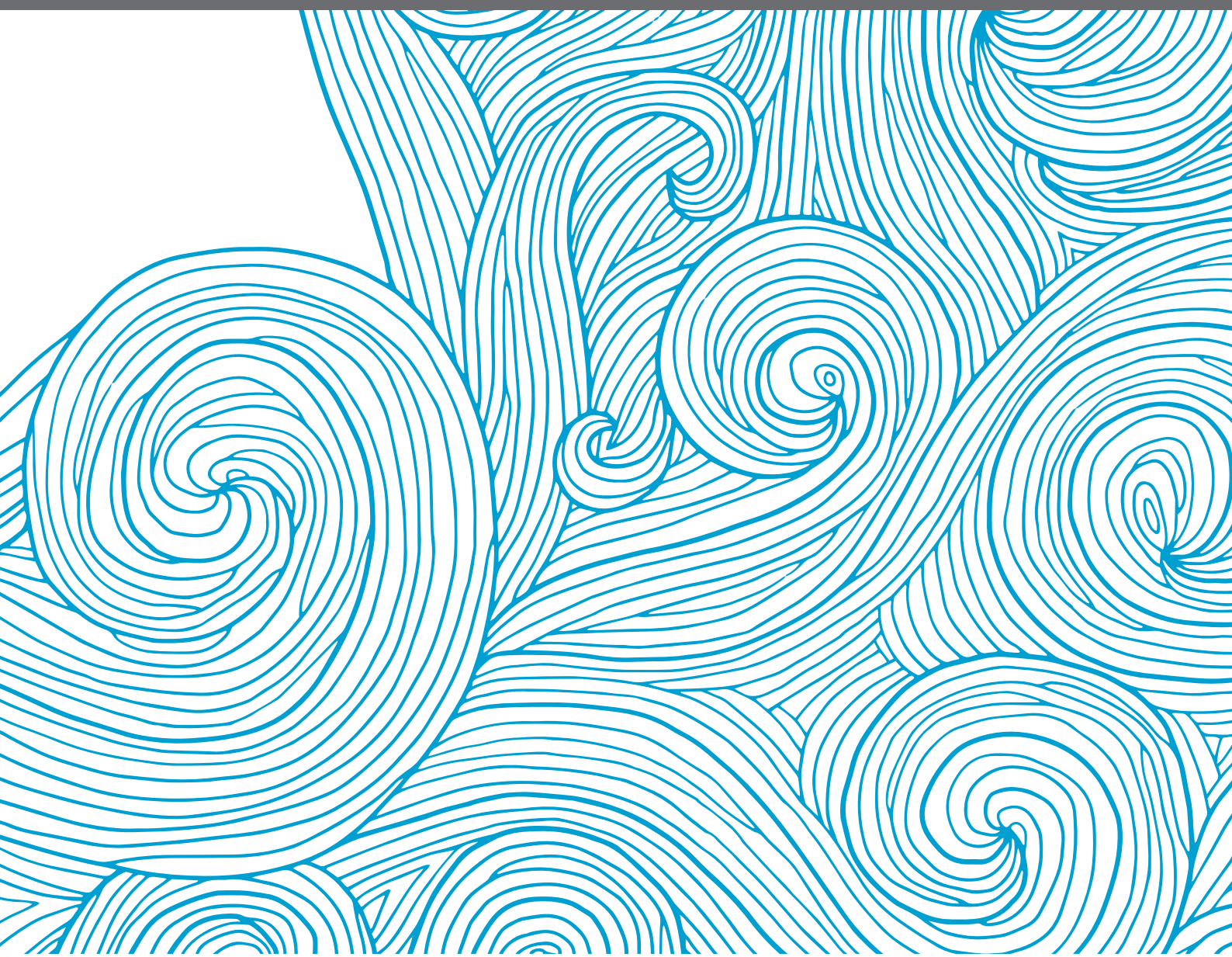
NEURAL COMPUTING AND APPLICATIONS TO MARINE DATA ANALYTICS

EDITED BY: Jun Li, Zhengguang Zhang, Lulu Qiao and Andrei Herdean
PUBLISHED IN: Frontiers in Marine Science



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-780-1
DOI 10.3389/978-2-88974-780-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

NEURAL COMPUTING AND APPLICATIONS TO MARINE DATA ANALYTICS

Topic Editors:
Jun Li, University of Technology Sydney, Australia
Zhengguang Zhang, Ocean University of China, China
Lulu Qiao, Ocean University of China, China
Andrei Herdean, University of Technology Sydney, Australia

Citation: Li, J., Zhang, Z., Qiao, L., Herdean, A., eds. (2022). Neural Computing and Applications to Marine Data Analytics. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-780-1

Table of Contents

frontiers in Marine Science

ORIGINAL RESEARCH
published: 03 May 2021
doi: 10.3389/fmars.2021.637759

Efficacy of Feedforward and LSTM Neural Networks at Predicting and Gap Filling Coastal Ocean Timeseries: Oxygen, Nutrients, and Temperature

Steefan Contractor and Moninya Roughan**

Coastal and Regional Oceanography Lab, School of Mathematics and Statistics, The University of New South Wales, Sydney, NSW, Australia

OPEN ACCESS

Edited by:
Jun Li,
University of Technology Sydney, Australia

Reviewed by:
Oliver Zielinski,
University of Oldenburg, Germany
Laurent Coppola,
UMR7093 Laboratoire d'Océanographie de Villefranche (LOV), France

***Correspondence:**
Moninya Roughan
mroughan@unsw.edu.au
orcid.org/0000-0003-3825-7533
Steefan Contractor
s.contractor@unswalumni.com
orcid.org/0000-0002-3987-2311

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *04 December 2020*
Accepted: *22 March 2021*
Published: *03 May 2021*

Citation:
Contractor S and Roughan M (2021) Efficacy of Feedforward and LSTM Neural Networks at Predicting and Gap Filling Coastal Ocean Timeseries: Oxygen, Nutrients, and Temperature. Front. Mar. Sci. 8:637759. doi: 10.3389/fmars.2021.637759

Ocean data timeseries are vital for a diverse range of stakeholders (ranging from government, to industry, to academia) to underpin research, support decision making, and identify environmental change. However, continuous monitoring and observation of ocean variables is difficult and expensive. Moreover, since oceans are vast, observations are typically sparse in spatial and temporal resolution. In addition, the hostile ocean environment creates challenges for collecting and maintaining data sets, such as instrument malfunctions and servicing, often resulting in temporal gaps of varying lengths. Neural networks (NN) have proven effective in many diverse big data applications, but few oceanographic applications have been tested using modern frameworks and architectures. Therefore, here we demonstrate a "proof of concept" neural network application using a popular "off-the-shelf" framework called "TensorFlow" to predict subsurface ocean variables including dissolved oxygen and nutrient (nitrate, phosphate, and silicate) concentrations, and temperature timeseries and show how these models can be used successfully for gap filling data products. We achieved a final prediction accuracy of over 96% for oxygen and temperature, and mean squared errors (MSE) of 2.63, 0.0099, and 0.78, for nitrates, phosphates, and silicates, respectively. The temperature gap-filling was done with an innovative contextual Long Short-Term Memory (LSTM) NN that uses data before and after the gap as separate feature variables. We also demonstrate the application of a novel dropout based approach to approximate the Bayesian uncertainty of these temperature predictions. This Bayesian uncertainty is represented in the form of 100 monte carlo dropout estimates of the two longest gaps in the temperature timeseries from a model with 25% dropout in the input and recurrent LSTM connections. Throughout the study, we present the NN training process including the tuning of the large number of NN hyperparameters which could pose as a barrier to uptake among researchers and other oceanographic data users. Our models can be scaled up and applied operationally to provide consistent, gap-free data to all data users, thus encouraging data uptake for data-based decision making.

Keywords: East Australian Current, machine learning, statistical modeling, depth profile observations, nitrate, phosphate, silicate, coastal oceanography

1. INTRODUCTION

Oceans play a pivotal role in the global weather and climate systems and support a multi-billion dollar blue economy, hence continuous monitoring of ocean conditions, in coastal marine areas in particular, is important to a wide range of stakeholders (UNESCO, 2019). Although many different types of observation platforms can measure the same ocean variable (e.g., temperature), there can be fundamental differences between them due to the instrumentation and the spatiotemporal sampling making comparison difficult (Hemming et al., 2020). For example, satellite observations are restricted to the surface and represent area-averages (often over many kilometers) while data from *in situ* observations represents a point-based observation in space (Lee et al., 2018). Even within the *in situ* observations, differences exist in terms of the observation depth, spatial coverage and temporal sampling frequency (Bailey et al., 2019). Furthermore, regardless of the type of observation, there are inevitably gaps in ocean timeseries due to various reasons including instrument loss, servicing and repairs, biofouling, deployment schedules, loss of funding etc. (Morello et al., 2014).

These data gaps and inconsistencies make ocean observations inaccessible to non-expert users, and even expert users (e.g., ocean modelers) typically take the path of least resistance when accessing ocean data. It is not a straightforward exercise to compare, for example, recent high frequency (e.g., 5 min) ocean temperature observations from a mooring with sparse (e.g., monthly) bottle data collected in the 1950s (Hemming et al., 2020). Furthermore, timeseries gaps can significantly affect trend analysis (Wynn and Wickwar, 2007).

Savvy data users by comparison, suffer the consequences of non-standardized methods of interpolating these gaps, leading to unaccountable differences in analysis. Furthermore, oceanographers often need gap-free timeseries for a particular analysis, for example, for determining physical mechanisms or teleconnections using empirical orthogonal function (EOF) analysis (e.g., Ashok et al., 2007 or for researching marine heatwaves Schaeffer and Roughan, 2017) since most common heatwave definitions involve comparing temperature anomalies of consecutive days with a climatology constructed from 30 years of daily data (Schlegel et al., 2019). Non-specialists may require gap-free observations concurrent with other timeseries, such as ecological or biological variables (Lee et al., 2019). Finally, measurement of climate impacts (e.g., trends in ocean warming) also necessitates continuous and consistent monitoring of variables over a long period (Malan et al., 2020). However, although ocean observations can date back many decades, there are statistical discontinuities that could be related to changes in ocean observation platforms and practices.

Some ocean observing systems, such as Australia's Integrated Marine Observing System (IMOS, www.imos.org.au) are maturing so as to allow for "learning" of relationships between ocean variables to fill inevitable gaps. Australia has a network of "National Reference Stations" with sampling dating back to the 1940s and 50s (Lynch et al., 2014). Initially sampling was boat based "bottle" sampling, but at the inception of IMOS, this was augmented with moored temperature (electronic sensor) timeseries at 8 m intervals through the water column, since 2008 (Roughan et al., 2010, 2013, 2015), and monthly vertical profiling with an electronic CTD (conductivity temperature and depth meter) sampling every meter. Here we use data from the Port Hacking National Reference Station in 100 m of water off Sydney (34°S) Australia as a case study to demonstrate the use of statistical models for prediction of oxygen and nutrient concentrations and gap filling temperature timeseries data.

Statistical models present an opportunity to fill gaps in observational records using data based approaches that do not involve making assumptions about the underlying physical processes. One such statistical model that has gained popularity in the last decade is the artificial neural network or neural network model (NN) (Emmert-Streib et al., 2020). An NN feeds the input features through numerous neurons arranged in multiple layers to generate an output layer that can be used for solving classification or regression problems (see LeCun et al., 2015 and Emmert-Streib et al., 2020 for more details). Although the first model of a single neuron was proposed in the 1950s (Rosenblatt, 1957), NNs have seen a large resurgence in the last decade partly due to breakthroughs in training efficiency (Hinton et al., 2006). The recent rise in popularity is also due to the development of new models capable of taking advantage of big data to solve real world problems, e.g., involving image recognition (Rawat and Wang, 2017), natural language processing (Young et al., 2018), and timeseries and text analysis (Lipton et al., 2015). See Schmidhuber (2015) and Emmert-Streib et al. (2020) for a detailed account of the historical development of neural networks. These recent breakthroughs in NNs have led to the development of modern open-source programming platforms, such as TensorFlow (developed by Google Brain) (Abadi et al., 2016), PyTorch (developed primarily by Facebook's AI Research Lab), etc. These libraries allow for quick and scalable implementations of state-of-the-art yet "off-the-shelf" NN models.

Due to their success in learning complex relationships, it stands to reason that NNs could be useful for learning the complex spatiotemporal relationships between physical, chemical and biological ocean variables. One of the first applications of NNs in oceanography was by Tangang et al. (1997) to forecast sea surface temperature anomalies of the Niño3.4 climate index. Since then they have been used to assist in forecasting wind generated ocean waves (Makarynskyy, 2005; Tolman et al., 2005), predict sea level fluctuations (Makarynskyy et al., 2004; Han and Shi, 2008), calculate Pacific Ocean heat content (Tang and Hsieh, 2003), statistical downscaling of ocean model output (Bolton and Zanna, 2019), predicting subsurface ocean temperature timeseries (Su et al., 2018; Han et al., 2019; Lu et al., 2019), and ocean eddy detection (Lguensat et al., 2018). Of particular interest to this study is the prediction of water column nutrient concentrations by Sauzède et al. (2017) and Bittig et al. (2018), and creation of virtual marine sensors by Oehmcke et al. (2018). For a comprehensive historical overview of NN applications in oceanography, see Hsieh (2009) and Krasnopolsky (2013).

Despite, their widespread use in Oceanography, examples of NNs that predict nutrients or gap-fill temperatures in the water column (i.e., below the surface) are scarce (exceptions

include Sauzède et al., 2017, Bittig et al., 2018, and Fourrier et al., 2020) especially after the breakthroughs in training efficiency in 2006 (Emmert-Streib et al., 2020), and the few studies that do accomplish this do not take advantage of modern frameworks, such as TensorFlow. TensorFlow is an open-end ecosystem/framework for neural network modeling. It contains a suite of tools, libraries/packages, and community resources that are constantly updated with the latest advances in the field of machine learning. It is language agnostic with implementations in Python, R, and Julia, with the Python implementation being the most popular. Furthermore, it allows for the development of models with various levels of abstraction, which means that programmers and researchers new to modeling with NNs can spin up models quickly and easily, and add complexity as they require. Thanks to its large community of machine learning practitioners, researchers in adjacent fields who wish to take advantage of statistical modeling but do not have the knowledge of, or expertise in computer science, can take advantage of latest breakthroughs without worrying about being aware of and downloading latest libraries/packages, and working out compatibility issues. Users must, however, judge the risks of applying such black-box models for themselves and determine the model's suitability to their use case individually. Furthermore, progress has recently been made on the interpretability of neural networks (Lundberg and Lee, 2017; Kwon et al., 2019).

In this study, we demonstrate the efficacy of using "off-the-shelf" NN models built using TensorFlow for modeling dissolved oxygen, nutrient concentrations and temperature throughout the water column at a valuable long-term observational site. We use a simple feedforward neural network to model oxygen and nutrients, however due to the larger amount of data available, we treat temperature modeling as a timeseries prediction problem. For this we use a type of neural network that is adept at modeling timeseries called recurrent neural networks (RNN) (Lipton et al., 2015). RNNs include connections between adjacent timesteps in addition to connections between the input, hidden and output layers. However, typical RNNs struggle to remember long-term dependencies (Graves, 2013) and are also difficult to train (Rosindell and Wong, 2018). Hence we gap-fill temperature using a special type of RNN called a Long Short Term Memory (LSTM) NN (Hochreiter and Schmidhuber, 1997; Adikane et al., 2001) that overcomes both of these shortcomings. As this is a proof of concept study, we outline the NN model design and training in detail. We show the high degree of accuracy that can be obtained when predicting and gap filling using these modern multi-layered NNs. The models developed in this study serve as a proof of concept for applications to other such long term ocean timeseries datasets aiming to create a suite of easy-to-handle, continuous and filled, derived observational products.

This study demonstrates two types of models befitting two datasets with varying characteristics and is hence divided in two parts. First, section 2 describes the two types of datasets, then section 3 details the first model that uses concurrent observations from related oceanic variables to predict dissolved oxygen and nutrient concentrations. The second suite of models detailed in section 4 are used to gap-fill temperature timeseries observations at a single depth over multiple length temporal gaps from days to months using LSTMs.

2. DATA

2.1. Prediction Based on Co-variates

Data from discrete depths ("bottle data," see below) was used for training a simple NN to predict other variables. The "bottle data" consists of measurements of pressure, temperature, salinity, dissolved oxygen and dissolved nutrient concentrations (nitrate, phosphate, and silicate) with timestamps at a long term observation site. The Port Hacking National Reference Station (151.2°E, 34.1°S) is located in 100 m of water, ∼8.5 km off the coast of Sydney Australia where tidal influences are weak. See Roughan et al. (2010, 2013, 2015), and Hemming et al. (2020) for more information on the long term sampling at Port Hacking. The water samples are collected in bottles at a range of depths through the water column, nominally, at 10, 20, 30, 40, 50, 75, and 100 m in 100 m of water. The data collection began in 1953 with a frequency of weekly to monthly but due to the nature of manual collection consists of many gaps and irregular sampling intervals. See Hemming et al. (2020) their Figure 3 and Table 1 for a full description of the data collection.

Due to irregularity of the bottle data collection, the prediction problem could not be treated as a timeseries prediction problem but rather as a simple co-variate based prediction problem where time was treated as another co-variate. The timestamps were converted into three co-variates; year, day of the year and hour of the day. This resulted in six predictor variables: pressure, temperature, salinity and the three temporal co-variates. **Figures 1**, **2** show the relationship between the six co-variates and the Oxygen and Nitrate concentrations, respectively, colored by the pressure at which the measurement was recorded. The Oxygen concentration (**Figure 1**) is in the range ∼160–275 $\mu mol\,l^{-1}$ whereas the nutrients are bounded by 0 $\mu mol\,l^{-1}$ and are heavily skewed to the left. In general, the nutrient concentrations increase with depth and oxygen decreases with depth. A seasonal cycle is evident in the Oxygen concentrations and for the high Nitrogen concentrations. Some inter-annual variability is also visible for both Oxygen and Nitrogen concentrations, albeit the pattern seems less predictable. Since the data were typically collected in the morning Australian Eastern Time (close to midnight UTC time), little useful information is added by the hour-of-the-day co-variate. Models that do not use the hour-of-the-day as an additional co-variate were also tested but the model accuracy remained unchanged. Although most measurements were documented as having been taken at the nominal pressure depths, there are many examples of measurements recorded outside of these nominal depths (**Figures 1**, **2**). As a result, the pressure is treated as a continuous variable.

Data was collected following standard IMOS procedures to ensure data is of high precision and accuracy. IMOS provides guidelines on pre-deployment planning, data collection, post-deployment data processing, and quality control, see Sutherland et al. (2017) for IMOS QA procedures, and Ingleton et al. (2014) and Morello et al. (2014) regarding IMOS quality control (QC).

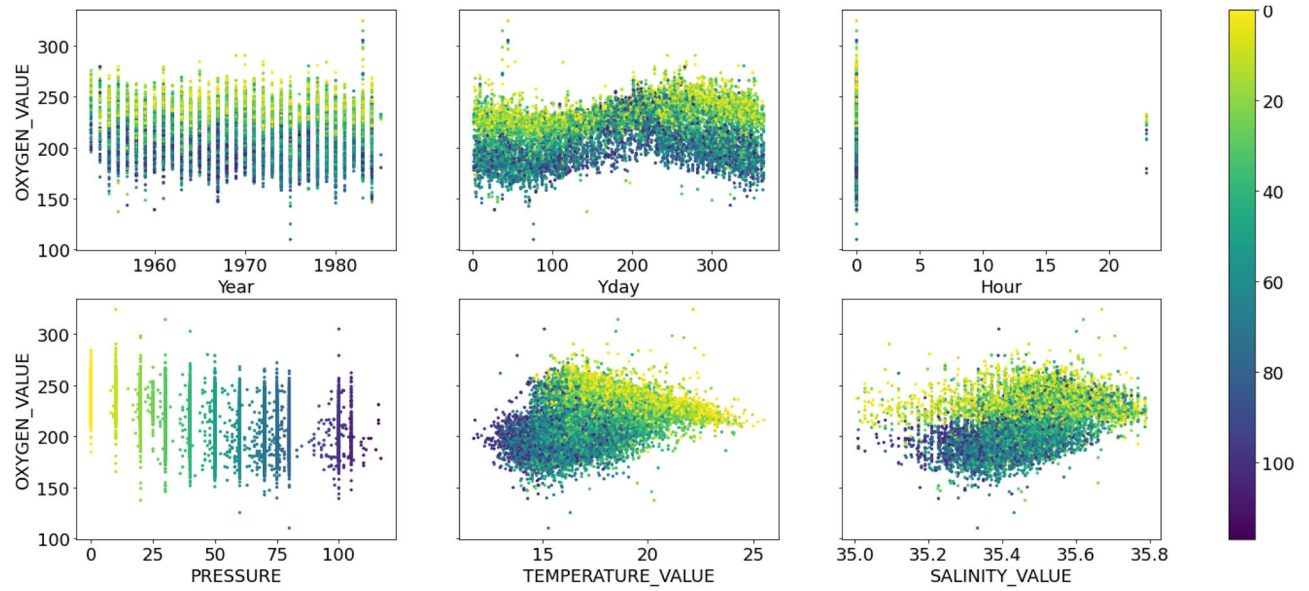


FIGURE 1 | Scatter plots of dissolved oxygen concentration (μmol l^{-1}) (y-axis) against the six co-variates; (from left to right, top to bottom) year, day of the year, hour of the day, pressure (dbar), temperature (°C), and salinity (psu). All points are colored by the pressure value.

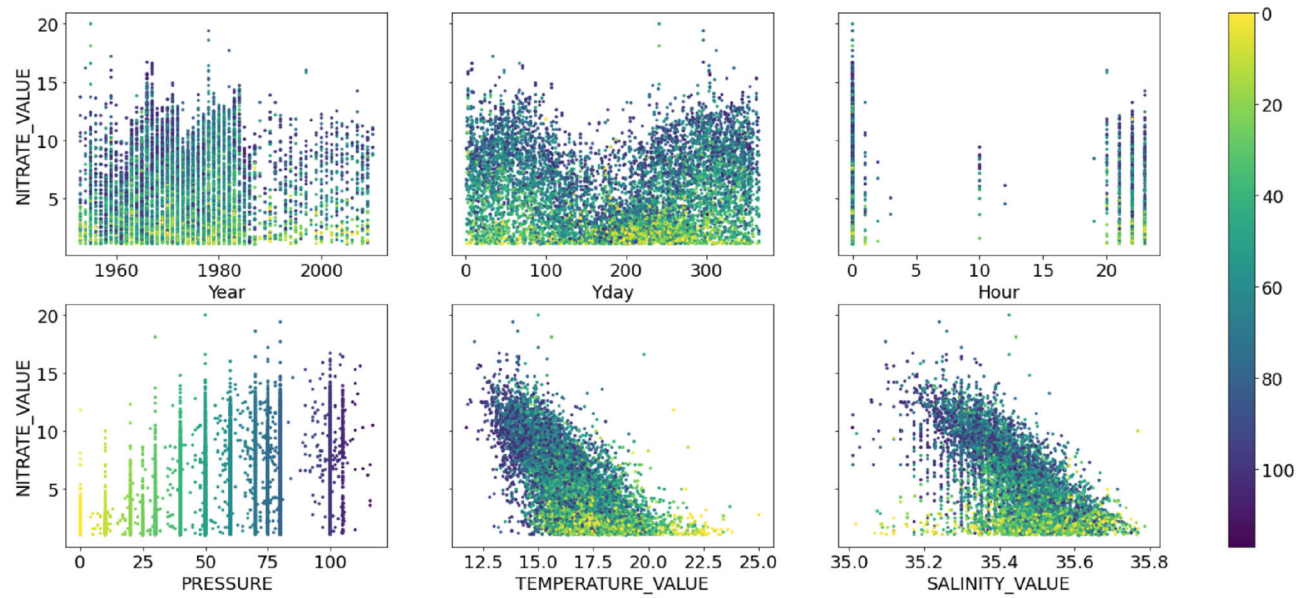


FIGURE 2 | Same as **Figure 1** but showing the scatter plots of nitrate concentration (μmol l^{-1}) against the six co-variates.

QC includes but is not limited to checking for impossible date, position, depth range (based on global and regional values) correlation with adjacent data (above, below, before, and after). All steps including the use of the IMOS data toolbox for QC are described in Hemming et al. (2020). The data used here were flagged as either "good data" or "probably good data" following UNESCO Intergovernmental Oceanographic Commission (IOC) protocol (flags 1 and 2, respectively) (Lynch et al., 2014; Morello et al., 2014). In addition to the IMOS QA/QC procedures a visual quality control was conducted to remove outliers to aid the model learning process. For example, records corresponding to salinity values below 35 parts per thousand (psu) and above 35.8 psu, pressure above 120 dbar, and nitrate values above 24 μmol l^{-1} were removed, resulting in 587 records removed for nitrate concentration modeling and 621 records removed for oxygen modeling.

2.2. Gap-Filling Temperature Timeseries

A recurrent neural network model was trained on a "gappy" temperature timeseries obtained from a sensor on a fixed mooring located near the Port Hacking NRS (PH100). The mooring was deployed with AQUATec Aqualogger 520T temperature or 520TP temperature and pressure sensors at 8 m intervals in 2009 to augment the historic bottle data. Temperature data have been acquired using automated electronic senors at 5 min intervals since 2009. Gaps in the temperature data occur through loss of the mooring or sensor failure. The mooring is manually serviced every quarter and has gaps in the timeseries ranging from a few hours, days, to up to 3 months (**Table 1**). The PH100 mooring was deployed in 2009 and was still operational as of 2020 providing over a decade of temperature observations. Due to the greater number of data samples available, this gap-filling problem was treated as a timeseries prediction problem where the model was trained to predict purely based on the timeseries of the variable of interest, instead of a prediction based on co-variates, as was the case for oxygen and nutrients.

Since a daily timeseries is adequate for many oceanographic applications, we resample the 5-min observations to a daily frequency using an arithmetic mean. Due to the movement of the sensors, observations are available continuously throughout the water column, with the highest density of measurements available at the optimal depths (the depths at which the temperature sensors are deployed). To find these optimal depths, the daily observations were binned in 1 m intervals and the bins with the highest density of observations were considered as the optimal depths. Since Hemming et al. (2020) showed that the temperature observations at PH100 were highly correlated within 8–9 m, we combined observations from 2 m below and 3 m above the optimal depths and labeled them by the average depth of the resulting range. This results in a regular daily frequency timeseries. For this study, we chose to gap-fill the 24.5 m bin timeseries as a proof-of-concept, which as explained, contains observations between 22 and 27 m. In total, the 24.5 m timeseries contains 3,459 days, with 256 days (7.4% of total timeseries) of missing temperature observations as depicted in **Figure 3**. These 256 missing days are spread across 9 gaps with the longest gap being 91 days and the shortest being 2 days (**Table 1**).

TABLE 1 | Length of gaps in the order they are encountered in the 24.5 m bin temperature timeseries from the Port Hacking 100 mooring.

Gap number	Gap length (days)
1	91
2	4
3	3
4	6
5	19
6	29
7	21
8	2
9	79
Total (days)	256 (7.4%)

Bottom row shows the total number of days with missing data and percentage of total days that are missing. The length of the entire timeseries (with gaps) is 3,459 days.

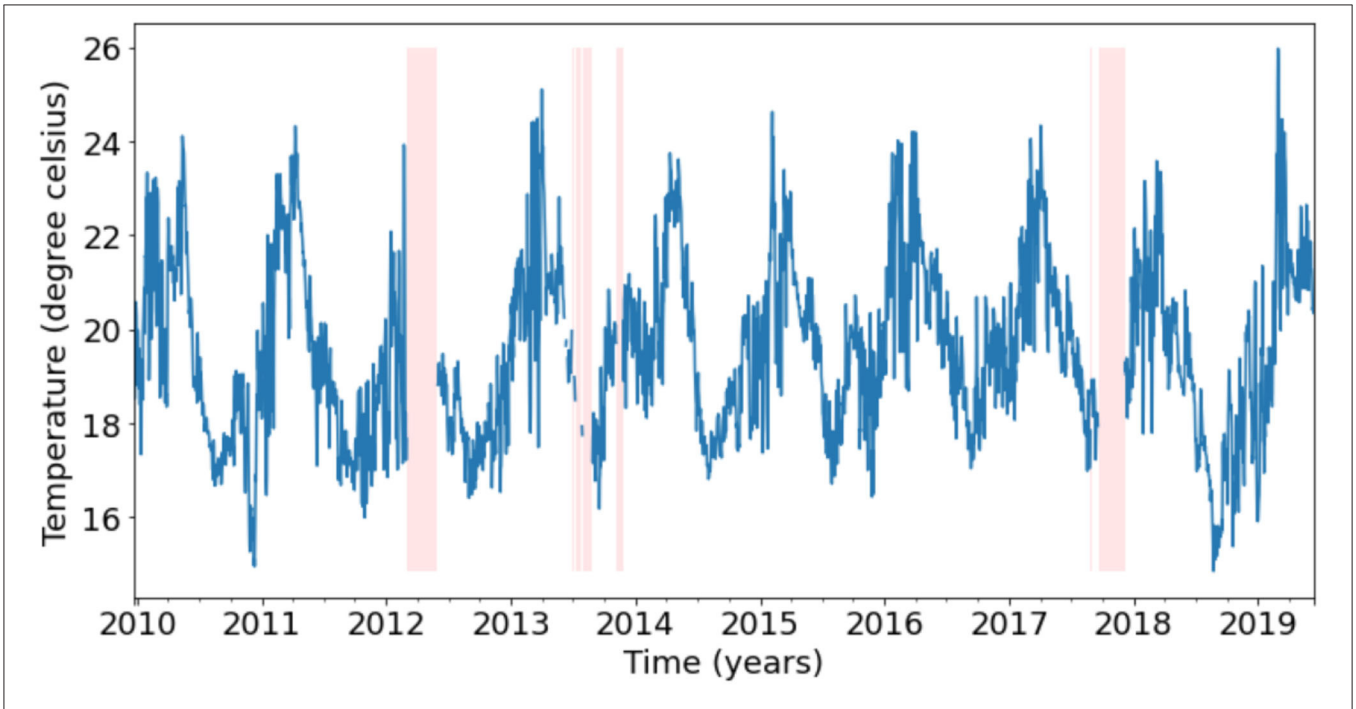


FIGURE 3 | Depiction of the 24.5 m bin temperature timeseries with gaps represented by pink vertical bands. Refer to **Table 1** for the lengths of the gaps.

3. PREDICTION BASED ON CO-VARIATES

3.1. Model

The simplest, most "vanilla," neural network known as a feedforward NN consists of many connected nodes or neurons arranged in one or more "hidden" layers (Schmidhuber, 2015). The output of each neuron **y** consists of a non-linear activation function ϕ applied to a linear function with a weight matrix W and bias vector **b**.

$$\mathbf{y} = \phi(W \cdot \mathbf{x} + \mathbf{b}) \quad (1)$$

Here **x** is the input vector for the neuron. Training such a NN involves minimizing the loss, which is a function representing how close the NN output is to the true value. Common choices for loss functions include mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) for regression problems. There are many optimization algorithms or "optimizers" used to minimize the loss function. One of the most basic but still highly used optimization algorithms is stochastic gradient descent which updates the neuron weights and biases in the direction of the steepest gradient of the loss function during each training step. The size of these parameter updates is determined by a learning rate. All other optimizers are in some sense a variation of stochastic gradient descent. The two most popular variations are RMSprop and Adam which both introduce concepts of adaptive learning rates. See Ruder (2016) and Ketkar (2017) for a complete list of optimizers and their details.

We used a feedforward neural network with four dense, fully connected layers for predicting oxygen, nitrate, phosphate and silicate concentrations, as seen in **Table 2**. The first three layers contain 64, 64, and 8 nodes, respectively, and the output layer contains one node, which gives the predicted concentration. A Rectified Linear Unit (ReLU) (Nair and Hinton, 2010) activation function was used for the first three layers. The ReLU activation function is a piecewise linear function that returns values equal to its input for positive inputs and zero otherwise. No activation was used for the output layer as no activation is necessary for regression problems. The mean absolute percentage difference (MAPE) between the predicted and the true concentrations was used as the loss function for oxygen prediction and the mean squared error (MSE) was used for nutrient prediction. The Adam optimization algorithm (Kingma and Ba, 2015) was used to minimize this loss function by tuning the node weight parameters. These design choices have been justified further below.

A total of 12,970, 8,633, 12,639, and 6,303 records were used for oxygen, nitrate, phosphate, and silicate concentration modeling, respectively. In all cases 80% of records were chosen at random for the training and validation datasets, and the remaining 20% were used for evaluation of the trained model. The training and validation datasets were also split randomly at an 80–20 ratio.

The model architecture was tested by experimenting with more layers (up to 6) and nodes per layer (up to 1,024), however, more complex models provided little to no improvement to the final model accuracy. Similarly, the hyperparameters were also chosen by experimenting with different values. One of the most important architectural design decisions when tuning an NN is the activation function. As described earlier each node of an NN consists of a linear estimator which is passed through a non-linear activation function to create non-linear estimates (Emmert-Streib et al., 2020). Hence an activation function can influence the decision boundaries and convergence of the NN model. See Table 1 of Emmert-Streib et al. (2020) for a list of the most common activation functions. We tested the ReLU, tanh and sigmoid functions as activations. Although the final validation loss was similar for all three activation choices, ReLU resulted in the smoothest and quickest convergence, with minimal validation dataset loss variability. During training with large datasets, it is often more efficient to take many small, quick steps based on subsets of training data instead of one large step based on the entire dataset. These data subsets are called batches and the batch size determines the variability of the losses during training (smaller batch sizes mean more variable convergence and vice versa). For oxygen we used the default Keras (TensorFlow wrapper) batch size of 32 and settled on batch sizes of 1,024, 64, and 64 for nitrate, phosphate, and silicate models, respectively. The size of the training steps is determined by the learning rate of the "optimizer" and also affects the convergence characteristics during training. We used the Adam optimizer (Kingma and Ba, 2015) with the default learning rates of 0.001 for all models (oxygen, phosphate, and silicate) except for nitrate modeling for which we used a learning rate of 0.0001. Sometimes, a model overfits to the training data resulting in a gap between the final training and validation losses as seen in **Figures 4a,b**. This can be reduced by handicapping the models using "regularization" techniques that make it harder for models to learn too quickly. No regularization was necessary for oxygen, however, we used L2 regularization (Krogh and Hertz, 1992) for the nutrients.

The models were trained using "early-stopping," meaning the training was stopped when the validation dataset error flattened out for 50 epochs, where an epoch is one pass through all training examples. All features for all models were standardized by subtracting the mean and dividing by the standard deviation before training so all features carried equal emphasis for the models.

Finally, the loss metric used to train the models also influences the predictions. Since we are primarily interested in reducing

TABLE 2 | Description of the four layer model architecture used to predict dissolved oxygen and nutrient concentrations.

	Layer type	Activation function	# Nodes
Layer 1	Dense, fully connected	ReLU	64
Layer 2	Dense, fully connected	ReLU	64
Layer 3	Dense, fully connected	ReLU	8
Layer 4	Output layer		1

No activation function was used for the output layer.

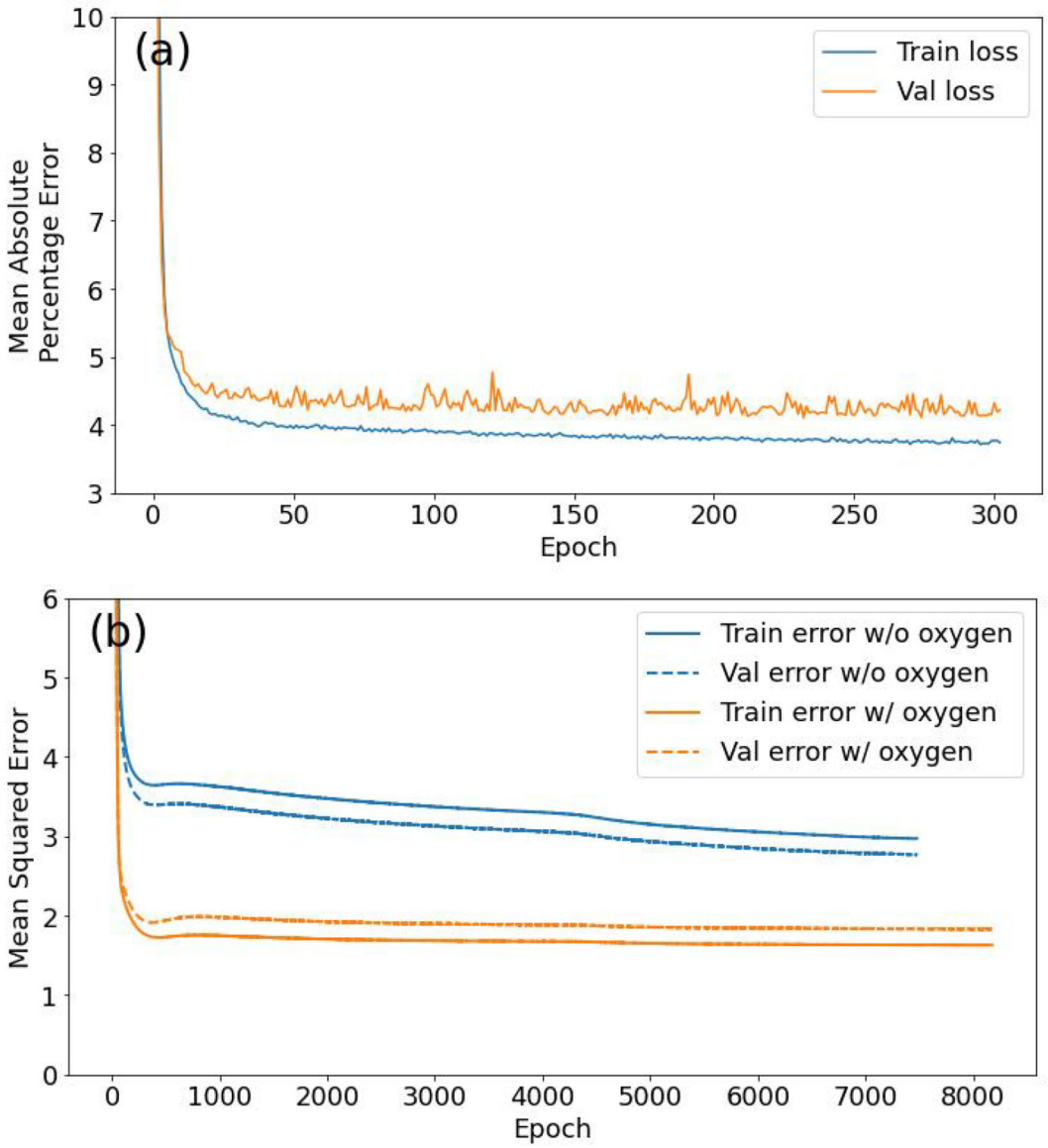


FIGURE 4 | (a) Oxygen model training, showing the reduction of the mean absolute percentage error (loss) as a function of epochs (302 in total). Losses over both training (blue) and validation (orange) datasets are shown. **(b)** Comparison of training history of two different nitrate models: first without oxygen as an additional feature variable (blue lines), and second with oxygen as an extra feature (orange lines). Both the training (solid) and validation (dashed) losses are shown. An epoch is a single pass through all training examples.

the mean absolute percentage error (MAPE) between the true and predicted values, we trained the models with MAPE loss. Although this resulted in good results, for oxygen, it produced misleading results for the nutrient predictions. This is because the absolute errors ($true - predicted$) can translate into large percentage errors for $true \approx 0$ (which is the case for most nutrient observations), relative to percentage errors for larger true values. As a result, we used MSE loss for nutrient models.

The aim of the nutrient prediction models was to apply them to mooring data to create a mooring timeseries. Since no reliable dissolved oxygen data is available, we do not use oxygen as a feature variable during prediction. However, to gauge the model improvement, alternative models with oxygen as an extra feature were also trained for predicting nitrates, phosphates, and silicates. The training curves for the 6 feature and $6 + 1$ feature nitrate models are shown in **Figure 4b**.

3.2. Results

The final test dataset mean absolute percentage error of the oxygen model was 3.97% which equates to an accuracy of 96.03%.

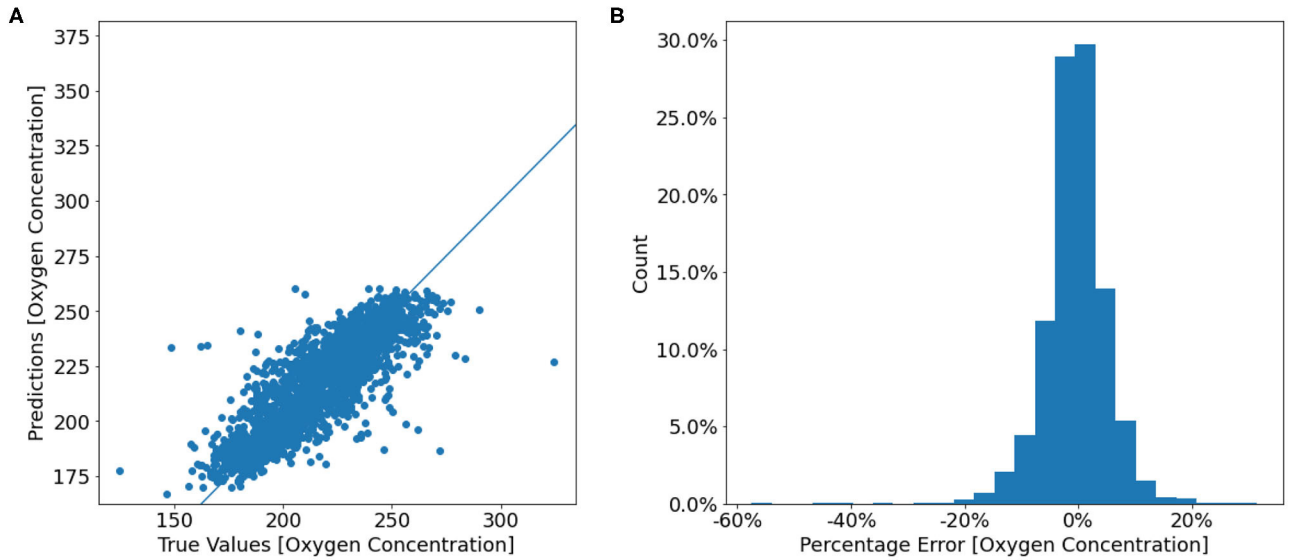


FIGURE 5 | (A) Scatter plot of the true oxygen concentrations from the test dataset against the model predicted oxygen concentrations. **(B)** A histogram of the prediction error percentage [(*true* − *predicted*)/*true*]. Counts are displayed as a percentage of the total test data.

