Separated Value (*.CSV) or *.TXT is required to open and visualize data for further processing and analysis utilizing either Python, MATLAB, or R programming language. Additional algorithm for data pre-processing can be made for instance for quality control, data cleaning, data normalization, feature selection, and data normalization. Another additional approach can involve activities such as centralizing data collection and marking up data sets to make the data easily accessible.

4.2 Comparison of artificial intelligence model for prediction/forecast

Predicting the behaviour of a dynamic system is really challenging in general, and it becomes far more difficult when the prediction needs to be accomplished in real time. There are a lot of AI algorithmic models out there, but nobody can agree on which one would work best for a certain maritime data mining issue. Therefore, different training strategies, including tweaks to parameters and training procedures, are commonly utilized, and comprehensive comparisons between traditional model, statistical model, and soft computing model are required for model creation through a large number of experimental comparisons to build a more accurate model.

4.3 Development of a Model based on a specific purpose

With regard to various uses, novel deep neural network architectures need to be created. Dataset preparation is also required for many specific uses, such as particle tracking, eddy detection, storm surge detection, and seasonal predictions. The selection of predictive modelling techniques and the preparation of prediction models would differ based on the specific purpose of the prediction. For example, a model developed for particle tracking may require different input data and modelling techniques than a model developed for storm surge detection.

4.4 Incorporation from various sources of data

Missing or incomplete data can reduce the performance of operational forecasts/predictions. Incorporating data from various sources can complement/fill the missing gap in time-series data observations. For example, in-situ observations such as current meter, current profiler, and research vessels were usually taken to complement the data observation from remote sensing instruments such as synthetic aperture radar (SAR), satellite altimetry (SA), and coastal high-frequency radars. The temporal resolution of both SAR and SA with the broader coverage is typically only on the order of days to weeks. In comparison, the temporal resolution of the current meter with the narrow coverage is typically on the order of seconds to minutes. However, coastal High-Frequency Radar is usually preferred over other instruments because of its advantage of offering vast coverage up to

200 kilometers with a high spatial resolution of 1-2 kilometers and providing the observation data within minutes.

Incorporation of data sources from other observing instruments, not only can be used for improvement of the accuracy and reliability of the prediction, but also can be used for data evaluation and calibration.

4.5 Combination of different model / coupled model

While applying one single model can have several factors that reduce the accuracy of the predictions, using a combined model/coupled model, such as a physics-based numerical model combined with an artificial intelligence model, can improve the predictions/ forecast of ocean surface currents.

As a result of the wide variety of datasets, time-steps, prediction-ranges, settings, and performance metrics, it might be challenging to optimize a single ML/DL model for better forecasting results. Predictions of ocean surface currents may benefit from the development of hybrid machine-learning or deep-learning models or overarching prediction methodologies.

4.6 Implementation of explainable AI (xAI) model

The applicability of classical neural networks to problems of ocean prediction is constrained by the absence of uncertainty analysis. Since neural networks in ocean science must be trained on historical data before being applied to a changing ocean situation where the dynamics governing an area may have fundamentally changed, uncertainty measures are particularly critical for out-of-sample predictions. Decisions based on the predictions of neural networks may have far-reaching consequences, making the quantification of uncertainty within ocean applications crucial. In addition, the ocean scientific community may be sceptical about neural network forecasts due to the possibility of misleading correlations leading to nonphysical forecasts. If the reasoning behind a network's prediction is clear to those working in the field of climate science, then that network's prediction is more credible. Adding explainability methods to uncertainty analysis, however, is a somewhat unexplored field.

5. Discussions of key challenges and future prospects

There were challenges and limitations that researchers face when predicting ocean surface currents. One of the major challenges is the accumulation of errors from various sources in numerical models. These errors can arise from formulation and discretization problems, uncertainties in flow forcing and bathymetric data, defining initial and boundary conditions, and more. The errors tend to multiply in highly dynamic coastal regions where conditions shift frequently. Addressing these errors and improving the accuracy of numerical models will be crucial for enhancing the performance of ocean current predictions.