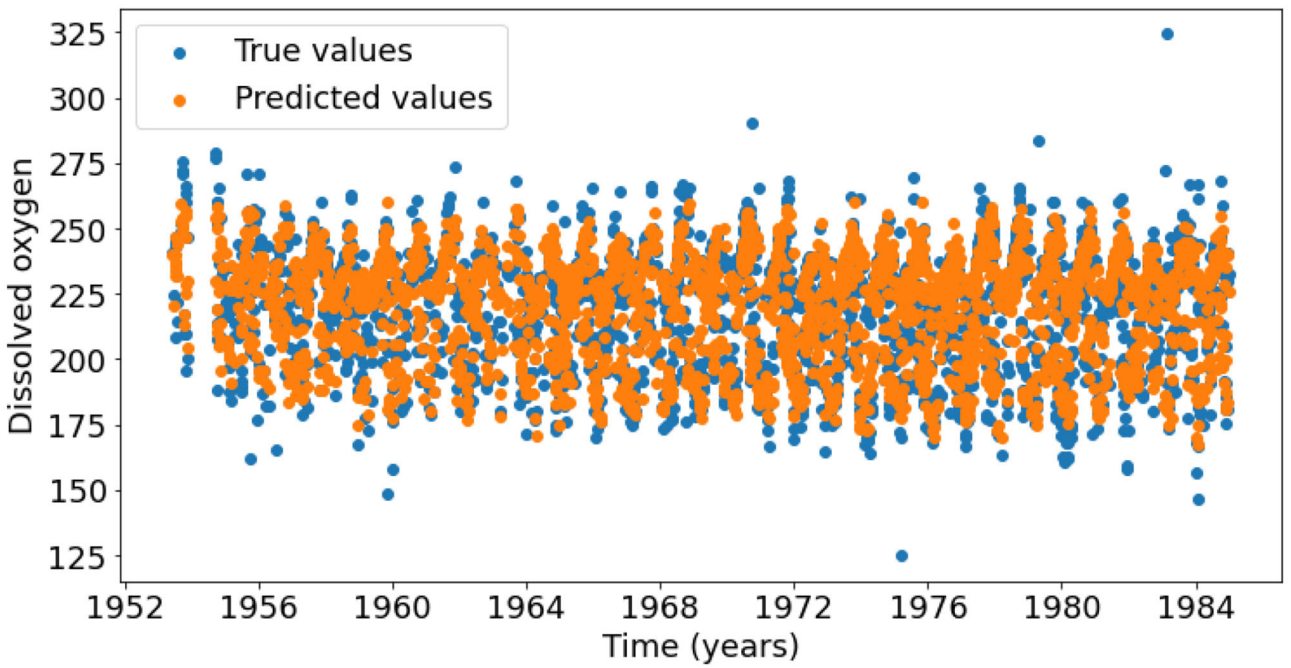


FIGURE 6 | Timeseries of the true (blue points) and predicted (orange points) dissolved oxygen concentrations.

Figures 5A,B show that most predictions fall close to the true oxygen concentration, however there are some large outliers with errors up to 60% (**Figure 5B**). Furthermore, the model seems to under-predict the oxygen concentrations on the higher end of the distribution, and over-predict the concentrations on the lower end (**Figure 5**). This is also seen in **Figure 6**, which shows that the range of the predicted values is smaller than that of the true Oxygen concentrations. Finally, **Figure 7A** shows that the large errors occur randomly throughout the temperature-salinity distribution instead of being concentrated in a particular range.

The final mean squared error (MSE) difference between the true and predicted values over the nitrate test dataset loss is 2.63 $(\mu mol/l)^2$ which equates to a root mean squared error

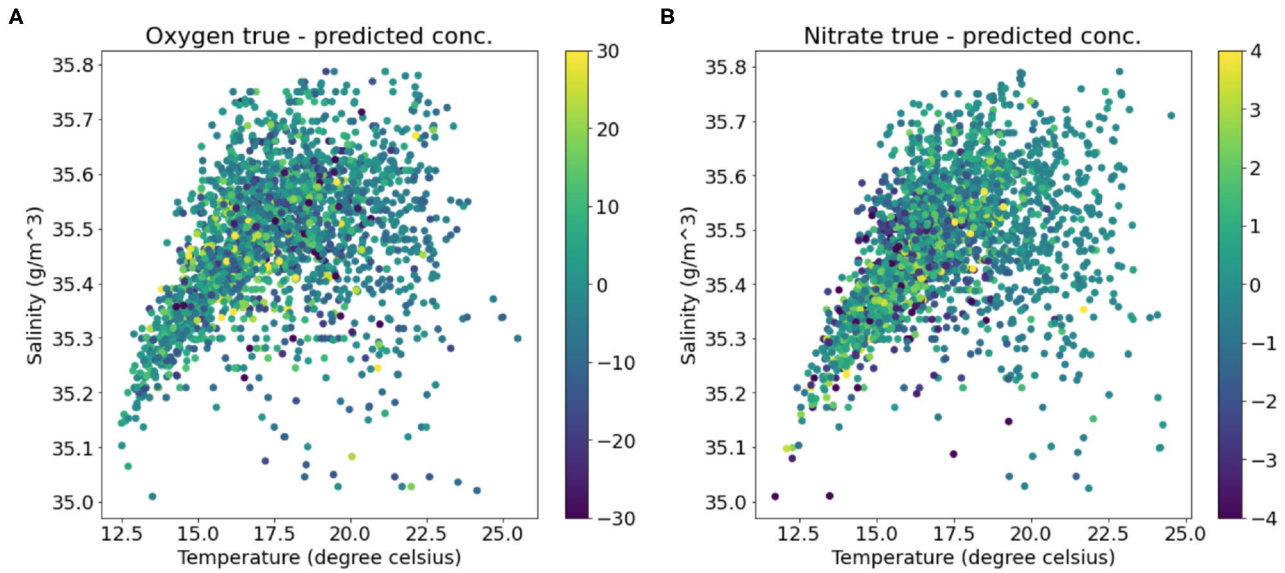


FIGURE 7 | Temperature-Salinity diagrams colored by **(A)** the error in the predicted oxygen concentrations and **(B)** the error in the predicted nitrate concentrations.

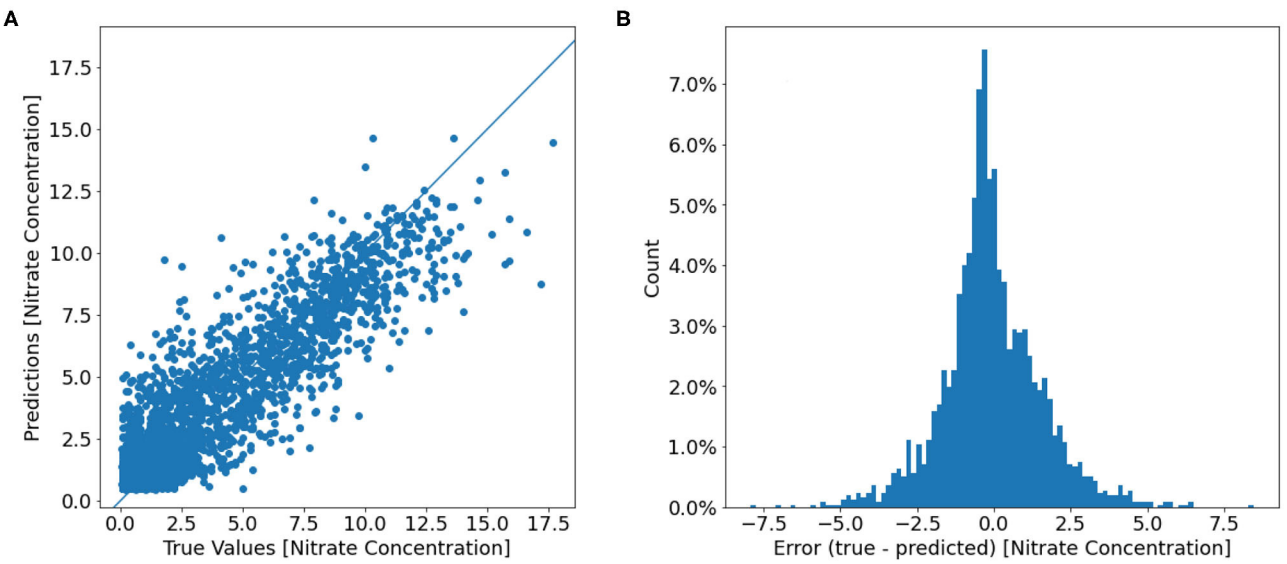


FIGURE 8 | (A) Scatter plot of the true nitrate concentrations (μmol l^{-1}) from the test dataset against the model predicted oxygen concentrations. **(B)** A histogram of the prediction error [(*true* − *predicted*)]. Counts are displayed as a percentage of the total test data.

(RMSE) of 1.62 μmol l^{-1}. Two models were trained with identical model architectures and hyperparameters, however, the second model include oxygen as an additional feature variable besides the six feature variables used by the first model. The training curves for these two models are shown in **Figure 4b**. The model with oxygen as an extra feature had a lower loss compared to the model with the original six variables, with the final MSE (RMSE) of 1.83 $(\mu mol/l)^2$ (1.35 μmol l^{-1}). Temperature-Salinity diagram (**Figure 7B**) indicates that the large errors are located throughout the entire temperature-salinity range with the majority close to the mean of the temperature-salinity distribution (**Figure 8B**).

The final MSE and RMSE over the test dataset for the six feature phosphate model were 0.0099 $(\mu mol/l)^2$ and 0.0995 μmol l^{-1}, respectively, and for the seven feature (including oxygen) were 0.0058 $(\mu mol/l)^2$ and 0.0765 μmol l^{-1}, respectively. For comparison, the phosphate data ranged between 0.02 and 1.80 μmol l^{-1}. Similarly, the MSE and RMSE over the test dataset for the six feature phosphate model were 0.78 $(\mu mol/l)^2$ and 0.89 μmol l^{-1}, respectively, and for the seven feature (including

oxygen) were 0.42 $(\mu mol/l)^2$ and 0.65 $\mu mol\, l^{-1}$, respectively. The silicate concentration ranged between 0.03 and 15.20 $\mu mol\, l^{-1}$.

4. GAP-FILLING TIMESERIES

4.1. Model

Long Short-term Memory (LSTM) NNs have been used to predict (fill) gaps in temperature timeseries. These models have proved extremely effective in predicting sequence data, such as timeseries and language problems (Lipton et al., 2015). The nodes in LSTMs are enclosed within memory cells along with an input gate, an internal state, a forget gate and an output gate (see Lipton et al., 2015 for a detailed explanation). These elements inside the memory cell are connected to each other in a specific way, and also contain recurrent connections with adjacent timesteps as is typical of an RNN. It is through the inclusion of this internal state that LSTMs remember long-term information and also overcome the vanishing and exploding gradients problem (Hochreiter and Schmidhuber, 1997). LSTMs use a short past window to predict multiple timesteps into the future and average loss over all the predicted timesteps.

Three successive models were trained in order to infill gaps of different lengths. Since three out of nine gaps were <6 days long (**Table 1**), the first model used a history window 30 days long to fill gaps up to 6 days long. This history window length was chosen since it resulted in the lowest loss compared to history lengths of 15 and 6. To increase the number of training examples, the second model uses samples from the original timeseries and the filled gaps based on the first model to infill gaps of up to 30 days based on 30 days of history. The three gaps that ranged 6–30 days long are thus filled with model 2. The third model was trained based on training examples from the timeseries with gaps up to 30 days long filled with models 1 and 2, and was used to fill the longest remaining gaps (30–91 days long) based on 91 days of history. We avoided using history windows longer than the prediction windows for models 2 and 3, since this would reduce the number of training examples extracted from the timeseries. Furthermore, our choices of history window lengths results in high model accuracy as discussed in the results section (section 4.2). Thus, our approach is somewhat auto-regressive as predictions from previous models are used to generate training examples for the next model. Note, that such an auto-regressive approach results in compounding errors from previous models. As a result, the true model losses for models 2 and 3 could be higher than as noted in section 4.2.

We used 2,217, 2,130, and 1,793 training examples to train models 1, 2, and 3, respectively with each example consisting of the history window and the prediction window. These made up 90% of the total pairs of history and prediction windows available for each model with the remaining 10% of examples used for validation. We used a smaller percentage of examples for validation compared to the oxygen and nutrient models since we have more data available and hence fewer validation examples were necessary.

Temperature timeseries were standardized before training similar to the feature standardization for the oxygen and nutrient models as this has been shown to improve training efficiency and model convergence for LSTMs (Laurent et al., 2016). This meant that although MAPE is the metric of interest for us, it could not be used as a loss function because of the scaled temperature values close to zero. Thus, a modified MAPE loss that used the training history mean and standard deviation to reverse the standardization before calculating the MAPE was used. Similar to the co-variate modeling, the Adam optimizer with the default learning rate of 0.001 was used.

TABLE 3 | Description of the four layer model architecture used to fill temperature gaps.

	Layer type	Activation function	# Units
Layer 1	Bidirectional LSTM	ReLU	64
Layer 2	Bidirectional LSTM	ReLU	32
Layer 3	Bidirectional LSTM	ReLU	8
Layer 4	Output layer, dense		Variable (6, 30, or 91)

No activation function was used for the output layer.

The architecture for the three models used for filling temperature gaps is described **Table 3**. The bidirectional layer is a wrapper that takes the layer of hidden nodes and connects them in the opposite direction so the hidden state of the first layer "remembers" information from the past while the hidden state variable of the second layer "remembers" information from the future. Thus, a bidirectional LSTM is able to learn contextual information which is important for gap-filling problems. However, since model 3 had to predict a much longer window compared to models 1 and 2, the approaches used for models 1 and 2 resulted in larger losses of around 4%. As a result, model 3 was treated as a multivariate timeseries forecasting problem where the 91 days after the gap were chronologically reversed (i.e., the vector of values for day 1, 2, 3, …, 91 after the gap became the vector of values corresponding to days 91, 90, 89, …, 1 after the gap) and given to the model as another feature variable. This way contextual information was more explicitly encoded into the "multi-variate" model and the resulting losses were lower.

To assess the uncertainty of the long window predictions, we include dropout in each LSTM layer. Dropout is a regularization technique where a given percentage of nodes are randomly turned of during each training step to reduce model overfitting. Traditionally no dropout is included during the evaluation stage. However, it has recently been shown that predictions with dropout act as monte carlo estimates and represent Bayesian approximations (Gal and Ghahramani, 2016b). Such dropout based uncertainty is much less computationally expensive compared to fully Bayesian approaches, such as Bayesian NNs. In LSTMs dropout is typically applied to the input and recurrent connections as opposed to the nodes themselves (Gal and Ghahramani, 2016a). We drop 25% of input and recurrent connection randomly in model 3. We apply dropout only to model 3 as a demonstration of deriving modeling uncertainty using NNs.

Similar to the oxygen and nutrient models, the training was automatically stopped when there was no improvement in validation loss for 20 epochs. The performance of model 1

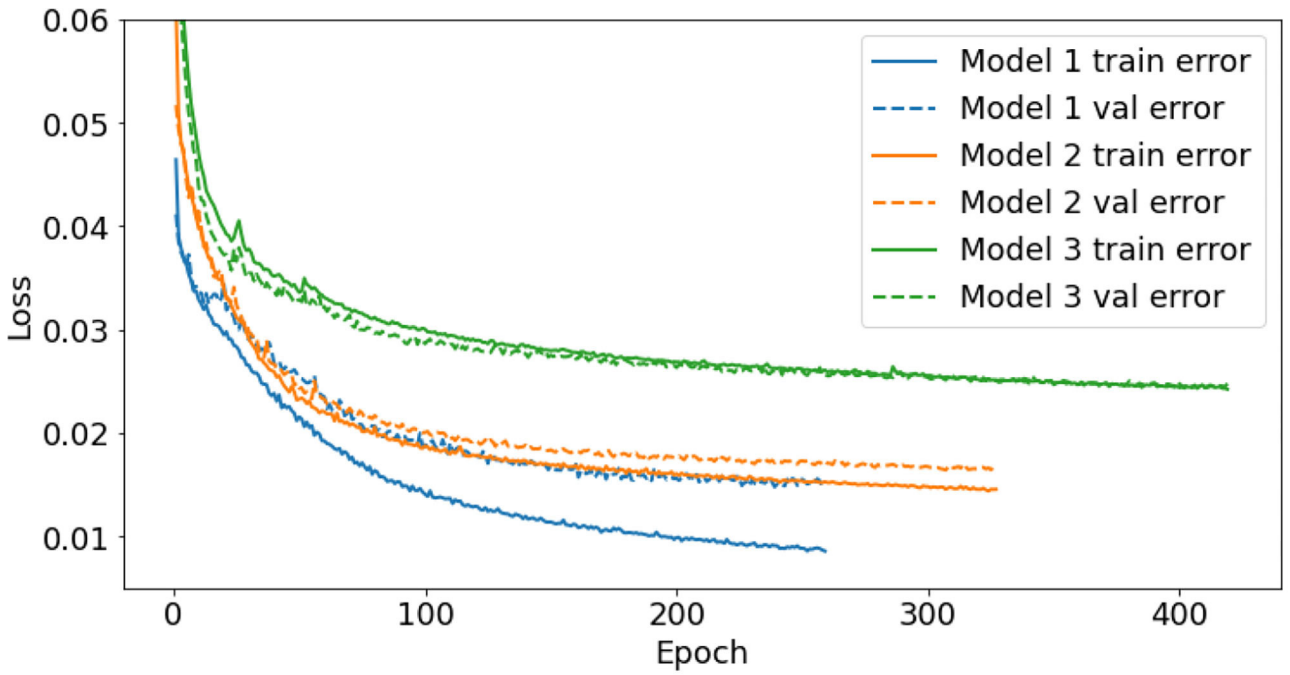


FIGURE 9 | Training history of the three models used to fill temperature gaps. Model 1 (blue) fills gaps up to 6 days long, model 2 (orange) fills fill gaps 6–30 days long, and model 3 (green) fills the remaining 2 gaps, 79 and 91 days long. Both the training (solid) and validation (dashed) losses are shown. An epoch is a single pass through all training examples.

was compared with three other models; a single layer model with 64 nodes, a single layer model with 128 nodes, and finally a model with the same three layer architecture as model 1, except without the bidirectional layers. All three alternative models were trained with the same training data as model 1, however, the final losses in all three cases were worse than the original model 1. Furthermore, there was no noticeable differences between the single layer model with 64 nodes and the single layer model with 128 nodes. **Figure 9** shows the training history of the final three models trained for the short, medium and long gaps.

4.2. Results

Sample predictions from model 1, 2, and 3 reveal that the models predict the general shape of the true observations well-based on 30, 30, and 91 days of history, respectively, but lose progressively more day-to-day variability as the prediction window lengthens (6, 30, and 91, respectively) (**Figure 10**). Although **Figure 10** only shows a single example from the test dataset for each model, similar behavior is observed for other examples as well. Overall, however, the models (1, 2, and 3) perform well with the final validation dataset losses of only 1.53, 1.65, and 2.42%, respectively (resulting in accuracies of 98.47, 98.35, and 97.58%, respectively). Note that the validation loss was calculated by averaging the differences between the true and predicted values over all days of the prediction window of the validation examples.

The NN predictions of the largest gaps contain much less variability when compared to the temperature observations before and after **Figure 11**. A hundred monte carlo dropout estimates are shown in **Figure 11**. The prediction from model 3 without any dropout is approximately in the middle of the Monte Carlo estimates which is reasonable since it can be thought of as averaging a hundred different model architectures (Hinton et al., 2012).

The entire timeseries from 2011 to 2018 (**Figure 3**) was filled using predictions from the three chosen LSTM models and is shown in **Figure 12**. The linear interpolation does not preserve any of the seasonal or shorter timescale variability that the NN prediction does, however, as mentioned earlier the NNs struggle to reproduce even shorter timescale variability present in the observations. At first glance the jumps between the filled gaps and the original data may not seem smooth in some cases (e.g., the last gap in 2013, **Figure 12B**). However, in all cases the jumps at the start or end of the gaps are similar to the day to day variability in the vicinity of the gaps.

5. DISCUSSION

A simple artificial neural network was shown to fit oxygen data with a high degree of accuracy. The model contained three hidden layers with 64, 64, and 8 nodes, respectively and was able to achieve accuracy of over 96%. Although most predictions were comfortably close to the true values, a small number of predictions presented large biases. **Figure 6** indicates that the predicted values struggles to predict some extreme concentrations. This is further confirmed by **Figure 5A** that shows that the distribution of true to predicted oxygen

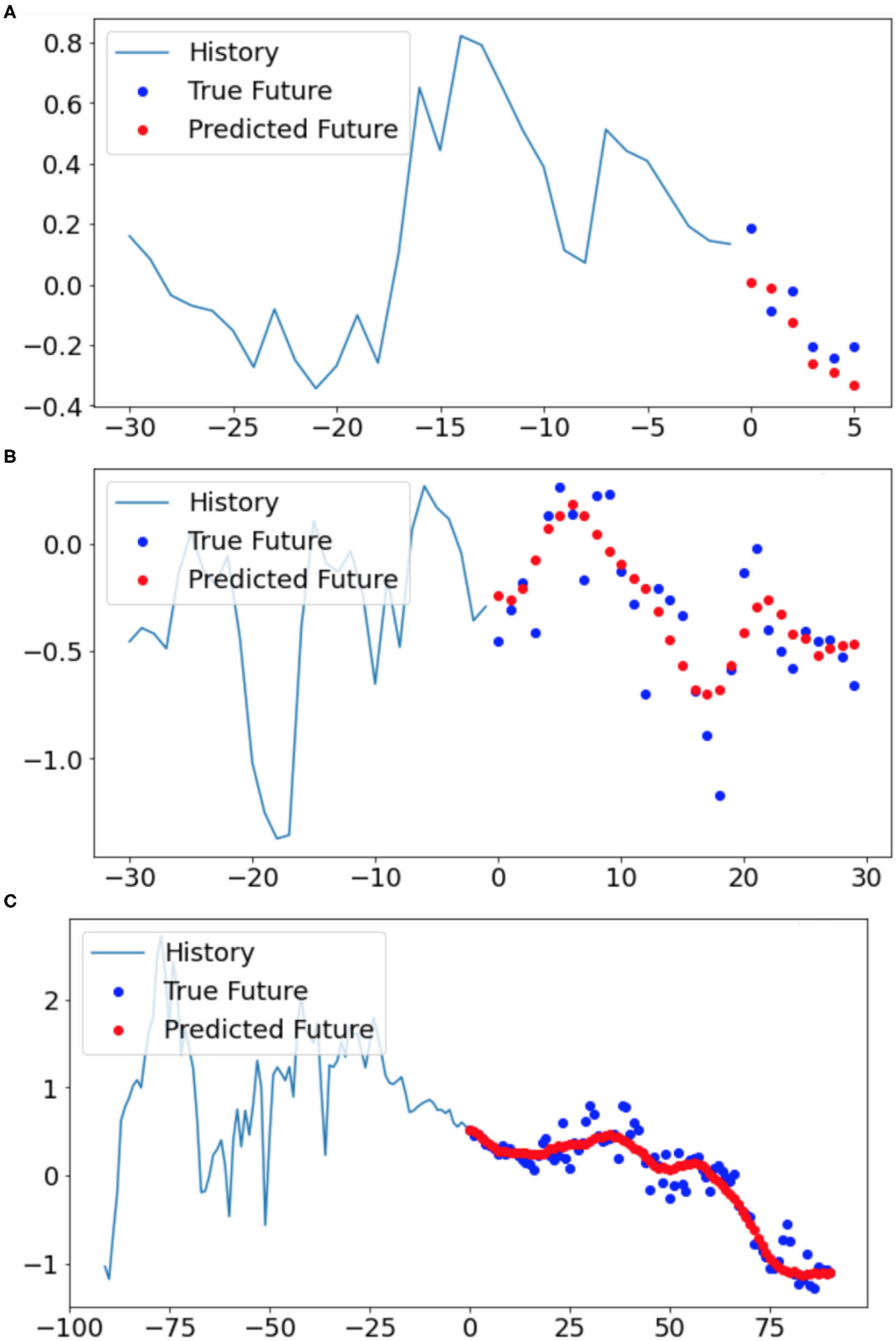


FIGURE 10 | Three samples of the model filled gaps from the validation dataset based on **(A)** model 1, **(B)** model 2, and **(C)** model 3, respectively. The blue lines show the 30, 30, and 90-day history timeseries (normalized units) which were used as input for the LSTM models 1, 2, and 3, respectively. The blue points demonstrate the observed temperatures whereas the red points indicate the model predicted temperatures over the next **(A)** 6, **(B)** 30, and **(C)** 91 days after the history window.

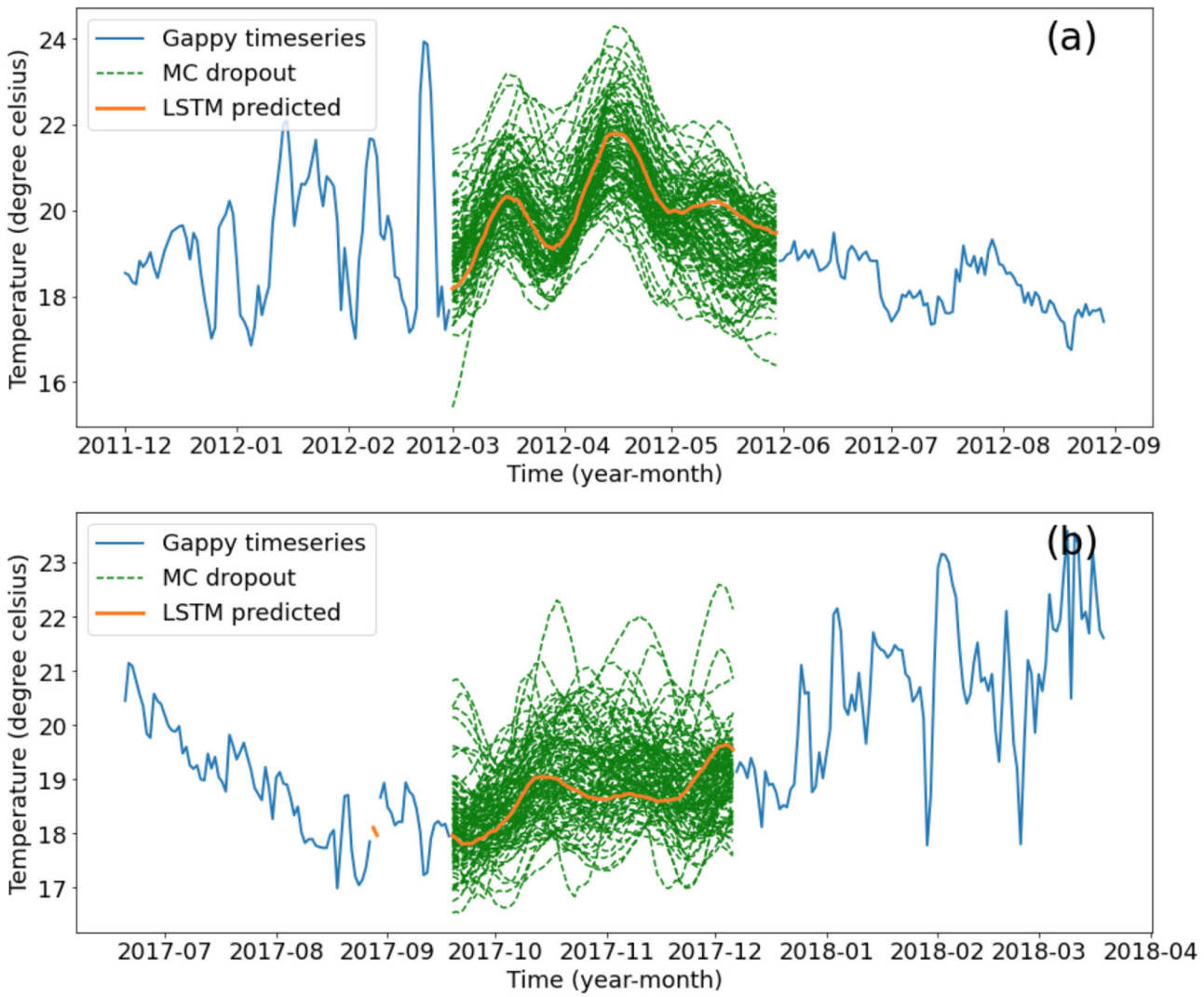


FIGURE 11 | Gap-filling of the **(a)** 91 day gap and **(b)** 79 day gap based on the final LSTM model (orange) and the five models with dropout layers (where 40% of nodes in each of the three layers are randomly switched off in the training process) in green, representing model uncertainty. Original timeseries is shown in blue. Units: °C.

concentrations is wide rather than tall, meaning the NN model is predicting concentrations in the 170–260 $\mu mol\,l^{-1}$ range even though the true values are outside this range. Note that, some of these true extreme concentrations may be a result of stochastic variability or due to variance that cannot be explained based on the chosen variables.

Sauzède et al. (2017) has also developed feedforward NN models with three layers called CANYON to predict subsurface nutrient concentration using the GLODAPv2 database (Olsen et al., 2016) that contains data from 37,863 stations globally. The CANYON models predict nutrients from latitude, longitude, DOY, year, pressure, temperature, salinity, and oxygen as features using a two hidden layer NN. Our nutrient models compare well with the CANYON models. The nitrate model RMSE with and without oxygen as a feature are 1.35 and 1.62 $\mu mol\,l^{-1}$ compared to 0.93 $\mu mol\,l^{-1}$ for CANYON. The phosphate model RMSE with and without oxygen as a feature are 0.077 and 0.010 $\mu mol\,l^{-1}$ compared to 0.066 $\mu mol\,l^{-1}$ for CANYON. Finally, the silicate model RMSE with and without oxygen as a feature are 0.65 and 0.89 $\mu mol\,l^{-1}$ compared to 3.0 $\mu mol\,l^{-1}$ for CANYON. Note, that the GLODAPv2 silicate values are much larger, in the range of 0–200 $\mu mol\,l^{-1}$, compared to the range of values used here (0–10 $\mu mol\,l^{-1}$) likely due to the depth of the samples in the water column. We suggest that the slightly lower RMSE obtained using CANYON to predict nitrate and phosphate can be explained by the significantly larger dataset used by Sauzède et al. (2017). Furthermore, our largest errors are comparable if not smaller than the CANYON models. Finally, we note that, unlike the CANYON models, our models have successfully learned to not predict concentrations below zero (**Figure 8A**).

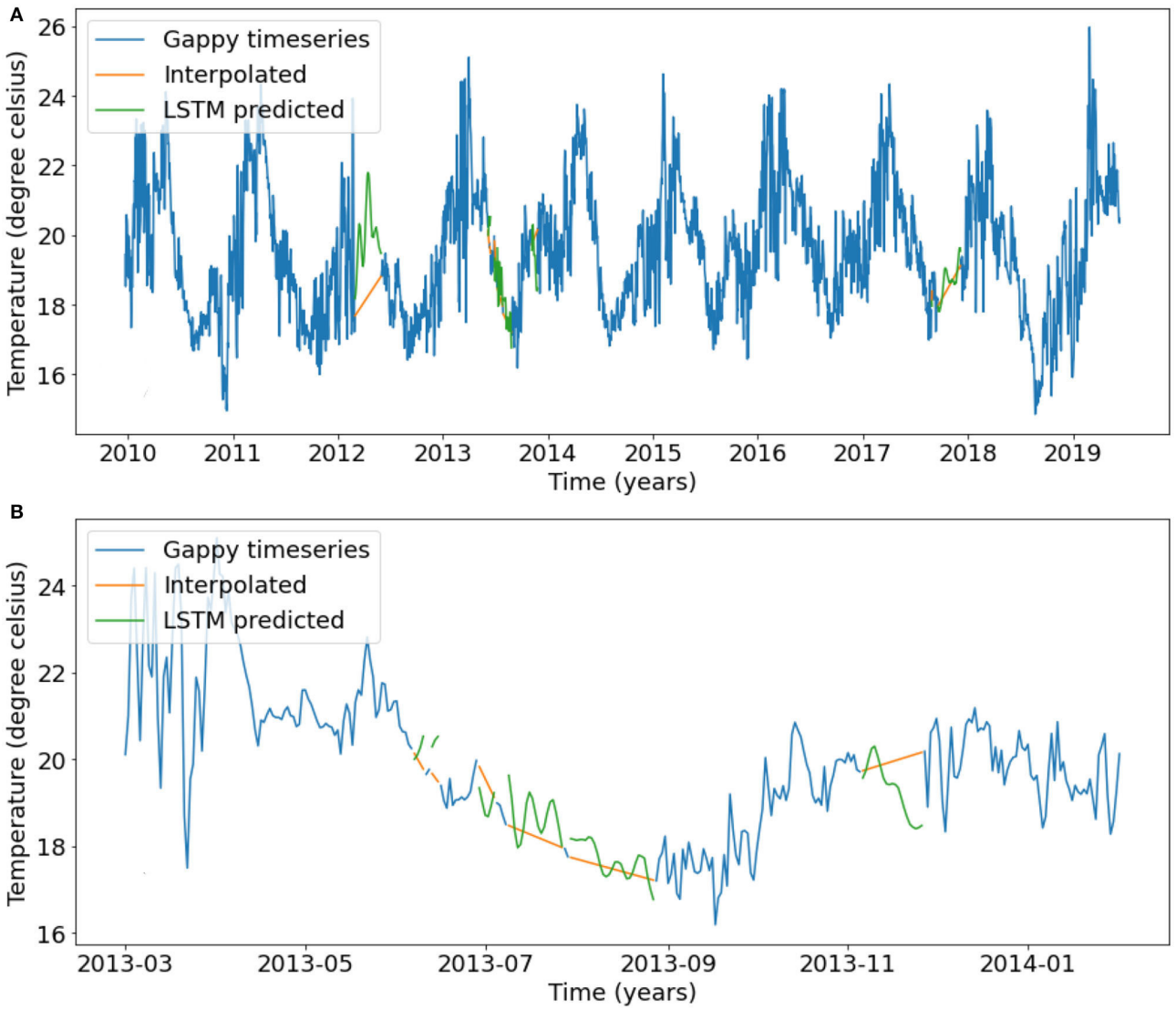


FIGURE 12 | (A) The final filled temperature timeseries at the 24.5 m bin of the Port Hacking mooring. Blue line represents the original timeseries with gaps, green lines represent the predictions from the LSTM models and the orange lines represent the predictions from linear interpolation. The long 2012 and 2017 gaps were filled using the 91-day model 3, the gaps longer than 6 days were filled using the 30 day model 2, and the remaining gaps were filled using model 1. **(B)** Same as **(A)** but zoomed in on the series of gaps in 2013.

The CANYON models are based on stations located in varying regions globally including regional coastal and open ocean sites making them much more generalized than the models presented in this study. This means that the results presented here, while not fully comparable do provide some indication of the suitable performance of our models.

The oxygen data distribution is unbounded and symmetric about its mean, whereas the distributions of the nutrients are bounded by zero and skewed with most observations close to zero. Since MAPE is defined as $(true - predicted)/true \times 100$, when the true distribution is dominated by true value close to zero (<1), the errors for these values carry much greater importance than larger true values. This means that models that use MAPE as a loss can artificially decrease their final mean loss by paying more attention to training examples which correspond to true values (labels) close to zero. This can be seen in **Supplementary Figure 1**, which shows the error distribution using a NN model with MAPE loss is skewed toward the right compared to **Figure 8**. A similar issue was encountered with the timeseries gap-filling models where the output variables needed to be standardized rendering them unsuitable to be trained with MAPE loss. In this case, we ended up implementing a modified MAPE loss that removed the standardization by multiplying by the training dataset standard deviation and adding the training dataset mean. Thus, the metric used to calculate loss proved pivotal when training the nutrient and temperature gap-filling models.

The training variance (difference between the training and validation error) for the temperature timeseries gap-filling models decreases with longer observations (**Figure 9**). However, for the purposes of a long, continuous, filled timeseries, such as creating climatologies or calculating trends, i.e., calculating long-term average statistics, short gaps are less influential on the result compared to the long gaps. Hence, the higher overfitting in model 1 compared to models 2 and 3 is less concerning. An explanation for the higher training variance of model 1 could be that models 2 and 3, which predict longer windows, learn the underlying long-term variability better than model 1, and do not learn the stochastic variability which is specific to the training dataset.

Although the predicted sequences based on the validation examples look plausible, it seems impractical to gap-fill with prediction from a single model without uncertainty estimates. There are two sources of uncertainty here: the dataset uncertainty or sampling uncertainty/bias and the modeling uncertainty. As demonstrated, dropout in hidden layers can account for the modeling uncertainty. Gal and Ghahramani (2016b) showed that such a dropout based approach can act as a Bayesian approximation of a Gaussian process, thus making it unnecessary to use Bayesian NN (Blundell et al., 2015) which can be difficult to train and computationally expensive [e.g., CANYON-B (Bittig et al., 2018)]. Furthermore, dropout in the input layers can also account for dataset uncertainty (especially useful when training on small datasets) (Hinton et al., 2012).

Since the oxygen model fails to predict some extremes (**Figure 6**) and the temperature gap-filling model predictions resemble smoothed estimates (**Figure 10**), some smoothing is apparent in NN predictions. Depending on the use case this might be worth keeping in mind. For example, our predictions are still useful for estimating long term variability.

Since ocean observing has evolved significantly over the past century, so have the ocean observing practices. It is not possible to train a neural network to use the long term, low frequency bottle data alongside shorter record, higher frequency data, such as from satellites and moorings. Hence it is not possible to train a single model that takes advantage of all the data. Furthermore, the size of the bottle dataset used for the nitrate and oxygen concentration modeling is limited by its manual collection. It has a nominal collection frequency of 2 weeks but, due to the manual nature of the data collection, sampling frequency varies greatly, with gaps up to a few months at times. Additionally, there are a lot more observations at certain depths compared to others. Automated sampling can also create consistent observations at all depths. This emphasizes the requirement for automated instruments that can collect data at regular intervals and at high frequencies, and hence has implications for ocean observing system design.

6. CONCLUSIONS

The application of neural network models was demonstrated on typical oceanographic datasets using off-the-shelf programming libraries. Point based observations at a single location (Port Hacking, Australia) were used to accurately model oxygen, nutrients, and temperatures. The oxygen model loss was below 4% resulting in an accuracy of over 96%. The nitrate, phosphate and silicate RMSE were 1.62, 0.0995, and 0.89 $\mu mol\,l^{-1}$, respectively. Finally, the temperature gap-filling model losses were 1.53, 1.65, and 2.42%, respectively (resulting in accuracies of 98.47, 98.35, and 97.58%, respectively).

Such NN based approaches have the advantage of being computationally inexpensive to both train and run. The models can be used to generate realtime predictions suitable for web-based data visualization and outreach. Furthermore, all modeling in this study was done with popular, open source, off-the-shelf frameworks, allowing easy implementation by non-experts. To further facilitate the uptake of NN modeling in oceanography, we provided details on the training process and the architectural design of the NN used in the context of fit-for-purpose modeling throughout this study.

The models developed in this study are site specific and likely do not generalize to other ocean regions, however our results do provide suitable proof of concept. Unfortunately, due to the limited data the nutrient models do not yet perform well-enough to be used as a virtual sensor or as a replacement for *in situ* sampling. As such future work will include training the oxygen and nutrient models at other locations where long-term observations are available to see if the models generalize and/or are able to improve their predictive skill.

The temperature gap-filling procedure can be applied at all depths to create a suite of temperature products. However, distinct univariate models for each depth may not preserve the correlation between adjacent temperature sensors. This could potentially be helped by training multivariate models that employ correlated sensors above and below the desired depth as additional feature variables based on our multivariate gap-filling approach. Our approach paves the way to create a whole suite of data products using the long term IMOS data.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://catalogue-imos.aodn.org.au/geonetwork/srv/eng/catalog.search#/metadata/fa93c66e-0e56-7e1d-e043-08114f8c1b76, https://catalogue-imos.aodn.org.au/geonetwork/srv/eng/catalog.search#/metadata/7e13b5f3-4a70-4e31-9e95-335efa491c5c.

AUTHOR CONTRIBUTIONS

SC and MR conceived the study and wrote the manuscript. MR led the data collection. SC conducted the analysis.

FUNDING

SC was partially funded by UNSW Australia as a co-investment to IMOS and an Australian Research Council Linkage Project #LP170100498 to MR. This research includes computations

using the computational cluster Katana supported by Research Technology Services at UNSW Sydney.

ACKNOWLEDGMENTS

Data was sourced from Australia's Integrated Marine Observing System (IMOS)–IMOS was enabled by the National Collaborative Research Infrastructure Strategy (NCRIS). It was operated by a consortium of institutions as an unincorporated joint venture, with the University of Tasmania as Lead Agent.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.637759/full#supplementary-material

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). "Tensorflow: a system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (Savannah, GA), Vol. 101, 265–283.

Adikane, H. V., Singh, R. K., Thakar, D. M., and Nene, S. N. (2001). Single-step purification and immobilization of penicillin acylase using hydrophobic ligands. *Appl. Biochem. Biotechnol*. 94, 127–134. doi: 10.1385/ABAB:94:2:127

Ashok, K., Behera, S. K., Rao, S. A., Weng, H., and Yamagata, T. (2007). El Niño Modoki and its possible teleconnection. *J. Geophys. Res. Oceans* 112, 1–27. doi: 10.1029/2006JC003798

Bailey, K., Steinberg, C., Davies, C., Galibert, G., Hidas, M., McManus, M. A., et al. (2019). Coastal Mooring observing networks and their data products: recommendations for the next decade. *Front. Mar. Sci*. 6:180. doi: 10.3389/fmars.2019.00180

Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., et al. (2018). An alternative to static climatologies: robust estimation of open ocean CO2 variables and nutrient concentrations from T, S, and O_2 data using bayesian neural networks. *Front. Mar. Sci*. 5:328. doi: 10.3389/fmars.2018.00328

Blundell, C., Cornebise, J., Kavukcuoglu, K., and Wierstra, D. (2015). Weight uncertainty in neural networks. *arXiv [Preprint]. arXiv:*1505.05424.

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

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., and Dehmer, M. (2020). An introductory review of deep learning for prediction models with big data. *Front. Artif. Intell*. 3:4. doi: 10.3389/frai.2020.00004

Fourrier, M., Coppola, L., Claustre, H., D'Ortenzio, F., Sauzède, R., and Gattuso, J. P. (2020). A regional neural network approach to estimate water-column nutrient concentrations and carbonate system variables in the Mediterranean sea: CANYON-MED. *Front. Mar. Sci*. 7:620. doi: 10.3389/fmars.2020.00620

Gal, Y., and Ghahramani, Z. (2016a). "A theoretically grounded application of dropout in recurrent neural networks," in *Advances in Neural Information Processing Systems*, eds D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Barcelona: Barcelona NIPS), 1027–1035. Available online at: https://proceedings.neurips.cc/paper/2016

Gal, Y., and Ghahramani, Z. (2016b). "Dropout as a bayesian approximation: representing model uncertainty in deep learning," in *International Conference on Machine Learning*, 1050–1059.

Graves, A. (2013). Generating Sequences With Recurrent Neural Networks. *arXiv [Preprint]*. arXiv:1308.0850.

Han, G., and Shi, Y. (2008). Development of an Atlantic Canadian coastal water level neural network model. *J. Atmos. Ocean. Technol*. 25, 2117–2132. doi: 10.1175/2008JTECHO569.1

Han, M., Feng, Y., Zhao, X., Sun, C., Hong, F., and Liu, C. (2019). A convolutional neural network using surface data to predict subsurface temperatures in the Pacific Ocean. *IEEE Access* 7, 172816–172829. doi: 10.1109/ACCESS.2019.2955957

Hemming, M. P., Roughan, M., and Schaeffer, A. (2020). Daily subsurface ocean temperature climatology using multiple data sources: new methodology. *Front. Mar. Sci*. 7:485. doi: 10.3389/fmars.2020.00485

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput*. 18, 1527–1554. doi: 10.1162/neco.2006.18.7.1527

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. *arXiv [Preprint]*. arXiv:1207.0580.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput*. 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hsieh, W. W. (2009). *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*. New York, NY: Cambridge University Press.

Ingleton, T., Morris, B., Pender, L., Darby, I., Austin, T., and Galibert, G. (2014). *Standardised Profiling CTD Data Processing Procedures V2.0*. Technical Report March, Australian National Mooring Network.

Ketkar, N. (2017). "Stochastic gradient descent BT–deep learning with python: a hands-on introduction," in *Deep Learning With Python*, ed N. Ketkar (Berkeley, CA: Apress), 113–132. doi: 10.1007/978-1-4842-2766-4_8

Kingma, D. P., and Ba, J. L. (2015). "Adam: a method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015–Conference Track Proceedings* (San Diego, CA), 1–15.

Krasnopolsky, V. M. (2013). *The application of neural networks in the earth system sciences. Neural Networks Emulations for Complex Multidimensional Mappings*. Dordrecht: Springer.

Krogh, A., and Hertz, J. A. (1992). "A simple weight decay can improve generalization," in *Advances in Neural Information Processing Systems*, Vol. 517, eds J. Moody, S. Hanson, and R. P. Lippmann (Denver, CO: Denver NIPS), 4698. Available online at: https://proceedings.neurips.cc/paper/1991

Kwon, B. C., Choi, M. J., Kim, J. T., Choi, E., Kim, Y. B., Kwon, S., et al. (2019). RetainVis: visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Trans. Vis. Comput. Graph*. 25, 299–309. doi: 10.1109/TVCG.2018.2865027

Laurent, C., Pereyra, G., Brakel, P., Zhang, Y., and Bengio, Y. (2016). "Batch normalized recurrent neural networks," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing–Proceedings* (Shanghai), 2657–2661. doi: 10.1109/ICASSP.2016.7472159

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Lee, K. A., Roughan, M., Malcolm, H. A., and Otway, N. M. (2018). Assessing the use of area- and time-averaging based on known de-correlation scales to provide satellite derived sea surface temperatures in coastal areas. *Front. Mar. Sci*. 5:261. doi: 10.3389/fmars.2018.00261

Lee, K. A., Smoothey, A. F., Harcourt, R. G., Roughan, M., Butcher, P. A., and Peddemors, V. M. (2019). Environmental drivers of abundance and residency of a large migratory shark, *Carcharhinus leucas*, inshore of a dynamic western boundary current. *Mar. Ecol. Prog. Ser*. 622, 121–137. doi: 10.3354/meps13052

Lguensat, R., Sun, M., Fablet, R., Tandeo, P., Mason, E., and Chen, G. (2018). "EddyNet: a deep neural network for pixel-wise classification of oceanic eddies," in *IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium* (Valencia), 1764–1767. doi: 10.1109/IGARSS.2018.8518411

Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. *arXiv [Preprint]*. arXiv:1506.00019

Lu, W., Su, H., Yang, X., and Yan, X. H. (2019). Subsurface temperature estimation from remote sensing data using a clustering-neural network method. *Rem. Sens. Environ*. 229, 213–222. doi: 10.1016/j.rse.2019.04.009

Lundberg, S. M., and Lee, S. I. (2017). "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Long Beach, CA: Long Beach NIPS), 4766–4775. Available online at: https://proceedings.neurips.cc/paper/2017

Lynch, T. P., Morello, E. B., Evans, K., Richardson, A. J., Rochester, W., Steinberg, C. R., et al. (2014). IMOS national reference stations: a continental-wide physical, chemical and biological coastal observing system. *PLoS ONE* 9:e0113652. doi: 10.1371/journal.pone.0113652

Makarynskyy, O. (2005). Artificial neural networks for wave tracking, retrieval and prediction. *Pacific Oceanogr*. 3, 21–30.

Makarynskyy, O., Makarynska, D., Kuhn, M., and Featherstone, W. E. (2004). "Using artificial neural networks to estimate sea level in continental and island coastal environments," in *Hydrodynamics IV: Theory and Applications*, ed Y. Goda (Yokohama: Yokohama ICHD), 451–457. doi: 10.1201/b16815-66

Malan, N., Roughan, M., and Kerry, C. (2020). The rate of coastal temperature rise adjacent to a warming western boundary current is nonuniform with latitude. *Geophys. Res. Lett*. 48:e2020GL090751. doi: 10.1029/2020GL090751

Morello, E. B., Galibert, G., Smith, D., Ridgway, K. R., Howell, B., Slawinski, D., et al. (2014). Quality Control (QC) procedures for Australia's National Reference Station's sensor data–comparing semi-autonomous systems to an expert oceanographer. *Methods Oceanogr*. 9, 17–33. doi: 10.1016/j.mio.2014.09.001

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted Boltzmann machines," in *ICML* (Haifa).

Oehmcke, S., Zielinski, O., and Kramer, O. (2018). Input quality aware convolutional LSTM networks for virtual marine sensors. *Neurocomputing* 275, 2603–2615. doi: 10.1016/j.neucom.2017.11.027

Olsen, A., Key, R. M., Van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., et al. (2016). The global ocean data analysis project version 2 (GLODAPv2)–an internally consistent data product for the world ocean. *Earth Syst. Sci. Data* 8, 297–323. doi: 10.5194/essd-8-297-2016

Rawat, W., and Wang, Z. (2017). Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput*. 29, 2352–2449. doi: 10.1162/neco_a_00990

Rosenblatt, F. (1957). *The Perceptron–A Perceiving and Recognizing Automaton*. New York, NY: Cornell Aeronautical Laboratory.

Rosindell, J., and Wong, Y. (2018). "Biodiversity, the tree of life, and science communication," in *Phylogenetic Diversity: Applications and Challenges in Biodiversity Science*, eds Rosa Scherson and D. Faith (New York, NY: Springer International publishing), 41–71. doi: 10.1007/978-3-319-93145-6_3

Roughan, M., Morris, B. D., and Suthers, I. M. (2010). NSW-IMOS: an integrated marine observing system for Southeastern Australia. *IOP Conf. Ser. Earth Environ. Sci*. 11:12030. doi: 10.1088/1755-1315/11/1/012030

Roughan, M., Schaeffer, A., and Kioroglou, S. (2013). "Assessing the design of the NSW-IMOS moored observation array from 2008–2013: recommendations for the future," in *2013 OCEANS* (San Diego, CA), 1–7.

Roughan, M., Schaeffer, A., and Suthers, I. M. (2015). "Sustained ocean observing along the coast of southeastern australia: NSW-IMOS 2007–2014," in *Coastal Ocean Observing Systems*, eds Y. Liu, H. Kerkering, and R. H. B. T. Weisberg (Boston, MA: Academic Press), 76–98. doi: 10.1016/B978-0-12-802022-7.00006-7

Ruder, S. (2016). An Overview of Gradient Descent Optimization Algorithms. *arXiv [Preprint]*. arXiv:1609.04747.

Sauzède, R., Bittig, H. C., Claustre, H., Pasqueron de Fommervault, O., Gattuso, J.-P., Legendre, L., et al. (2017). Estimates of water-column nutrient concentrations and carbonate system parameters in the global ocean: a novel approach based on neural networks. *Front. Mar. Sci*. 4:128. doi: 10.3389/fmars.2017.00128

Schaeffer, A., and Roughan, M. (2017). Subsurface intensification of marine heatwaves off southeastern Australia: the role of stratification and local winds. *Geophys. Res. Lett*. 44, 5025–5033. doi: 10.1002/2017GL073714

Schlegel, R. W., Oliver, E. C. J., Hobday, A. J., and Smit, A. J. (2019). Detecting marine heatwaves with sub-optimal data. *Front. Mar. Sci*. 6:737. doi: 10.3389/fmars.2019.00737

Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw*. 61, 85–117. doi: 10.1016/j.neunet.2014.09.003

Su, H., Li, W., and Yan, X. H. H. (2018). Retrieving temperature anomaly in the global subsurface and deeper ocean from satellite observations. *J. Geophys. Res. Oceans* 123, 399–410. doi: 10.1002/2017JC013631

Sutherland, M., Ingleton, T., and Morris, B. (2017). *Pre-Run Check and Field Sampling CTD Procedural Guide v. 3.0*. Technical Report August, Coasts and Marine Team, New South Wales Office of Environment and Heritage, Sydney and Sydney Institute of Marine Science, Sydney, NSW.

Tang, Y., and Hsieh, W. W. (2003). Nonlinear modes of decadal and interannual variability of the subsurface thermal structure in the Pacific Ocean. *J. Geophys. Res. Oceans* 108, 1–12. doi: 10.1029/2001JC001236

Tangang, F. T., Hsieh, W. W., and Tang, B. (1997). Forecasting the equatorial Pacific sea surface temperatures by neural network models. *Clim. Dyn*. 13, 135–147. doi: 10.1007/s003820050156

Tolman, H. L., Krasnopolsky, V. M., and Chalikov, D. V. (2005). Neural network approximations for nonlinear interactions in wind wave spectra: direct mapping for wind seas in deep water. *Ocean Modell*. 8, 253–278. doi: 10.1016/j.ocemod.2003.12.008

UNESCO (2019). *United Nations Decade of Ocean Science For Sustainable Development*. UNESCO. Available online at: https://www.oceandecade.org/ (accessed March 03, 2021).

Wynn, T. A., and Wickwar, V. B. (2007). "The effects of large data gaps on estimating linear trend in autocorrelated data," in *Annual Fellowship Symposium of the Rocky Mountain NASA Space Grant Consortium* (Salt Lake City, UT).

Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag*. 13, 55–75. doi: 10.1109/MCI.2018.2840738

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 07 May 2021
doi: 10.3389/fmars.2021.646926

Global Oceanic Eddy Identification: A Deep Learning Method From Argo Profiles and Altimetry Data

Xiaoyan Chen[1], Ge Chen[1,2], Linyao Ge[3], Baoxiang Huang[3] and Chuanchuan Cao[1]*

[1] Department of Marine Technology, Institute for Advanced Ocean Study, Ocean University of China, Qingdao, China, [2] Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, [3] College of Computer Science and Technology, Qingdao University, Qingdao, China

The inadequate spatial resolution of altimeter results in low identification efficiency of oceanic eddies, especially for small-scale eddies. It is well known that eddies can not only induce sea surface signal but more importantly have typical vertical structure characteristics. However, although the vertical structure characteristics are usually used for statistical analysis, they are seldom considered in the process of eddy recognition. This study is devoted to identifying eddies from the perspective of their vertical signal derived from the 18-year Argo data. Due to the irregular and noisy profile pattern, the direct identification of eddy core from Argo profile is deemed to be a challenge. With the popularity of artificial intelligence, a new hybrid method that combines the advantages of convolutional neural network (CNN) with extreme gradient boosting (XGBoost) is proposed to extract the representative vertical feature and identify eddy from a profile. First, CNN is employed as a feature extractor to automatically obtain vertical features from the input profile at the bottom of the network. Second, the obtained high-dimensional feature vectors are inputted into the XGBoost model, combined with other profile features for classifying profiles that are outside altimeter-identified eddies (Alt eddy). Finally, extensive experiments are implemented to demonstrate the efficiency of the proposed method. The results show that the classification accuracy of CNN-XGBoost model can reach 98%, and about 36% eddies are recaptured. These eddies, dubbed CNN-XGB eddies, are benchmarked against Alt eddies for the vertical structure and geographical distribution, demonstrating a similar or even stronger vertical signal and a prominent eddy belt in the tropical ocean. Within the proposed theory framework, there are various potentials to obtain a better outlook for eddy identification and *in situ* float observations.

Keywords: oceanic eddy identification, Argo profile, convolutional neural network, extreme gradient boosting, eddy core

OPEN ACCESS

Edited by:
Jun Li,
University of Technology Sydney, Australia

Reviewed by:
Antoine De Ramon N'Yeurt, University of the South Pacific, Fiji
Taoping Liu, University of Technology Sydney, Australia

****Correspondence:***
Baoxiang Huang
hbx3726@163.com

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *28 December 2020*
Accepted: *29 March 2021*
Published: *07 May 2021*

Citation:
Chen X, Chen G, Ge L, Huang B and Cao C (2021) Global Oceanic Eddy Identification: A Deep Learning Method From Argo Profiles and Altimetry Data. Front. Mar. Sci. 8:646926. doi: 10.3389/fmars.2021.646926

INTRODUCTION

Ocean eddies are omnipresent and play a significant role in transporting water mass, heat, and nutrients since their capacity of trapping fluid parcels and generate vertical movements within their cores, effectively impacting the ocean's circulation, large-scale water distribution, and biology (Bryden and Brady, 1989; Chelton et al., 2011a,b; Faghmous et al., 2012; Zhang et al., 2014; Amores et al., 2018).

Oceanic eddy detection is a key step in promoting the development of eddy science, which has been widely studied based on the variety of remote sensing data. The current mainstream eddy recognition method based on an automatic algorithm from gridded maps of sea level anomaly (SLA) is proved to be an effective method (Chelton et al., 2011b; Faghmous et al., 2015; Le Vu et al., 2018). Although the merged dataset is enabling observational studies of mesoscale features [with a spatial resolution of (1/4)° and a temporal resolution of daily] that were not previously possible using altimetry data (e.g., ∼315 km at the equator of the TOPEX/Poseidon), the limitation of satellite altimeter dataset in spatial resolution is still proven to be too large to resolve the two-dimensional of oceanic eddies (Fu et al., 2010), especially the oceanic eddies with a diameter smaller than 50 km. Amores et al. (2018) concluded that eddy recognition method based on sea surface height is largely underestimating the density of eddies; capturing only between 6 and 16% of the total number of eddies due to the limited resolution of the altimeter gridded products is not enough to capture the small-scale eddies that are the most abundant. Besides, some studies have tried to identify eddies from other sea surface characteristics. For instance, Gonzalez-Silvera et al. (2004) effectively identified and tracked 18 eddies in the tropical Pacific Ocean using 5 months of SeaWiFS and AVHRR data. D'Alimonte (2009) used sea surface temperature (SST) data to iterate the SST isotherm for automatic eddy detection. However, eddy detection using SST is prone to false positives because many other ocean phenomena also impact the sea surface temperature and surface ocean color (chlorophyll concentration) (Du et al., 2018). While the eddy detection based on the SAR principle has a high resolution and a small coverage area that can identify submesoscale or even small-scale eddies, it is easily affected by the wind field on the sea surface. Besides, many studies have shown that eddies also sometimes present as subsurface phenomena that are not noticeable by snapshots from satellite altimetry or other satellite sensors, leading to detecting uncertainties (Jeronimo and Gomez-Valdes, 2007; Andrade et al., 2014; Zhang et al., 2014; Gordon et al., 2017).

The sign of temperature/salinity anomalies inside individual eddies occurs due to geostrophic uplift and depression of the background pycnocline vertical profiles associated with eddies (Dong et al., 2012). Compared with 1–30 cm amplitude of the sea surface caused by eddies, the ∼1,000 m seawater anomaly amplitude below the sea surface caused by them cannot be ignored. The main part of the eddy should be considered as the underwater position where the largest seawater density anomaly occurs or in other words, the eddy core. It is more stable than the eddy features mapped to the sea surface and is not limited by the spatial resolution of satellite remote sensing. Its strength and depth directly determine the eddy energy and is the key to the study of eddy dynamics and thermodynamics. In this way, eddies, especially small eddies, weak eddies, subsurface eddies, etc., which cannot be detected by remote sensing, can be identified by their three-dimensional structure characteristics. With the development of the Argo global network, studies on the vertical structure of eddy have been enriched and developed. Recently, it has been effectively proved that we can derive the vertical distribution caused by eddies as criteria for anticyclonic eddies (AE)/cyclonic eddies (CE) identification through a training–learning process using concurrent altimeter-Argo measurements at a given location and then scanning all Argo profiles outside altimetrically derived eddies according to the setup criteria to locate missing eddies (Chen et al., 2020). However, the shape of an eddy's vertical structure is complex, regionally susceptible, and polarity controlled. For example, by collocating historical records of Argo profiles and satellite altimetry data, Chaigneau et al. (2011) reconstructed the mean 3-D eddy structure in the eastern South Pacific Ocean and suggested that the core (maximum temperature and salinity anomalies in the vertical direction) of CE is centered at ∼150 m below the surface, while the core of AE is centered at ∼400 m. Dong et al. (2012) shed light on three types of eddies' vertical shapes: bowl shaped (with the largest radius at the surface), lens shaped (with the largest radius at the middle), and cone shaped (with the largest radius at the bottom) based on high-resolution numerical model product. Pegliasco et al. (2015) analyzed different eddy vertical shapes by using clustering analysis in four major Eastern Boundary Upwelling Systems. These existing studies have a common conclusion, that is, the vertical shapes of oceanic eddies are not spatially and timely uniform, and the eddy cores of AE and CE are asymmetric even if they are in the same region. Moreover, these eddy structures were obtained through the composite superposition or multiprofile averaged, which were relatively smooth and regular. But in general, an original eddy profile will present a more irregular and noisy profile pattern making it a challenge to seek out the eddy core directly. Consequently, the application of artificial intelligence will be the expecting method to extract the representative feature of each profile for eddy identification.

Deep-learning algorithms, which learn the representative and discriminative features hierarchically from the data, have been becoming a hotspot and have been introduced into the geoscience community for big data analysis. Considering the low-level features as the bottom level, the output feature representation from the top level of the network can be directly fed into a subsequent classifier for pixel-based classification (Zhang et al., 2016). Convolutional neural network (CNN) is an efficient deep learning model with hierarchical structure to learn high-quality features at each layer. Since the model can reduce the complexity of network structure and the number of parameters through local receptive fields, weight sharing, and pooling operation, and can actively extract high-dimensional features from big data as well, it is a suitable model that can be used to extract profile feature information. Although CNN has been recognized as one of the most powerful and effective mechanisms for feature extraction, traditional classifiers connected to CNN cannot fully understand the extracted features (Ren et al., 2017). Boosting is one of the most prominent classification techniques in the state-of-the-art, providing the best accuracy levels at many problems. However, the most known boosting algorithm, AdaBoost, has already been proven to be sensitive against noise. Since Chen's proposal of the extreme gradient boosting (XGBoost) model (Chen and Guestrin, 2016), the theory and application of the decision tree method have developed significantly. This model is proven to be the most robust algorithm in both binary and multiclass datasets for its strong resolution of data noise, fast calculation speed,

and high accuracy and has been widely employed in various classification applications. Inspired by the above facts, we aimed to develop a novel hybrid CNN-XGBoost model for fully learning the profile features. This composite model can provide more accurate classification results by regarding CNN as a trainable feature extractor to automatically obtain features from the input profile and XGBoost as a recognizer and combined with other remarkable profile features in the top level of the network to produce results.

This paper aims to propose a new perspective to identify eddies from vertical structure characteristics with deep learning architecture. The objective of this paper is mainly twofold: one is the hybrid CNN-XGBoost architecture construction for vertical characteristics extraction, and the other is the comparison and verification of the obtained eddy dataset. This is a significant step forward compared to previous studies that focused only on "sea surface characteristics of eddies" for identification. Such an unprecedented improvement in state-of-the-art spatial–temporal sampling and coverage of the Argo project allows us to obtain a unique and massive dataset of subsurface T/S records. The rest of the paper is organized as follows: In *Data*, we describe the two datasets (altimetry and Argo data) used in this work, and the eddy identification algorithm based on satellite data is briefly presented. *Methods* introduces the methods, including the dataset preprocessing principles, CNN model, and XGB model principles and architecture along with the experimental environment. The results are presented in *Results*. The global eddy vertical structure characteristics are described in *Global Eddy Vertical Structure Characteristics*. *Model Training Process* explains the training process and accuracy evaluation of the model. In *Eddy Identification Based on CNN-XGBoost Model*, both the vertical structure and geographical distribution characteristics of identified eddies by our method are analyzed in detail. Additionally, a concise result on the relationship between eddy property and its vertical structure is presented in *Vertical Structure of CNN-XGB Eddy*. Finally, *Conclusions* contains a brief discussion and our conclusions.

DATA

Satellite Altimeter Data

The sea level anomaly (SLA) data used in this study is delayed time products generated by Archiving, Validation, and Interpretation of Satellite Oceanographic (AVISO) from a combination of T/P, Jason-1, Jason-2, Jason-3, and Envisat missions. The SLA dataset spanned 18 years, from January 2002 to October 2019, and had a daily temporal resolution and a (1/4)° × (1/4)° spatial resolution. In the present analysis, a four-step scheme has been optimized for eddy identification based on our earlier work by Liu et al. (2016). First, a high-pass filtering is applied to the global SLA data using a Gaussian filter with a zonal radius of 10° and a meridional radius of 5° before seed points are effectively determined. Second, the global SLA fields are divided into regular blocks with a zonal spacing of 45° and a meridional spacing of 36°. Third, SLA contours are computed with a 0.25-cm interval, and eddy boundaries corresponding to maximum geostrophic velocity are subsequently extracted. Finally, all blocks are merged seamlessly into a global map with duplicated eddies eliminated. Following the identification schemes, a comprehensive eddy dataset has been created for the global ocean, which is available at http://coadc.ouc.edu.cn/tfl/ and http://data.casearth.cn/ (Data ID: XDA19090202), and relevant technical details (including models and procedures) can be found in Liu et al. (2016), Sun et al. (2017), and Tian et al. (2020).

Argo Floats

The Array for Real-time Geostrophic Oceanography (Argo) data also spanned from January 2002 to October 2019. The Argo project is the first global observation system for the subsurface ocean and is one of the best sources for *in situ* temperature measurements. Argo floats observe large temporal (seasonal and longer) and spatial (thousand kilometers and larger) scale subsurface ocean variability worldwide (Roemmich et al., 2009). At present, Argo is collecting ∼12,000 data profiles each month (∼400 a day). This greatly exceeds the amount of data that can be collected from below the ocean surface by any other method. By September 21, 2020, there are as many as 3910 active floats disseminated around the global ocean spacing nominally at every 3° of longitude and latitude. In this analysis, the Argo floats data are provided by the Coriolis Global Data Acquisition Center of France through their website: www.coriolis.eu.org. The quality control and processing of Argo data are conducted automatically by the Argo data center, and only profiles flagged as "good" or "probably good" are downloaded. Meanwhile, additional data filtering is applied to profiles with first measurement shallower than 10 m and last measurement deeper than 1,000 m and meanwhile having at least 30 valid data points within the 0–1,000 m depth range. Finally, linear interpolation with an interval of 1 m is carried out for all edited high-quality profiles in the global ocean.

Since density is a combination of temperature and salinity, it can comprehensively evaluate the vertical structure anomalies caused by eddies. Therefore, in this paper, the potential density anomaly (PDA) is chosen as the property to represent the vertical structure of eddies for eddy identification. After preprocessing and filtering the profile data, the potential density is calculated based on International Thermodynamic Equation of Seawater 2010 from Argo T/S profile data using the Gibbs Seawater Oceanographic toolbox (McDougall et al., 2009), and anomalies of every profile are computed by removing climatological profiles. Here, the climatological T/S profiles are obtained by interpolating the CSIRO Atlas of Regional Seas to floats' positions and times (Dunn and Ridgway, 2002; Ridgway et al., 2002).

MATERIALS AND METHODS

Data Preprocessing for Different Types of Eddies

We established our dataset starting from the Argo profiles data. Based on the positions and times of eddies and Argo floats, Argo profiles are classified into three categories depending on whether

the floats are inside the effective boundaries (i.e., the outermost enclosed contour of SLA surrounding the eddy centroid) of the altimeter recognized eddies (Alt eddy) or outside eddies. The numbers of profiles that fall inside anticyclonic eddies (Alt AE) or cyclonic eddies (Alt CE) in this study are 343,802 and 328,936, respectively, and 1,271,784 profiles are outside Alt eddies. **Figure 1** shows the geographical distribution of Argo profiles. In general, the patterns of these figures are determined by the locations where the Argo floats are gathered such as the Kuroshio extension, the Arabian Sea, and the North Atlantic. However, compared with **Figures 1b,c**, it can be found that the Argo profile inside Alt eddy only accounts for about one-third of the total profile (**Figure 1b**). Two-thirds of profiles are outside the Alt eddy (**Figure 1c**). However, whether these profiles are actually outside the eddy or actually inside the eddy but just not captured by the altimeter needs to be determined by their vertical structure characteristics.

There is a part of the eddies that are identified by the altimeter as cyclonic (anticyclonic) eddies, but there are negative (positive) PDA; that is, their positive and negative polarities of the eddy core are opposite to the polarity detected by the altimeter (Chen et al., 2020). Therefore, before performing eddy recognition, we first need to purify the profile data within the eddy and pick out abnormal eddies to ensure the data quality. Theoretically, the CE (AE) should have a positive (negative) anomaly, so we calculated the sum of the CE (AE) profile from 20 to 1,000 m (the near-surface layer depth is set to −20 m to reduce the effect of noise near the ocean surface), and eliminated the profiles with negative (positive) values. At present, there is no clear definition of which profiles are determined outside eddies. But generally, we suppose the profiles that are located outside the effective eddy boundary identified by the existing altimeter resolution and causes weak vertical structure signal to have a great probability of actually being outside eddies. Therefore, based on our research foundation and in order to keep a balance of the sample dataset, the out eddy profiles (OEs) dataset is established for the profiles that are outside the effective boundary of eddies and cause 15% of the weakest vertical structure signal in every 5° × 5° grid. After data preprocessing, the numbers of profiles inside AE or CE as the modeling dataset are 207,798 and 188,809, respectively. The datasets of these profiles are randomly sampled and divided into a training dataset (60%) and a testing dataset (40%). The statistical results of the training and testing datasets are shown in **Table 1**, including the total numbers of sample profiles and the mean ± standard deviation of the PDA values in each dataset.

Models

In this paper, a new hybrid model that combines the advantages of convolutional neural network (CNN) and extreme gradient boosting (XGBoost) is proposed. The proposed model is combined with two parts: feature extraction and classification. CNN is used to extract and select the features of the profile data at the bottom of the network, and the obtained high-dimensional feature vectors are inputted into the XGBoost model for profile classification. The experimental environment of model building is performed on a computer with an Intel i7-9700F CPU @3.00 GHz with 32 GB memory, Windows 10 OS, and Anaconda 3. Python is the main programming language, and CNN is implemented under the neural network framework Pytorch 1.5.1.

TABLE 1 | Sample statistics of training and testing datasets.

Sample datasets	All samples	Training samples	Test samples	Training mean (kg/m^3)	Testing mean (kg/m^3)
AE	207,798	124,679	83,119	−0.128 ± 0.178	−0.129 ± 0.180
CE	188,809	113,285	75,524	0.083 ± 0.147	0.082 ± 0.149
OE	183,841	110,305	73,536	−0.0004 ± 0.062	−0.0003 ± 0.062

Convolutional Neural Network

A supervised CNN model is proposed to extract features from profiles. CNN model goes through a training process with inputted training profiles, and the cross-entropy loss is computed for back propagation to optimize the model parameters. The detailed process of the model is described as follows.

For our dataset with m profiles and k-dimensional vertical features, let $p^l_{train} = (d_{20}, d_{21}, d_{22}, \ldots, d_{k-1}, d_k, k = 1,000)$ be an input vector of the CNN, where d_z represents the PDA at the depth of z (m). The features extracted by the first hidden layer can be described as **Equation (1)**.

$$Z_1 = W_1 \otimes P^l_{train} + b_1 \tag{1}$$

where $\bigotimes$ represents the convolution process. W_1 and b_1 are the weight and bias of the first hidden layer, respectively, and Z_1 is the output of the first hidden layer.

Then, other hidden layers can be described as **Equation (2)**.

$$Z_i = W_i \otimes Z_{i-1} + b_{i-1}, i = 2, ..., n \tag{2}$$

where n is the number of the hidden layer. Suppose $q\left(f_1, f_2, , f_j, , j = 64\right) = C(p^l)$ is the obtained feature vector from CNN, the process of XGBoost model can be defined as $P = X(q, features_{vertical})$, where P represents the output probability of XGBoost model, which is computed by softmax function. $features_{vertical}$ represents the additional vertical structure features (see details in *CNN-XGBoost Model*). Therefore, the proposed model can be defined as **Equation (3)**.

$$P = X\left[C(p^l_{train})\right] \tag{3}$$

The cross-entropy loss is defined as **Equation (4)**.

$$L = -\sum_{a=1}^{c} y_a \log P_a \tag{4}$$

where y_a is the real label, P_a is the probability of class a, and c is the class number. The training process finished when the loss convergence. Then, the testing dataset $[p^l{}_{test} = (d_{20}, d_{21}, d_{22}, \ldots, d_{k-1}, d_k, k = 1,000)]$ are inputted into the model to evaluate accuracy.

Extreme Gradient Boosting

XGBoost is a highly effective scalable machine learning model developed by Chen and Guestrin (2016), which has been widely used in many fields to achieve state-of-the-art results on data

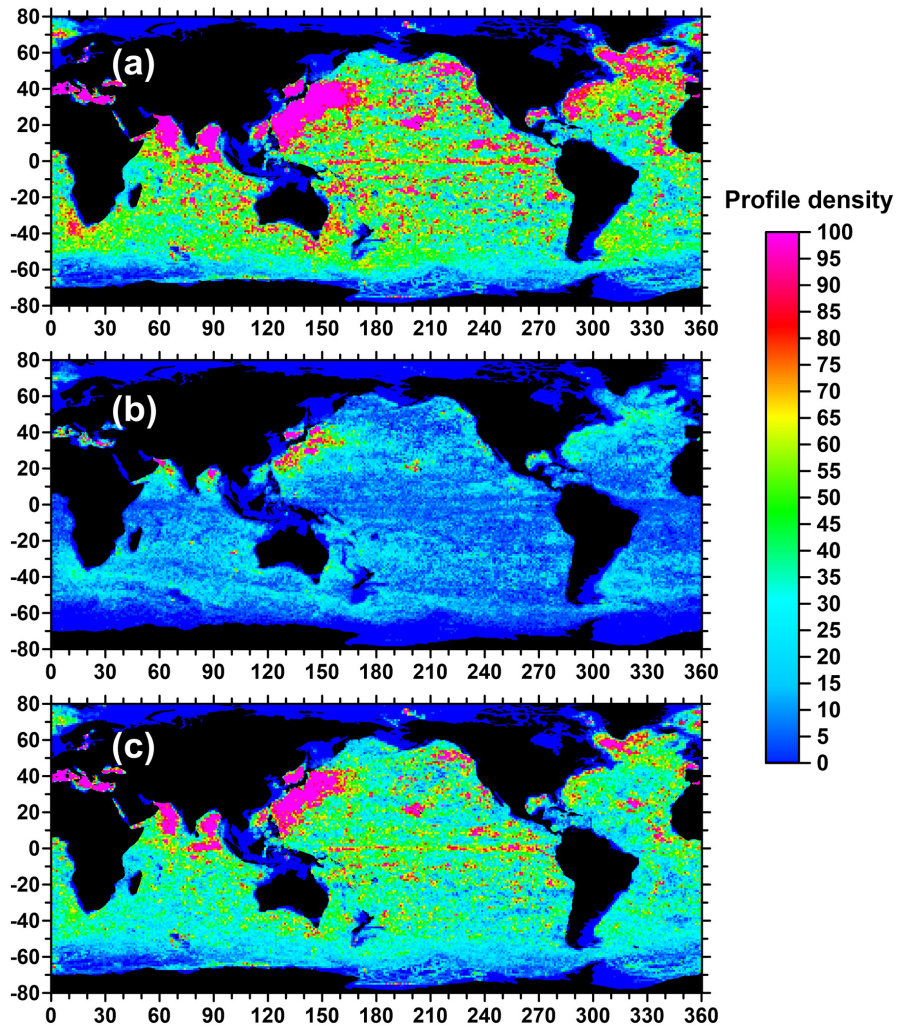


FIGURE 1 | Global geographic distribution of Argo profiles during 2002–2019 under 1° × 1° grid. **(a–c)** Cumulative number of all profiles, profiles inside Alt eddies, and profiles outside Alt eddies, respectively.

challenges. It is greatly improved and optimized on the traditional Gradient Boosting Decision Tree, mainly including using the second-order Taylor expansion and introducing the regular term to prevent overfitting, which greatly promotes the accuracy, efficiency, and flexibility of the model. After the training of CNN model, a 64-dimensional feature vector is obtained and inputted into the XGB model. The principle of the XGBoost model is described as follows. For a dataset with n samples and m features $D = \{(x_i, y_i)\}$ $(|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R}^n)$, $\hat{y}_i^k$ is the predicted result of x_i in round k. A tree boosting model output $\hat{y}_i^k$ with K trees is defined as follows:

$$\widehat{y_i} = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F \tag{5}$$

where $F = \{f(x) = w_{q(x)}\}$ $(q : \mathbb{R}^m \to T, w \in \mathbb{R}^T)$ is the space of regression or classification trees, w is the weight of the leaf child nodes of the regression tree, and q is the structure of the regression tree. T is the number of leaf nodes. In order to learn the parameters of this model, it is necessary to minimize the objective function:

$$Obj = \sum_i l(\widehat{y_i}, y_i) + \sum_k \Omega(f_k) \tag{6}$$

where l in **Equation (6)** is a training loss function that measures the distance between the prediction $\widehat{y_i}$ and the object y_i. The second term in **Equation (6)** is defined as follows, which represents the penalty term of the tree model complexity.

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{7}$$

where γ is the regularization term (based on the number of leaves of the tree and the scores of each leaf), and T is the number of leaves. λ is the L2 regularization parameter, and the final prediction by summing up the score in the corresponding leaves

is given by ω. XGBoost is a forward step-by-step algorithm, that is, in the round t iteration, adding a new model f_t needs to minimize the following target function:

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \widehat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{8}$$

XGBoost algorithm uses second-order Taylor expansion for optimization. After the second-order Taylor expansion, the following formula is obtained:

$$Obj^{(t)} = \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega\left(f_t\right) \tag{9}$$

Denote $I_j = \left\{ i | q(x_i) = j \right\}$ as the instance set of leaf. Where $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$, $h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$ are first- and second-order gradient statistics on the loss function. After substituting g_i, h_i, and into **Equation (9)**, the formula can be rewritten as:

$$Obj^{(t)} = \sum_{j=1}^{T} \left[G_j w_j + \frac{1}{2}\left(H_j + \lambda\right) w_j^2 \right] + \gamma T \tag{10}$$

where $G_j = \sum_{i \in I_j} g_i$, $H_j = \sum_{i \in I_j} h_i$. For a regression tree with a specific structure, the solution weight $w_j^* = \frac{G_j}{H_j + \lambda}$ is used to evaluate the quality of the tree structure by deriving w_j. A greedy algorithm is used to determine the tree structure that starts from a single leaf and iteratively adds branches to grow the tree structure. Whether adding a split to the existing tree structure can be decided by the following function:

$$O_{split} = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G^2}{H + \lambda} \right] - \gamma \tag{11}$$

where I_L and I_R are the instance sets of the left and right nodes after the split and $I = I_L \cup I_R$.

CNN-XGBoost Model

In this paper, the specific structure CNN-XGBoost model is used for eddy identification. **Figure 2** visualizes the architecture of the model. We use six CNN building blocks with convolutional, batch normalization, activation layers, and a pooling layer. Because of the strong feature expression ability of CNN, the accuracy of the model will decrease with the increase in convolution layer, but restricting the depth of the model may reduce the classification accuracy. Therefore, we adopt a doubleconvpool structure. Doubleconvpool structure includes two convolution layers, two batch normalization layers, and one pooling layer. That is, there is no pooling layer between two consecutive convolution layers to realize the retention and transmission of feature information and balance between the high-dimensional characteristics of profile and model depth. Meanwhile, the model uses batch normalization and dropout to avoid overfitting. Then, the ReLU activation function is set after convolution to add nonlinear factors to increase the expression ability of the model. Each profile is inputted into the model as a one-dimensional feature vector (the depth range is 20–1,000 m). After six layers of CNN learning, a 64-dimensional feature vector is obtained and input into the XGB model. For the XGBoost classifier, in addition to the profile feature vector learned through the CNN model, the input features also include the spatiotemporal features of the eddy (location: longitude and latitude, month in which the eddy is recognized) as well as the vertical structure characteristics of eddy (including profile minimum, maximum and its corresponding depth, integral area of the entire profile, integral areas of 300 m interval, extreme values of 300 m intervals, and the corresponding depth) to comprehensively evaluate the characteristics of the eddy profile. After training and verifying the XGBoost model, the outputs are the probabilities of the classes "AE," "CE," or "non-eddy."

RESULTS

Global Eddy Vertical Structure Characteristics

The vertical PDA profiles of Alt eddy are shown in **Figure 3**. The thick lines in **Figure 3A** represent the global average profiles over the entire time series. The thin lines are two randomly selected profiles in Alt eddy without smoothing. They show notable eddy vertical structures and accompanied by large noise and burr signal. Meanwhile, there is a dual-core in the AE (thin red line in **Figure 3A**). Even if the anomaly caused by the lower eddy core is smaller, the main eddy core should be located here with a wider eddy core. For CE (thin blue line in **Figure 3A**), obviously, in addition to a small anomaly in the sea surface, this eddy presents a single core structure with a depth ranging from about 50 m to nearly 400 m. The seasonal mean profiles from 2002 to 2019 are shown in **Figure 3B**, suggesting that the position and intensity of the eddy core will change along with the change in seasons, which is a reason that we add months as a feature to the model. **Figures 3C,D** present the latitude average vertical structure of the eddy PDA to show the variation in the depth and intensity of the eddy core on the latitude zone. On the whole, the eddy intensity in the northern hemisphere is stronger than that in the southern hemisphere, and the intensity of AE is stronger than that of CE. Along with the increase in latitude, the depth of the eddy core gradually increases, and a dual-core structure may appear in mid-latitude regions. Regarding the eddy intensity, we take about 200 m as the boundary. With the increase in latitude, the PDA intensity above 200 m becomes weaker and below 200 m becomes stronger. From the latitude-averaging profiles, we can see the general change trend of the eddy core, but for different small areas, the pattern of the eddy profiles will be greatly different (see **Figure 6** for details). Thus, it is essential to consider the location (longitude and latitude of the eddy) when determining whether a float is located inside the eddy.

Model Training Process

In order to reveal the effectiveness of the CNN-XGB hybrid model, we test the accuracy of CNN model, XGB model, and the hybrid model for profile classification. The input feature of CNN and XGBoost model is the PDA profile vector, and softmax is

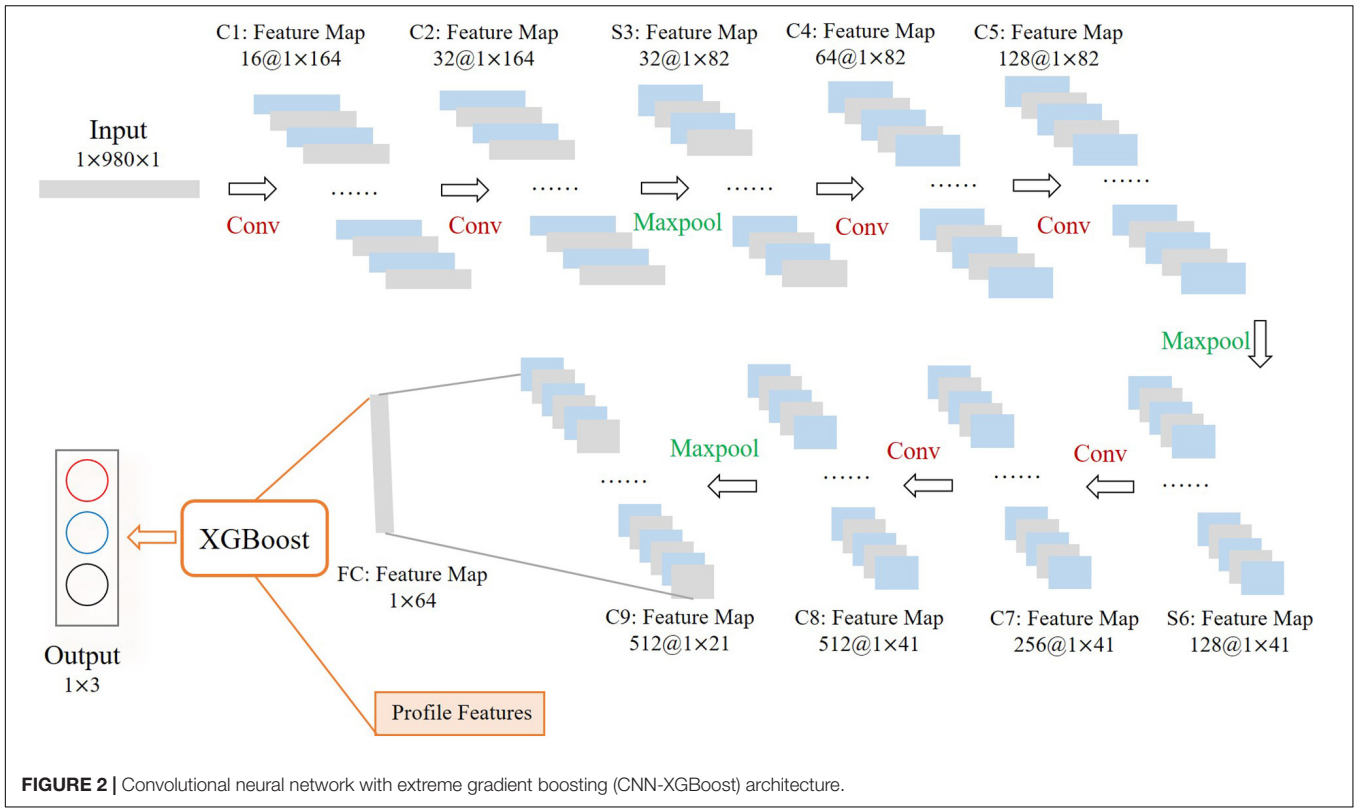


FIGURE 2 | Convolutional neural network with extreme gradient boosting (CNN-XGBoost) architecture.

used as the classifier. The running time, accuracy, and sensitivity assessment of these three models are summarized in **Table 2**. Note that the sensitivity assessment refers to the proportion of a kind of eddies that are mistakenly classified into the other two types (i.e., AE is misjudged as CE or OE), so as to evaluate the specificity of the proposed algorithm to eddies in different polarity. It can clearly reflect that the classification accuracy of CNN-XGBoost model is better than that of CNN or XGBoost alone, and the accuracy can be improved by ∼4 and ∼6% on CNN and XGBoost model, respectively. The specificity analysis of the three models shows that the probability of AE being misjudged is less than that of CE. The reason for this result is that compared with the PDA induced by AE, CE is weaker (see **Table 1** and **Figure 3** for the mean PDA values and profiles of AE and CE, respectively). This result is also consistent with the conclusion that the proportion of abnormal CE is larger (see Chen et al., 2020 for details). In addition, AE (CE) is more likely to be judged as OE than CE (AE). It is comprehensible that eddies with weak signals are easier to be judged as OE (for its critical PDA signal close to zero) rather than the classification of the opposite polarity.

To train CNN-XGBoost model, the first is the training of the CNN model. We used cross-entropy as the loss function, and the optimization algorithm of the training process adopts Adam algorithm. A 3 × 3 convolution kernel size is used in every convolutional layer, and the stride is 5. Furthermore, maximum pooling using a kernel of size 2 × 2 is added to prevent overfitting. We choose different learning rates (0.0005, 0.001, 0.003, 0.005, and 0.01) to experiment on the dataset. It shows that when the learning rate is 0.0005, the convergence speed is the fastest, and the loss function is the smallest when it converges, so this experiment uses 0.0005 as the value of the learning rate. After 100 epochs of training, the test accuracy of the CNN model can achieve ∼94%. **Figure 4** presents the accuracy and loss variation on validation sets, showing that the loss and accuracy of the model have been well converged.

For the XGBoost classifier, the Bayesian optimization is applied to obtain the best configuration of the hyperparameters. This optimization method is obtained from a Gaussian process prior and constantly updates the prior knowledge by considering the previous parameter information, whereas a conventional grid search or random search considers no prior parameter information. In addition, the Bayesian optimization process uses a small number of iterations and has a rapid running speed, allowing it to optimize algorithms with multiple parameters such as XGBoost. After parameters optimization, we select the best parameters including the *eta*, *gamma*, *max_depth*, and *min_child_weight* values of 0.4, 0.8, 10, and 1, respectively. The other parameters are set to the default values. The objective function is softmax for multiclassification, and the multiclass error rate is selected for the evaluation metric. After the 100 iterations' training of XGBoost model, the CNN-XGB model accuracy reaches 98%.