Another challenge is the high computational requirements of numerical models and data assimilation techniques. High-precision global forecasts often require supercomputers, limiting the accessibility and scalability of these models. Additionally, data assimilation relies on assimilating massive amounts of data gathered by monitoring systems in real-time, further increasing the computational burden. Future research should focus on developing advanced computing environments and optimization techniques to make the computation more efficient and enable long-term ocean predictions. Moreover, data assimilation introduces uncertainties, which can arise from incomplete ocean model dynamics, measurement sensors, and various parameters and inputs. Future research should aim to reduce these uncertainties through improved model calibration and validation techniques, as well as the development of more accurate and precise measurement sensors.

Addressing the limitations of observing instruments is also a significant challenge for researchers. Delays in receiving data from ocean observation instruments like satellites, moored buoys, drifters, and research ships, as well as versioning errors, missing data, and undocumented data processing techniques are common. Due to a lack of metadata, the accuracy of observational data is typically unknown or underestimated, which hinders efficient bias correction and affects data quality. Each observing instrument also has its own disadvantages, such as limited coverage, high operational costs, and challenges in data interpretation. Improving data availability and quality will be essential for enhancing the accuracy of predictions. Future research should focus on developing methods for effective bias correction, a series of data pre-processing procedures, and utilizing combined models and xAI models to incorporate data from various sources. Advancements in predicting ocean surface currents will benefit various applications such as maritime operations, climate studies, and ecosystem management.

Acknowledgments

The study for the research was fully funded by The Indonesian Agency of Meteorology Climatology and Geophysics (BMKG).

References

Allen, M., Dube, O.P., Solecki, W., Aragón-Durand, F., Cramer, W., Humphreys, S. and Kainuma, M. (2018), "Special Report: Global Warming of 1.5 C", *Intergovernmental Panel on Climate Change (IPCC)*.

Bao, Q., Dong, X., Zhu, D., Lang, S. and Xu, X. (2015), "The Feasibility of Ocean Surface Current Measurement Using Pencil-Beam Rotating Scatterometer", *IEEE J. Selec. Topics Appl. Earth Observ. Remote Sens.*, **8**(7), 3441-3451, https://doi.org/10.1109/JSTARS.2015.2414451.

Bayindir, C. (2019), "Predicting the Ocean Currents using Deep Learning", *ArXiv Preprint ArXiv:1906.08066*.

Bleck, R. (2002), "An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates", *Ocean Model.*, **4**(1), 55-88.

Bolaños, R., Tornfeldt Sørensen, J.V., Benetazzo, A., Carniel, S. and Sclavo, M. (2014), "Modelling ocean currents in the northern Adriatic Sea", *Continental Shelf Res.*, **87**, 54-72. https://doi.org/10.1016/j.csr.2014.03.009.

Booij, N., Holthuijsen, L.H. and Ris, R.C. (1996), "The" SWAN" wave model for shallow water", *Coast. Eng.,* 668-676.

Bouttier, F. and Courtier, P. (2002), "Data assimilation concepts and methods March 1999", *Meteorol. Training Course Lect. Series,* **718**, 59.

Chen, C., Beardsley, R.C., Cowles, G., Qi, J., Lai, Z., Gao, G., Stuebe, D., Xu, Q., Xue, P., and Ge, J. (2012), *An Unstructured-Grid, Finite-Volume Community Ocean Model: FVCOM User Manual*, Sea Grant College Program, Massachusetts Institute of Technology Cambridge….

Chen, Y., Oliver, D.S. and Zhang, D. (2009), "Data assimilation for nonlinear problems by ensemble Kalman filter with reparameterization", *J. Petroleum Sci, Eng.*, **66**(1-2), 1-14.

Cook, T., Hazard, L., Otero, M. and Zelenka, B. (2008), "Deployment and maintenance of a High-frequency Radar (HFR) for ocean surface current mapping: Best practices", University of California San Diego, Scripps Institution of Oceanography.

Dauji, S., Deo, M.C. and Bhargava, K. (2015), “Prediction of ocean currents with artificial neural networks”, *ISH J. Hydraulic Eng.*, **21**(1), 14-27.