Eddy Identification Based on CNN-XGBoost Model

The CNN-XGBoost model is used to learn and classify almost 120 million profiles that are not in the Alt eddy. It took only

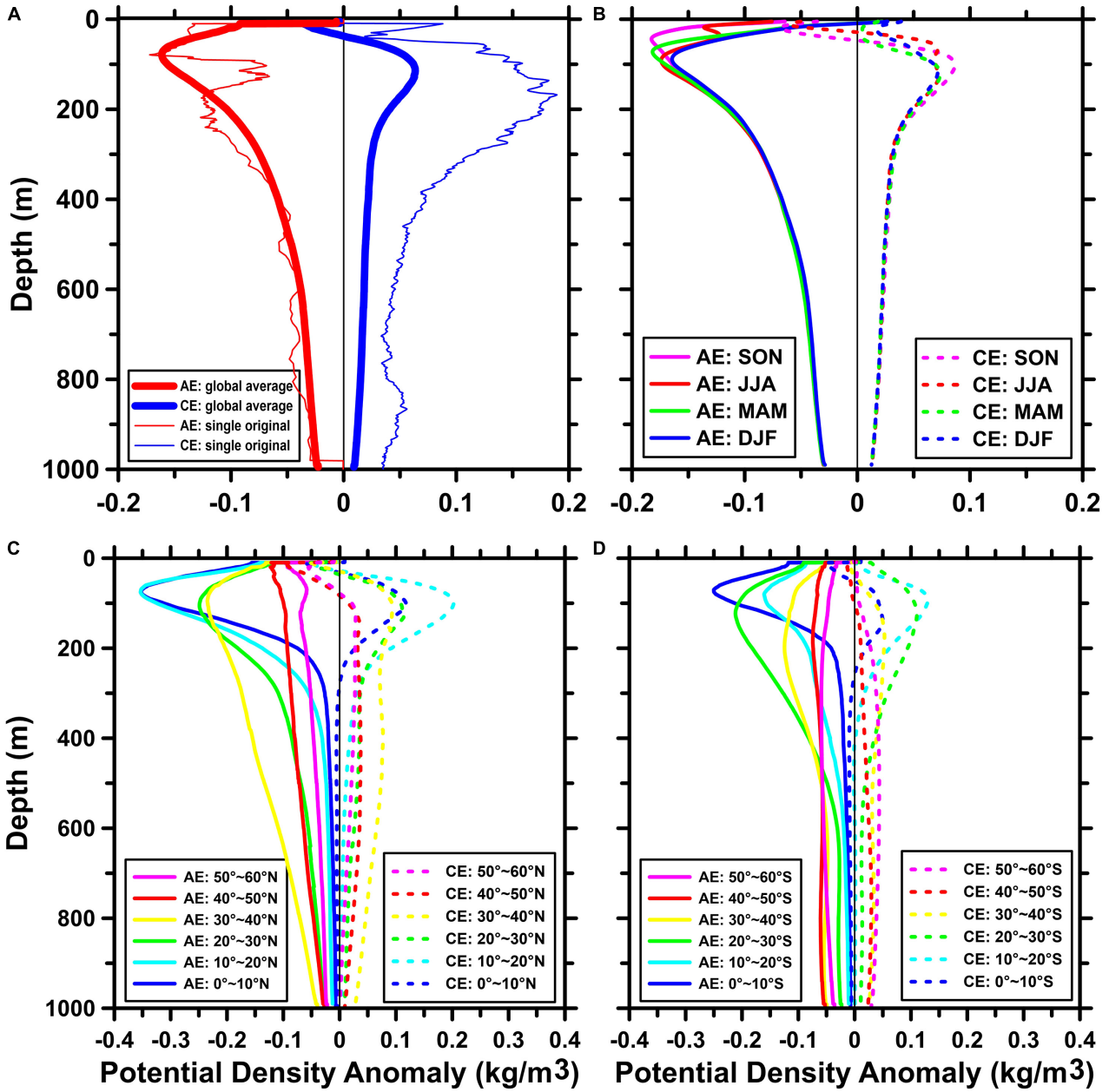


FIGURE 3 | Vertical profiles of the potential density anomaly of Alt eddy. **(A)** Global average profiles from 2002 to 2019 (thick lines). The thin red (blue) line is the original eddy profile inside Alt anticyclonic eddies (AE) [cyclonic eddies (CE)], and the buoy number is 1901621 on March 26, 2018 and 1901817 on October 9, 2018, respectively. **(B)** Global average profiles in different seasons. **(C)** Latitudinal average profiles in the Northern hemisphere and **(D)** latitudinal average profiles in the Southern hemisphere.

~17 min to classify the entire dataset, and then, we got the profiles that are outside Alt eddy but inside the eddy identified by the CNN-XGB method (CNN-XGB eddy). Recently, we proposed a methodology to effectively combine the vertical structure signals and sea surface topological structure of eddies based on a mathematical PDA algorithm (PDA eddy, Chen et al., 2020). Compared with this method, the CNN-XGB model has two significant advantages. First is higher computing efficiency. Based on the CNN-XGB method, we can scan the global profiles in < 20 min, while the mathematical method takes several days or even longer to complete the calculation, which reflects the unique advantages of AI model. Second is better spatial continuity. In order to take into account both the regional characteristics of eddy vertical structure and the number of training samples, we use 5° grid as the unit for eddy identification in the PDA algorithm, which limits the resolution of the algorithm to 5°. However, The AI model is not limited by the spatial grid and thus has stronger spatial continuity. In this section, we compare eddy

TABLE 2 | Comparison of three different models.

Model	Running time	Accuracy	Sensitivity and specificity			
			AE ($\varphi_{CE}/\varphi_{OE}$)		CE ($\varphi_{AE}/\varphi_{OE}$)	
CNN	138 min	93.8%	0.43%	0.98%	1.12%	2.88%
XGBoost	43 min	92.1%	0.65%	1.56%	1.62%	3.19%
CNN-XGBoost	221 min	98.3%	0	0.06%	0	1.41%

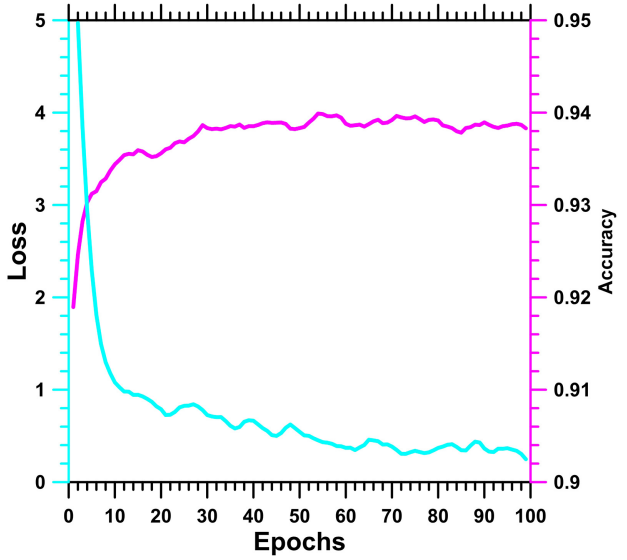


FIGURE 4 | Accuracy and loss variation on validation sets of convolutional neural network (CNN) model.

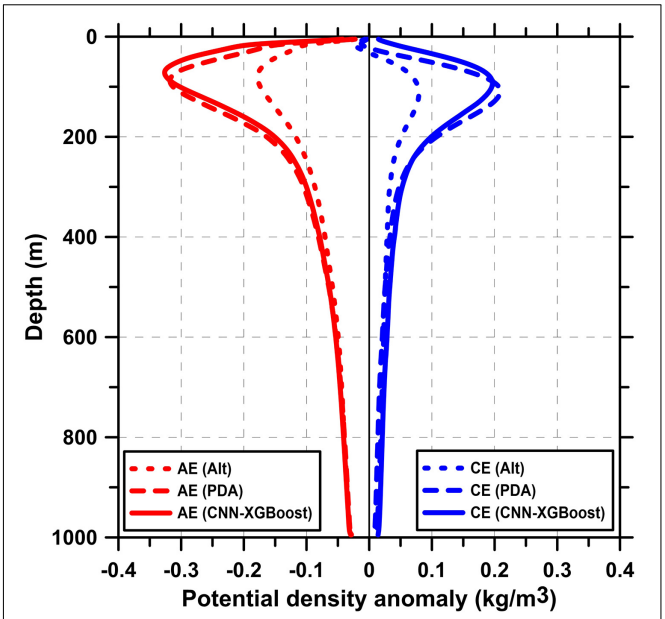


FIGURE 5 | Global mean potential density anomaly profiles for different types of eddies from 2002 to 2019. The short dashed line is the profile of Alt-eddy. The long dashed line is the profile of potential density anomaly (PDA) eddy and the solid line is the vertical structure that is inside eddies through the convolutional neural network with extreme gradient boosting (CNN-XGBoost) model but not recognized by the altimeter. Red lines represent anticyclonic eddies (AE), and blue lines are cyclonic eddies (CE).

features identified by the CNN-XGB model with those captured by the altimeter and obtained based on mathematical PDA methods, so as to verify the effectiveness of our model algorithm.

Vertical Structure of CNN-XGB Eddy

We first examined the vertical signals of the obtained CNN-XGB eddies. **Figure 5** reflects the global average vertical structures of Alt eddy (short dashed line), PDA eddy (long dashed line), and CNN-XGB eddy (solid line). By and large, the global mean CNN-XGB eddy profiles have very significant PDA signals, which are much stronger than those of the Alt eddy and are almost as strong as PDA eddy signals, implying that the eddy identification based on CNN-XGB model is effective and robust. We can find out in detail that for these profiles, the anomaly increases with depth from the sea surface, reaches its maximum at ∼90 m and decreases thereafter. Although they present a similar depth position of eddy core, the magnitude of eddy core intensity varies greatly. For Alt eddy, the magnitude of maximum anomaly is ∼−0.18 kg/m^3 for AE and only ∼0.08 kg/m^3 for CE, while a maximum anomaly inside CNN-XGB AE (CE) is ∼−0.33 kg/m^3 (∼0.20 kg/m^3) for centered at ∼72 m (∼93 m). The eddy core intensity of CNN-XGB eddy is almost equal to PDA eddy (∼−0.32 kg/m^3 for PDA AE and ∼0.21 kg/m^3 for PDA CE, respectively) and present a slightly shallower eddy core depth compared to PDA eddy, which may due to the fact that the CNN-XGB model can find more eddies in the equatorial region where eddy cores are shallower (see **Figure 7**). The anomaly intensity of CNN-XGB eddy is almost twice as strong as Alt eddy, while the eddy core depth is close to each other, proving that some of the Argo profiles that are not captured by the altimeter show typical eddy vertical structure signals, which can be successfully extracted by the CNN-XGB algorithm.

Since the vertical structure of eddies will have obvious inconsistencies with regional and latitudinal differences, we randomly selected four small areas in different latitude zones (white boxes in **Figure 6**) as well as the corresponding latitude zones (banded colors in **Figure 6**) to compare the vertical structures of the Alt eddy, PDA eddy, and CNN-XGB eddy to further verify the reliability of our results. Examining each panel, an overall impression is that CNN-XGB eddies show a stronger vertical signal than Alt eddies and a close signal with PDA eddies, which further proved that our eddy identification model is not only globally fitted but also regionally sensitive. The geographical distribution at the top of **Figure 6** shows the location of the selected grid points and the division of latitude zone. Areas A and B are within the latitude range of 20N–20S (L1), which is covered by the light blue area. In area A, the signal at the eddy core (∼120 m) of CNN-XGB eddy is much stronger than that of Alt eddy, displaying a maximum of ∼−0.76 kg/m^3 (∼−0.19 kg/m^3) of CNN-XGB (Alt) AE and ∼0.66 kg/m^3 (∼0.08 kg/m^3) of CNN-XGB (Alt) CE. Area B locates in the Arabian Sea where the characteristics of the eddy core are consistent with **Figure 7**

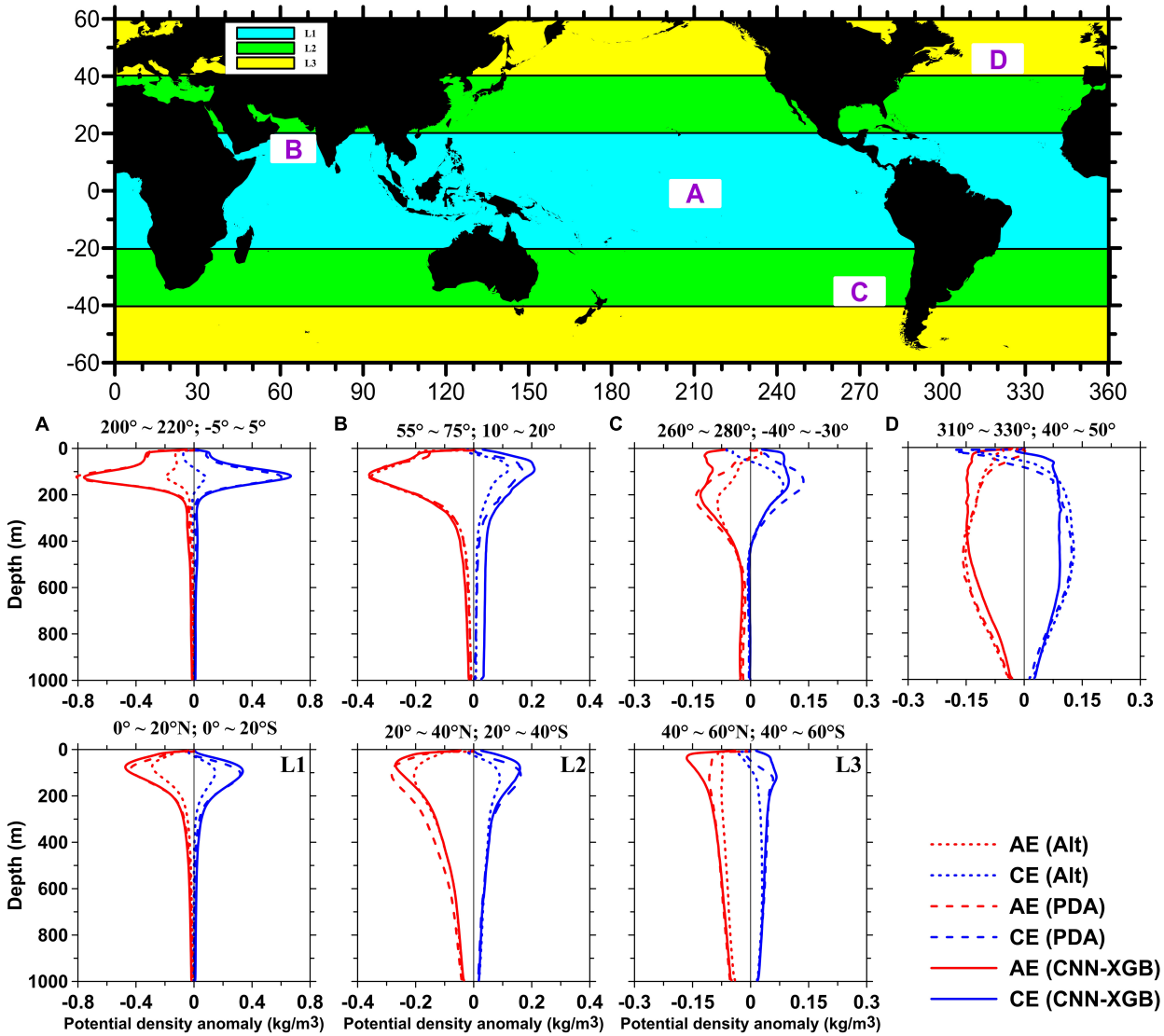


FIGURE 6 | Mean potential density anomaly profiles of four areas in different latitude zones and average profiles of the different latitude zones. The white boxes marked with capital letters in the geographic distribution map represent the location of the corresponding areas, and the bands of different colors represent the corresponding latitude zones. **(A,B)** Average profiles in areas A and B in latitude 1 (L1: 0–20N and 0–20S), **(C)** Average profiles in area C in latitude 2 (L2: 20–40N and 20–40S), and **(D)** Average profiles in area D in latitude 3 (L3: 40–60N and 40–60S). The short dashed red (blue) lines are the profiles of Alt anticyclonic eddies (AE) [cyclonic eddies (CE)], long dashed red (blue) lines are the profiles of potential density anomaly (PDA) AE (CE), and the solid red (blue) lines are the profiles of CNN-XGB AE (CE).

in de Marez et al. (2019), showing that eddy-induced ocean anomalies in this area are mainly confined in the upper 300 m. The shift increases rapidly for both AE and CE, and the AE anomalies are slightly greater than those of CE. Looking at the mid-latitude, ranging from 20N–40N to 20S–40S (L2), which is covered by the green area, the eddy core goes deeper, and a double core structure may appear. A similar conclusion can be found in area C, which can be compared with the eddy structure in Pegliasco et al. (2015). Results show that the maximum PDA inside composite Alt CE is 0.08 kg/m^3 at ∼170 m, while that inside the Alt AE is −0.09 kg/m^3 at ∼240 m. In the CNN-XGB eddy profiles, a stronger PDA appears; the maximum inside the AE is ∼−0.13 kg/m^3 and the maximum inside the CE is ∼0.1 kg/m^3. A similar pattern is also found in the mean PDA profile of the latitudinal zone (L1–L3). Along with the increase in latitude, the magnitude of maximum anomaly remarkably decreases, but the signal of CNN-XGB eddy is always stronger than that of Alt eddy.

The vertical structure can intuitively reflect the strength of the eddy signal of the two identification methods. We suppose that such a large proportion of profiles with strong vertical structure signals are not recognized by the altimeter and may mainly include three aspects. First, a considerable part of CNN-XGB eddies is found distributed at the equator where the altimeter

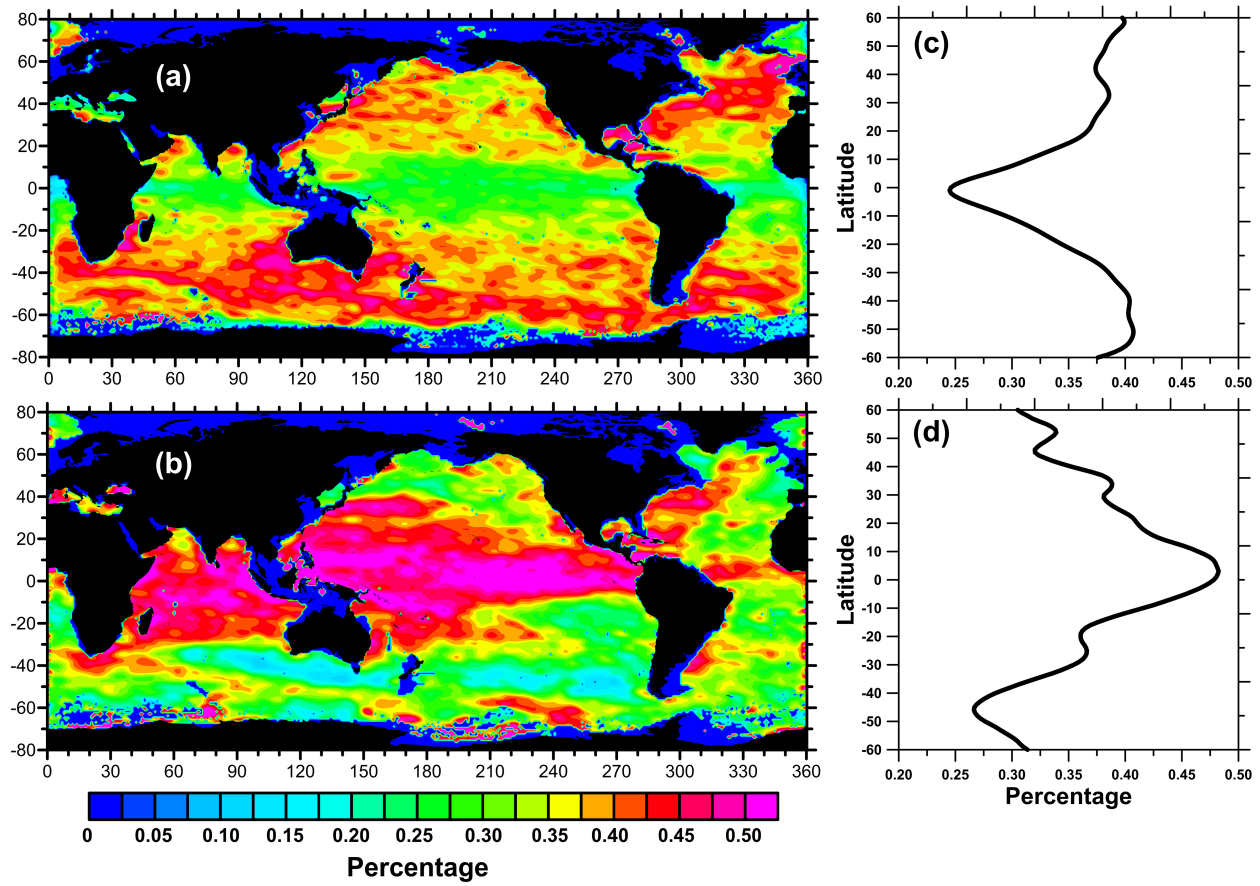


FIGURE 7 | Global geographic distribution of profiles inside Alt eddy and convolutional neural network with extreme gradient boosting (CNN-XGBoost) eddy during 2002 to 2019 under a 1° grid. **(a,b)** The frequencies of Argo profiles in Alt eddy and CNN-XGBoost eddy occurrence with respect to all Argo profiles within the same grid, respectively. **(c,d)** Corresponding zonal distributions of **Panels (a,b)**.

sampling interval is the largest (see **Figure 7**) so that even a part of eddies having notable sea surface characteristics is missing. Second, a part of the floats inside Alt eddy is relatively far away from the eddy center, which makes the vertical structure signal weaker because the vertical structure of the eddy weakens with the increase in the distance between the float and the eddy center. Third, even if the sea surface signal of the eddy is too small to be distinguished by the altimeter in higher latitude, it will have an evident vertical signal, and the existence of subsurface-intensified eddy that lacks strong surface signals cannot be eliminated.

Geographical Distribution of CNN-XGB Eddy

Another perspective to validate the effectiveness of eddy recognition is its global distribution characteristics. **Figure 7** shows the global geographic distribution (left column) and the corresponding zonal distribution (right column) of Alt eddies and CNN-XGB eddies. In order to eliminate the influence of Argo floats location on the geographical distribution, the percentage of floats per 1° grid is calculated and shown. The percentage refers to the profiles falling into eddies divided by the total number of profiles in the corresponding grid. Comparing **Figures 7a,b**, the complementarity between the Alt eddy and CNN-XGB eddy is remarkable, showing that almost 50% of the profiles are inside eddies but are missed by the altimeter in the tropical ocean. The equatorial low latitude area is also the area where altimeter sampling is relatively sparse, and more eddies may not be detected. With the increase in latitude, the percentage of the profiles in Alt eddy increases gradually, while that in CNN-XGB eddy decreases. Since the geographic distance corresponding to the 1° spatial distance on the earth decreases with the increase in latitude, the spatial resolution of an altimeter with the same orbital spacing near the equator is much lower than that of high-latitude regions. The space distance of 1° latitude on the earth drops from about 111 km at the equator to about 55 km at the poles; that is, the corresponding geographic scale in the equatorial region is larger, and the spatial resolution of the altimeter is lower than that in the middle and high latitudes, so more eddies are missed. Through the vertical structure of the eddy, a large number of eddies in the equatorial region have been identified where the spatial resolution of the gridded altimetric products is not enough to capture the small-scale eddies. In addition, the geographical distribution of profiles inside CNN-XGB eddy has a

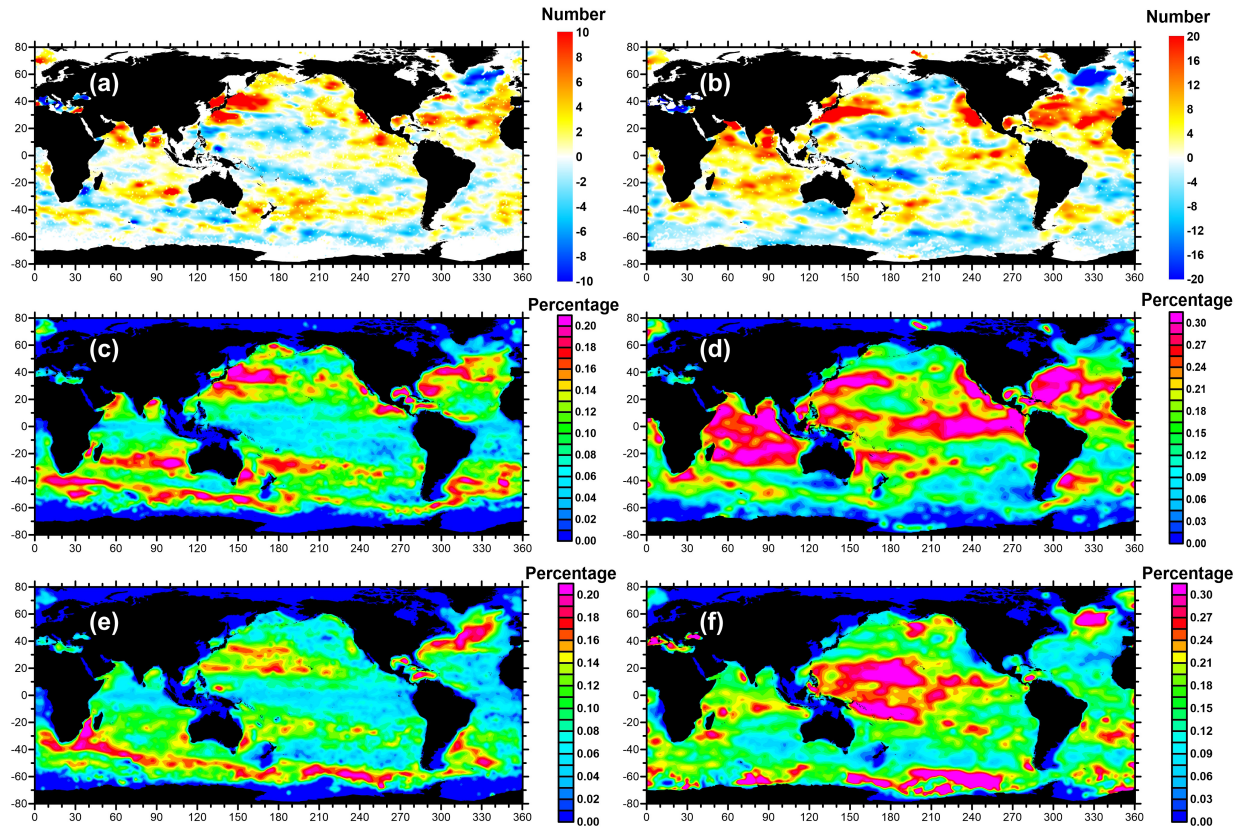


FIGURE 8 | Global geographic distribution of Argo profile density per 1° × 1° grid during 2002 to 2019. **(a)** Distribution of numbers of Argos in Alt anticyclonic eddies (AE) minus the numbers of Argos in Alt cyclonic eddies (CE) in the corresponding grid. **(b)** Same as **Panel (a)** but is the Argos in CNN-XGB eddies. **(c,d)** The frequencies of Argo profiles in Alt AE and CNN-XGB AE occurrence with respect to all Argo profiles within the same grid, respectively. (**e,f**) same as **Panels (c,d)** but for CEs.

consistency with that of the short-lived eddy proposed by Chen and Han (2019). It is also proved that eddy properties such as amplitude, vorticity, and kinetic energy are positively correlated with the eddy's lifetime, and this can further confirm that there are plenty of eddies with a smaller radius and lower energy in the tropical equatorial region that can hardly be captured through the altimeter.

Further, both Alt eddies and CNN-XGB eddies are classified into AEs and CEs, as shown in **Figure 8**. **Figures 8a,b** are the difference between the Alt (CNN-XGB) AE and Alt (CNN-XGB) CE. Both of these two figures show a strip distribution of alternating positive and negative values, which further imply that our model has no eddy polarity preference. A belt of maximum in the tropical ocean also appears for both CNN-XGB AE and CE where altimeter eddy identification is known to be ineffective (**Figures 8d,f**). **Figures 8c,d** show the percentage of Alt (CNN-XGB) AE. It is obvious that in addition to picking up a large proportion of eddies in low latitudes, more missing AEs are found in AE accumulation areas such as the Kuroshio area, eastern Australia, and Southwest Atlantic. CE is the same, seeing in the South Pacific and North Atlantic for example (**Figures 8e,f**).

Table 3 summarizes the global average profile rate picked up by Alt eddy and CNN-XGB eddy. For the Alt eddy, the percentage of Argo captured is about 34.60%. After the CNN-XGB algorithm, the percentage of Argo inside eddy is increased by 36.04%, and the remaining 29.36% of Argo is indeed outside eddies. We further confirm the percentage inside CNN-XGB AE and CE for 18.69 and 17.35%, respectively, increases by ∼1% compared with Alt eddy.

Relationship Between Eddy Property and Its Vertical Structure

At present, we have realized the recognition of the eddy missed by the altimeter through artificial intelligence and the vertical structure of the eddy but only the recognition of the eddy point. It is well-known that an eddy is not only a vertical profile point but also a body with a certain radius and can cause the amplitude

TABLE 3 | Comparison of the statistics of Argo floats identified by altimeter and CNN-XGB.

Identify rate of Argo floats	AE	CE	ALL
Alt eddies	17.68%	16.92%	34.60%
CNN-XGB eddies	18.69%	17.35%	36.04%
Outside eddies	–	–	29.36%

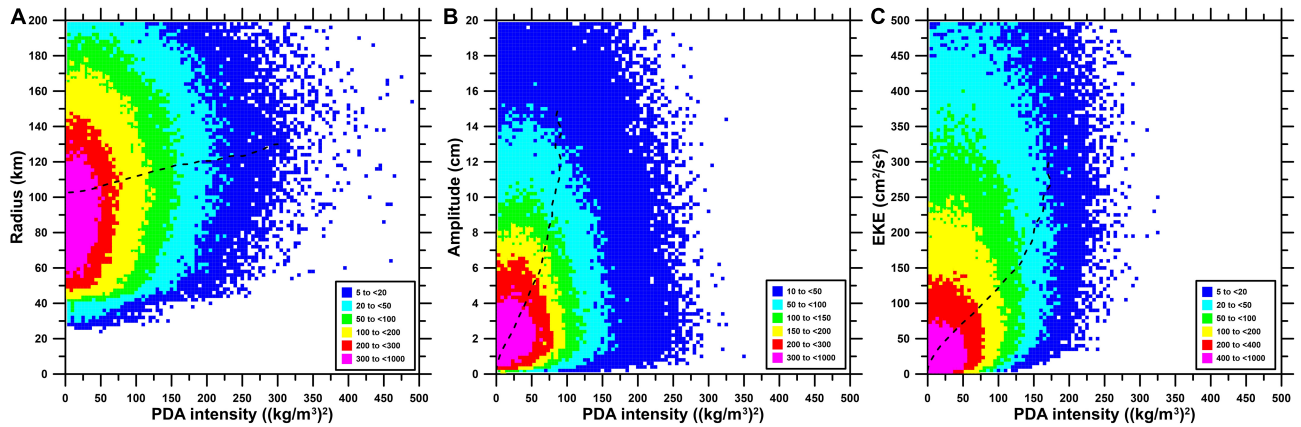


FIGURE 9 | Scatter diagrams of eddy property versus potential density anomaly intensity during 2002–2019. **(A)** Radius, **(B)** amplitude, and **(C)** eddy kinetic energy (EKE). The color bars denote eddy numbers. The black dashed curves depict the most likelihood of eddy properties with respect to potential density anomaly intensity.

of the sea surface with a certain kinetic energy. Thus, the next step is to predict the surface characteristics of the eddy based on deep learning regression analysis, such as the radius, amplitude, and kinetic energy. Scatter diagrams of eddy radius, amplitude, and eddy kinetic energy (EKE) as a function of PDA intensity are illustrated in **Figure 9**. Basically, the three properties all have the most likelihood line in terms of data density, which displays a slowly increasing trend with PDA intensity: from ∼100 to ∼130 km for radius (**Figure 9A**), from ∼1 to ∼15 cm for amplitude (**Figure 9B**), and from ∼0 to ∼300 cm^2/s^2 for EKE (**Figure 9C**). This result proves that there is an intrinsic correlation between the eddy surface properties and its vertical structure. It is feasible and reliable to invert the eddy surface parameters through the vertical structure characteristics.

CONCLUSION

In this paper, a new hybrid model that combines the advantages of CNN and XGBoost is proposed to achieve oceanic eddy detection from 18-year Argo profiles. The major conclusions of the study can be summarized as follows.

First, the proposed CNN-XGBoost model has promising performance in extracting eddy vertical profile features. The CNN model has a design of six convolution layers, and a doubleconvpool structure is applied to improve the accuracy and efficiency of feature extraction. Moreover, the method can get the highest convergence when the learning rate is 5×10^{-4}. Furthermore, the 64-dimensional feature vectors learned from CNN are inputted into the XGBoost model combined with the profile position, date, and profile features. After 100 iterations, the final model is obtained, with a classification accuracy of 98%. Compared with CNN or XGBoost, the accuracy is improved by 4 and 6%, respectively. We further measured the sensitivity of the proposed model by calculating the erroneous judgment of AE and CE. Results show that compared with AE, CE is more likely to be misjudged (1.41 versus 0.06%) since the abnormal signal caused by CE is smaller than that of AE. Nevertheless, the CNN-XGBoost model presents the lowest error percentage among the three models with the highest misclassification rate of <2% (see **Table 2** for details). The results provide an insight that the proposed deep learning method can be used as an effective methodology for eddy identification.

Second, vertical profiles of CNN-XGB eddies show a remarkable consistency with that of the Alt eddies. After carrying out the global average, latitudinal average, and grid average, we found that these eddies' mean vertical shapes are homologous, and the vertical anomaly signal of the CNN-XGB eddy is similar or even stronger than that of the Alt eddy. For the global average, the magnitude of the maximum anomaly of Alt eddy is ∼−0.18 kg/m^3 for AE and only ∼0.08 kg/m^3 for CE, while the anomaly intensity of the CNN-XGB eddy is almost twice as strong as that of the Alt eddy, indicating ∼−0.33 kg/m^3 for AE and ∼0.20 kg/m^3 for CE. The reason why such a large proportion of profiles with strong vertical structure signals are not recognized by the altimeter may due to the larger sampling interval of the altimeter at the equator, the farther distance between the float and eddy center, and the existence of subsurface eddies.

Third, among Argo profiles from 2002 to 2019, ∼34.6% of the profiles are captured by Alt eddies, and about 36% of the profiles (∼19% for AE and ∼17% for CE) are actually inside eddies but missed by the altimeter and captured through our eddy identification method; the other ∼29% profiles are indeed outside eddies. The geographical distribution of the profiles inside CNN-XGB eddies is complementary to that inside Alt eddies. A prominent eddy belt with more than 50% of the profiles inside the CNN-XGB eddies is found in the tropical ocean. There is also the area abundant in short-lived eddy, which further proved that weak and small eddy may be lost with a greater probability. Consequently, the vertical profile datasets can be extended to improve the identification ability of the altimeter.

The present research shows that there is a positive correlation between eddy properties (e.g., radius, amplitude, kinetic energy)

and the abnormal strength of eddy's vertical structure. Therefore, based on the deep learning method, the inversion of eddy properties from its vertical structure should be further studied, so as to establish a more complete Argo-eddy identification dataset. In the spirit of reproducibility, the Python code will be available at https://github.com/, and we will also share the training and testing data used for this work to encourage competing methods and, especially, other deep learning architectures.

DATA AVAILABILITY STATEMENT

The data used to support the findings of this study are available as follows. The altimeter all-sat merged sea level anomaly datasets generated for eddy identification can be found in the Archiving, Validation, and Interpretation of Satellite Oceanographic (AVISO, https://www.aviso.altimetry.fr/index.php?id=3122). The eddy identification dataset can be obtained from http://coadc.out.edu.cn/tfl/ and http://data.casearth.cn/ (Data ID: XDA19090202). Argo profile data can be obtained from ftp://ftp.ifremer.fr/ifremer/argo/.

AUTHOR CONTRIBUTIONS

XC performed methodology, software, validation, data curation, visualization, and writing of the original draft. GC contributed to conceptualization, investigation, writing review and editing, and supervision. LG performed the model framework construction. BH performed validation and investigation. CC performed validation and review. All authors contributed to the article and approved the submitted version.

FUNDING

This research was jointly supported by the National Natural Science Foundation of China (No. 42030406), Marine Science & Technology Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (No. 2018SDKJ0102), and ESA-NRSCC Scientific Cooperation Project on Earth Observation Science and Applications: Dragon 5 (No. 58393).

REFERENCES

Amores, A., Jordà, G., Arsouze, T., and Le Sommer, J. (2018). Up to what extent can we characterize ocean eddies using present-day gridded altimetric products? *J. Geophys. Res.* 123, 7220–7236. doi: 10.1029/2018JC014140

Andrade, I., Hormazabal, S., and Combes, V. (2014). Intrathermocline eddies at the Juan Fernandez Archipelago, southeastern Pacific Ocean. *Lat. Am. J. Aquat. Res.* 42, 888–906. doi: 10.3856/vol42-issue4-fulltext-14

Bryden, H. L., and Brady, E. C. (1989). Eddy momentum and heat fluxes and their effect on the circulation of the equatorial Pacific Ocean. *J. Mar. Res.* 47, 55–79. doi: 10.1357/002224089785076389

Chaigneau, A., Texier, M. L., Eldin, G., Grados, C., and Pizarro, O. (2011). Vertical structure of mesoscale eddies in the eastern South Pacific Ocean: a composite analysis from altimetry and Argo profiling floats. *J. Geophys. Res.* 116:C11025. doi: 10.1029/2011JC007134

Chelton, D. B., Gaube, P., Schlax, M. G., Early, J. J., and Samelson, R. M. (2011a). The influence of nonlinear mesoscale eddies on near-surface oceanic chlorophyll. *Science* 334, 328–332. doi: 10.1126/science.1208897

Chelton, D. B., Schlax, M. G., and Samelson, R. M. (2011b). Global observations of nonlinear mesoscale eddies. *Prog. Oceanogr.* 91, 167–216. doi: 10.1016/j.pocean.2011.01.002

Chen, G., Chen, X., and Huang, B. (2020). Independent eddy identification with profiling Argo as calibrated by altimetry. *J. Geophys. Res.* 125:e2020JC016729. doi: 10.1029/2020JC016729

Chen, G., and Han, G. (2019). Contrasting short-lived with long-lived mesoscale eddies in the global ocean. *J. Geophys. Res.* 124, 3149–3167.

Chen, T., and Guestrin, C. (2016). "XGBoost: a scalable tree boosting system," in *Proceedings of the Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*, San Francisco, CA, 13–17.

de Marez, C., L'Hégaret, P., Morvan, M., and Carton, X. (2019). On the 3D structure of eddies in the Arabian Sea. *Deep Sea Res. Part I.* 150:103057. doi: 10.1016/j.dsr.2019.06.003

Dong, C., Lin, X., Liu, Y., Nencioli, F., Chao, Y., Guan, Y., et al. (2012). Three-dimensional oceanic eddy analysis in the Southern California Bight from a numerical product. *J. Geophys. Res. Oceans* 117:46218. doi: 10.1029/2011JC007354

Du, Y., Song, W., He, Q., Huang, D., Liotta, A., and Su, C. (2018). Deep learning with multi-scale feature fusion in remote sensing for automatic oceanic eddy detection. *Inf. Fusion* 49, 89–99. doi: 10.1016/j.inffus.2018.09.006

Dunn, J. R., and Ridgway, K. R. (2002). Mapping ocean properties in regions of complex topography. *Deep Sea Res. Part I* 49, 591–604. doi: 10.1016/S0967-0637(01)00069-3

D'Alimonte, D. (2009). Detection of mesoscale eddy-related structures through iso-sst patterns. *IEEE Geosci. Remote Sens. Lett.* 6, 189–193. doi: 10.1109/lgrs.2008.2009550

Faghmous, J. H., Frenger, I., Yao, Y., Warmka, R., Lindell, A., and Kumar, V. (2015). A daily global mesoscale ocean eddy dataset from satellite altimetry. *Sci. Data* 2:150028. doi: 10.1038/sdata.2015.28

Faghmous, J. H., Styles, L., Mithal, V., Boriah, S., Liess, S., Kumar, V., et al. (2012). "Eddyscan: a physically consistent ocean eddy monitoring application," in *Proceedings: 2012 Conference on Intelligent Data Understanding; CIDU 2012*, eds K. Das, N. V. Chawla, and A. N. Srivastava (Piscataway, NJ: IEEE), 96–103. doi: 10.1109/cidu.2012.6382189

Fu, L. L., Chelton, D. B., LeTraon, P. Y., and Morrow, R. (2010). Eddy dynamics from satellite altimetry. *Oceanography* 23, 14–25. doi: 10.5670/oceanog.2010.02

Gonzalez-Silvera, A., Santamaria-del-Angel, E., Millán-Nuñez, R., and Manzo-Monroy, H. (2004). Satellite observations of mesoscale eddies in the Gulfs of Tehuantepec and Papagayo (Eastern Tropical Pacific). *Deep Sea Res. Part II* 51, 587–600. doi: 10.1016/J.DSR2.2004.05.019

Gordon, A. L., Shroyer, E., and Murty, V. (2017). An intrathermocline eddy and a tropical cyclone in the Bay of Bengal. *Sci. Rep.* 7:46218.

Jeronimo, G., and Gomez-Valdes, J. (2007). A subsurface warm-eddy off northern Baja California in July 2004. *Geophys. Res. Lett.* 34:L06610. doi: 10.1029/2006GL028851

Le Vu, B., Stegner, A., and Arsouze, T. (2018). Angular momentum eddy detection and tracking algorithm (ameda) and its application to coastal eddy formation. *J. Atmos. Ocean. Technol.* 35, 739–762. doi: 10.1175/JTECH-D-17-0010.1

Liu, Y., Chen, G., Sun, M., Liu, S., and Tian, F. (2016). A parallel SLA-based algorithm for global mesoscale eddy identification. *J. Atmos. Ocean. Tech.* 33, 2743–2754. doi: 10.1175/jtech-d-16-0033.1

McDougall, T. J., Feistel, R., Millero, F. J., Jackett, D. R., Wright, D. G., Kin, B. A., et al (2009). *The International Thermodynamic Equation of Seawater 2010 (TEOS-10): Calculation and Use of Thermodynamic Properties.* Global Ship-based Repeat Hydrography Manual, IOCCP Report No, 14. París: UNESCO.

Pegliasco, C., Chaigneau, A., and Morrow, R. (2015). Main eddy vertical structures observed in the four major Eastern Boundary Upwelling Systems. *J. Geophys. Res. Oceans* 120, 6008–6033. doi: 10.1002/2015JC010950

Ren, X., Guo, H., Li, S., Wang, S., and Li, J. (2017). "A novel image classification method with CNN-XGBoost model," in *Digital Forensics and Watermarking. IWDW 2017*. Lecture Notes in Computer Science, Vol. 10431, eds C. Kraetzer, Y. Q. Shi, J. Dittmann, and H. Kim (Cham: Springer).