Dohan, K. (2017), “Ocean surface currents from satellite data”, *J. Geophys. Res.: Oceans*, John Wiley & Sons, Ltd, **122**(4), 2647-2651. https://doi.org/10.1002/2017JC012961.

Dotto, T.S., Naveira Garabato, A., Bacon, S., Tsamados, M., Holland, P.R., Hooley, J., Frajka-Williams, E., *et al.* (2018), “Variability of the Ross Gyre, Southern Ocean: Drivers and responses revealed by satellite altimetry”, *Geophys. Res. Lett.*, **45**(12), 6195-6204.

Durazo, R. and Baumgartner, T.R. (2002), “Evolution of oceanographic conditions off Baja California: 1997–1999”, *Progress in Oceanography*, **54**(1), 7-31. https://doi.org/10.1016/S0079-6611(02)00041-1.

Egger, A.E. and Carpi, A. (2008), “Data: Analysis and interpretation”, *Visionlearning*, **1**.

Fringer, O.B., Dawson, C.N., He, R., Ralston, D.K. and Zhang, Y.J. (2019), “The future of coastal and estuarine modeling: Findings from a workshop”, *Ocean Model.*, **143**, 101458. https://doi.org/10.1016/j.ocemod.2019.101458.

Frolov, S., Paduan, J., Cook, M. and Bellingham, J. (2012), “Improved statistical prediction of surface currents based on historic HF-radar observations”, *Ocean Dynam.*, **62**, 1111-1122.

Griffies, S.M. (2012), “Elements of the modular ocean model (MOM)”, *GFDL Ocean Group Tech. Rep*, **7**(620), 47.

Günther, H., Hasselmann, S. and Janssen, P.A.E.M. (1992), “The WAM model cycle 4”.

Hays, G.C. (2017), “Ocean currents and marine life”, *Current Biology*, **27**(11), 470-473.

Hernandez-Lasheras, J., Mourre, B., Orfila, A., Santana, A., Reyes, E. and Tintoré, J. (2021), “Evaluating high-frequency radar data assimilation impact in coastal ocean operational modelling”, *Ocean Sci.*, **17**(4), 1157-1175. https://doi.org/10.5194/os-17-1157-2021.

Hoteit, I., Luo, X., Bocquet, M., Kohl, A. and Ait-El-Fquih, B. (2018), “Data assimilation in oceanography: Current status and new directions”, *New Frontiers in Operational Oceanography*, 465-512.

Houssein, E.H., Hosney, M.E., Emam, M.M., Younis, E.M.G., Ali, A.A. and Mohamed, W.M. (2023), “Soft computing techniques for biomedical data analysis: open issues and challenges”, *Artif. Intell. Rev.*, https://doi.org/10.1007/s10462-023-10585-2.

Jan Ackmann. (2017), *Error Estimation Algorithms for Ocean Models*, Universität Hamburg, Hamburg. https://doi.org/ 10.17617/2.2383652.

Jirakittayakorn, A., Kormongkolkul, T., Vateekul, P., Jitkajornwanich, K. and Lawawirojwong, S. (2017), “Temporal kNN for short-term ocean current prediction based on HF radar observations”, *Proceedings of the 14th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, https://doi.org/10.1109/JCSSE.2017.8025921.

Kärnä, T., Ljungemyr, P., Falahat, S., Ringgaard, I., Axell, L., Korabel, V., Murawski, J., *et al.* (2021), “Nemo-Nordic 2.0: Operational marine forecast model for the Baltic Sea”, *Geoscientific Model Development*, **14**(9), 5731-5749.

Karri, R.R., Badwe, A., Wang, X., El Serafy, G., Sumihar, J., Babovic, V. and Gerritsen, H. (2013), “Application of data assimilation for improving forecast of water levels and residual currents in Singapore regional waters”, *Ocean Dynam.*, **63**(1), 43-61. https://doi.org/10.1007/s10236-012-0584-y.

Khan, A.I. and Al-Habsi, S. (2020), “Machine learning in computer vision”, *Procedia Computer Science*, **167**, 1444-1451.

Kowalik, Z. and Murty, T.S. (1993), *Numerical Modeling of Ocean Dynamics*, **5**, World Scientific.