Ridgway, K. R., Dunn, J. R., and Wilkin, J. L. (2002). Ocean interpolation by four-dimensional weighted least squares: application to the waters around Australasia. *J. Atmos. Ocean. Technol.* 19, 1357–1375. doi: 10.1175/1520-0426(2002)019<1357:oibfdw>2.0.co;2

Roemmich, D., Johnson, G., Riser, S., Davis, R., Gilson, J., Owens, W., et al. (2009). The Argo program observing the global ocean with profiling floats. *Oceanography* 22, 34–43. doi: 10.5670/oceanog.2009.36

Sun, M., Tian, F., Liu, Y., and Chen, G. (2017). An improved automatic algorithm for global eddy tracking using satellite altimeter data. *Remote Sens.* 9:206. doi: 10.3390/rs9030206

Tian, F., Wu, D., Yuan, L., and Chen, G. (2020). Impacts of the efficiencies of identification and tracking algorithms on the statistical properties of global mesoscale eddies using merged altimeter. *Int. J. Remote Sens.* 41, 2835–2860. doi: 10.1080/01431161.2019.1694724

Zhang, Z., Wang, W., and Qiu, B. (2014). Oceanic mass transport by mesoscale eddies. *Science* 345, 322–324. doi: 10.1126/science.1252418

Zhang, L., Zhang, L., and Du, B. (2016). Deep learning for remote sensing data: a technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* 4, 22–40. doi: 10.1109/MGRS.2016.2540798

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 07 June 2021
doi: 10.3389/fmars.2021.665716

The Identification and Prediction in Abundance Variation of Atlantic Cod *via* Long Short-Term Memory With Periodicity, Time–Frequency Co-movement, and Lead-Lag Effect Across Sea Surface Temperature, Sea Surface Salinity, Catches, and Prey Biomass From 1919 to 2016

Rui Nian[1,2], Qiang Yuan[1], Hui He[1], Xue Geng[1], Chi-Wei Su[3], Bo He[1] and Amaury Lendasse[4]*

[1] College of Information Science and Engineering, Ocean University of China, Qingdao, China, [2] Harvard Medical School, Harvard University, Cambridge, MA, United States, [3] School of Economics, Qingdao University, Qingdao, China, [4] Department of Information and Logistics Technology, University of Houston, Houston, TX, United States

OPEN ACCESS

Edited by:
Andrei Herdean, University of Technology Sydney, Australia

Reviewed by:
Xinjun Chen, Shanghai Ocean University, China
Tsangyao Chang, Feng Chia University, Taiwan
Xin Li, Shanghai Jiao Tong University, China

***Correspondence:**
Rui Nian
nianrui_80@163.com

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: 08 February 2021
Accepted: 05 May 2021
Published: 07 June 2021

Citation:
Nian R, Yuan Q, He H, Geng X, Su C-W, He B and Lendasse A (2021) The Identification and Prediction in Abundance Variation of Atlantic Cod via Long Short-Term Memory With Periodicity, Time–Frequency Co-movement, and Lead-Lag Effect Across Sea Surface Temperature, Sea Surface Salinity, Catches, and Prey Biomass From 1919 to 2016. Front. Mar. Sci. 8:665716. doi: 10.3389/fmars.2021.665716

The population of Atlantic cod significantly contributes to the prosperity of fishery production in the world. In this paper, we quantitatively investigate the global abundance variation in Atlantic cod from 1919 to 2016, in favor of spatiotemporal interactions over manifold impact factors at local observation sites, and propose to explore the predictive mechanism with the help of its periodicity, time–frequency co-movement, and lead-lag effects, *via* long short-term memory (LSTM). We first integrate evidences yielded from wavelet coefficients, to suggest that the abundance variation potentially follows a 36-year major cycle and 24-year secondary cycle at the time scales of 55 years and 37 years. We further evaluate the responses of Atlantic cod abundance to the external impact factors, including sea surface temperature (SST), catches, prey biomass, and sea surface salinity (SSS), in aid of the wavelet coherence and phase difference, which allows us to identify the dominantly correlative factors and capture the leading roles along the time domain and then divide the responses around the recent 60 years into three stages: before 1985, 1985–1995, and after 1995. At the first stage, the reason for the decline in abundance could be mainly attributed to the rapid rise of fish catches. At the second stage, the impact of SST and SSS also provides significant indices, besides overfishing; meanwhile, the mortality of primary producers and forced migration of fish species indirectly cause the decline. At the third stage, warming SST and growing SSS directly led to the decrease of abundance. Finally, we establish one ensemble of LSTM-SAE architecture to comprehensively reflect the predictive patterns at each stage. It has been demonstrated from experimental results that the models behaved better

when intentionally feeding with the dominantly correlative multivariate inputs, instead of either all factors or only the abundance. The proposed scheme provides opportunities to symmetrically identify the underlying predictive attributes of Atlantic cod abundance and potentially perform as the quantitative references in reasonably making fishing decision. With the rapid development in deep learning capabilities, it is hopeful to expect better predictions of the responses to global changes, not only for Atlantic cod but also for other fish species and the ecosystem as a whole.

Keywords: wavelet analysis, Atlantic cod abundance, sea surface temperature, Atlantic cod catches, prey biomass, sea surface salinity

INTRODUCTION

Atlantic cod (*Gadus morhua*), one of the most consumed edible fish species with great economic values, is fecund with reproductive potentials, but has long been overfished and highly associated with marine environment changes, causing dramatic declines for decades. The symmetrical investigation on the known abundance variability in Atlantic cod could integrate evidences on its spatiotemporal responses to manifold intrinsic and external indices and help provide opportunities to explore the underlying patterns and mechanism of Atlantic cod abundance to be predicted.

There might exist quite a few possible reasons behind the abundance variation of Atlantic cod, in combination with the life history characteristics (Jordaan, 2002; Drinkwater, 2005; Kristiansen et al., 2011; Árnason et al., 2013). First of all, from the perspective of marine ecology, the adaptability of Atlantic cod in both the early stage of life and adulthood depends much on the food supply and seawater temperature (Rose and Leggett, 1989; Miller et al., 1995). Sea surface temperature (SST) behaves as one of the most essential and fundamental indices (Fullard et al., 2000; O'Brien et al., 2000; Pörtner et al., 2001), indirectly delivering influences to the physical and biological systems of the deeper ocean for the benthic Atlantic cod habitat (Parmesan and Yohe, 2003; Rao and Sivakumar, 2003; Rosenzweig et al., 2008), altering both benthic and pelagic ecosystems from phytoplankton to zooplankton to higher trophic levels (Meyer et al., 1999; Richardson and Schoeman, 2004; Wanless et al., 2005; Behrenfeld et al., 2006). SST dictates movement and habitat connectivity of Atlantic cod (Staveley et al., 2019). Warming sea temperature causes Atlantic cod to move out of the habitat, either entering deeper waters or migrating to more northerly waters. The higher sea temperature also brings huge varieties in the number of primary producers, as well as their seasonal growth changes, which will definitely disrupt the normal food supply and nutrition that feed Atlantic cod (Schwalme and Chouinard, 1999; Drinkwater et al., 2000). Atlantic cod's diet ranges from phytoplankton and small zooplankton as the larvae, to crustaceans and shrimps when they gradually grow up after the larval stage, to capelin, sand lance, herring, red fish, etc., and even other cod (Scott and Scott, 1988; Bogstad et al., 1994; Sundby, 2000), where the structure or abundance modification of the primary nutrition may propagate through higher trophic levels at the basis (Kirby and Beaugrand, 2009).

Meanwhile, ocean acidification driven by rising global CO_2 concentrations increases the sensitivity of Atlantic cod embryos to extreme temperatures, which results in tissue damage to cod larvae and changes in their behavior (Dahlke et al., 2017; Sswat, 2017); the decrease of seawater salinity inhibits the activity of sodium and potassium ions and decreases plasma osmotic pressure, influencing the feeding conditions of Atlantic cod, resulting in decreased recruitment (Lambert et al., 1994; Conover et al., 1995; Beaugrand et al., 2011). In addition, ocean currents, with their vertical or horizontal movement of both surface and deep water throughout the ocean, are also considered to contribute to the propagation and redistribution of marine thermocline and marine nutrients during the growth of Atlantic cod (Wroblewski et al., 2000; Hays, 2017). It is also argued that the rapid development of fishing technology from the 1980s has played a major role in the collapse of the Atlantic cod abundance in the North Atlantic (Baird et al., 1991; Hutchings and Myers, 1994; Hutchings, 1996). Although a series of restricted fishing and even fishing bans have been issued since 1992, the scale of local Atlantic cod like Canada has not shown significant signs of rebound (Myers et al., 1997). The variation of Atlantic cod abundance is undoubtedly not determined by a single impact factor but attributed to the outcome systemically and integrally generated from the series of intrinsic and extrinsic relevant indices that directly or indirectly interact with the abundance from the perspective of environment, physiology, migration, reproduction, and so on.

The relationship between Atlantic cod abundance and manifold impact factors is always not a simple causal function. It has been noted that most commonly applied approaches, such as bootstrapping regression models (Freedman, 1981), rolling window regression (Oehlert and Swart, 2019), and auto-regressive distributed lag (ADL) model (Haseeb et al., 2019), often only consider one single or several independent predictive elements and tend to ignore the combined effects of the complex marine climates and ecosystem structures, which is too idealized, while the classic correlation calculation is only suitable for studying relatively short-term time series and severely limited to stationary series. Therefore, on the basis of the historic spatiotemporal responses to manifold intrinsic and external indices, how to establish a comprehensive predictive framework that potentially explores underlying mechanism for its variation of Atlantic cod abundance is actually a pioneering topic. Firstly, the abundance variation in Atlantic cod may

inherently follow its own specific patterns that probably present elementary periodic tendencies or illustrate some time–frequency characteristics (Berberidis et al., 2002). Secondly, no matter the time series of Atlantic cod abundance itself or the extrinsic correlative indices in study, most of them are non-stationary for a quite long period, where the possible mutations may occur in any individual stage, so that there may be time–frequency co-movements at a specific moment or relatively short period (Vacha and Barunik, 2012). Finally, there exist lead-lag effects diversified between Atlantic cod abundance and manifold impact factors, indicating a variety of modifications in the leading roles that will dominate co-movements in every certain time period, at a certain influencing direction of either positive or negative correlation (Dajcman, 2013).

To address the potential periodicity, time–frequency co-movement, and lead-lag effects of non-stationary long-term time series, wavelet analysis is a mathematically basic tool to discover how to approximate time series to a certain frequency band within a certain time period simultaneously, which could optimally describe the occurrence of transient events, and well adapt to conditions where the amplitude of the response varies significantly in non-stationary time series (Steel and Lange, 2007; Cloern and Jassby, 2010; Rhif et al., 2019). As the extended usage of wavelet transform, wavelet coherence (Maraun and Kurths, 2004; Su et al., 2018, 2019) and phase difference (Aguiar-Conraria and Soares, 2011) can be utilized to recognize whether two time series are quantitatively linked by a certain correlation, even causality relationship. Kogovšek et al. (2010) have applied continuous wavelet transform to investigate the long-term fluctuation of time series of jellyfish abundance in the past 200 years and further confirmed a major and minor period of its proliferation. Carey et al. (2016) adopted continuous wavelet transform to identify the periodic changes of phytoplankton at multiple time scales and determined the dominant scale of its periodic changes. Ménard et al. (2007) has quantified the non-stationary association between the Indian Ocean climate and tuna populations through wavelet analysis and further explored the dependency between them. Li et al. (2015) have adopted wavelet analysis to explore the multi-temporal scale characteristics and interrelationships between algal biomass and selected aquatic environmental parameters and tried to explain the complex non-linear eutrophication dynamics. All of the above wide range of applications reflect to varying degrees the efficacy of wavelet analysis in non-stationary long-term multivariate time series at multiple time scales.

In addition, in order to comprehensively investigate the mutual influences between manifold intrinsic and external indices, we tried to develop multivariate prediction model for the variation tendency of Atlantic cod abundance in a longer-term move. Traditional time series prediction methods, like hidden Markov model (HMM) (Rabiner and Juang, 1986), Auto-regressive integrated moving average (ARIMA) (Zhang, 2003), moving average model (MA) (Box and Pierce, 1970), etc., often have a complicated modeling process, and when it comes to multi-step forward prediction problem, most of them do not have ideally consistent performances in the long run. For decades, machine learning (ML) has become one of the most powerful tools in the field of multivariate multi-step time series prediction (Joy and Death, 2004; Yáñez et al., 2010; Miller et al., 2019), while deep learning can be regarded as one of the hottest topics developed recently in the context, which drives predominant forces in modeling, functioning, forecasting the structural, long-term, dynamic properties in time series, and all kinds of most emerging and advanced algorithms have been put forward and made progresses (Hinton and Salakhutdinov, 2006; Krizhevsky et al., 2012; Huang et al., 2017).

Among them, the deep architecture of the recurrent neural network (RNN), which could model the sequences at the arbitrary length by applying a transition function to all its hidden layer states in a recursive manner, and capture all information stored in sequence in the previous element (Jagannatha and Yu, 2016; Fan et al., 2017), has been proven to be superior in prediction. Long short-term memory (LSTM) has been revolutionarily designed by adjusting the structure of the hidden neurons in RNN, based on a series of memory cells recurrently connected through layers to capture and retain the long-term dependencies, not only suppressing the disappearing gradient problem, but also enhancing the capability in predicting multi-step ahead into the future (Duan et al., 2016; Zhang et al., 2017; Park et al., 2018).

Despite the fact that LSTM models have outstanding advantages in prediction accuracy, its performance remains unsatisfactory, particularly when attempting to process highly non-linear and long-interval time series. There are hereby quite a lot of variant LSTM algorithms newly developed, including bidirectional LSTM (Bi-LSTM) (Habernal and Gurevych, 2016), gated recurrent unit (GRU) (Dey and Salem, 2017), LSTM-based stacked autoencoder (LSTM-SAE) (Sagheer and Kotb, 2019), etc. Among them, LSTM-SAE improves the random initialization strategies in an unsupervised learning fashion and circumvents the expression limitations of the conventional shallow LSTM model, which could hereby, to some extent, alleviate the convergence to local minimum and perform better in multi-step prediction.

In this article, in order to explore the potential response mechanism of Atlantic cod abundance changes, we make an attempt to establish a comprehensive prediction model of Atlantic cod abundance, with the help of its periodicity, time–frequency co-movement, and lead-lag effects, which can quantify spatiotemporal interactions with manifold intrinsic and external indices. We systematically retrieve the annual abundance of Atlantic cod, the catches, the biomass of primary producers, SST, and sea surface salinity (SSS), respectively, from BioTIME, FAOSTAT, ERSST.v5, and EN4, and select local observation sites with relatively complete abundance records for our study, i.e., ID 119 near the Gulf of Maine and ID 428 on the southeastern coast of Norway, to accordingly attain standardized time series at specific geographical coordinates, by spatial and time averaging. We employ the modulus and variance of wavelet coefficients to reveal the potential periodicity of global Atlantic cod abundance, to generate the wavelet coherence map by the ratio between the wavelet cross-spectrum and power spectrum, to evaluate the responses of the abundance variability toward the external impact factors, and to produce the wavelet phase difference by

the ratio between the imaginary and the real part of the wavelet cross-spectrum, to reflect leading roles that dominate time–frequency co-movements in every certain time period. This will provide a complete framework for the quantitative correlation and even causality among multivariate time series at multiple stages. Finally, in view of the historical time–frequency responses between Atlantic cod abundance and manifold impact factors, we attempt to determine the comprehensive prediction patterns of Atlantic cod abundance, by establishing the ensemble of LSTM-SAE architecture at each stage, and respectively, feed into three types of input vectors, namely, Atlantic cod abundance itself, all the external impact factors, and the dominantly correlative impact factors, to help with the estimation.

The remainder of the paper is organized as follows: Section "Materials and Methods" describes the access of multivariate time series referred, and outlines the basics in periodicity, wavelet coherence, phase difference, as well as the LSTM-SAE architecture in our study. Section "Results" quantitatively and systematically investigates the periodicity, time-frequency co-movement, and lead-lag effects in abundance variation of Atlantic cod, regarding the external impact factors, SST, catches, prey biomass, SSS, and exhibits the experimental results. Section "Discussion" suggests potentially mutual relationships along the time domain, together with the known events and its consequences, divides the responses of Atlantic cod abundance into stages, and establishes the ensemble of LSTM-SAE architecture for multivariate prediction. Finally, the conclusions are drawn in Section "Conclusion".

TABLE 1 | Annual average global abundance of Atlantic cod.

Year	Abundance	Year	Abundance	Year	Abundance	Year	Abundance
1919	17	1944	19	1969	13	1994	25
1920	16	1945	37	1970	23	1995	23
1921	27	1946	7	1971	25	1996	30
1922	40	1947	6	1972	25	1997	17
1923	28	1948	11	1973	27	1998	23
1924	37	1949	6	1974	25	1999	9
1925	11	1950	10	1975	17	2000	8
1926	7	1951	6	1976	21	2001	7
1927	18	1952	15	1977	19	2002	9
1928	33	1953	20	1978	24	2003	7
1929	22	1954	15	1979	28	2004	4
1930	6	1955	24	1980	27	2005	4
1931	9	1956	16	1981	29	2006	9
1932	8	1957	42	1982	28	2007	12
1933	6	1958	14	1983	42	2008	10
1934	6	1959	17	1984	25	2009	3
1935	15	1960	18	1985	44	2010	9
1936	5	1961	25	1986	50	2011	27
1937	5	1962	9	1987	51	2012	20
1938	77	1963	14	1988	32	2013	5
1939	9	1964	38	1989	39	2014	5
1940	19	1965	19	1990	35	2015	4
1941	44	1966	44	1991	35	2016	5
1942	14	1967	12	1992	26		
1943	49	1968	9	1993	30		

MATERIALS AND METHODS

Data

Atlantic Cod Abundance

The abundance of Atlantic cod has been retrieved *via* BioTIME database, which is composed of more than 12 million records of the raw data on species identities and abundances in ecological assemblages of more than 20 biomes at 361 observation sites from 1858 to 2018, where 5,640,610 records collected from six communities have been applied here from fish to marine invertebrates, consisting of the most comprehensive database for marine species abundance and richness statistics. Atlantic cod mainly inhabit most waters overlying the continental shelves of the Northwest and the Northeast Atlantic Ocean, such as the Gulf of St. Lawrence (Lambert and Dutil, 1997) and the Gulf of Maine (Ames, 2004). In our study, the global abundance of Atlantic cod records at all sites all over the world has first been selected from 1919 to 2016, and then we screened out records below 10 years or the singular values, such as the duplicates, zero-abundance counting, and non-organismal records, and deleted them prior to statistical analysis. **Table 1** is the annual average global abundance of Atlantic cod from 1919 to 2016, calculated by merging the local abundance at the remaining sites. Furthermore, two specific observation sites 428 (latitude: 58.958, longitude: 9.768) and 119 (latitude: 41.987425, longitude: −66.669701), respectively, located around the colder sea area of Gulf of Maine and North Sea, have been particularly taken for the local study. Considering the possible singular measurement values of Atlantic cod abundance at each site, we first eliminate them by defining the superior and inferior bounds *via* mathematical expectation calculation and then collect the annual records of Atlantic cod abundance at each site by time averaging. **Figure 1** shows the annual average global distribution of Atlantic cod, with sites 119 and 428 specifically marked by stars in color red and in color blue. The midpoint of the geographic grid has been used to merge the abundance records associated to it averagely, with yellow dot representing the spatial coordinates and the spatial distance less than 2° in latitude and longitude, and the radius refers to the corresponding local abundance in a certain proportion. **Tables 2**, **3** list the annual abundance of Atlantic cod at site 119 from 1970 to 2010 and at site 428 from 1952 to 2011.

Sea Surface Temperature

Sea surface temperature can be considered as one of the most essential climate indicators of abundance variation in Atlantic cod. The global monthly SST available from the SST version 5 (ERSST.v5) dataset of the NOAA Climate Center is produced by a monthly extended reconstruction of SST observations from the Comprehensive Ocean-Atmosphere Data Set (COADS) and can be tracked back to the mid-19th century. Since SST adopts smooth local and short-term changes, with many options in quality control, deviation adjustment, and interpolation having been substantially revised, it is suitable for long-term global studies. In order to focus on the local annual SST at both site

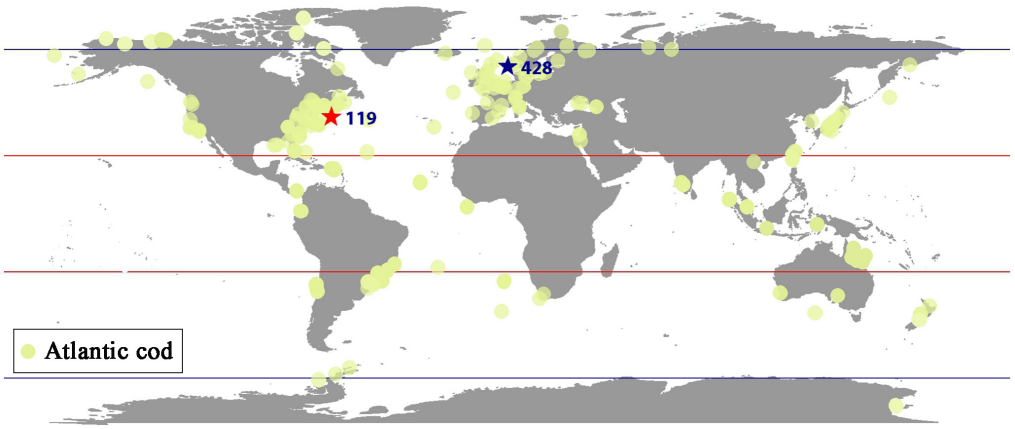


FIGURE 1 | Global distribution of Atlantic cod.

119 and site 428, we made use of the geographic midpoint and selected the closer temperature points associated with it, with the spatial distance less than 2° in latitude and longitude. The Atlantic cod abundance at each site is only recorded in certain months of the year, i.e., February, March, July, and October at site 119, and September and October at site 428. Therefore, we finally retrieved the corresponding annual SST through the spatial averaging of the relevant nearby points and the time averaging of a specific month, for site 119 from 1970 to 2010 and site 428 from 1952 to 2011.

TABLE 2 | Annual average abundance of Atlantic cod at site 119.

Year	Abundance	Year	Abundance	Year	Abundance	Year	Abundance
1970	22	1981	38	1992	14	2003	7
1971	19	1982	31	1993	27	2004	8
1972	16	1983	55	1994	21	2005	14
1973	20	1984	25	1995	24	2006	17
1974	16	1985	32	1996	21	2007	12
1975	20	1986	33	1997	15	2008	9
1976	17	1987	18	1998	12	2009	13
1977	23	1988	27	1999	13	2010	33
1978	26	1989	27	2000	12		
1979	21	1990	30	2001	14		
1980	30	1991	24	2002	12		

TABLE 3 | Annual average abundance of Atlantic cod at site 428.

Year	Abundance	Year	Abundance	Year	Abundance	Year	Abundance
1952	20	1967	12	1982	28	1997	17
1953	15	1968	9	1983	42	1998	23
1954	24	1969	13	1984	25	1999	16
1955	16	1970	23	1985	44	2000	8
1956	42	1971	25	1986	50	2001	7
1957	14	1972	25	1987	51	2002	7
1958	17	1973	27	1988	32	2003	4
1959	18	1974	25	1989	39	2004	4
1960	25	1975	17	1990	35	2005	9
1961	9	1976	21	1991	35	2006	12
1962	14	1977	19	1992	26	2007	10
1963	38	1978	24	1993	30	2008	3
1964	19	1979	28	1994	25	2009	9
1965	44	1980	27	1995	23	2010	13
1966	20	1981	29	1996	30	2011	20

Sea Surface Salinity

The global monthly SSS is procurable by the quality-controlled ocean salinity profiles in the EN4 database from Met Office Hadley Centre, covering the time period from 1900 to present. We retrieved the annual SSS at both site 119 and site 428 in the same process as the annual SST, namely, by spatial averaging and time averaging after collating the relevant salinity points around the geographical midpoint of each site.

Atlantic Cod Catches

The global annual catches of Atlantic cod have been extracted *via* Food and Agriculture Organization Statistical Database (FAOSTAT). The Atlantic cod catch records at all observation sites since 1952 to 2011 have first been retrieved and then we selected the observation stations relevant to geographic midpoint of each site in study, with relatively complete records. Due to fishing prohibition in specific months, e.g., the spawning season of Atlantic cod, we could not collect the catches for every month, but the total number in each year. The annual average catches could be calculated by spatial averaging of neighboring observation stations.

Prey Biomass

For the prey biomass of Atlantic cod at each site, we first compiled the primary prey abundance *via* BioTIME database, such as herring, hairscales, capelin, and various zooplankton and invertebrates, and carried out the same screening process as Atlantic cod abundance, i.e., deleting duplicates, zero abundance counts, and abiotic records, as well as eliminating singular values through mathematical expectations. Finally, the annual

prey biomass of each site has been derived by accumulating the abundance of all prey generated by time averaging.

Methods

Basic Wavelet Analysis

The basic wavelet analysis decomposes the abundance of Atlantic cod in the form of time series into a superposition of wavelet functions, all of which are derived from a mother wavelet function through translation and scaling. Let $A(t)$ be the abundance of Atlantic cod time series, the irregularity of the wavelet function can approximately reflect the sharp non-stationary changes in Atlantic cod abundance $A(t)$, and hereby more realistically describe the original dynamics at a certain time scale:

$$\phi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\phi\left(\frac{t-\tau}{a}\right)(a, \tau \in R, a > 0) \quad (1)$$

$$\text{with} \quad \int_{-\infty}^{+\infty} \phi(t)dt = 0 \quad (2)$$

where $\phi(t)$ is a mother wavelet function, $\phi(t) \in L^2(R)$, which refers to a type of oscillation function that can quickly decay to zero and could take many forms, such as Haar wavelet, Morlet wavelet (Goupillaud et al., 1984), Daubechies wavelet (Daubechies, 1991), Meyer wavelet (Meyers et al., 1993), Mexican straw hat wavelet (Torrence and Compo, 1998), and so on. $\phi_{a,\tau}(t)$ is a sub-wavelet, a is a scale factor, representing the period length of the wavelet, and τ is a shift factor, reflecting the shift in time. In view of the abundance variation of Atlantic cod at multiple time scales, we try to focus on the principle of continuous wavelet transform over time:

$$W_A(a, \tau) = \frac{1}{\sqrt{a}}\int_R A(t)\overline{\phi}\left(\frac{t-\tau}{a}\right)dt \quad (3)$$

where $W_A(a, \tau)$ is the wavelet transform coefficient.

Periodicity

In order to discover the potential periodicity embedded in Atlantic cod abundance $A(t)$, it is more desirable to achieve smooth wavelet amplitude by a non-orthogonal wavelet function, while the complex wavelet transform could provide both phase and amplitude to help with it (Torrence and Compo, 1998). We apply Morlet wavelet as our mother wavelet which not only has non-orthogonality but also has exponential complex wavelet adjusted by Gaussian. The Morlet wavelet function can be denoted as follows:

$$\phi_0(t) = \pi^{-1/4}e^{iw_0t}e^{-t^2/2} \quad (4)$$

where t represents the time variable and w_0 is a dimensionless frequency.

The modulus and modulus square of wavelet coefficients can first determine certain range of periodicity of global Atlantic cod abundance $A(t)$ at multiple time scales. The modulus of the Morlet wavelet coefficient can be considered to be a reflection of the energy density distribution in the time domain. The greater the modulus, the stronger the periodicity of the global Atlantic cod abundance. We could hereby take it as one historic evidence to evaluate the future trend of global Atlantic cod abundance $A(t)$ at multiple time scales. The wavelet spectrum of Atlantic cod abundance $A(t)$ can be defined as:

$$|W_A(a, \tau)|^2 = W_A(a, \tau)W_A^*(a, \tau) \quad (5)$$

where $|W_A(a, \tau)|$ and $W_A^*(a, \tau)$ are the modulus and the complex conjugate of the wavelet coefficient W_A (a, τ).

The variance of the wavelet coefficients could help determine the periodicity in a more exact way, so here we further integrate all squared values of the wavelet coefficients with scale a in the time domain as follows:

$$Var(a) = \int_{-\infty}^{+\infty} |W_A(a, \tau)|^2 \, d\tau \quad (6)$$

Wavelet Coherence

Let $T(t)$ be the corresponding SST time series. The cross-wavelet transform between Atlantic cod abundance $A(t)$ and SST $T(t)$ can be represented as:

$$W_{AT}(a, \tau) = W_A(a, \tau)W_T^*(a, \tau) \quad (7)$$

where $W_A(a, \tau)$ and $W_T(a, \tau)$ are, respectively, the wavelet transforms of $A(t)$ and $T(t)$, and $W_T^*(a, \tau)$ indicates the complex conjugation of $W_T(a, \tau)$. The covariance between the cross-wavelet power spectrum is then defined as follows:

$$|W_{AT}(a, \tau)| = \sqrt{|W_A(a, \tau)|\left|W_T^*(a, \tau)\right|} \quad (8)$$

Assuming the potential periodicity of Atlantic cod abundance, we try to expand the total finite length of time series to deal with the edge effect; there will exist the discontinuities at the edges with the decrease of the amplitude in the wavelet power spectrum, which is specified as the Cone of Influence (COI). Since the results in COI will be affected by boundary distortion, reliable information cannot be provided and should be removed.

The correlation between Atlantic cod abundance and SST could be further measured *via* wavelet coherence (Torrence and Webster, 1999), as follows:

$$C_{AT}(a, \tau) = \frac{|S(W_{AT}(a, \tau))|}{\sqrt{S(|W_{AA}(a, \tau)|^2) \bullet S(|W_{TT}(a, \tau)|^2)}} \quad (9)$$

where $S(W_{AT}(a, \tau))$ is the cross spectral density between $A(t)$ and $T(t)$. The wavelet coherence coefficient reflects the synchronization similarity, i.e., co-movement between two time series, or whether the variation law takes on a linear relationship. The higher wavelet coherence is, the stronger the correlation is.

Meanwhile, the wavelet phase differences capture negative and positive correlation directions and depict the possible lead-lag effects among two time series:

$$\psi_{AT} = \tan^{-1}\left(\frac{I\{S(W_{AT}(a, \tau))\}}{R\{S(W_{AT}(a, \tau))\}}\right), \psi_{AT} \in [-\pi, \pi] \quad (10)$$

where $I$and R are the imaginary and real parts of the smoothed cross-wavelet transform, respectively. If ψ_{AT} is zero, the two series co-move together (in-phase), and if it is π (or

$-\pi$), they co-move in opposite directions (out of phase). If $\psi_{AT} \in (0, \pi/2)$, they positively co-move, and $A_{(t)}$ leads $T_{(t)}$; if $\psi_{AT} \in (\pi/2, \pi)$, they negatively co-move, and $T_{(t)}$ leads $A_{(t)}$; if $\psi_{AT} \in (-\pi, -\pi/2)$, they negatively co-move, and $A_{(t)}$ leads $_{T(t)}$; if $\psi_{AT} \in (-\pi/2, 0)$, they positively co-move, and $T_{(t)}$ leads $A_{(t)}$. The abovementioned wavelet coherence can also be extended to evaluate the mutual relationship between Atlantic cod abundance and other extrinsic indices, including Atlantic cod catches $C_{(t)}$, prey biomass $B_{(t)}$, and SSS $S_{(t)}$.

LSTM-SAE

We employ LSTM as an optimized recurrent neural network that allows to capture and retain long-term dependencies between Atlantic cod abundance $A_{(t)}$ and other extrinsic indices, and meanwhile concludes the mutual relationship between the historic and current linkage effectively. Let the multivariate input vector be the combination of all or parts of time series from Atlantic cod abundance, SST, catches and prey biomass, and SSS in the time step $_t$, and the output vector be the label from Atlantic cod abundance itself in the next time step $_{t+1}$. The advantage of the LSTM structure is that three types of gates, i.e., the input gate, output gate, and forget gate, are included, which could allow the storage of time series of Atlantic cod abundance and the external impact factors over a long-term period and also solve the vanishing gradient problem. Let $_{X^t=[A^t,T^t,C^t,B^t,S^t]}$ and $_{y^t=A^{t+1}}$ be the input and output of LSTM unit at a certain time point $_t$, then the following calculation is performed,

$$\begin{aligned} i^t &= \sigma(G_i \times X^t + R_i \times y_L^{t-1} + b_i) \\ f^t &= \sigma(G_f \times X^t + R_f \times y_L^{t-1} + b_f) \\ o^t &= \sigma(G_o \times X^t + R_o \times y_L^{t-1} + b_o) \\ z^t &= \sigma(G_z \times X^t + R_z \times y_L^{t-1} + b_z) \\ c^t &= i^t \odot z^t + c^{t-1} \odot f^t \\ y^t &= o^t \odot \tan(c^t) \end{aligned} \tag{11}$$

where G_i, G_o, G_f, G_z, R_i, R_o, R_f,R_z, b_i, b_o, b_f, and b_z are, respectively, the weights of the input, the shared recurrent weights, and the bias weights for the three types of gates and input unit; i^t, f^t, o^t, and z^t are, respectively, the output of the three types of gates and input unit; c^t is the cell state of LSTM unit; σ represents the sigmoid activation function; and $\odot$ is the Hadamard product, which refers to the element-wise product among matrices.

LSTM-based stacked autoencoder replaces random initialization strategies by applying stacked auto-encoder (SAE) into deep LSTMs in an unsupervised pre-training manner. As is shown in **Figure 2**, LSTM-SAE consists of three conventional shallow LSTM-AEs (LSTM-based autoencoder), which not only enhances the non-linear characteristics and compresses the input dimension, but also circumvents the expression limitations in multivariate prediction.

The LSTM-AE module comes from the basic autoencoder (AE), which is originally sequentially connected as three layers in an unsupervised learning paradigm. The training procedure consists of two phases: encoding, in which the input is mapped into the hidden layer, and decoding, in which the input is reconstructed from the compressed representation. Given an unlabeled input X^t, it can be formulated as follows:

$$h_1 = f(w_1 X^t + b_1) \tag{12}$$

$$\hat{X^t} = w_2 h_1 + b_2 \tag{13}$$

where h_1 represents the hidden encoder vector calculated from the input vector X^t, and $\hat{X^t}$ is the decoder vector of the output layer. Additionally, f is the activation function, w_1 and w_2 are the weight matrix of the encoder and decoder, and b_1 and b_2 are the bias vectors in each phase, respectively. The process of training is to minimize the reconstruction error; i.e., the cost function of each LSTM-AE module can be written as the difference between the original input X^t and the reconstructed time series $\hat{X^t}$.

LSTM-based stacked autoencoder sequentially combined and superposed shallow LSTM-AEs into deep learning networks through layer-by-layer greedy methods. The learning process consists of a pre-training phase and a fine-tuning phase. In the pre-training stage, we train the first LSTM-AE module in the stack and save its LSTM encoder layer as the input of the second LSTM-AE module, the second LSTM-AE module in the stack is trained with the encoded version of inputs loaded, and its LSTM encoder layer is saved as the input of the third LSTM-AE module. The purpose of reconstructing the original input instead of using the coded input of the previous layer is to strengthen the encoder to learn the characteristics of the original input. The original input is encoded twice by loading the encoder saved in the first two layers to train the third LSTM-AE module. We reconstruct the original inputs, not the encoded inputs, in order to enforce the encoder to learn the features of the inputs. Finally, we manage to initialize the three saved LSTM encoders and feed the real multivariate input in the time step t, together with the expected output of Atlantic cod abundance in the next time step $t + 1$, to learn in a supervised learning fashion during the fine-tuning stage.

RESULTS

Periodicity

The periodicity estimation has been performed from 1919 to 2016 on the global annual average abundance of Atlantic cod derived by BioTIME *via* wavelet coefficients, on the basis of Morlet wavelet base. In order to offset the boundary effect of the specified region of COI, the extended time series at both ends has been employed for the wavelet transform coefficient. The symmetrical expansion around both ends for Atlantic cod abundance was first done within the time domain. After eliminating the expansion, the real part of the retained wavelet coefficient was made up of the wavelet coefficient contour map, as demonstrated in **Figure 3**, with the abscissa representing time and the ordinate referring to the time scale.

It has been found from the wavelet coefficient contour map in **Figure 3** that there exist three types of period variation

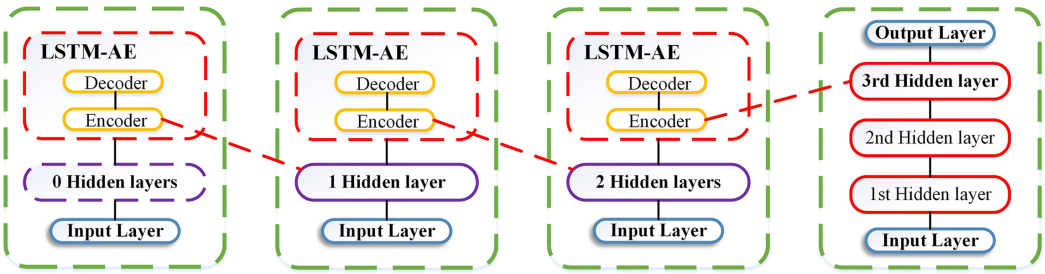


FIGURE 2 | LSTM-SAE model.

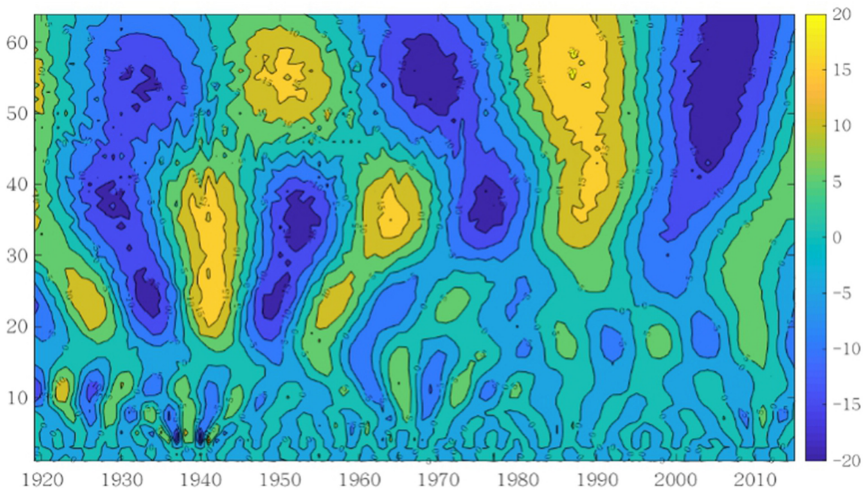


FIGURE 3 | Contour map of wavelet coefficient.

of global Atlantic cod abundance at multiple time scales, i.e., 17–30 years, 30–45 years, and 45–64 years. Among them, at the time scale of 17–30 years, the global Atlantic cod abundance fluctuates abundantly, but the corresponding real value of wavelet coefficients is smaller, indicating that the periodicity is less obvious. At the time scale of 30–45 years, Atlantic cod abundance has experienced four fluctuations from rich to poor period according to the reflection of the oscillation intensity. Considering that there is a hidden period, the real value of wavelet coefficients is medium. At the time scale of 45–64 years, the global Atlantic cod abundance has the most obvious periodic variation in rich and poor periods, and the real value of wavelet coefficients is relatively large. Three poor centers of the oscillation intensity, 1930, 1970, and 2005, and two rich centers, 1950 and 1985, have been discovered. The periodic variation of the last two time scales behaved relatively stably throughout the time domain of the Atlantic cod abundance time series in study, while the first time scale was relatively weak and only stable before the 1960s.

We also applied the wavelet energy spectrum to analyze the evolution of oscillation energy at different periods in global Atlantic cod abundance. The modulus and modulus square maps of the wavelet coefficients are illustrated in **Figure 4**, where the largest values correspond to the 45- to 64-year time scale, especially since 1960, indicating that there is a major period and the periodicity of global Atlantic cod abundance is more stable. At the time scale of 30–45 years, there is a minor period from 1919 to 1990 accordingly, with relatively small values of the modulus and modulus square. Similarly, the periodicity emerges at the time scale of 17–30 years from 1919 to 1960.

The wavelet variance map proportionally reflects the distribution of fluctuating energy in Atlantic cod abundance time series at multiple time scales and can be employed to derive the time periods in the evolution process, as shown in **Figure 5A**. There exist four obvious peaks in the wavelet variance map, which in turn correspond to the time scales of 55, 37, 24, and 11 years, i.e., the highest of the first major period in Atlantic cod abundance variation, as well as the second to third, fourth highest peak. The fluctuations of the above four periods influence the variation characteristics of Atlantic cod abundance in the entire time domain, where the first and the second periods play dominant roles, and the third and fourth periods only have minor effects for reference because of their small value in variance curves.

We further characterized and developed the trend graph of the first and second periods for global Atlantic cod abundance

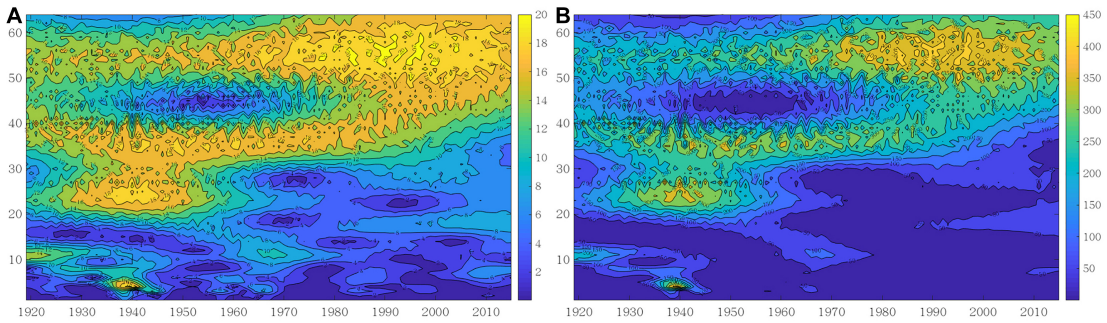


FIGURE 4 | Wavelet contour map with Morlet wavelet. **(A)** Modulus of wavelet coefficient. **(B)** Modulus square of wavelet coefficient.

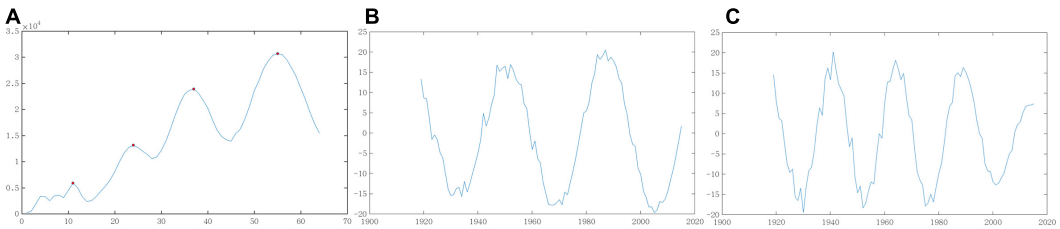


FIGURE 5 | Variance and periodic map of wavelet coefficients. **(A)** Wavelet variance. **(B)** The first cycle. **(C)** The second cycle.

at two time scales. From the test results of the wavelet variance in **Figures 5B,C**, it has been illustrated that there appear the characteristics of global Atlantic cod abundance lasting about 36 years within an average period at the 55-year time scale, with roughly 2.7 periods in rich and poor, respectively. At the 37-year time scale, the average period of global Atlantic cod lasts around 24 years with four periods; the stability of this time period is relatively poor, but the overall periodicity is still very obvious. We could hereby make use of the inherent periodicity tendency of global Atlantic cod abundance to explore its potential alternating fluctuation patterns along time, providing a basis for the abundance prediction itself.

Wavelet Coherence

We utilized the wavelet coherence to search for how the co-movement between Atlantic cod abundance and the extrinsic indices performs over time and described the lead-lag effects through the wavelet phase difference. The wavelet coherence and phase difference map of Atlantic cod abundance with SST, catches, prey biomass, and SSS at sites 119 and 428 are, respectively, shown in **Figures 6**, **7**, where wavelet correlation is affected by discontinuity, and the color bar corresponds to the relative intensity of each frequency (the stronger correlation tends to be in red, while the weaker correlation tends to be in blue), the edge COI is denoted in thick black curve, and 5% and 10% significance levels are, respectively, represented by a black thin line and a black dashed line.

For Atlantic cod abundance vs. SST, in **Figure 6A.1**, from the perspective of long-term frequency, the correlation coefficient at site 119 in the 7- to 13-year band is greater than 0.95 after 1987, indicating that there exists a significant long-term correlation and then the degree of correlation coefficients begins to decline after 2003. In **Figure 7A.1**, there is a significant correlation at site 428 in the 10- to 15-year band from 1985 to 2000, and the correlation coefficient reached more than 0.95. In the short-term frequency band, the correlation at both sites is not obvious. Meanwhile, in the long-term frequency band, the phase difference at both sites is in the range of $\pi/2$ to π throughout the period when the correlation is strong, which implies that there is a significant negative correlation between the two time series at both sites 119 and 428.