Kwon, D.S., Kim, S.J., Jin, C., Kim, M., Guha, A., Esenkov, O.E. and Ryu, S. (2023), “Inverse estimation of a vertical current velocity profile using motions of an FPSO and artificial neural network”, *Ocean Eng.*, **285**, 115343.

Lammers, W.J. and Badia, P. (2005), “Fundamentals of behavioral research”.

Li, T., Xuan, A. and Shen, L. (2021), “Study of nonlinear interaction between waves and ocean currents using high-fidelity simulation and machine learning”, *ArXiv Preprint ArXiv:2101.03439*.

Li, Z., McWilliams, J.C., Ide, K. and Farrara, J.D. (2015), “Coastal ocean data assimilation using a multi-scale three-dimensional variational scheme”, *Ocean Dynam.*, **65**, 1001-1015.

Lumpkin, R., Özgökmen, T. and Centurioni, L. (2017), "Advances in the application of surface drifters", *Annu. Rev. Mar. Sci.*, **9**, 59-81.
Miller, R.N. (2007), *Numerical Modeling of Ocean Circulation*, Cambridge University Press.
Mishra, P., Pandey, C.M., Singh, U., Keshri, A. and Sabaretnam, M. (2019), "Selection of appropriate statistical methods for data analysis", *Annals of Cardiac Anaesthesia*, Wolters Kluwer--Medknow Publications, **22**(3), 297.
Mjolsness, E. and DeCoste, D. (2001), "Machine learning for science: state of the art and future prospects", *Science*, American Association for the Advancement of Science, **293**(5537), 2051-2055.
Moore, A.M., Arango, H.G., Broquet, G., Powell, B.S., Weaver, A.T. and Zavala-Garay, J. (2011), "The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: Part I – System overview and formulation", *Progress in Oceanography*, **91**(1), 34-49. https://doi.org/10.1016/j.pocean.2011.05.004.
Moreno, F.M., Neto, L.A.S., Cozman, F.G., Dottori, M. and Tannuri, E.A. (2022), "Enhancing the forecast of ocean physical variables through physics informed machine learning in the Santos Estuary, Brazil", *OCEANS 2022 - Chennai*, 1-7, https://doi.org/10.1109/OCEANSChennai45887.2022.9775449.
Muhamed Ali, A., Zhuang, H., VanZwieten, J., Ibrahim, A.K. and Chérubin, L. (2021), "A deep learning model for forecasting velocity structures of the loop current system in the gulf of mexico", *Forecasting*, **3**(4), 934-953.
Nerger, L., Schulte, S. and Bunse-Gerstner, A. (2014), "On the influence of model nonlinearity and localization on ensemble Kalman smoothing", *Q. J. Roy. Meteorol. Soc.*, **140**(684), 2249-2259.
Ness, K.L., Paul, A., Sun, L. and Zhang, Z. (2022), "Towards a generic physics-based machine learning model for geometry invariant thermal history prediction in additive manufacturing", *J. Mater. Process. Tech.*, **302**, 117472. https://doi.org/10.1016/j.jmatprotec.2021.117472.
Oke, P.R., Brassington, G.B., Griffin, D.A. and Schiller, A. (2010), "Ocean data assimilation: a case for ensemble optimal interpolation", *Aust. Meteorol. Oceanographic J.*, **59**(1), 67-76.
Otter, D.W., Medina, J.R. and Kalita, J.K. (2020), "A survey of the usages of deep learning for natural language processing", *IEEE T. Neural Networks Learning Syst.*, **32**(2), 604-624.
Pongto, R., Wiwattanaphon, N., Lekpong, P., Lawawirojwong, S., Srisonphan, S., Kee, K.F. and Jitkajornwanich, K. (2020), "The grid-based spatial ARIMA model: An innovation for short-term predictions of ocean current patterns with big HF radar data", *Recent Advances in Information and Communication Technology 2019: Proceedings of the 15th International Conference on Computing and Information Technology (IC2IT 2019) 15*, 26-36.
Pugliese, R., Regondi, S. and Marini, R. (2021), "Machine learning-based approach: global trends, research directions, and regulatory standpoints", *Data Sci. Management*, **4**, 19-29. https://doi.org/10.1016/j.dsm.2021.12.002.
Rahmstorf, S. (2003), "Thermohaline circulation: The current climate", *Nature*, **421**(6924), 699. https://doi.org/ 10.1038/421699a.
Reichle, R.H. (2008), "Data assimilation methods in the Earth sciences", *Adv. Water Resour.*, **31**(11), 1411-1418. https://doi.org/10.1016/j.advwatres.2008.01.001.
Rotor, M.A., Hefazi, H. and Enano Jr., N. (2023), "Review on tidal stream energy and blade designs for tropical site conditions and a look at Philippines' future prospects", *Ocean Syst. Eng.*, **13**(3), 247-268. https://doi.org/10.12989/ose.2023.13.3.247.
Sarker, I.H. (2021), "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions", *SN Computer Science*, **2**(6), 420.
Sarker, I.H. (2022), "Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems", *SN Computer Science*, **3**(2), 158.
Schultz, M.G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L.H., Mozaffari, A., *et al.* (2021), "Can deep learning beat numerical weather prediction?", *Philos. T. Roy. Society A*, **379**(2194), 20200097.
Shchepetkin, A.F. and McWilliams, J.C. (2005), "The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model", *Ocean Model.*, **9**(4), 347-404.