For Atlantic cod abundance vs. catches, in **Figure 6B.1**, with reference to the wavelet coherence map, we found that it exceeded 0.95 at site 119 in the 1- to 5-year band in the 1990s, indicating significant correlation in the short term. In the long-term frequency band of 7–14 years, this significant correlation lasted from 1977 to 1995, with the coefficient greater than 0.95, and there was no obvious correlation thereafter. In **Figure 7B.1**, the two time series is significantly correlated at site 428 in the 9- to 15-year frequency band almost throughout the sampling period, especially in the period from 1966 to 2000 when the significance ratio reached its peak, with the correlation coefficient exceeding 0.95. The coherence map displays a common significant correlation at both sites from the 1970s to the mid-1990s, while the correlation at site 428 lasted longer, until 2000. The phase difference described that, in the long-term frequency band, the two time series at site 428 are significantly negatively correlated throughout the time period, while the phase difference at site 119 in the frequency band of 7–14 years is close to zero during the entire period, indicating that the two time series are approximately moving in synchronization, and at

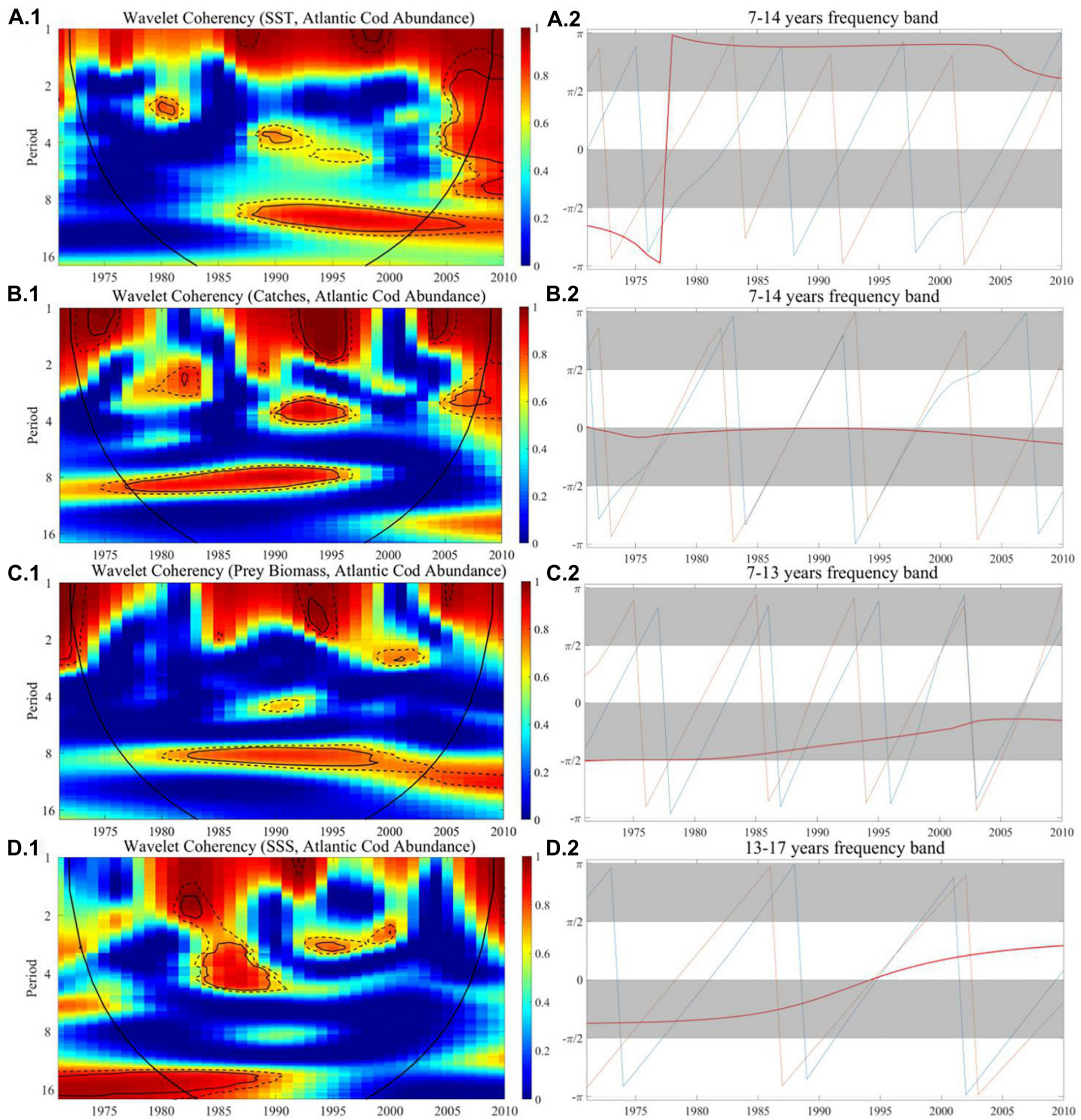


FIGURE 6 | The wavelet coherence and phase difference between Atlantic cod abundance and SST, catches, prey biomass, and SSS at site 119. **(A–D.1)** Wavelet coherence and **(A–D.2)** phase difference. The *y*-axis refers to frequency (in years), and the *x*-axis refers to the time period.

both sites, Atlantic cod caches tend to dominantly lead its changes in abundance.

For Atlantic cod abundance vs. prey biomass, in **Figure 6C.1**, the correlation coefficient at site 119 in the long-term frequency band of 7–13 years was about 0.9 from 1982 to 1998, but in the short term, there is no obvious connection being displayed. In **Figure 7C.1**, the coherence map shows significant correlation at site 428 in the long-term frequency band from 7 to 13 years during almost the entire time period. The correlation coefficient is greater than 0.9 around the 1990s, and began to decline after 2005, indicating that the correlation was most significant from the mid-1980s to the end of the last century. The phase difference demonstrates that there is a significantly positive correlation between two time series at both sites during the entire co-movement period, and the Atlantic cod prey biomass tends to dominantly lead its changes in abundance.

For Atlantic cod abundance vs. SSS, in **Figure 6D.1**, from the coherence map, we could find that there is a significant correlation in the 3- to 5-year band at site 119 from 1983 to 1989, but after that, it turned to a weak correlation. In the long run, this correlation becomes more significant in the frequency band of 13–17 years, with the coefficient close to 1, and lasts from 1980 to 1988, but after that, this correlation was not obvious. In **Figure 7D.1**, the correlation coefficient at site 428 in the long-term frequency band of 10–17 years exceeded 0.9 from 1990 to 2000, while before 1990, the correlation coefficient was generally less than 0.6. In the short term, this correlation is not obvious. The phase difference shows that the two time series at site 119 are

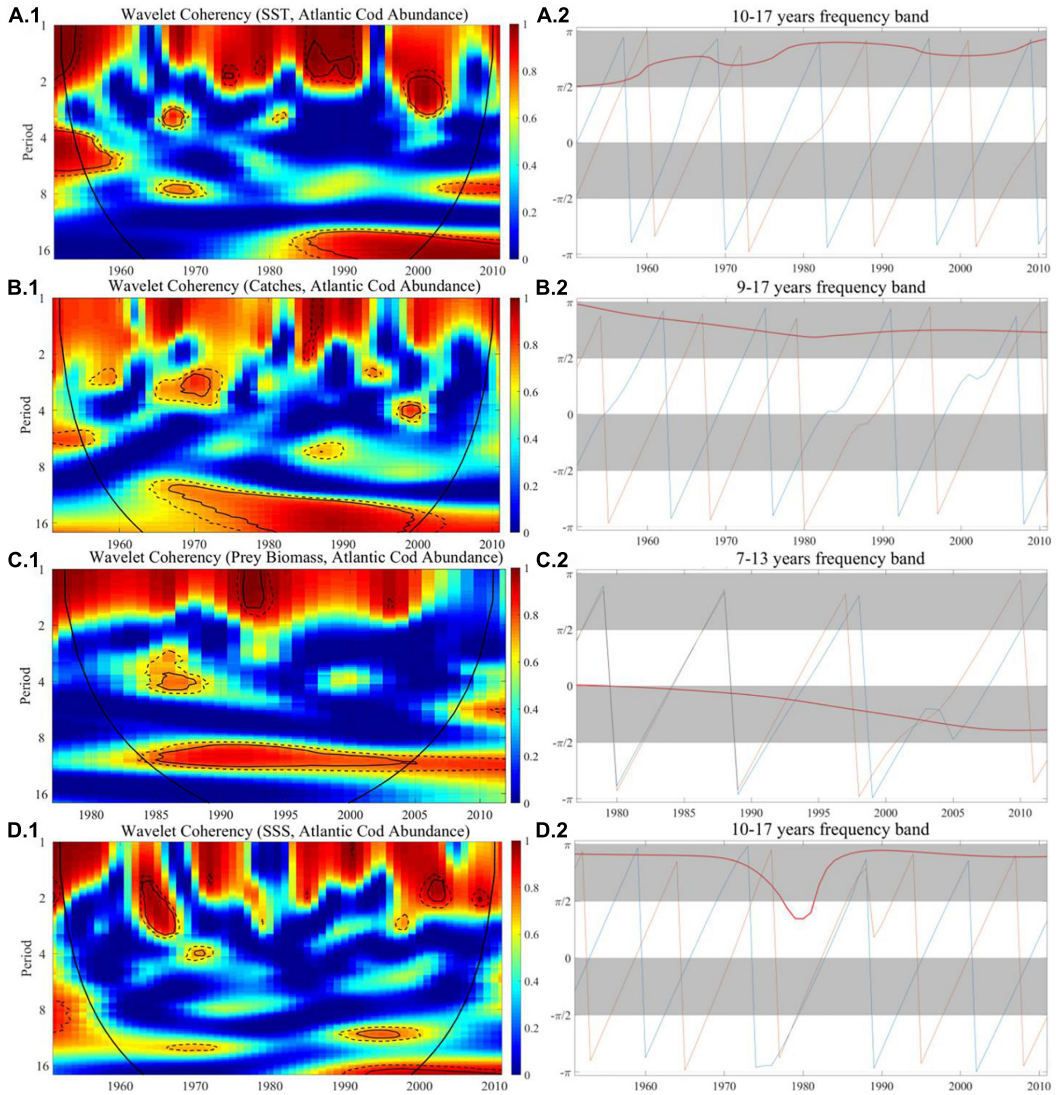


FIGURE 7 | The wavelet coherence and phase difference between Atlantic cod abundance and SST, catches, prey biomass, and SSS at site 428. **(A–D.1)** Wavelet coherence and **(A–D.2)** phase difference. The *y*-axis refers to frequency (in years), and the *x*-axis refers to the time period.

positively correlated throughout the period of co-movement, while at site 428, they are more probably negatively correlated. However, at both sites, SSS tends to dominantly lead its changes in abundance.

DISCUSSION

First, it has been basically revealed from our experimental results that at both sites 119 and 428, SST began to have dominantly negative effects on the abundance variation of Atlantic cod since the 1980s. The increase of sea temperature actually exerted dramatic influences to the Atlantic cod abundance. For example, it has been investigated that from the 1980s to the beginning of the 21st century, the average SST and bottom temperature of the North Sea around site 428 have increased by 0.7°C and 0.15°C per decade (Walther et al., 2002; Sheppard, 2004). It has also been reported by NOAA that the North Atlantic Oscillation (NAO) index, i.e., the oscillation of atmospheric mass between the Arctic and the subtropical Atlantic, remains a positive phase since then, indicating the indirect consistency in its variation pattern on Atlantic cod (Brander and Mohn, 2004), e.g., the changes in the recruitment and spawning of Atlantic cod with the local environmental variables (Stige et al., 2006). At site 428, the annual SST from 1980 to 2000 was 10.5°C on average, an annual increase of 0.6°C from the 1970s, corresponding to the average decrease of Atlantic cod abundance by roughly 40%. In particular, due to the El Niño events in 1982–1983 and 1997–1998, the annual SST at site 428 both exceeded 12°C in the year 1983 and 1997, which exactly corresponds to the sharp drop in the abundance of Atlantic cod by 45% and 49%, respectively, compared to the entire 1970s. Similarly, the annual SST at site 119 in 1997 have reached

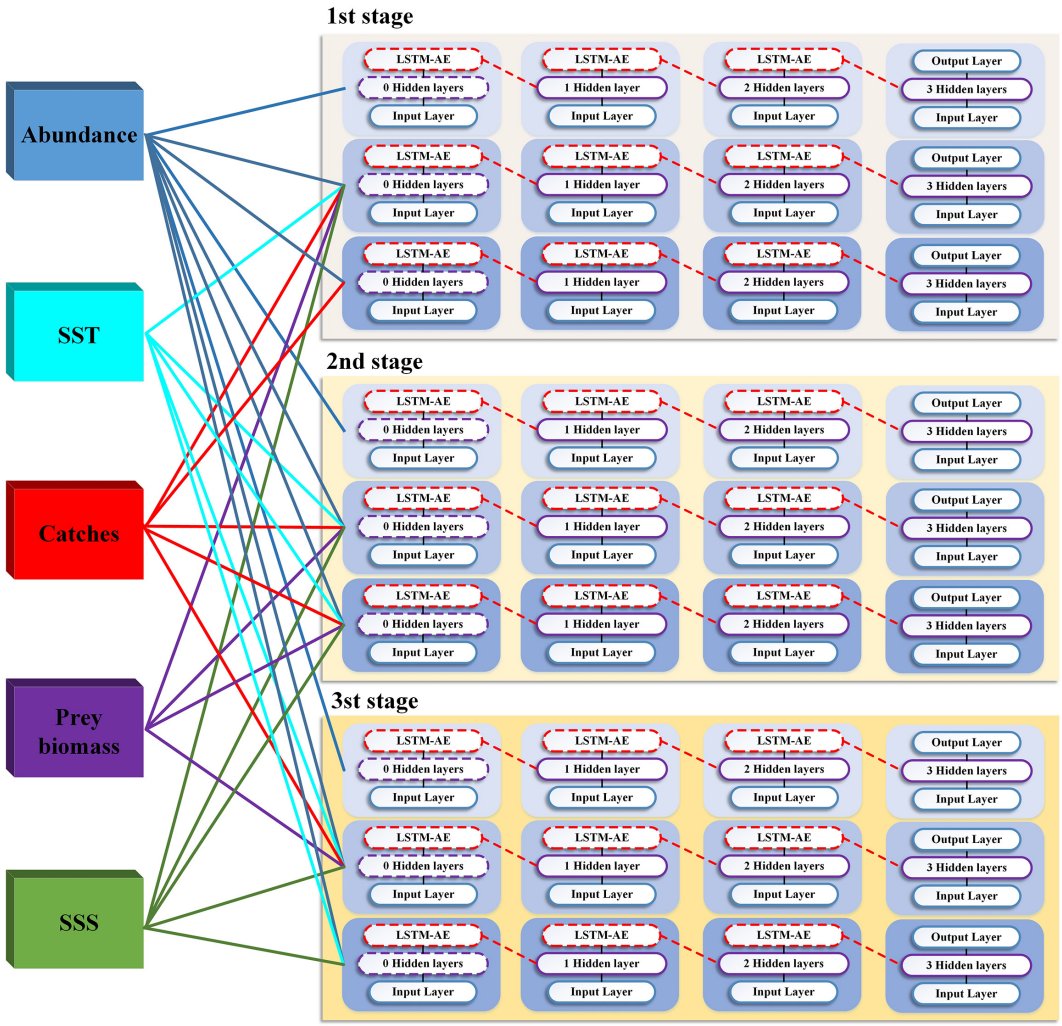


FIGURE 8 | LSTM-SAE predictive ensemble.

13.1°C, far exceeding the average values 10.3°C in the 1970s, and the corresponding abundance of Atlantic cod dropped by 56%. In total, 13 El Niño events have occurred globally from 1950 to 2000, 69% of them after 1980, demonstrating that the abundance variation of Atlantic cod is closely related to the changes of SST due to frequent activities of El Niño since the 1980s, which could be one of the significant reasons for the decline of Atlantic cod abundance after 1980.

Second, a strong positive correlation between the Atlantic cod abundance and its prey richness has been demonstrated from our experimental results at site 119 around 1980–2000 and at site 428 around 1985–2008. While the richness of marine primary producers decreased by 26% and 34% during the above time period, compared to the average values in the whole 1970s, the Atlantic cod abundance dropped by 40% and 47% correspondingly. We believe that there are two major reasons for the descent in the richness of primary producers. (1) The influence of the marine environment changes, such as the rising SST and the ocean acidification. Whether it is smaller zooplankton, invertebrates, or herring and capelin, most of them are more sensitive to changes in the marine environment. The abundance of small pelagic species, such as shrimp and capelin, decreases with increasing temperature, and even some zooplankton may have disappeared. For example, it has been investigated that the abundance of capelin in 1998 at Newfoundland around site 119 has particularly made a reduction of 55.3% compared to that in 1978 (Rose and O'Driscoll, 2002), and the corresponding abundance of Atlantic cod dropped by roughly 42% during the same period. (2) The increase in abundance of Atlantic cod competitors, such as those migrant predators in deep open waters, like moray, perch, tuna, etc. It has been reported that the abundance of these large fish species in the North Atlantic was not significantly affected by the increase in temperature, and even showed an increasing trend because of migration. For example, the SST of Celtic Sea to the south of Ireland had exceeded 12°C in 1992, causing some tuna to migrate northward to North Sea near site 428 (Hiddink and Ter Hofstede, 2008). The increase in the number and types of advanced species

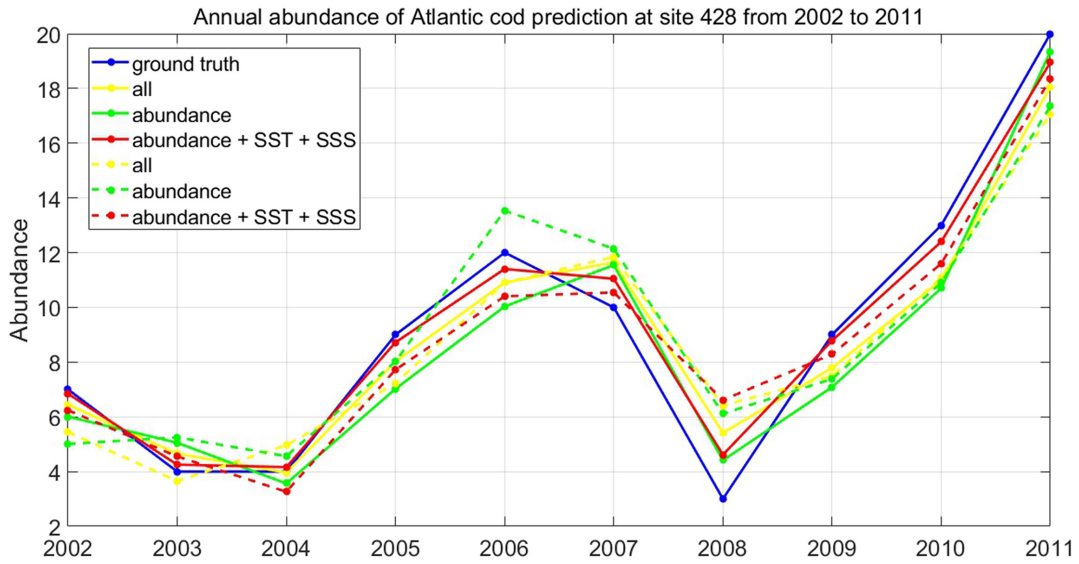


FIGURE 9 | Prediction of Atlantic cod abundance.

could be attributed to accelerate the rate of decrease of primary species richness. As the abundance variation of Atlantic cod basically produced responses to both the prey biomass and SST since the 1980s and continued to the present, the effect of prey biomass also indirectly proves the influence of SST, indicating that SST could not only cause a direct response, but may be an indirect response to the changes of Atlantic cod abundance.

Third, it is obviously shown from our experimental results that overfishing imposed tremendous influences on the abundance variation of Atlantic cod. The occurrence of highly correlative responses to Atlantic cod catches has been manifested from the 1960s to the end of the last century, at both sites 119 and 428; the strongest happened in the 1970s to the 1990s. For example, at site 119, the abundance of Atlantic cod has strongly followed the catches synchronously during 1975–1995, indicating that the catches played an actively guiding role in the abundance variation during this time period. In 1976, the catches of Atlantic cod at sites 119 and 428, respectively, increased by 32.2% and 19.1%, compared to that in 1960s, and the abundance of Atlantic cod decreased by 28.6% and 12.1% correspondingly. The dramatic increase in Atlantic cod catches since the early 1970s was mainly due to the rapid development of fishery technology. High-efficiency equipment such as ocean trawl and seine nets have been widely adopted (Valdemarsen, 2001). As a result, the world's annual fish catches were more than quadrupled in the late 1970s. It has been investigated that the average annual catches of Atlantic cod in the Gulf of Maine near site 119 in the 1980s increased to 40,000 tons, more than four times that of the 1960s. However, there was a cliff-like decline of Atlantic cod catches after 1992, and the abundance of Atlantic cod increased by 18%, and then the average annual catches had returned to 10,000 tons after 2000 (Fogarty et al., 2008). There have been a series of ban policies issued from the 1990s in many countries, looking forward to seeing the abundance of Atlantic cod recover. Although, after years of limited fishing or even a ban on fishing, the scale of the local Atlantic cod showed no significant signs of rebound as expected in the end, it coincided with strong negative correlation at the beginning of its implementation.

Fourth, it has been demonstrated from our experimental results that SSS made great differences regarding the responses of the abundance variation in Atlantic cod at sites 119 and 428, respectively. On the one hand, at site 428, there mainly occurred a strong negative lead to Atlantic cod abundance after 1990. Since the increase in SSS is assumed to affect the activity of sodium and potassium ions in Atlantic cod, it might hereby cause the slow growth of Atlantic cod. In 2003, the SSS value at site 428 reached 36%, an increase of 7% compared to 1970, and at an average annual increase rate of 0.24% since then, while the abundance of Atlantic cod fell 47% during the corresponding time period. It has been reported that ocean evaporation (EVP) increased from the trough in 1977 to the peak in 2003 (Yu, 2007), which exactly coincides with the increase in SSS at site 428. In addition, the time–frequency co-movement of both SSS and SST at site 428 had the same tendency with the abundance variation of Atlantic cod after the 1990s. We believe that this is mainly because the rising SST causes the sea surface to evaporate more easily, so their correlation with Atlantic cod abundance was similar. On the other hand, at site 119, we have noticed that there was a strong positive correlation between SSS and Atlantic cod abundance mainly in the 1980s. For example, the SSS value at site 119 dropped from 32% in 1980 to 29% in 1983, and the abundance of Atlantic cod decreased by 27% during the corresponding time. We suppose that the decline in SSS in the corresponding period may be highly correlative to the influences of North American hurricanes in 1982. Humid climate, especially frequent rainfall during that time period, made it possible to lead to further SSS changes. It has been reported that Hurricane Alicia moved northward along the West Coast of the United States and

Canada near site 119 in 1982 at an average wind speed of 95 mph, causing a 13-foot storm surge in the Gulf of Maine, with 6–12 inches of heavy rain.

The above external responses of the abundance variability run through the whole life history of Atlantic cod, and could be further attributed to the consistency with the simulation model or hypothesis that have been put forward. The warming and SSS reached deeper ocean due to strong mixing and convection associated to thermohaline circulation (Rahmstorf, 1995), which implied that the change in sea surface would also be transmitted to deeper habitat of Atlantic cod. The individual-based mechanistic model (Kristiansen et al., 2011) once incorporates the physical characteristics and the biological properties of the Atlantic cod larva and its surroundings, to assess the seasonal pattern of size variation of Atlantic cod eggs through adjusting environmental conditions of the mode. It has been shown from our experimental result that when SST at site 428 increased greatly in the 1990s, there was a significant negative correlation with Atlantic cod abundance. We could hereby presumably infer from the model that the size of Atlantic cod eggs might be reduced during that period and thus prolonged their hatching time and increased the possibility to be preyed. Combining with the Regional Ocean Modeling System (ROMS) (Shchepetkin and McWilliams, 2005), Vikebø et al. (2007) simulated the drift and distribution of Atlantic cod eggs and larvae. It has been shown from our experimental result that there was a dramatic decline in Atlantic cod abundance at site 428 during 2007–2008, and we could hereby infer that this might be related to the strong Norway current with large resolution motion in 2008 (Skagseth et al., 2008, 2011); the survival rate of Atlantic cod larvae was relatively reduced due to the long-term drift time. Árnason et al. (2013) have evaluated the growth rate of Atlantic cod under different salinity in a radioimmunoassay procedure and a protein assay kit, respectively. It has been shown from our experimental result that when SSS at site 428 increased obviously in 1997, there was a negative correlation with Atlantic cod abundance during this period. We could hereby assume that the continuous increase of SSS might have reduced the plasma cortisol level and the total plasma protein of Atlantic cod, resulting in the decrease of growth rate and even death of Atlantic cod. The probit regression model of Atlantic cod (Hutchings and Myers, 1994) classified maturing and spawning females and spent females, defined the time of spawning on which 50% of mature females were in a spent state, and described how the probability that a female will be spent was related to age at different seawater temperatures. It has been shown from our experimental result that at both site 119 and site 428, SST increased significantly around 1988–1992, and there was a negative correlation with the abundance. We could hereby estimate from the model that the higher seawater temperature might result in earlier Atlantic cod spawning through faster gonad development, which would indirectly lead to the drop in abundance. Wroblewski et al. (2000) monitored the activity of the tagged Atlantic cod in 1992 through an acoustic model and showed that the movement of Atlantic cod was related to the acoustically determined position of its prey and event structure in the local ocean currents. It has been shown from our experimental result that both the biomass of prey and the abundance of Atlantic cod exhibited a dramatically downward trend at site 119 in 1992. We could hereby suppose that in response to the Gulf Stream that moved northward through the Gulf of Maine, the prey shoal moved with the current (Johns et al., 1995), indirectly leading to the similar migration of Atlantic cod.

In general, in view of spatiotemporal interactions with SST, prey biomass, catches, and SSS around the recent 60 years, we conclude that the external responses of the abundance variability in Atlantic cod can be roughly divided into three distinguished stages: namely, the time period before 1985, 1985–1995, and after 1995. At the first stage, the reason for the decline in the abundance of Atlantic cod can be mainly attributed to the rapid rise of fish catches, due to the breakthrough of emerging advanced fishing technologies and its expansion in the fishery industry. In addition, in some observation sites, such as site 119, there were also strongly positive correlation effects from the relatively low SSS, while the impact of SST has not been highlighted, and even hardly observed. At the second stage, the abundance and catches of Atlantic cod still showed strong simultaneous moves, indicating that overfishing remained as one of the main factors. At the same time, warming SST began to be a significant index for the abundance variability in Atlantic cod, imposing both direct and indirect forces to the abundance, together with the significant impact of prey biomass changes and SSS. At the third stage, due to the beginning of fishing restrictions, protective measures, and policies, the impact of overfishing became relatively insignificant. The warming SST at sites 119 and 428 and the growing SSS in some areas such as site 428 directly lead to the decrease of Atlantic cod abundance, implying the increased sensitivity to changes in such marine environment factors.

The long-term dependency in the time series analysis above could quantitatively and systematically reveal the fundamental roles that those external elements responses to the abundance variability in Atlantic cod, and allow the identification of dominantly correlative factors for each stage. Concerning the self-correlated feedback that is always inherently underlined in the abundance itself, together with its periodicity, time–frequency co-movement, and lead-lag effects, we further capture the potentially predictive patterns and mechanism for abundance variation in Atlantic cod at each stage by constructing LSTM architecture. One group of LSTM-SAE units of diverse model selection have been employed to establish the predictive ensemble across stages, as shown in **Figure 8**. For each stage, we, respectively, pick only one input variable, i.e., Atlantic cod abundance itself, all the five input variables of correlative time series in our study, i.e., the Atlantic cod abundance, SST, SSS, prey biomass, and catches, as well as the particularly dominant indices at each stage, for training.

Due to the scarcity of the records in all time series, one type of data augmentation is used, i.e., the Time-Conditional Generative Adversarial Network (T-CGAN), in which both the generator and the discriminator are conditioned on the sampling timestamps, to learn the hidden relationship between data and timestamps and consequently to generate new time series (Ramponi et al.,

TABLE 4 | Prediction performance comparison.

Model	LSTM-SAE			DLSTM		
	All	Abundance	Abundance + SST + SSS	All	Abundance	Abundance + SST + SSS
RMSE	1.439	1.544	**0.958**	1.736	1.936	**1.544**
MAE	1.254	1.424	**0.796**	1.536	1.794	**1.284**
SMAPE	15.933	19.160	**10.547**	20.658	23.607	**18.293**

The bold values in table 4 indicate the best performance of RMSE, MAE and SMAPE evaluation parameters under three ensembles.

2018). At the same time, in order to eliminate the differences among dimensions, the common Min-Max normalization has also been taken to project the five input variables and speed up the convergence in the LSTM-SAE unit. In addition, the length of all the five types of time series, no matter the original and the augmented ones, remains identical, corresponding to the period 1952–2011. The total number of records is 2950 after the standardization and amplification process, including all of the five input variables, and then we divided the original time series dataset into three types, 60% for training, 15% for verification, and 25% for testing, to facilitate learning in the LSTM-SAE unit for a higher prediction accuracy.

Taking the prediction of Atlantic cod abundance at the observation site 428 as an example, we, respectively, select the dominantly correlative factors at three stages, i.e., Atlantic cod catches itself, all the four external impact factors, SST, and SSS, as the auxiliary input variables to help with the estimation, in a time domain of totally 59 observation years, with 43 years for training and verification, and 16 years for test. For each stage, we select three types of input vectors for training and testing, namely, Atlantic cod abundance, the additive dominantly correlative factors, and all the external impact factors. We applied both the classic LSTM and LSTM-SAE strategies in the ensemble into the time series of Atlantic cod abundance at site 428, and the corresponding prediction results at the third stage of 10-year duration from 1995 to 2011 are listed in **Figure 9**, where the ground truth is represented in blue solid line, and the solid and dashed lines in other colors stand for the prediction results of two models with respect to three types of input variables concerned, respectively. The prediction performances of the two competitive models have also been evaluated by means of RMSE, MAE, and symmetric mean absolute percentage error (SMAPE) measure, as listed in **Table 4**.

As it is mentioned above, we have made an initial attempt to explore the abundance prediction in Atlantic cod at site 428 for 1995–2011, and it could be seen from the experimental results that both LSTM-SAE and DLSTM models behaved better when applying multivariate prediction instead of only taking Atlantic cod abundance as the univariate input, with RMSE, MAE, and SMAPE of 0.958, 0.796, and 10.547 in LSTM-SAE and 1.544, 1.284, and 18.293 in the DLSTM model, respectively. It has also proved that the prediction process might perform slightly better when we intentionally select and feed with the dominantly correlative time series as the additional inputs, rather than all the referred external indices, indicating that there might exist relatively higher priority among all the other concerning variables at site 428 in the last decade; i.e., SST and SSS play relatively leading roles in determining the trend in Atlantic cod abundance to be predicted. Meanwhile, LSTM-SAE obviously exhibit better prediction performances compared to the classic DLSTM model for whatever type of training inputs, indicating that the impact of stacking LSTM in an auto-encoder fashion as one kind of pre-trained DLSTM has more advantages in Atlantic cod abundance prediction than the original DLSTM, which depends on most randomized initialization for LSTM.

The recent advances of deep neural network-powered, machine intelligence techniques provide a hopeful solution with the promise of end-to-end prediction modeling for the Atlantic cod abundance variation. First, we quantitatively and systematically reveal the fundamental roles of these external element responses to the abundance variability in Atlantic cod by analyzing the long-term dependencies in time series and allow them to identify dominantly correlative factors for each stage. Then, concerning the self-correlated feedback that is always inherently underlined in the abundance itself, together with its periodicity, time–frequency co-movement, and lead-lag effects, we further capture the potentially predictive hypothesis and mechanism for abundance variation in Atlantic cod at each stage. In the future, we could transfer such a comprehensive prediction mechanism to other sea areas, with the increase of the other influencing factors, and make an effective assessment of Atlantic cod abundance, which could hopefully provide a reference baseline for future fishery management and reasonable fishing decision-making.

CONCLUSION

As one of the economic fish species, Atlantic cod contributes significantly to the world's total fishery production. Atlantic cod distributes in a wide range of habitats, and its survival mode has been influenced by multivariate factors, while most of the existing literature often concentrates on the discussion about the major univariate effects in the abundance variation of Atlantic cod. In this paper, we have quantitatively identified the spatiotemporal responses of Atlantic cod abundance across manifold intrinsic and extrinsic indices, on the basis of its periodicity, co-movement, and lead-lag effects, and then proposed to explore the potentially predictive mechanism about how Atlantic cod abundance would probably evolve over time in favor of LSTM. We have systemically retrieved the annual global abundance variation of Atlantic cod from 1919 to 2016

via BioTIME, and then selected the local observation sites, ID 119 and 428, with relatively complete abundance records for our study, and accordingly accessed and attained the other four external impact factors, i.e., the catches, the prey biomass, SST, and SSS, respectively, at specific geographical coordinates, by spatial and time averaging. We have provided evidences by integrating the modulus and variance of wavelet coefficients that the abundance variation of Atlantic cod will be influenced by its potential 36-year major cycle and 24-year secondary cycle at the time scales of 55 years and 37 years, respectively. We have generated the wavelet coherence map by the ratio between the wavelet cross-spectrum and power spectrum, and the phase difference by the ratio between the imaginary and the real part of the wavelet cross-spectrum, to evaluate the time–frequency co-movement and lead-lag effects over the external impact elements, and allow the identification of the dominantly correlative factors and capture the leading roles at each stage in the time domain. In view of spatiotemporal interactions, we further divided the responses of the abundance variation in Atlantic cod in the recent 60 years into three stages, namely, before 1985, 1985–1995, and after 1995. At the first stage, the reason for the decline in abundance can be mainly attributed to the rapid rise of fish catches. At the second stage, overfishing remained as one of the main factors, and the impact of SST and SSS also provides significant indices for the abundance variability; meanwhile, the mortality of primary producers and forced migration of fish species indirectly led to the decline. At the third stage, the impact of overfishing became relatively insignificant, and warming SST and growing SSS directly led to the decrease of abundance. It should be noted that we primarily consider the above critical external impact factors on the abundance variation of Atlantic cod from a long-term perspective, but we do not rule out the influences of the other factors, such as ocean acidification and marine microorganisms, which have not been involved in our study but could be taken into account in the future. Finally, with the help of historical spatiotemporal responses to manifold intrinsic and extrinsic indices, we have established the ensemble of LSTM-SAE architecture to comprehensively capture the predictive patterns possibly underlined in the variation of Atlantic cod abundance at each stage, by three types of input vectors, namely, Atlantic cod abundance, the additive dominantly correlative factors, and all the external impact factors. It has been demonstrated from the experimental results that the models behaved better when we intentionally select and feed with the dominantly correlative multivariate input, instead of all factors or only taking Atlantic cod abundance as the univariate input, with RMSE, MAE, and SMPE of 0.958, 0.796, and 10.547, respectively, when taking the prediction of Atlantic cod abundance during 1995–2011 at site 428 as an example. Meanwhile, LSTM-SAE obviously exhibits better prediction performances compared to the classic LSTM for whatever type of training inputs, indicating that the impact of stacking LSTM in an auto-encoder fashion has more advantages in Atlantic cod abundance prediction than the original LSTM. The proposed prediction strategies could not only symmetrically investigate the predictive attributes of the Atlantic cod abundance, potentially providing the quantitative references in reasonably making fishing decision, but also integrate evidences on its spatiotemporal responses of abundance variability and help explore the changes of biomass statistics, distribution, and biodiversity. Although the developed prediction scheme in our study still tends to be an initial attempt to understand the potential variation tendency of Atlantic cod abundance, we believe that it is quite reasonable to approximate the changes based on the past observations, in accordance with the direct and indirect external responses. Due to the rapid development of deep learning capabilities, it is to be hoped that understanding the abundance variability toward marine environment and anthropic forces could be further improved in the near future, and the public, the politicians, and the fishing industry could expect better and more quantitative predictions of the responses to global changes of not only Atlantic cod but other fish species and the ecosystem as a whole.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: BioTIME database for this study is available from the website (http://biotime.st-andrews.ac.uk) or through the Zenodo repository (https://zenodo.org/record/1095627); FAOSTAT database is retrieved in the website (http://www.fao.org/fishery/statistics/global-capture-production/query/en); ERSST.v5 database is available from NOAA website (https://www1.ncdc.noaa.gov/pub/data/cmb/ersst/v5), EN4 database can be available in the Met Office Hadley Centre observations website (https://www.metoffice.gov.uk/hadobs/en4/).

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because all the data used in the study are publicly available.

AUTHOR CONTRIBUTIONS

RN conceived and designed the study, contributed to analysis and interpretation of data, drafted the manuscript, and made final consent about the manuscript to be published. QY performed the main experimental work, contributed to analysis and interpretation of data, and drafted the manuscript. HH assisted in the experimental work and acquisition of data. XG helped with data statistics. CS provided guidance on experimental methods and data analysis. BH discussed and helped with the experiments. AL suggested the idea of the prediction mechanism. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key R&D Program (2019YFC1408304), the National Program of International S&T

Cooperation (2015DFG32180), the National High-Tech R&D 863 Program (2014AA093410), the National Natural Science Foundation of China (31202036), the National Science & Technology Pillar Program (2012BAD28B05), the National Key R&D Program (2016YFC0301400), and the National Natural Science Foundation of China (41376140).

ACKNOWLEDGMENTS

We would like to acknowledge team members Yu Cai and Jie Wang for their guidance in data analysis. We would also like to acknowledge Anne Magurran, Laura Antão, and Maria Dornelas for their support in availability of BioTIME.

REFERENCES

Aguiar-Conraria, L., and Soares, M. J. (2011). Oil and the macroeconomy: using wavelets to analyze old issues. *Empir. Econ.* 40, 645–655. doi: 10.1007/s00181-010-0371-x

Ames, E. P. (2004). Atlantic cod stock structure in the Gulf of Maine. *Fisheries* 29, 10–28.

Árnason, T., Magnadóttir, B., Björnsson, B., Steinarsson, A., and Björnsson, B. T. (2013). Effects of salinity and temperature on growth, plasma ions, cortisol and immune parameters of juvenile Atlantic cod (*Gadus morhua*). *Aquaculture* 380, 70–79. doi: 10.1016/j.aquaculture.2012.11.036

Baird, J. W., Bishop, C. A., and Murphy, E. F. (1991). Sudden changes in the perceptions of stock size and reference catch levels for cod in northeastern Newfoundland Shelves. *Sci. Counc. Stud.* 16, 111–119.

Beaugrand, G., Lenoir, S., Ibanez, F., and Manté, C. (2011). A new model to assess the probability of occurrence of a species, based on presence-only data. *Mar. Ecol. Prog. Ser.* 424, 175–190. doi: 10.3354/meps08939

Behrenfeld, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., et al. (2006). Climate-driven trends in contemporary ocean productivity. *Nature* 444, 752–755.

Berberidis, C., Aref, W. G., Atallah, M., Vlahavas, I., and Elmagarmid, A. K. (2002). "Multiple and partial periodicity mining in time series databases," in *Proceedings of the 15th Eureopean Conference on Artificial Intelligence, ECAI'2002*, Vol. 2, Lyon, 370–374.

Bogstad, B., Lilly, G. R., Mehl, S., Palsson, O. K., and Stefánsson, G. (1994). Cannibalism and year-class strength in Atlantic cod (*Gadus morhua* L.) in arcto-boreal ecosystems (Barents Sea, Iceland, and eastern Newfoundland). *ICES Mar. Sci. Symp.* 198, 576–599.

Box, G. E., and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Stat. Assoc.* 65, 1509–1526. doi: 10.1080/01621459.1970.10481180

Brander, K., and Mohn, R. (2004). Effect of the North Atlantic Oscillation on recruitment of Atlantic cod (*Gadus morhua*). *Can. J. Fish. Aquat. Sci.* 61, 1558–1564.

Carey, C. C., Hanson, P. C., Lathrop, R. C., and St. Amand, A. L. (2016). Using wavelet analyses to examine variability in phytoplankton seasonal succession and annual periodicity. *J. Plankton Res.* 38, 27–40. doi: 10.1093/plankt/fbv116

Cloern, J. E., and Jassby, A. D. (2010). Patterns and scales of phytoplankton variability in estuarine–coastal ecosystems. *Estuaries Coasts* 33, 230–241. doi: 10.1007/s12237-009-9195-3

Conover, R. J., Wilson, S., Harding, G. C. H., and Vass, W. P. (1995). Climate, copepods and cod: some thoughts on the long-range prospects for a sustainable northern cod fishery. *Clim. Res.* 5, 69–82.

Dahlke, F. T., Leo, E., Mark, F. C., Pörtner, H.-O., Bickmeyer, U., Frickenhaus, S., et al. (2017). Effects of ocean acidification increase embryonic sensitivity to thermal extremes in Atlantic cod, *Gadus morhua*. *Glob. Change Biol.* 23, 1499–1510. doi: 10.1111/gcb.13527

Dajcman, S. (2013). Interdependence between some major European stock markets-a wavelet lead/lag analysis. *Prague Econ. Pap.* 22, 28–49. doi: 10.18267/j.pep.439

Daubechies, I. (1991). "Ten lectures on wavelets," in *Proceedings of the CBMS-NSF Regional Conference Series in Applied Mathematics*, (Philadelphia, PA: SIAM). doi: 10.1137/1.9781611970104

Dey, R., and Salem, F. M. (2017). "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, (Piscataway, NJ: IEEE), 1597–1600.

Drinkwater, K. F. (2005). The response of Atlantic cod (*Gadus morhua*) to future climate change. *ICES J. Mar. Sci.* 62, 1327–1337. doi: 10.1139/f94-155

Drinkwater, K. F., Frank, K., and Petrie, B. (2000). "The effects of *Calanus* on the recruitment, survival, and condition of cod and haddock on the Scotian Shelf," in *Proceedings of the ICES CM, 1000, 07* (Brugge: 2000 ICES Annual Science Conference).

Duan, Y., Yisheng, L. V., and Wang, F. Y. (2016). "Travel time prediction with LSTM neural network," in *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, (Piscataway, NJ: IEEE), 1053–1058. doi: 10.1109/ITSC.2016.7795686

Fan, J., Li, Q., Hou, J., Feng, X., Karimian, H., and Lin, S. (2017). A spatiotemporal prediction framework for air pollution based on deep RNN. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 4, 15–22. doi: 10.5194/isprs-annals-IV-4-W2-15-2017

Fogarty, M., Incze, L., Hayhoe, K., Mountain, D., and Manning, J. (2008). Potential climate change impacts on Atlantic cod (*Gadus morhua*) off the northeastern USA. *Mitig. Adapt. Strateg. Glob. Change* 13, 453–466. doi: 10.1007/s11027-007-9131-4

Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Stat.* 9, 1218–1228.

Fullard, K. J., Early, G., Heide-Jørgensen, M. P., Bloch, D., Rosing-Asvid, A., and Amos, W. (2000). Population structure of long-finned pilot whales in the North Atlantic: a correlation with sea surface temperature? *Mol. Ecol.* 9, 949–958. doi: 10.1046/j.1365-294x.2000.00957.x

Goupillaud, P., Grossmann, A., and Morlet, J. (1984). Cycle-octave and related transforms in seismic signal analysis. *Geoexploration* 23, 85–102. doi: 10.1016/0016-7142(84)90025-5

Habernal, I., and Gurevych, I. (2016). "Which argument is more convincing? Analyzing and predicting convincingness of web arguments using bidirectional lstm," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Vol. 1: Long Papers, (Berlin: Association for Computational Linguistics), 1589–1599. doi: 10.18653/v1/P16-1150

Haseeb, M., Abidin, I. S. Z., Hye, Q. M. A., and Hartani, N. H. (2019). The impact of renewable energy on economic well-being of Malaysia: fresh evidence from auto regressive distributed lag bound testing approach. *Int. J. Energy Econ. Policy* 9:269. doi: 10.32479/ijeep.7229

Hays, G. C. (2017). Ocean currents and marine life. *Curr. Biol.* 27, R470–R473. doi: 10.1016/j.cub.2017.01.044

Hiddink, J. G., and Ter Hofstede, R. (2008). Climate induced increases in species richness of marine fishes. *Glob. Change Biology* 14, 453–460. doi: 10.1111/j.1365-2486.2007.01518.x

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi: 10.1126/science.1127647

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 4700–4708.