Siedler, G., Grieffies, S., Gould, J. and Church, J. (2013), *Ocean Circulation and Climate*, 2nd Ed., *A 21st Century Perspective*, Academic Press.

Sinha, A. and Abernathey, R. (2021), "Estimating ocean surface currents with machine learning", *Front. Mar. Sci.*, **8.**

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., *et al.* (2019), "A description of the advanced research WRF model version 4", *National Center for Atmospheric Research: Boulder, CO, USA*, **145**(145), 550.

Song, T., Li, Y., Meng, F., Xie, P. and Xu, D. (2022), "A novel deep learning model by BiGRU with attention mechanism for tropical cyclone track prediction in the northwest pacific", *J. Appl. Meteorol. Climatology*, American Meteorological Society, Boston MA, USA, **61**(1), 3-12. https://doi.org/10.1175/JAMC-D-20-0291.1.

Song, T., Pang, C., Hou, B., Xu, G., Xue, J., Sun, H. and Meng, F. (2023), "A review of artificial intelligence in marine science", *Front. Earth Sci.*, **11**, 1090185.

Stanev, E. and Schulz-Stellenfleth, J. (2014), "Methods of data assimilation", *Die Küste, 81 Modelling*, 81, 133-151.

Storto, A., Falchetti, S., Oddo, P., Jiang, Y. and Tesei, A. (2020), "Assessing the impact of different ocean analysis schemes on oceanic and underwater acoustic predictions", *J. Geophys. Res.: Oceans*, **125**(7), e2019JC015636.

Sugiura, N., Awaji, T., Masuda, S., Mochizuki, T., Toyoda, T., Miyama, T., Igarashi, H., *et al.* (2008), "Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations", *J. Geophys. Res.: Oceans*, **113**(10).

Szczepanski, M. (2019), "Economic impacts of artificial intelligence (AI)", EPRS: European Parliamentary Research Service.

Tapeh, A.T.G. and Naser, M.Z. (2023), "Artificial intelligence, machine learning, and deep learning in structural engineering: A scientometrics review of trends and best practices", *Arch. Comput. Method. Eng.*, **30**(1), 115-159. https://doi.org/10.1007/s11831-022-09793-w.

Thongniran, N., Vateekul, P., Jitkajornwanich, K., Lawawirojwong, S. and Srestasathiern, P. (2019), "Spatio-temporal deep learning for ocean current prediction based on HF radar data", *Proceedings of the 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)*.

Tokmakian, R. and Challenor, P. (2019), "Near surface ocean temperature uncertainty related to initial condition uncertainty", *Climate Dynam.*, **53**(7), 4683-4700. https://doi.org/10.1007/s00382-019-04872-4.