Hutchings, J. A. (1996). Spatial and temporal variation in the density of northern cod and a review of hypotheses for the stock's collapse. *Can. J. Fish. Aquat. Sci.* 53, 943–962. doi: 10.1139/f96-097

Hutchings, J. A., and Myers, R. A. (1994). What can be learned from the collapse of a renewable resource? Atlantic cod, *Gadus morhua*, of Newfoundland and Labrador. *Can. J. Fish. Aquat. Sci.* 51, 2126–2146. doi: 10.1139/f94-214

Jagannatha, A. N., and Yu, H. (2016). "Structured prediction models for RNN based sequence labeling in clinical text," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing Conference on Empirical Methods in Natural Language Processing*, Vol. 2016 (NIH Public Access), 856.

Johns, W. E., Shay, T. J., Bane, J. M., and Watts, D. R. (1995). Gulf Stream structure, transport, and recirculation near 68 W. *J. Geophys. Res. Oceans* 100, 817–838. doi: 10.1029/94JC02497

Jordaan, A. (2002). *The Effect of Temperature on the Development, Growth and Survival of Atlantic cod (Gadus morhua) During Early Life-Histories*, The University of Maine.

Joy, M. K., and Death, R. G. (2004). Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. *Freshw. Biol.* 49, 1036–1052. doi: 10.1111/j.1365-2427.2004.01248.x

Kirby, R. R., and Beaugrand, G. (2009). Trophic amplification of climate warming. *Proc. R. Soc. B Biol. Sci.* 276, 4095–4103. doi: 10.1098/rspb.2009.1320

Kogovšek, T., Bogunović, B., and Malej, A. (2010). "Recurrence of bloom-forming scyphomedusae: wavelet analysis of a 200-year time series," in *Jellyfish Blooms: New Problems and Solutions*, eds J. E. Purcell, and D. L. Angel (Dordrecht: Springer), 81–96. doi: 10.1007/s10750-010-0217-8

Kristiansen, T., Drinkwater, K. F., Lough, R. G., and Sundby, S. (2011). Recruitment variability in North Atlantic cod and match-mismatch dynamics. *PLoS One* 6:e17456. doi: 10.1371/journal.pone.0017456

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105. doi: 10.1145/3065386

Lambert, Y., and Dutil, J. D. (1997). Condition and energy reserves of Atlantic cod (*Gadus morhua*) during the collapse of the northern Gulf of St. Lawrence stock. *Can. J. Fish. Aquat. Sci.* 54, 2388–2400. doi: 10.1139/f97-145

Lambert, Y., Dutil, J. D., and Munro, J. (1994). Effects of intermediate and low salinity conditions on growth rate and food conversion of Atlantic cod (*Gadus morhua*). *Can. J. Fish. Aquat. Sci.* 51, 1569–1576. doi: 10.1139/f94-155

Li, W., Qin, B., and Zhang, Y. (2015). Multi-temporal scale characteristics of algae biomass and selected environmental parameters based on wavelet analysis in Lake Taihu, China. *Hydrobiologia* 747, 189–199. doi: 10.1007/s10750-014-2135-7

Maraun, D., and Kurths, J. (2004). Cross wavelet analysis: significance testing and pitfalls. *Nonlinear Process. Geophys.* 11, 505–514. doi: 10.5194/npg-11-505-2004

Ménard, F., Marsac, F., Bellier, E., and Cazelles, B. (2007). Climatic oscillations and tuna catch rates in the Indian Ocean: a wavelet approach to time series analysis. *Fish. Oceanogr.* 16, 95–104. doi: 10.1111/j.1365-2419.2006.00415.x

Meyer, J. L., Sale, M. J., Mulholland, P. J., and Poff, N. L. (1999). Impacts of climate change on aquatic ecosystem functioning and health 1. *JAWRA J. Am. Water Resour. Assoc.* 35, 1373–1386.

Meyers, S. D., Kelly, B. G., and O'Brien, J. J. (1993). An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of Yanai waves. *Mon. Weather Rev.* 121, 2858–2866. doi: 10.1175/1520-04931993121<2858:AITWAI<2.0.CO;2

Miller, T. H., Gallidabino, M. D., MacRae, J. I., Owen, S. F., Bury, N. R., and Barron, L. P. (2019). Prediction of bioconcentration factors in fish and invertebrates using machine learning. *Sci. Total Environ.* 648, 80–89. doi: 10.1016/j.scitotenv.2018.08.122

Miller, T. J., Herra, T., and Leggett, W. C. (1995). An individual-based analysis of the variability of eggs and their newly hatched larvae of Atlantic cod (*Gadus morhua*) on the Scotian Shelf. *Can. J. Fish. Aquat. Sci.* 52, 1083–1093. doi: 10.1139/f95-106

Myers, R. A., Hutchings, J. A., and Barrowman, N. J. (1997). Why do fish stocks collapse? The example of cod in Atlantic Canada. *Ecol. Appl.* 7, 91–106. doi: 10.2307/2269409

O'Brien, C. M., Fox, C. J., Planque, B., and Casey, J. (2000). Climate variability and North Sea cod. *Nature* 404:142. doi: 10.1038/35004654

Oehlert, A. M., and Swart, P. K. (2019). Rolling window regression of δ13C and δ18O values in carbonate sediments: implications for source and diagenesis. *Depos. Rec.* 5, 613–630. doi: 10.1002/dep2.88

Park, S. H., Kim, B., Kang, C. M., Chung, C. C., and Choi, J. W. (2018). "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture," in *Proceedngs of the 2018 IEEE Intelligent Vehicles Symposium (IV)*, (Piscataway, NJ: IEEE), 1672–1678. doi: 10.1109/IVS.2018.8500658

Parmesan, C., and Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421, 37–42. doi: 10.1038/nature01286

Pörtner, H. O., Berdal, B., Blust, R., Brix, O., Colosimo, A., De Wachter, B., et al. (2001). Climate induced temperature effects on growth performance, fecundity and recruitment in marine fish: developing a hypothesis for cause and effect relationships in Atlantic cod (*Gadus morhua*) and common eelpout (*Zoarces viviparus*). *Cont. Shelf Res.* 21, 1975–1997. doi: 10.1016/S0278-4343(01)00038-3

Rabiner, L., and Juang, B. (1986). "An introduction to hidden Markov models," in *Proceedings of the IEEE ASSP Magazine*, Vol. 3, (Piscataway, NJ: IEEE), 4–16. doi: 10.1109/MASSP.1986.1165342

Rahmstorf, S. (1995). Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* 378, 145–149. doi: 10.1038/378145a0

Ramponi, G., Protopapas, P., Brambilla, M., and Janssen, R. (2018). T-cgan: conditional generative adversarial network for data augmentation in noisy time series with irregular sampling. *arXiv* [Preprint]. arXiv:1811.08295

Rao, R. R., and Sivakumar, R. (2003). Seasonal variability of sea surface salinity and salt budget of the mixed layer of the north Indian Ocean. *J. Geophys. Res. Oceans* 108:3009. doi: 10.1029/2001JC000907

Rhif, M., Ben Abbes, A., Farah, I. R., Martínez, B., and Sang, Y. (2019). Wavelet transform application for/in non-stationary time-series analysis: a review. *Appl. Sci.* 9:1345. doi: 10.3390/app9071345

Richardson, A. J., and Schoeman, D. S. (2004). Climate impact on plankton ecosystems in the Northeast Atlantic. *Science* 305, 1609–1612. doi: 10.1126/science.1100958

Rose, G. A., and Leggett, W. C. (1989). Interactive effects of geophysically-forced sea temperatures and prey abundance on mesoscale coastal distributions of a marine predator, Atlantic cod (*Gadus morhua*). *Can. J. Fish. Aquat. Sci.* 46, 1904–1913. doi: 10.1139/f89-240

Rose, G. A., and O'Driscoll, R. (2002). Capelin are good for cod: can the northern stock rebuild without them? *ICES J. Mar. Sci.* 59, 1018–1026. doi: 10.1006/jmsc.2002.1252

Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., et al. (2008). Attributing physical and biological impacts to anthropogenic climate change. *Nature* 453, 353–357.

Sagheer, A., and Kotb, M. (2019). Unsupervised pre-training of a deep LStM-based stacked autoencoder for multivariate time series forecasting problems. *Sci. Rep.* 9:19038. doi: 10.1038/s41598-019-55320-6

Schwalme, K., and Chouinard, G. A. (1999). Seasonal dynamics in feeding, organ weights, and reproductive maturation of Atlantic cod (*Gadus morhua*) in the southern Gulf of St Lawrence. *ICES J. Mar. Sci.* 56, 303–319.

Scott, W. B., and Scott, M. G. (1988). *Atlantic Fishes of Canada: Canadian Bulletin of Fisheries and Aquatic Science, 219*. Toronto, Ont: The University of Toronto Press.

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, 347–404. doi: 10.1016/j.ocemod.2004.08.002

Sheppard, C. (2004). Sea surface temperature 1871-2099 in 14 cells around the United Kingdom. *Mar. Pollut. Bull.* 49, 12–16. doi: 10.1016/j.marpolbul.2004.05.011

Skagseth, Ø, Drinkwater, K. F., and Terrile, E. (2011). Wind-and buoyancy-induced transport of the Norwegian coastal current in the Barents Sea. *J. Geophys. Res. Oceans* 116:C08007. doi: 10.1029/2011JC006996

Skagseth, Ø, Furevik, T., Ingvaldsen, R., Loeng, H., Mork, K. A., Orvik, K. A., et al. (2008). "Volume and heat transports to the Arctic Ocean via the Norwegian and Barents Seas," in *Arctic–Subarctic Ocean Fluxes*, eds R. R. Dickson, J. Meincke, and P. Rhines (Dordrecht: Springer), 45–64.

Sswat, M. (2017). *Direct and Indirect Effects of Ocean Acidification on Early Life Stages of Atlantic Cod and Herring*. Doctoral dissertation. Kiel: Christian-Albrechts Universität Kiel.

Staveley, T. A., Jacoby, D. M., Perry, D., van der Meijs, F., Lagenfelt, I., Cremle, M., et al. (2019). Sea surface temperature dictates movement and habitat connectivity of Atlantic cod in a coastal fjord system. *Ecol. Evol.* 9, 9076–9086. doi: 10.1002/ece3.5453

Steel, E. A., and Lange, I. A. (2007). Using wavelet analysis to detect changes in water temperature regimes at multiple scales: effects of multi-purpose dams in the Willamette River basin. *River Res. Appl.* 23, 351–359. doi: 10.1002/rra.985

Stige, L. C., Ottersen, G., Brander, K., Chan, K. S., and Stenseth, N. C. (2006). Cod and climate: effect of the North Atlantic Oscillation on recruitment in the North Atlantic. *Mar. Ecol. Prog. Ser.* 325, 227–241. doi: 10.3354/meps325227

Su, C. W., Khan, K., Tao, R., Nicoleta-Claudia, M. (2019). Does geopolitical risk strengthen or depress oil prices and financial liquidity? evidence from Saudi Arabia. *Energy* 187:116003. doi: 10.1016/j.energy.2019.116003

Su, C. W., Yin, X. C., and Tao, R. (2018). How do housing prices affect consumption in China? New evidence from a continuous wavelet analysis. *PLoS One* 13:e0203140. doi: 10.1371/journal.pone.0203140

Sundby, S. (2000). Recruitment of Atlantic cod stocks in relation to temperature and advection of copepod populations. *Sarsia* 85, 277–298. doi: 10.1080/00364827.2000.10414580

Torrence, C., and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 79, 61–78. doi: 10.1175/1520-04771998079<0061:APGTWA<2.0.CO;2

Torrence, C., and Webster, P. J. (1999). Interdecadal changes in the ENSO-monsoon system. *J. Clim.* 12, 2679–2690. doi: 10.1175/1520-04421999012<2679:ICITEM<2.0.CO;2

Vacha, L., and Barunik, J. (2012). Co-movement of energy commodities revisited: evidence from wavelet coherence analysis. *Energy Econ.* 34, 241–247. doi: 10.1016/j.eneco.2011.10.007

Valdemarsen, J. W. (2001). Technological trends in capture fisheries. *Ocean Coast. Manag.* 44, 635–651. doi: 10.1016/S0964-5691(01)00073-4

Vikebø, F., Jørgensen, C., Kristiansen, T., and Fiksen, Ø (2007). Drift, growth, and survival of larval Northeast Arctic cod with simple rules of behaviour. *Mar. Ecol. Prog. Ser.* 347, 207–219. doi: 10.3354/meps06979

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

Wanless, S., Harris, M. P., Redman, P., and Speakman, J. R. (2005). Low energy values of fish as a probable cause of a major seabird breeding failure in the North Sea. *Mar. Ecol. Prog. Ser.* 294, 1–8. doi: 10.3354/meps294001

Wroblewski, J. S., Nolan, B. G., Rose, G. A., and deYoung, B. (2000). Response of individual shoaling Atlantic cod to ocean currents on the northeast Newfoundland Shelf. *Fish. Res.* 45, 51–59. doi: 10.1016/S0165-7836(99)00096-X

Yáñez, E., Plaza, F., Gutiérrez-Estrada, J. C., Rodríguez, N., Barbieri, M. A., Pulido-Calvo, I., et al. (2010). Anchovy (*Engraulis ringens*) and sardine (*Sardinops sagax*) abundance forecast off northern Chile: a multivariate ecosystemic neural network approach. *Prog. Oceanogr.* 87, 242–250. doi: 10.1016/j.pocean.2010.09.015

Yu, L. (2007). Global variations in oceanic evaporation (1958–2005): the role of the changing wind speed. *J. Clim.* 20, 5376–5390. doi: 10.1175/2007JCLI1714.1

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50, 159–175. doi: 10.1016/S0925-2312(01)00702-0

Zhang, Q., Wang, H., Dong, J., Zhong, G., and Sun, X. (2017). Prediction of sea surface temperature using long short-term memory. *IEEE Geosci. Remote Sens. Lett.* 14, 1745–1749. doi: 10.1109/LGRS.2017.2733548

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 09 June 2021
doi: 10.3389/fmars.2021.672477

Estimating Ocean Surface Currents With Machine Learning

Anirban Sinha[1]* *and Ryan Abernathey*[2]

[1] *Environmental Science and Engineering, California Institute of Technology, Pasadena, CA, United States,* [2] *Department of Earth and Environmental Sciences, Lamont Doherty Earth Observatory, Columbia University, Palisades, NY, United States*

Global surface currents are usually inferred from directly observed quantities like sea-surface height, wind stress by applying diagnostic balance relations (like geostrophy and Ekman flow), which provide a good approximation of the dynamics of slow, large-scale currents at large scales and low Rossby numbers. However, newer generation satellite altimeters (like the upcoming SWOT mission) will capture more of the high wavenumber variability associated with the unbalanced components, but the low temporal sampling can potentially lead to aliasing. Applying these balances directly may lead to an incorrect un-physical estimate of the surface flow. In this study we explore Machine Learning (ML) algorithms as an alternate route to infer surface currents from satellite observable quantities. We train our ML models with SSH, SST, and wind stress from available primitive equation ocean GCM simulation outputs as the inputs and make predictions of surface currents (u,v), which are then compared against the true GCM output. As a baseline example, we demonstrate that a linear regression model is ineffective at predicting velocities accurately beyond localized regions. In comparison, a relatively simple neural network (NN) can predict surface currents accurately over most of the global ocean, with lower mean squared errors than geostrophy + Ekman. Using a local stencil of neighboring grid points as additional input features, we can train the deep learning models to effectively "learn" spatial gradients and the physics of surface currents. By passing the stenciled variables through convolutional filters we can help the model learn spatial gradients much faster. Various training strategies are explored using systematic feature hold out and multiple combinations of point and stenciled input data fed through convolutional filters (2D/3D), to understand the effect of each input feature on the NN's ability to accurately represent surface flow. A model sensitivity analysis reveals that besides SSH, geographic information in some form is an essential ingredient required for making accurate predictions of surface currents with deep learning models.

Keywords: deep learning-artificial neural network, surface current balance, geostrophic balance, Ekman flow, regression, predictive modeling

OPEN ACCESS

Edited by:
Zhengguang Zhang,
Ocean University of China, China

Reviewed by:
Lei Zhou,
Shanghai Jiao Tong University, China
Chunyong Ma,
Ocean University of China, China

***Correspondence:**
Anirban Sinha
anirban@caltech.edu

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *25 February 2021*
Accepted: *03 May 2021*
Published: *09 June 2021*

Citation:
Sinha A and Abernathey R (2021) Estimating Ocean Surface Currents With Machine Learning. Front. Mar. Sci. 8:672477. doi: 10.3389/fmars.2021.672477

1. INTRODUCTION

The most reliable spatially continuous estimates of global surface currents in the ocean come from geostrophic balance applied to the sea surface height (SSH) field observed by satellite altimeters. For the most part, the dynamics of slow, large-scale currents (up to the mesoscale) are well-approximated by geostrophic balance, leading to a direct relationship between gradients

of SSH and near-surface currents. However, current meter observations for the past few decades and some of the newer generation ultra-high-resolution numerical model simulations indicate the presence of an energized submesoscale as well as high-frequency waves/tides at smaller spatial and temporal scales (Rocha et al., 2016). In addition, the next generation of satellite altimeters like the upcoming Surface Water and Ocean Topography (SWOT) mission (Morrow et al., 2018) is going to capture the ocean surface at a much higher spatial resolution, but with a low frequency repeat cycle (21 days). This presents unique challenges for the estimation of surface currents from SSH using traditional balances like geostrophy or Ekman. The high-wavenumber SSH variability is likely to be strongly aliased in the temporally sub-sampled data and may represent an entirely different, ageostrophic regime, where geostrophy might not be the best route to infer velocities. Motivated by this problem, we explore statistical models based on machine learning (ML) algorithms for inferring surface currents from satellite observable quantities like SSH, wind and temperature in this study. These algorithms can offer a potential alternative to the traditional physics-based models. We should point out that resolving the issues pertaining to spatio-temporal sampling and interpolation in satellite altimetry or the separation of balanced and unbalanced flows, while being important problems, are beyond the scope of our present study. Our goal is to examine whether we can extract more information about the surface flow from the spatial maps of these quantities and make more accurate predictions of surface currents with ML than we can with traditional balances.

The traditional method of calculating surface currents from sea surface height relies on the following physical principles. Assuming 2D flow and shallow water pressure, the momentum equation at the ocean surface can be written as:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + f \times \mathbf{u} = -g\nabla\eta + \mathbf{F} \tag{1}$$

where $\mathbf{F}$ is the frictional term due to wind stress. For a sufficiently low Rossby number (acceleration terms small), the leading-order balances are geostrophy and Ekman flow. The surface flow can be split into a geostrophic and an ageostrophic, Ekman component ($\mathbf{u} = \mathbf{u_g} + \mathbf{u_e}$), and this leading-order force balance can be written as

$$f \times \mathbf{u_g} = -g\nabla\eta \tag{2}$$

$$f \times \mathbf{u_e} = \mathbf{F} \tag{3}$$

Satellite altimetery products typically provide the sea surface height relative to the geoid (SSH, η), with tidally driven SSH signals removed (Traon and Morrow, 2001). Since geostrophic balance does not hold at the equator ($f \approx 0$), typically (Ducet et al., 2000), a higher order "equatorial geostrophic" treatment is used to compute velocities near the equator (Lagerloef et al., 1999), which is matched to the geostrophic regime away from the equator. Usually, the data-assimilative processing algorithms used to map along-track SSH observations to gridded maps (e.g., AVISO Ducet et al., 2000) also involve some form of temporal smoothing. The process of combining measurements from multiple satellites and filtering can also lead to spurious physical signals (Arbic et al., 2012) leading to exaggerated forward-cascades of energy.

In addition to the geostrophic velocities, some products like OSCAR (Ocean Surface Current Analysis Real Time, Bonjean and Lagerloef, 2002), or GEKCO (Geostrophic and Ekman Current Observatory, Sudre and Morrow, 2008; Sudre et al., 2013) provide an additional ageostrophic component due to Ekman flow. The Ekman velocity is related to friction, which in the upper layer of the ocean is provided by wind stress ($\tau = (\tau_x, \tau_y)$) and since the Coriolis parameter f changes sign at the equator, the functional relationship between velocity and wind stress is different between the two hemispheres. In the Northern Hemisphere the Ekman velocities can be derived as:

$$u_e = \frac{1}{\rho\sqrt{2A_z|f|}}(\tau_x + \tau_y) \tag{4}$$

$$v_e = \frac{1}{\rho\sqrt{2A_z|f|}}(-\tau_x + \tau_y) \tag{5}$$

And in the Southern Hemisphere as:

$$u_e = \frac{1}{\rho\sqrt{2A_z|f|}}(\tau_x - \tau_y) \tag{6}$$

$$v_e = \frac{1}{\rho\sqrt{2A_z|f|}}(\tau_x + \tau_y) \tag{7}$$

where A_z is the linear drag coefficient representing vertical eddy viscosity ($\tau = \rho A_z \frac{\partial \mathbf{u}}{\partial z}$). Alternatively we can write these equations in terms of the Ekman layer depth h_{Ek} which is related to the eddy viscosity A_z as:

$$h_{Ek} = \sqrt{\frac{2A_z}{f}} \tag{8}$$

Both of these quantities (A_z,h_{Ek}) are largely unknown for the global ocean and are estimated based on empirical multiple linear regression from Lagrangian surface drifters (Lagerloef et al., 1999; Sudre et al., 2013). Typical values of Ekman depth h_{Ek} in the ocean range from 10 to 40 m.

So geostrophy + Ekman is the essential underlying physical/dynamical "model" currently used for calculating surface currents from satellite observations. This procedure, combining observations with physical principles, represents a top-down approach A more bottom-up approach would be a data driven regression model that extracts information about empirical relationships from data. Recently, machine learning (ML) methods have grown in popularity and have been proposed for a wide range of problems in fluid dynamics: Reynolds-averaged turbulence models (Ling et al., 2016), detecting eddies from altimetric SSH fields (Lguensat et al., 2017), reconstructing subsurface flow-fields in the ocean from surface fields (Chapman and Charantonis, 2017; Bolton and Zanna, 2019), sub-gridscale

modeling of PDEs (Bar-Sinai et al., 2018), predicting the evolution of large spatio-temporally chaotic dynamical systems (Pathak et al., 2018), data-driven equation discovery (Zanna and Bolton, 2020), parameterizing unresolved processes, like convective systems in climate models (Gentine et al., 2018), or eddy momentum fluxes in ocean models (Bolton and Zanna, 2019), to name just a few examples.

In this study we aim to tackle a simpler problem than those cited above: training a ML model to "learn" the empirical relationships between the different observable quantities (sea surface height, wind stress, etc.) and surface currents (u, v). The hypothesis to be tested is the following: Can we use machine learning to provide surface current estimates more accurately than geostrophy + Ekman balance? The motivation for doing this exercise is 2-fold:

1. It will help us understand how machine learning can be applied in the context of traditional physics-based theories. ML is often criticised as a "black box." But can we use ML to complement our physical understanding? This present problem serves as a good test-bed since the corresponding physical model is straightforward and well-understood.
2. It may be of practical value when SWOT mission launches.

While statistical models can often be difficult to explain due to lack of simple intuitive physical interpretations, several recent publications (Ling et al., 2016; Gentine et al., 2018; Bolton and Zanna, 2019; Zanna and Bolton, 2020) have demonstrated that data-driven approaches, used concurrently with physics-based models can offer various computational advantages over traditional methods, while still respecting physical principles. For our problem, which is much simpler in terms of its scope, we aim to mitigate the so called "black-box" ness of statistical models in general with physically motivated choices about inputs and training strategies, to ensure results that are physically meaningful. In this study, we mainly explore two types of regression models (multiple linear regression and artificial neural networks) as potential alternative approaches for predicting surface currents, using data from a primitive equation global general circulation model, and discuss their relative strengths and weaknesses. We see this work as a stepping stone to more complex applications of ML to ocean remote sensing of ocean surface currents.

This paper is organized as follows. In section 2, we introduce the dataset that was used, the framework of the problem and identify the key variables that are required for training a statistical model to predict surface currents. In section 3 we describe numerical evaluation procedure for baseline physics-based model that we are hoping to match/beat. In sections 4 and 5 we discuss the statistical models that we used. We start with the simplest statistical model—linear regression in section 4 before moving on to more advanced methods like neural networks in section 5. In section 6 we compare the results from the different models. In section 7 we summarize the findings, discuss some of the shortcomings of the present approach, propose some solutions as well as outline some of the future goals for this project.

2. DATASET AND INPUT FEATURES

To focus on the physical problem of relating currents to surface quantities, rather than the observational problems of spatio-temporal sampling and instrument noise, we choose to analyze a high-resolution global general circulation model (GCM), which provides a fully sampled, noise-free realization of the ocean state. The dataset used for this present study is the surface fields from the ocean component of the Community Earth System Model (CESM), called the Parallel Ocean Program (POP) simulation (Smith et al., 2010) which has a $\approx 0.1^\circ$ horizontal resolution, with daily-averaged outputs available for the surface fields. The model employs a B-grid (scalars at cell centers, vectors at cell corners) for the horizontal discretization and a three-time-level second-order-accurate modified leap-frog scheme for stepping forward in time. The model solves the primitive equations of motion, which, for the surface flow, are essentially (1). Further details about the model physics and simulations can be found in Small et al. (2014) and Uchida et al. (2017). We selected this particular model simulation because of the long time record of available data (∼40 years), although, in retrospect, we found that all our ML models can be trained completely with just a few days of output!

A key choice in any ML application is the choice of features, or inputs, to the model. In this paper, we experiment with a range of different feature combinations; seeing which features are most useful for estimating currents is indeed one of our aims. The features we choose are all quantities that are observable from satellites: SSH, surface wind stress (τ_x and τ_y), sea-surface temperature (SST, θ) and sea-surface Salinity (SSS). Our choice of features is also motivated by the traditional physics-based model: the same information that goes into the physics-based model should also prove useful to the ML model. Just like the physics-based model, all the ML models we consider are pointwise, local models: the goal is to predict the 2D velocity vector u, v at each point, using data from at or around that point.

Beyond these observable physical quantities, we also need to provide the models with geographic information about the location and spacing between the neighboring points. In the physics-based model, geography enters in two places: (1) in the Coriolis parameter f, and (2) in the grid spacing dx and dx, which varies over the model domain. Geographic information can be provided to the statistical models in a few different ways. The first method involves providing the same kind of spatial information that is provided to the physical models, i.e., f and local grid spacings—dx and dy. We can also encode geographic information (lat, lon) in our input features, using a coordinate transformation of the form:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} sin(lat) \\ sin(lon) \cdot cos(lat) \\ -cos(lon) \cdot cos(lat) \end{bmatrix} \tag{9}$$

to transform the spherical polar lat-lon coordinate into a homogeneous three dimensional coordinate (Gregor et al., 2017). This transformation gives the 3D position of each point in Euclidean space, rather than the geometrically warped lat/lon space (which has a singularity at the poles and a discontinuity at

the dateline). Note that one of the coordinates—X, that comes out of this kind of coordinate transformation, is functionally the same as the Coriolis parameter (f) normalized by 2Ω (Ω = Earth's rotation). Therefore, we will use X as proxy for f for all the statistical models throughout this study. We also explored another approach where the only geographic information provided to the models is $X\ (= \frac{f}{2\Omega})$.

Since geostrophic balance involves spatial derivatives, it is not sufficient to simply provide SSH and the local coordinates pointwise. In order to compute derivatives, we also need the SSH of the surrounding grid points as a local stencil around each grid point. The approach we used for providing this local stencil is motivated by the horizontal discretization of the POP model. Horizontal derivatives of scalars (like SSH) on the B grid requires four cell centers. At every timestep, each variable of the The 1° POP model ouput has $3,600 \times 2,400$ data points (minus the land mask). We can simply rearrange each variable as a $1,800 \times 1,200 \times 2 \times 2$ dataset or split it into four variables each with $1,800 \times 1,200$ data points, corresponding to the four grid cells required for taking spatial derivatives. The variables that require a spatial stencil for physical models, we will refer to as the stencil inputs. For the variables for which we do not need spatial derivatives for (like wind stress), we can simply use every alternate grid point resulting in a dataset of size $1,800 \times 1,200$. We will refer to these variables as point inputs. For the purpose of the statistical models the inputs need to be flattened and have all the land points removed. This means that each input variable has a shape of either $N \times 2 \times 2$ or N depending on whether or not a spatial stencil is used (where $N = 1,800 \times 1,200-$ the points that fall over land). Alternatively we can think of the stencilled variable as four features of length N. This kind of stencil essentially coarsens the resolution of the targets, and point variables. Similarly we can also construct a three point time stencil, by providing the values at preceding and succeeding time steps as additional inputs so that each variable that is stencilled in space and time has a shape of $N \times 2 \times 2 \times 3$ (or 12 features of length N). This data preparation leads to 10 potential features (for some of which we will use a stencil, which further expands the feature vector space) for predicting u, v at each point : τ_x, τ_y, SSH (η), SST (θ), SSS (S), the three transformed coordinates (X,Y,Z) and the local grid spacings (dx and dy).

For building any statistical/ML model, we need to split the dataset into two main parts, i.e., training and testing. For the purpose of training our machine learning models, the first step involves extracting the above mentioned variables from the GCM output as the input features and the GCM output surface velocities u, v as targets for the ML model. The data extracted from the GCM output for a certain date (or range of dates) is then used to fit the model parameters. This part of the dataset is called the training dataset. During training, the model minimizes a chosen cost function (we used mean absolute error for our experiments, but using mean squared error produced very similar results) and typically involves a few passes through this section of dataset. The trained models are then used to make predictions of u, v for a different date (or range of dates) where the model only receives the input variables. The model predictions are evaluated by comparing with the true (GCM output) velocity fields for that particular date (date range). This part of the dataset, which the model has not seen during training, that is used to evaluate model predictions is called the test dataset.

3. BASELINE PHYSICS-BASED MODEL: GEOSTROPHY + EKMAN

The two components of the physics-based model used as the baseline for our ML models are geostrophy and Ekman flow. In this section we describe how these two components are numerically evaluated for our dataset. For the sake of fair comparison, we evaluate the geostrophic and Ekman velocities from the same features that are provided to the regression models. With the POP model's horizontal discretization, finite-difference horizontal derivatives and averages are defined as (Smith et al., 2010):

$$\psi_x = \left[\psi(x + \Delta_x/2) - \psi(x + \Delta_x/2)\right]/\Delta_x \tag{10}$$

$$\overline{\psi^x} = \left[\psi(x + \Delta_x/2) + \psi(x + \Delta_x/2)\right]/2 \tag{11}$$

With the data preparation and stencil approach described in the previous section, η now has a shape of $N \times 2 \times 2$ and the f, u, v, dx, dy are all variables of length N. Following (2) the geostrophic velocities (u_g^j, v_g^j) are calculated on the stencil as:

$$v_g^j = g/f^j \left[\eta_i(1,1) + \eta_i(0,1) - \eta^j(1,0) - \eta^j(0,0)\right]/4dx^j \tag{12}$$

$$u_g^j = -g/f^j \left[\eta^j(1,1) + \eta^j(1,0) - \eta^j(0,1) - \eta^j(0,0)\right]/4dy^j \tag{13}$$

where $j \in [1, N]$. Similarly the Ekman velocity is calculated numerically from the τ_x^j, τ_y^j, and f^j as

$$u_e^j = \begin{cases} \frac{1}{\rho\sqrt{2A_z|f^j|}}(\tau_x^j + \tau_y^j), & \text{if } f^j > 0 \\ \frac{1}{\rho\sqrt{2A_z|f^j|}}(\tau_x^j - \tau_y^j), & \text{if } f^j < 0 \end{cases}$$

$$v_e^j = \begin{cases} \frac{1}{\rho\sqrt{2A_z|f^j|}}(-\tau_x^j + \tau_y^j, & \text{if } f^j > 0 \\ \frac{1}{\rho\sqrt{2A_z|f^j|}}(\tau_x^j + \tau_y^j), & \text{if } f^j < 0 \end{cases} \tag{14}$$

For calculating the Ekman velocity, we used constant values for vertical diffusivity ($A_z = 8 \times 10^{-3} m^2/s$) and density of water at the surface ($\rho = 1,027 kg/m^3$). It should be noted that both these quantities vary both spatially and temporally in the real ocean. For the vertical diffusivity we came up with this estimate by solving for A_z that provides the best fit between zonal mean ($(u, v)_{true} - (u, v)_g$) and $(u, v)_e$. In the CESM high res POP simulations, the parameterized vertical diffusivity was capped around 100 cm^2/s (Smith et al., 2010). For plotting spatial maps for both the physics based model predictions as well as the statistical model predictions, the velocity fields are then reshaped into $1,800 \times 1,200$ arrays, after inserting the appropriate land masks.

4. MULTIPLE LINEAR REGRESSION MODEL

The simplest of all statistical prediction models is essentially multiple linear regression, where an output or target is represented as some linear combination of the inputs. The input is characterized by a feature vector $\mathbf{x_i^j}$ where $\mathbf{i} \in [1, n_f]$; $\mathbf{j} \in [1, N]$, N being the number of samples, and n_f being the number of features. We can now write the linear regression problem as $\mathbf{U^j} = \mathbf{x_i^{jT}} \cdot \beta_\mathbf{i} + \delta^\mathbf{j}$. where $\beta_\mathbf{i}$ are the coefficients or weight vector. For our regression problem, the input features are wind stress, sea surface height and the three dimensional transformed coordinates. Of those features, η, X, Y, Z are stencil inputs (meaning four input columns per feature) and τ_x, τ_y are the point inputs, resulting in a total of 18 input features. The aim therefore is to find the coefficients $\beta_\mathbf{i}$ that minimize the loss (error) represented by $\delta^\mathbf{j}$ for a training set of $\mathbf{x_i^j}$ and $\mathbf{U^j}$ ($\mathbf{x^j_{i}}_{train}, \mathbf{U}_{train}$) and use these coefficients for a test set of $\mathbf{x_i^j}$ ($\mathbf{x^j_{i}}_{test}$) to make predictions for $\mathbf{U^j}$ ($\mathbf{U^j}_{pred}$). For implementing linear regression model as well as the deep learning models discussed subsequently in this study, we use the Python library Keras (https://keras.io) (Chollet, 2015), a high-level wrapper around TensorFlow (http://www.tensorflow.org).

Linear regression can be performed in one of two different ways

- The matrix method or Normal equation method (where we solve for the coefficients β that minimize the squared error $\|\delta\|^2 = \|\mathbf{U} - \mathbf{X^T} \cdot \beta\|^2$ and involves computing the pseudo-inverse of $\mathbf{X^T} \cdot \mathbf{X}$).
- A stochastic gradient descent (SGD) method (which represents a more general procedure that can be used for different regression algorithms with different choices for optimizers and is more scalable for larger datasets).

The normal-equation method is less computationally tractable for large datasets (large number of samples) since it requires loading the full dataset into memory for calculating the pseudoinverse of $\mathbf{x_i^{jT}} \cdot \mathbf{x_i^j}$, whereas the SGD method works well even for large datasets, but requires tuning of the learning rate. Due to the versatility offered by the gradient descent method we used that for performing the linear regression although the normal equation method also produced similar results. The essential goal for any regression problem is to minimize a predetermined cost/loss function (which for our experiments we chose as the mean absolute error):

$$J = MAE = \overline{\left(|u_{pred} - u_{true}| + |v_{pred} - v_{true}|\right)} \qquad (15)$$

where the overbar denotes the average over all samples. **Figure 1A** shows a schematic of the linear regression model. The number of trainable parameters for our example with 18 inputs and two outputs is 38 (18 × 2 weights + 2 biases). For the sake of consistency, we use the same optimizer (Adam; Kingma and Ba, 2017) and loss function (Mean absolute error, MAE) for this as well as all the subsequent models discussed here. All models are trained on 1 day of GCM output data and we use the same date of model output as the training data for all models.

5. DEEP LEARNING: ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (or neural networks for short) are machine learning algorithms that are loosely modeled after the neuronal structure of a biological brain but on a much smaller scale. A neural network is composed of layers of connected units or nodes called artificial neurons (LeCun et al., 2015; Nielsen, 2015; Goodfellow et al., 2016) that combine input from the data with a set of weights and passes the sum through the node's activation function along with a bias term, to the subsequent set of nodes, to determine to what extent that signal progresses through the network and how it affects the ultimate outcome. Neural nets are typically "feed-forward," meaning that data moves through them in only one direction. A layer is called densely connected when each node in that layer is connected to every node in the layers immediately above and below it. Deep learning, or deep neural networks is the name used for "stacked neural networks"—i.e., networks composed of several layers.

Our neural network code was written using the Python library Keras (https://keras.io) (Chollet, 2015), a high-level wrapper around TensorFlow (http://www.tensorflow.org). The feed-forward NNs consist of interconnected layers, each of which have a certain number of nodes. The first layer is the input layer, which in our case is a stacked vector containing the input variables just like in the linear regression example above. The last layer is the output layer, which is a stacked vector of the two outputs (U,V). All layers in between are called hidden layers. The activation function, i.e., the function acting on each node – is a weighted sum of the activations in all nodes of the previous layer plus a bias term, passed through a non-linear activation function. For our study, we used the Rectified Linear Unit (ReLU) as an activation function. The output layer is purely linear without an activation function. Training a NN means optimizing the weight matrices and bias vectors to minimize a loss function—in our case the MAE—between the NN predictions and the true values of (u, v).

The model reduces the loss, by computing the gradient of the loss function with respect to all weights and biases using a backpropagation algorithm, followed by stepping down the gradient—using stochastic gradient descent (SGD). In particular we use a version of SGD called Adam (Kingma and Ba, 2014, 2017). Although most neural network strategies involve normalizing the input variables, we did not use any normalization, since the normalization factors would be largely dependent on the choice of domain/ocean basin, given that the dynamical parameters (like SSH and wind stress) vary widely across the different ocean basins. Instead we wanted the NN to be generalizable across the whole ocean.

We construct a three-hidden-layer neural network to replace the linear regression model described in the previous section. A schematic model architecture for the neural network is presented in **Figure 1B**. Using the same basic model architecture, we

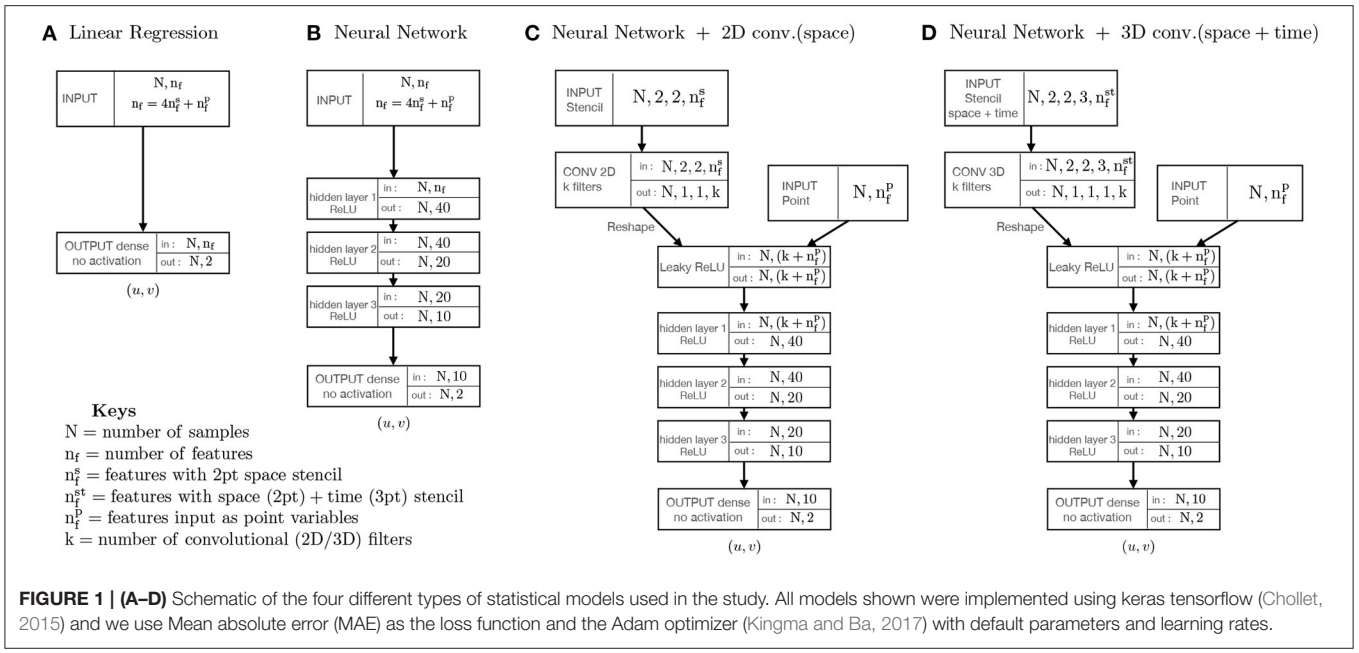


FIGURE 1 | (A–D) Schematic of the four different types of statistical models used in the study. All models shown were implemented using keras tensorflow (Chollet, 2015) and we use Mean absolute error (MAE) as the loss function and the Adam optimizer (Kingma and Ba, 2017) with default parameters and learning rates.

train three NNs on the same three subdomains (Gulf Stream, Kuroshio, ACC) along with one which is trained on the global ocean. Everything including batch size, the training data, the targets, the input features and the number of epochs the model is trained for in each region is kept exactly the same as what we used for the linear regression examples. The only thing that we changed is the model, where instead of one layer with no activation we now have three hidden layers with a total of 1,812 trainable parameters.

5.1. Neural Networks With Convolutional Filters

In section 2 we explained how we can use the local 2×2 stencil to expand the feature vector space by a factor of 4. We can further expand the feature vector space by passing all the stenciled input features through k convolutional filters of shape 2×2. If $k > 4n_f^s$ where n_f^s is the number of input features with a stencil, we end up with more input features that goes into the NN than before. There is very little functional difference between this kind of training approach and the one discussed previously, except that we end up with more trainable parameters, which we can potentially use to extract even more information from the data. We should point out that this is technically not the same as convolutional neural networks (CNN), where the convolutional layers serve to reduce the feature vector space without losing information. This is particularly important for problems like image classification where it is needed to scale down large image datasets without losing feature information. Typically in a CNN, the inputs would be in the form of an image or a set of stacked images on which multiple convolutional filters of varying sizes could be applied (followed by max-pooling layers) that effectively shrink the input size, before passing it on to the hidden layers, and the size of the convolutional filters determine the size of the stencil[1]. Whereas in this approach, we use convolutional filters (without max-pooling) to achieve the opposite effect, i.e., to expand the feature vector space from $N \times 2 \times 2 \times n_f^s$ to $N \times k$ (where k is always chosen to be $> 4n_f^s$). The reshaping of the input variables in the pre-processing stage fixes the stencil size, before the data is fed into the model.

A schematic of this subcategory of neural network is shown in **Figure 1C**. After applying the convolutional filter and passing it through a reshape layer in keras the point inputs and filtered stencil inputs are passed through a Leaky ReLU before being fed into a similar three-hidden layer NN framework as described before. Using a similar procedure, we can also apply k 3D convolutional filters of shape $2 \times 2 \times 3$ on the time and space stenciled inputs to effectively end up with k input features of length N for the stencil variables (**Figure 1D**). The goal with the time stenciled input being to potentially learn time derivatives and explore how the tendencies can affect the NN projections. In hindsight, this data set is probably not be the most suited for this kind of approach since the variables we used as input features are daily averaged and any fast-time scale/tendency effects that we hoped to capture from multiple snapshots of the same variable are probably filtered out by the time averaging. These two approaches are virtually identical with slightly different preprocessing of the input data.