Tolman, H.L. (2009), "User manual and system documentation of WAVEWATCH III TM version 3.14", *Technical Note, MMAB Contribution*, **276**(220).

Tonani, M., Balmaseda, M., Bertino, L., Blockley, E., Brassington, G., Davidson, F., Drillet, Y., *et al.* (2015), "Status and future of global and regional ocean prediction systems", *J. Operational Oceanography*, **8**(2), 201-220.

"Unravelling ENSO complexity". (2023), *Nature Geoscience*, **16**(2), 105. https://doi.org/10.1038/s41561-023-01134-1.

Villas Bôas, A.B., Ardhuin, F., Ayet, A., Bourassa, M.A., Brandt, P., Chapron, B., Cornuelle, B.D., *et al.* (2019), "Integrated observations of global surface winds, currents, and waves: requirements and challenges for the next decade", *Front. Mar. Sci.*, **6**, 425.

Wen, J., Yang, J., Jiang, B., Song, H. and Wang, H. (2021), "Big data driven marine environment information forecasting: A time series prediction network", *IEEE T. Fuzzy Syst.*, **29**(1), 4-18. https://doi.org/10.1109/TFUZZ.2020.3012393.

Wilkin, J.L. and Hunter, E.J. (2013), "An assessment of the skill of real-time models of Mid-Atlantic Bight continental shelf circulation", *J. Geophys. Research: Oceans*, **118**(6), 2919-2933.

Willard, J., Jia, X., Xu, S., Steinbach, M. and Kumar, V. (2020), "Integrating physics-based modeling with machine learning: A survey", *ArXiv Preprint ArXiv:2003.04919*, **1**(1), 1-34.

Willard, J., Jia, X., Xu, S., Steinbach, M. and Kumar, V. (2022), "Integrating scientific knowledge with machine learning for engineering and environmental systems", *ACM Computing Surveys*, ACM New York, NY, **55**(4), 1-37.

WMO, G. (2015), "Status of the global observing system for climate", World Metherological Organization.
Xie, C., Chen, P., Man, T. and Dong, J. (2023), "STCANet: Spatiotemporal coupled attention network for ocean surface current prediction", *J. Ocean Univ. China*, **22**(2), 441-451.
Xu, L., Chen, N., Chen, Z., Zhang, C. and Yu, H. (2021), "Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions", *Earth-Sci. Rev.*, **222**, 103828. https://doi.org/10.1016/j.earscirev.2021.103828.
Xu, S., Huang, X., Oey, L.-Y., Xu, F., Fu, H., Zhang, Y. and Yang, G. (2015), "POM. gpu-v1. 0: a GPU-based Princeton Ocean Model", *Geoscientific Model Development*, **8**(9), 2815-2827.
Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., *et al.* (2021), "Artificial intelligence: A powerful paradigm for scientific research", *The Innovation*, **2**(4), 100179. https://doi.org/10.1016/j.xinn.2021.100179.
Zandi, O., Zahraie, B., Nasseri, M. and Behrangi, A. (2022), "Stacking machine learning models versus a locally weighted linear model to generate high-resolution monthly precipitation over a topographically complex area", *Atmospheric Research*, **272**, 106159.
Zelenke, B.C. (2005), "An empirical statistical model relating winds and ocean surface currents: Implications for short-term current forecasts".
Zhang, Z., Yan, Y., Wang, Z., Song, T., Yin, J., He, X., Guo, P., *et al.* (2023), "SDPNet: A Novel DeepLearning Method for Ocean Surface Current Prediction", *J. Physics: Conference Series*, **2486**, 012066.
Zhao, Y., Yang, D., He, Z., Liu, C., Hao, R. and He, J. (2020), "Statistical methods in ocean prediction", *Global Oceans 2020: Singapore – U.S. Gulf Coast*, 1-7. https://doi.org/10.1109/IEEECONF38699.2020.9389485.
Życzkowski, M., Szłapczyńska, J. and Szłapczyński, R. (2019), "Review of Weather Forecast Services for Ship Routing Purposes", *Polish Maritime Res.*, **26**(4), 80-89. https://doi.org/10.2478/pomr-2019-0069.

MK