[1]For an input of size 500×500 for example, one can apply convolutional filters as small as 2×2 or as big as the entire image. However, since CNNs and other computer vision approaches rely on the property that nearby pixels are more strongly correlated than more distant pixels (Bishop, 2006), larger filters can be useful for reduction of data volume, but they often result in degradation of data quality and prediction accuracy, due to inclusion of non-local effects.

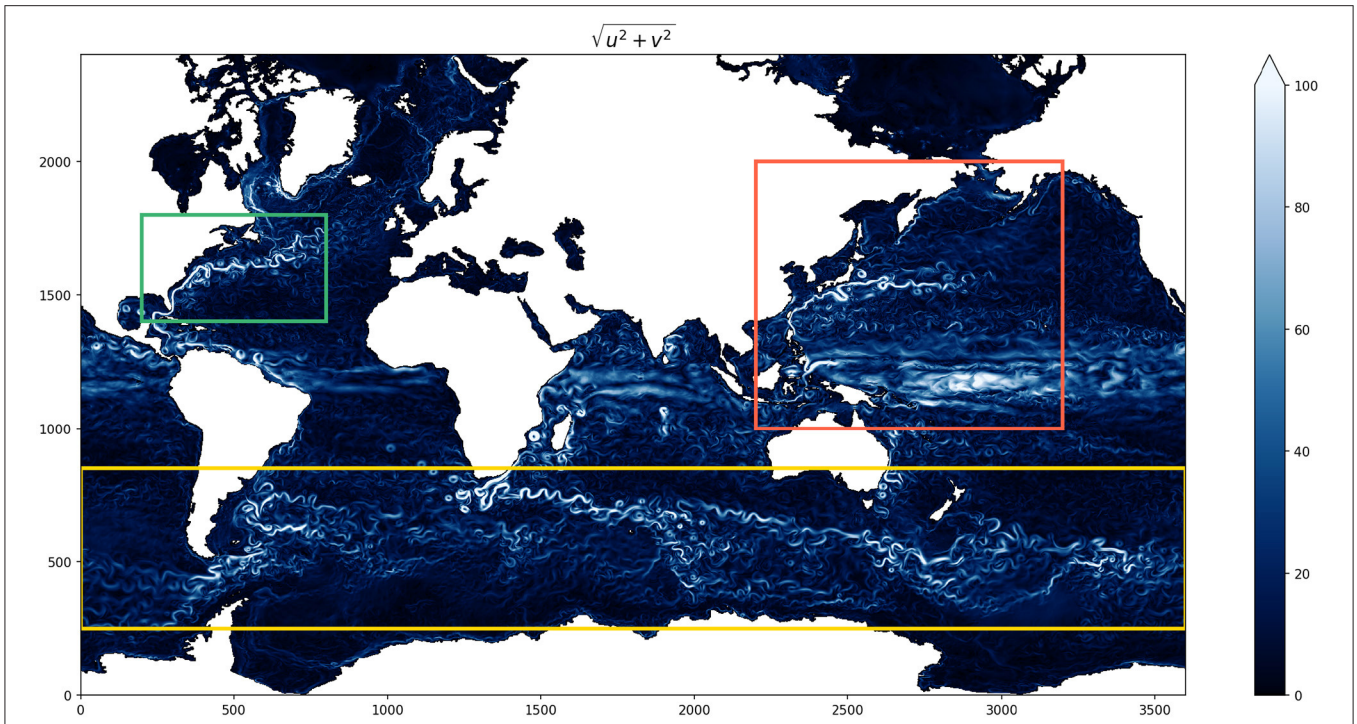


FIGURE 2 | Snapshot of the surface speed in the CESM POP model with the three boxes in different colors indicating the training regions chosen for the different regression models. The green box is chosen as the Gulf stream region, the red box is Kuroshio and the yellow box represents the Southern Ocean/Antarctic circumpolar current (ACC). The Kuroshio region extends slightly south of the equator to include the equatorial jets in the domain and to test the models' ability to generalize to large variations in f.

6. RESULTS

We start by splitting the global ocean into three boxes to zoom into three distinct regions of dynamical importance in oceanography, namely the Gulf stream, Kuroshio, and Southern ocean/Antarctic circumpolar current (ACC). The Kuroshio region is chosen to extend south of the equator to include the equatorial jets as well as to test whether the models can generalize to large variations in f. The daily averaged GCM output surface speed on a particular reference date, with the three regions (marked by three different colored boxes) is shown in **Figure 2**. We then train three different linear regression models with training data from these three sub-domains. We also trained a linear regression model for the whole globe using the same model architecture. During training, the models are fed a shuffled batch of the training data with 32 samples in each batch and the loss (MAE) is computed for the batch. For the linear regression model as well as for all the neural networks discussed in this study we present here we kept the batch size constant. Changing the batch size does not significantly alter the loss at the end of training, but smaller batch sizes generally help the model learn faster. The different models, the number of epochs (an epoch is defined as one pass through the training dataset) used for each, and losses at the end of training and during evaluation against a test dataset are summarized in **Table 1**. The evolution of model loss function during training for the 3 different models are presented in **Figure 3**. Linear regression is shown in the darker colors. The big jumps in the loss function correspond to the end of an epoch. We plot the models' training progress in the Gulf Stream region for 8 epochs, and for 5 epochs on the Kuroshio and ACC regions. The trained models are then evaluated for a test dataset (which the model has not seen, GCM output from a different point in time) and the evaluation loss is plotted as the horizontal dashed lines. The linear regression model trained on the whole globe is also evaluated for each subdomain (gulf stream, Kuroshio, ACC) and the global model evaluation loss is plotted as the dotted line. Comparing the model losses in the three different sections, we find that the linear regression model performs the most poorly for the Kuroshio region (i.e., the subdomain with the most variation in f). The model does progressively better for the gulf stream and the ACC in terms of MAE, where the variations in f are relatively smaller in comparison. However, the root mean squared error of predicted velocities is still quite large in all these regions (second panels of **Figures 4**–**6**). The linear regression model trained on the global ocean does even worse during evaluation. Since geostrophy relies on non-linear combination of the Coriolis parameter (f) with the spatial gradients, linear regression is ineffective at predicting velocities beyond localized regions with small variation of f or little mesoscale activity. This shows that a linear model fails to accurately represent surface currents in any region that includes significant variation in the Coriolis parameter f. Even in regions

far enough from the equator such that the variation in f is not significant (like the gulf stream or ACC), the performance of such a linear model does not improve with more training examples and/or starts overfitting. We also show that a lower MAE during training does not necessarily guarantee that the model is picking up on the small scale fluctuations in velocity, as can be seen from the relatively large squared errors especially in and around high surface current regions (**Figures 4–6**). We suspect that this failure is largely due to the fact that the linear model is trying to fit the velocities as a linear combination of the different features, whereas realistic surface current predictions should be based on non-linear combinations of features.

Neural networks on the other hand, due to the presence of multiple dense interconnected layers can be effectively used to extract these non-linear relationships in the data. Just like we did with the linear regression model, we tracked the evolution of the loss function as the model scans through batches of input data over multiple epochs (**Figure 3**, lighter colored lines in all panels). As we can see, in comparison to the linear regression model, the NNs perform significantly better at reducing the loss in all the ocean regions. What is even more striking is that the NN trained on the globe (dashed line) consistently outperforms the local models, predicting surface currents with lower MAE/MSE than the models trained on the local subdomains. This is especially noticeable for the Kuroshio region (**Figure 3**, second panel), where the NN trained on the globe manages to get the signature of the equatorial currents better than the NN trained specifically in that region (compare panels 3 and 4 of **Figure 5**) and gets the absolute error down to $\approx$5 cm/s. This shows, that in comparison to the linear model the neural network actually manages to learn the physics better when it receives a more spatially diverse input data, and is therefore more generalizable. Even though the linear regression models all manage to get the loss down to comparable magnitude, looking at the spatial plot of the predicted squared error. **Figures 4–6** gives us an idea how poorly it does at actually learning the physics of surface currents. In comparison, even a relatively shallow three-hidden-layer neural network performs significantly better with very few localized hotspots of large errors. This is to be expected since the largest order balance, i.e., geostrophy relies on non-linear combination of the Coriolis parameter (f) with the spatial gradients.

In **Figure 7** we plot the joint histogram of the zonal and meridional velocity predictions against the true (GCM output) values for the physical model, linear regression model (trained on the local subdomain) and the locally and globally trained neural networks in the ACC sector. From these joint histograms, it is obvious that the physical model, the local and global neural networks all predict velocities that are extremely well-correlated with the true velocities in this region. In addition the root mean squared (rms) errors normalized by the rms velocities are also very well-correlated between the physical model and neural network predictions. This provides us with reasonable confidence that the model is indeed learning the physics of surface geostrophy and Ekman flow.

We also plotted the squared errors in predicted velocity form the physical model (geostrophy + Ekman) and the local Rossby number (expressed as the ratio of the relative vorticity $\zeta = v_x - u_y$, to the planetary vorticity f) in the three domains (Gulf Stream—**Figure 4**; Kuroshio—**Figure 5**; and the ACC—**Figure 6**). It is interesting to note that the localized regions in large root squared errors in both the neural network and physical models coincide with regions where the local Rossby number is high. High Rossby numbers indicate unbalanced flow and the specific regions where we see high Rossby numbers are typically associated with heightened submesoscale activity. We speculate that the prediction errors in these locations are due to the NN's inability to capture higher order balances (e.g., gradient wind, cyclostrophic balance) that are necessary to fully capture the small scale variability associated with these motions and close the momentum budget.

The NN also generally predicts weaker velocities near the equator where the true values of the surface currents are quite large (due to strong equatorial jets). This can lead to large errors for the global mean, which get magnified when the differences are squared. However, we know that geostrophic and Ekman balance also doesn't hold near the equator. A fairer comparison would therefore involve masking out the near equatorial region ($5^\circ N - 5^\circ S$) for both the statistical model (i.e., NN predictions) as well as for the physical model ($geo + ekman$).

6.1. Model Dependence on Choice of Input Features

We then trained these NNs with varying combinations of input features to explore how the choice of input features can influence the model training rate and loss. Feeding the NN models varying combination of input features, either as stencilled or as point variables and by selectively holding out specific features for each training case allowed us to assess the relative importance of each physical input variable for the neural network's predictive capability. The different models with their corresponding input features and the number of trainable parameters for each case are summarized in **Table 2**. As with all previous examples, we chose mean absolute error as the loss function for all these experiments. We performed a few training exercises using the mean squared error instead and did not notice any significant difference. For models numbered 1–13, we used a two point space-stencil and for models 1t–10t, in addition to a stencil in space, we provide a three point time stencil with the intention of helping the neural network "learn" time derivatives. The different experiments listed in **Table 2**, can broadly be categorized into six groups based on their input features. In group 1, is model 1, where the model only sees η (stencil) and wind stress, τ (point) as input features. No spatial information is provided. In the second category, we have models that receive η (stencil) and spatial information $\mathbf{X}$ in some form, but no wind stress. This includes models 2, 5t, and 7t. The third category describes models that receive η, θ (stencil) and spatial information $\mathbf{X}$ and no wind stress and includes models 3, 6t, and 8t. The fourth category describes models that receive SSH (η), spatial information ($\mathbf{X}$) and wind stress (τ) but no SST and includes models 4, 6, 7, 10, 1t, and 3t. The fifth category of models receive SSH (η), SST (θ), spacial information ($\mathbf{X}$), and wind stress and the only input feature these models don't receive in any form is sea surface

TABLE 1 | Table summarizing model errors from the physics based model (geostrophy + Ekman flow) and the two types of regression models—linear regression and neural network (**Figures 1A,B**).

Model (training region)	Number of trainable parameters	Epochs	MAE (train) (cm/s)	MAE (eval) GS (cm/s)	MAE (eval) Kuroshio (cm/s)	MAE (eval) ACC (cm/s)
LR (Gulf Stream)	38	8	10.7	11.4	–	–
NN (GS)	1,812	8	2.3	3.7	–	–
LR (Kuroshio)	38	5	12.9	–	13.4	–
NN (Kuroshio)	1,812	5	5.8	–	7.0	–
LR (ACC)	38	5	7.5	–	–	7.5
NN (ACC)	1,812	5	1.9	–	–	4.5
NN (global)	1,812	4	3.0	2.4	5.1	1.8
geo + *Ek* (global)	–	–	–	6.1	29.2	3.9

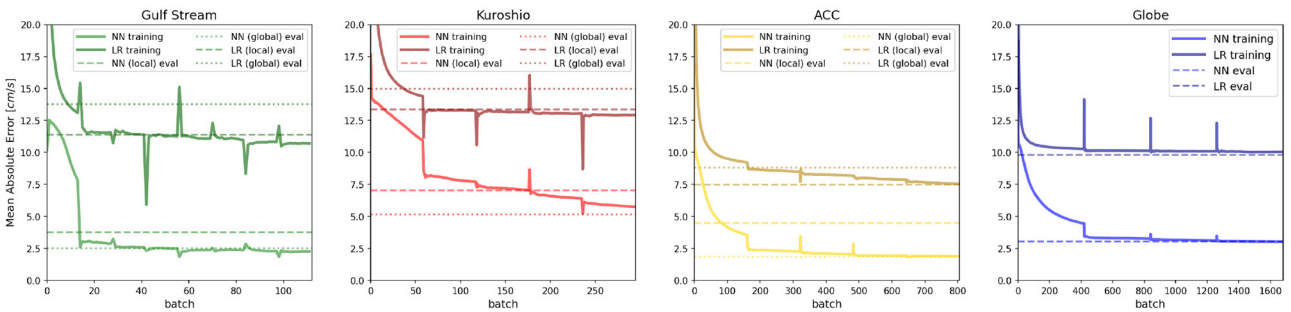


FIGURE 3 | Evolution of the loss function (mean absolute error; MAE) for Neural Networks and Linear regression models during training. Horizontal lines of the corresponding color denote the MAE for the model when evaluated at a different time snapshot. Dashed lines denote the evaluated (test data) MAE for the local model and dotted lines denote that for the model trained on the globe.

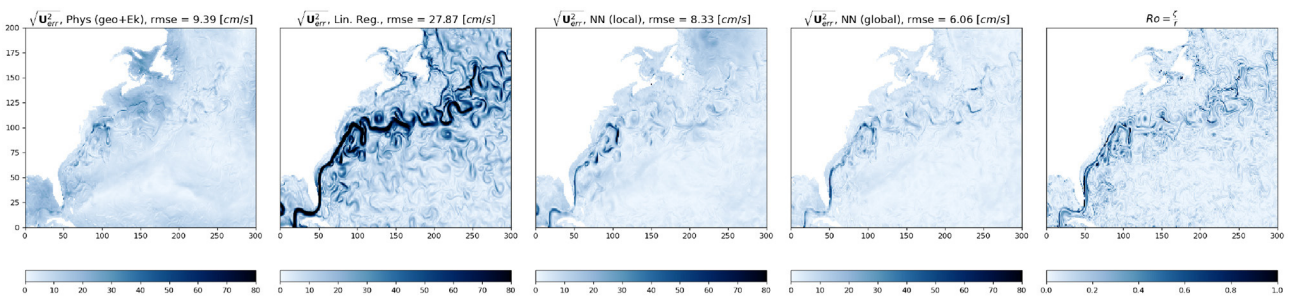


FIGURE 4 | Snapshot of model predicted root square errors for the physics based model (left) and the three different regression models—Linear regression (second from left), neural network, trained on this local domain (third panel) and neural network, trained on the globe (4th panel) compared side by side with the local Rossby Number (Ro, right panel) in the Gulf Stream region indicated by the green box in **Figure 2**.

salinity (S). This includes models 5, 8, 9, 11, 2t, 4t. The sixth and final category represents models tat receive all the input features $(\eta, \theta, S, \mathbf{X}, \tau)$ in some form or another and includes models 13, 9t, and 10t.

As mentioned previously, spatial information is provided in one of three ways, (a) in the form of three dimensional transformed coordinates (X, Y, Z), (b) just the Coriolis parameter (X here serves as a proxy for the Coriolis parameter) and (c) with both the Coriolis parameter and local dx and dy values. Barring a few examples (models 10, 11) windstress is always provided as a point variable and apart from models 6, 7, 8, 9, none of the models receive a stencil in the spatial coordinates. We also trained a few models without SSH as an input feature, but the loss in all these cases was much larger than those shown here (>50 cm/s) and the NNs fail to pick up any functional dependence on the input features. Those cases are therefore not presented. Each of these models are trained for 4 Epochs on the same day of data (or 3 consecutive days centered around that date for the time stencilled cases).

In **Figure 8** we summarize the findings from these experiments by plotting the rms error for all the model predictions along with the rms error for the physical model

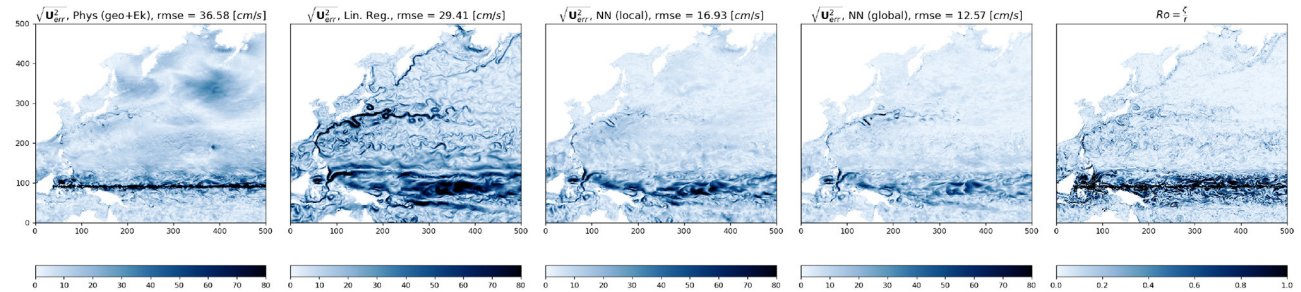


FIGURE 5 | Snapshot of model predicted root square errors for the physics based model (left) and the three different regression models—Linear regression (second from left), neural network, trained on this local domain (third panel) and neural network, trained on the globe (4th panel) compared side by side with the local Rossby Number (Ro, right panel) in the Kuroshio region indicated by the red box in **Figure 2**. Note the large errors in all the model predictions near the equator.

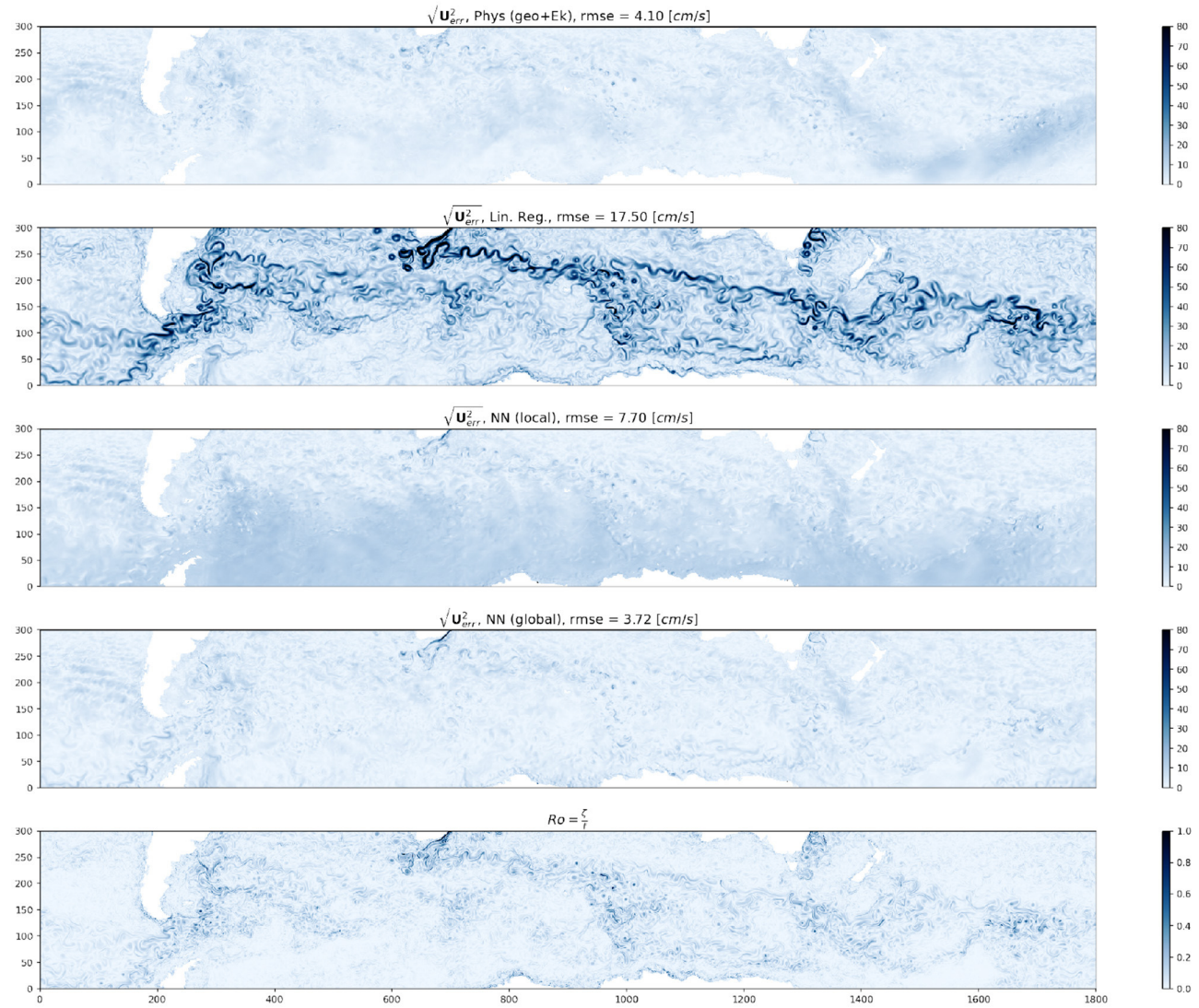


FIGURE 6 | Snapshot of model predicted root square errors for the physics based model (top) and the three different regression models—Linear regression (second panel), neural network, trained on this local domain (third panel) and neural network, trained on the globe (4th panel) compared side by side with the local Rossby Number (Ro, bottom panel) in the Southern Ocean/Antarctic circumpolar current region indicated by the yellow box in **Figure 2**.

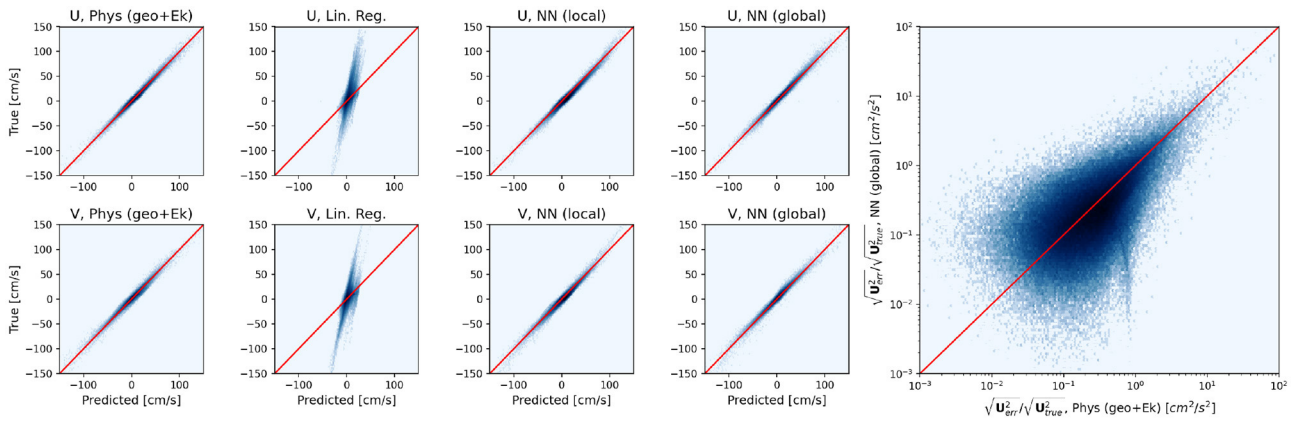


FIGURE 7 | Scatterplot of true v predicted zonal and meridional velocities for the different physical and regression models (eight panels on the left) in the ACC region. The right panel shows the scatterplot of the root mean squared errors (normalized by the root mean square velocities) for the physical and neural network model predictions.

TABLE 2 | Table summarizing the different CNNs and the training strategies explored.

Model No.	Stencil in space (2s)	Stencil in time (3t)	Stencil variables	Point variables	Number of trainable parameters
1	✓	×	η	τ_x, τ_y	4,772
2	✓	×	η	$X\ (=\frac{f}{2\Omega})$	4,732
3	✓	×	η, θ	X	5,052
4	✓	×	η	X, τ_x, τ_y	4,812
5	✓	×	η, θ	X, τ_x, τ_y	5,132
6	✓	×	η, X, Y, Z	τ_x, τ_y	5,732
7	✓	×	η, X	τ_x, τ_y	5,092
8	✓	×	η, θ, X, Y, Z	τ_x, τ_y	6,052
9	✓	×	η, θ, X	τ_x, τ_y	5,412
10	✓	×	η, τ_x, τ_y	X	5,372
11	✓	×	$\eta, \theta, \tau_x, \tau_y$	X	5,692
12	✓	×	η, θ	$X, dx, dy, \tau_x, \tau_y$	5,212
13	✓	×	η, θ, S	$X, dx, dy, \tau_x, \tau_y$	5,532
1t	✓	✓	η	$\tau_x, \tau_y, X, dx, dy$	5,532
2t	✓	✓	η, θ	$\tau_x, \tau_y, X, dx, dy$	6,492
3t	✓	✓	η	τ_x, τ_y, X	5,452
4t	✓	✓	η, θ	τ_x, τ_y, X	6,412
5t	✓	✓	η	X, dx, dy	5,452
6t	✓	✓	η, θ	X, dx, dy	6,412
7t	✓	✓	η	X	5,372
8t	✓	✓	η, θ	X	6,332
9t	✓	✓	η, θ, S	τ_x, τ_y, X	7,372
10t	✓	✓	η, θ, S	$\tau_x, \tau_y, X, dx, dy$	7,452

predictions side by side. With the exception of five models (model 1, 5t, 7t, 6t, and 8t) all our NN model predictions have lower domain mean squared errors than the physics-based models. In terms of features, model without spatial information has the largest error, followed by models without wind stress (The absolute largest error is for the model without SSH, which is too big to be considered here). This signifies that to accurately represent surface currents, apart from SSH, the most important pieces of information required by the neural networks to successfully learn the physics of surface currents are spatial information and wind stress. It is striking to see how much the model struggles without spatial information. This implies that latitude dependence is a critical component for a NN to be able to predict surface currents accurately. It is only expected since the dynamics of surface currents do depend very strongly on latitude and therefore it is impossible to construct a meaningful prediction model based on just snapshots without any knowledge of latitude.

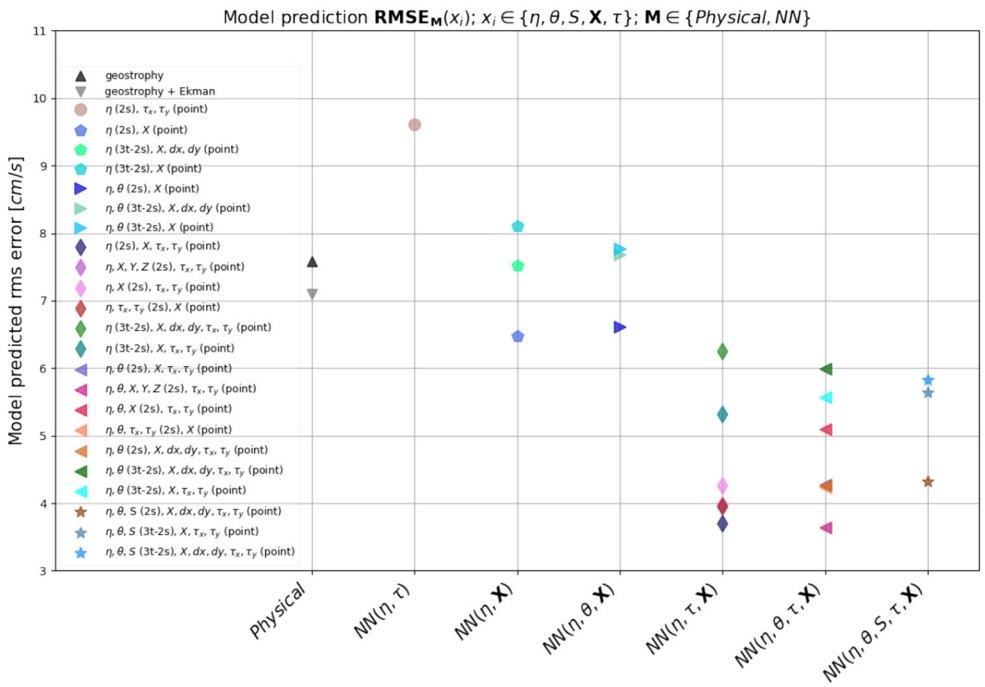


FIGURE 8 | Figure comparing the rms error of the different model predictions along with the rms error for the physical models as a function of features.

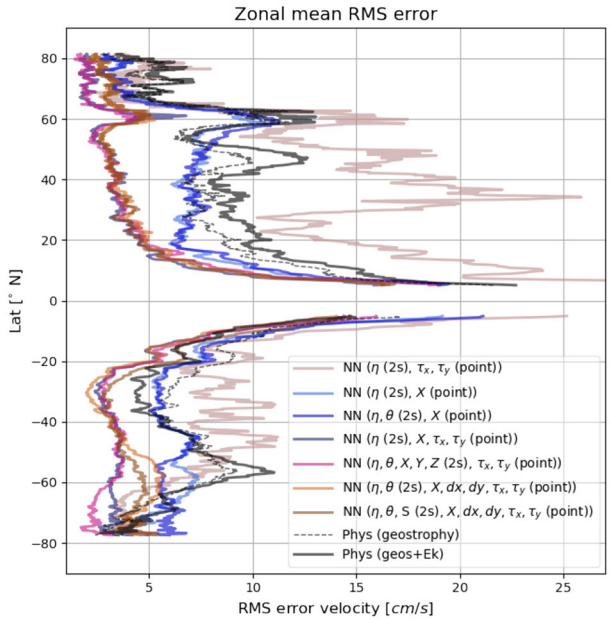


FIGURE 9 | Comparison of the zonal mean rms errors for the various NN predictions shown alongside the physical model (with and without Ekman flow).

The zonal mean rms error for the predictions from some of the representative models from the six categories described above are shown in **Figure 9**. The NNs all generally predict weaker velocities near the equator where the true values of the surface currents are quite large (due to strong equatorial jets). This can lead to large errors for the global mean, which get magnified when the differences are squared. However, we know that geostrophic and Ekman balance also doesn't hold near the equator. Therefore, to allow for a fair comparison between all the models, we mask out the rms errors in a 10° latitude band surrounding the equator ($5°N - 5°S$) for both the physical and statistical models. Out of all the models, model 1 which does not receive any spatial information ($\mathbf{X}$), has the highest mean squared errors throughout the globe. For the models that don't see wind stress (τ) as an input feature, the rms errors are comparable if not smaller at most latitudes when compared to the physics-based model where you only consider geostrophy (dashed black line). Additionally, all models that receive η, τ and $\mathbf{X}$ in some form perform consistently better than geostrophy + Ekman at all latitudes (except for near the equator where the physics-based models and the NN are all equally inadequate). We noticed that during training, the NN's minimize the loss function slightly faster when a stencil is provided for the spatial coordinates, but after a few epochs the differences in training loss between models that receive a spatial stencil and models that dont, diminish very rapidly. During prediction also, the models that receive stencils in spatial coordinates perform slightly better especially at the high latitudes than the ones where spatial information is provided pointwise.

Therefore among the various strategies tested, for this particular dataset, the models that perform the best in terms of prediction rms error are the models that receive SSH, wind stress

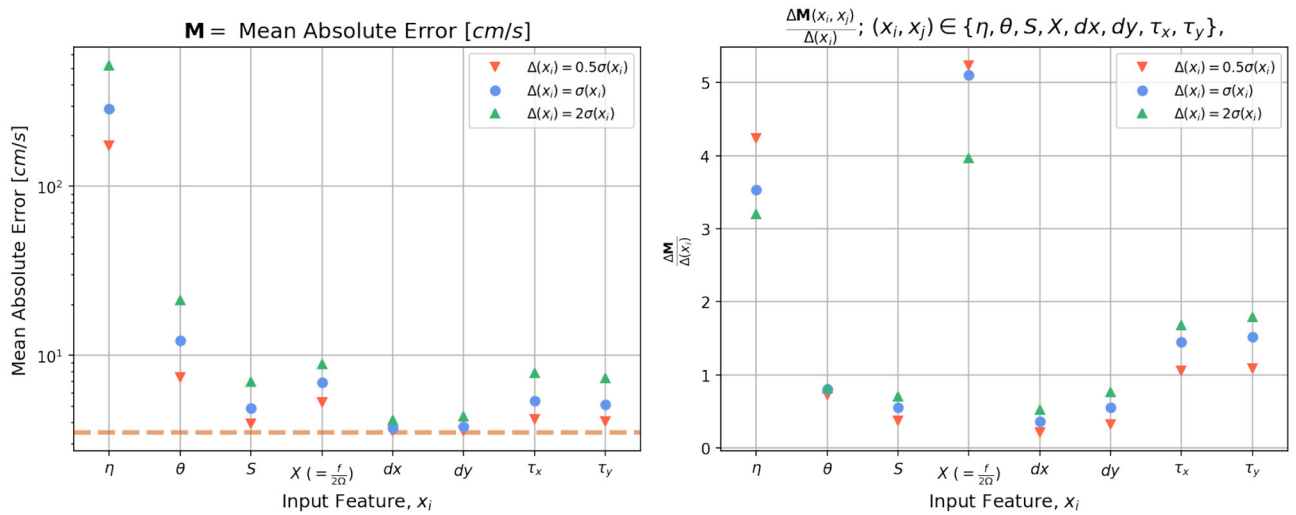


FIGURE 10 | Sensitivity of the neural networks to perturbations in the different input features. Each input feature is perturbed by three different Gaussian noise perturbations with standard deviations of 0.5σ, σ, and 2σ, where σ is the standard deviation of each variable, while keeping the remaining input variables fixed. The left panel shows the model loss (mean absolute error, MAE) evaluated for each of these perturbations. The horizontal dashed line represents the loss for the unperturbed/control case. The right panel shows the deviation in MAE for each of these perturbation experiments normalized by the amplitude of the perturbation.

and spatial information with a space stencil. The three point time stencil does not add anything meaningful and appears to hurt, rather than help the model overall, which was surprising, even though in hindsight we speculate that this might be due to the daily averaged nature of the POP model output. Variables like sea surface temperature and sea surface salinity have very little impact on the model as well.

In terms of choice of features, model 13 stands out as the most practical and physically meaningful training strategy for a few reasons.

- It is the most complete in terms of features
- It is the most straightforward to implement, since it does not involve calculating any transformed three dimensional coordinates. (All the input variables would be readily available for any gridded oceanographic dataset.)
- It is one of the models with the lowest prediction rms errors.

For these specific reasons we choose model 13 as the reference for performing a sensitivity analysis. The purpose of this analysis is to characterize the sensitivity of the model to perturbations in the different input features during testing/prediction. For the sensitivity tests we simply add a gaussian noise of varying amplitude to each of the input variables, while keeping the rest of the input variables fixed. For each of the input variables ($x_i \in \{\eta, \theta, X, dx, dy, \tau_x, tau_y\}$), we chose three different zero-mean gaussian noise perturbations with the standard deviations of $0.5\sigma(x_i)$, $\sigma(x_i)$, and $2\sigma(x_i)$, where $\sigma(x_i)$ is the standard deviation of the corresponding input variable x_i. The model loss is then evaluated for each of these perturbations and normalized by the amplitude of perturbations (right panel **Figure 10**). This normalization is done to level the playing field for all the input variables and allow for a one-to-one comparison since the different input variables vary in orders of magnitude [e.g., the amplitude of perturbations in SSH is $O(100)$, while the amplitude of perturbations in wind stress is $O(1)$ and therefore a perturbation of amplitude $\sigma(\eta)$ in η would lead to a much larger model error than a perturbation of $\sigma(\tau_x)$ in τ_x would, as can be seen from the log scaling of the y-axis in the left panel of **Figure 10**].

Given what we learned about the importance of spatial coordinates for NN training, it is not surprising to see that for prediction also, the NN is most sensitive to perturbations in the Coriolis parameter (or X). The input variables that the model is most sensitive to, arranged in descending order of model sensitivity are Coriolis parameter, SSH, and wind stress, followed by SST. The model is not particularly sensitive to perturbations in local grid spacing or salinity. The relative effect of the input variables, observed in the model sensitivity test closely matches what we saw in the different model training examples where we selectively held out these features. This again confirms that in order to train a deep learning model to make physically meaningful and generalizable predictions of surface currents it is not sufficient to simply provide it snapshots of dynamical variables like SSH as images. We also need to provide spatial information like latitude for the NN's to effectively "learn" the physics of surface currents.

7. SUMMARY AND FUTURE DIRECTIONS

The goal of this study was to use machine learning to make predictions of ocean surface currents from satellite observable quantities like SSH wind stress, SST etc. Our central question was: Can we train deep learning based models to learn physical models of estimating surface currents like geostrophy, Ekman flow and perhaps do better than the physical models themselves?

We used the output from the CESM POP model surface fields as our "truth" data for this study. As a first order example, we tested a linear regression model for a few of local subdomains extracted from the global GCM output. Linear regression works well only when the domains are small and far removed from the equator and gets progressively worse as the domain gets bigger and the variation in local Coriolis parameter gets large especially when f changes sign in the domain. This showed that unsurprisingly it is not possible to train a simple linear model to accurately predict surface currents. In addition, providing more data does not necessarily improve the predictive ability of a linear model and only made it worse as it starts overfitting. Whereas, for the same kind of domain, a neural network we can minimize the loss (MAE) with fewer data-points and still remain generalizable, since neural networks can learn functional relationships between regressors (input features) with only a small amount of data. The model's ability to make predictions is also shown to improve with more data. Furthermore, compared to a linear regression model, a NN even with a relatively small network of densely connected nodes, with a suitable non-linear activation function (like ReLU), allows us to have a large number of trainable parameters (weights, biases) that can be optimized to minimize the loss. The activation function is what allows the different non-linear combinations between the different regressors (input features). Somewhat surprisingly, a neural network trained on the entire globe is shown to predict surface currents more accurately in the sub-domains than neural networks trained in those specific sub domains. We suspect that is due to the fact that when trained locally with fewer data points, the neural network only sees part of the distribution of the input variables, as a result of which the stochastic gradient descent settles on a local minima and the model starts overfitting. Whereas, when trained globally, the neural network sees the full distribution of the input variables (not just a section) and SGD settles on a more realistic global minima. This can be observed in **Figure 3**, in all the local training examples, where model loss at the end of training is consistently lower than the model evaluation loss, when evaluated at a different date. In comparison, a similar approach with a linear regression model produces the opposite result, i.e., a globally trained linear regression model produces higher prediction errors than the one that's trained on each specific sub-domain. However, for linear regression, the evaluation losses are always consistently higher than the training losses, regardless of whether we consider a locally or a globally trained model. The fact that spatially diverse data actually makes the neural network perform better indicates that a neural network can actually "learn" the functional relationships needed for calculating surface currents, i.e., be more generalizable, instead of simply memorizing some target values for different combination of input features. By examining the dependence of the NNs on the choice of input features and by looking at the sensitivity of a NN model to perturbations in the input features, we established that apart from SSH, the physical location of the input features is one of the most crucial elements for the NN to "learn" the physics of surface currents. It is further demonstrated that with a careful and deliberate choice of input features the neural network can even beat the physics-based models at predicting surface currents accurately in most regions of the global ocean. A key ingredient for calculating the Ekman part of the flow using current physics based models is the vertical diffusivity, which is largely unknown for most of the global ocean. Most observational ocean current estimates that include the Ekman part of the flow relies on inferring the vertical diffusivity based on empirical multiple linear regressions with Lagrangian surface drifter data, The neural network approach, by comparison does not suffer from the same kind of limitation, since in this framework, we do not need to provide A_z as an input feature, which is one more added advantage for this method.

In this study, we wanted to see whether we can train a statistical model like a NN with data to essentially match or perhaps beat the baseline physics based models we currently use to estimate surface currents. By examining the errors in surface current predictions from our NN predictions and comparing them with predictions from physically motivated models (like geostrophy and Ekman dynamics), we showed that a relatively simple NN captures most of the large scale flow features equally well if not better than the physical models, with only 1 day of training data for the globe.

However, some key aspects of the flow, associated with mesoscale and sub-mesoscale turbulence are not reproduced. We speculate that this is possibly caused by the fact that the neural network framework can not capture the higher order balances (gradient wind) that are likely at play in these regions since these hotspots of high errors are collocated with regions of High Ro where balance breaks down (see **Figures 4**–**6**).

One of the biggest hurdles associated with these studies is figuring out efficient strategies to stream large volumes of earth system model data into a NN framework. So before diving headfirst into the highest resolution global ocean model (currently available), we wanted to test the feasibility of using a regression model based on deep learning as a framework for estimating surface currents with a lower resolution model data (smaller/more manageable dataset), while still being eddy resolving. Hence we chose the CESM POP model data for this present study. In the future, we propose to train a NN with data from a higher spatio-temporal resolution global ocean model like the MITgcm llc4320 model (Menemenlis et al., 2008; Rocha et al., 2016). As a further step, we could coarse-grain such a model to SWOT-like resolutions, or use the SWOT simulator, train NNs on that, and make predictions for global surface currents.

As for the weak surface currents predicted by our NN at the equator, we need to keep in mind that geostrophic balance (defined by the first order derivatives of SSH) only holds away from the equator and satellite altimetry datasets (e.g., AVISO, Ducet et al., 2000) typically employ a higher order balance (Lagerloef et al., 1999) at the equator, to match the flow regime with the geostrophic regime away from the equator. One possible way to train the NN to learn these higher order balances would be by increasing the stencil size around each point. Since our primary goal was for the model to learn geostrophy, we started with a spatial stencil in SSH. We also explored training approaches where we provided stencils in SST, wind and SSS, with the intention of helping the model learn about wind-stress curl and thermal wind balance. In practice, however these didn't payoff as much and these additional stencils did not significantly

improve model performance. In future approaches we can try to provide separate stencils of varying size to each of these input variables, to test whether we can further improve the model accuracy.

As another future step, we also aim to incorporate recursive neural networks (RNNs) in conjunction with convolutional filters of varying kernel sizes, to train the models on cyclostrophic or gradient wind balance. This recursive neural network approach would be analogous to iteratively solving the gradient wind equation (Knox and Ohmann, 2006), a technique which was originally developed for numerical weather prediction before advances in computing allowed for integrating the full non-linear equations.

The present work demonstrates that to a large extent, a simple neural network can be trained to extract functional relationships between SSH, wind stress, etc. and surface currents with quite limited data. The field of deep learning as of now is rapidly evolving. It remains to be seen, if with some clever choices of training strategies and by using some of the other more recently developed deep learning techniques, we can improve upon this. In this study, we propose a few approaches that can be implemented to improve upon our current results and would like to investigate this in further detail in future studies. In addition, we believe that data driven approaches, like the one shown in this present study, have strong potential applications for various practical problems in physical oceanography, and require further exploration. Insights gained from this type of analysis could be of great potential significance, especially for future satellite altimetry missions like SWOT.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://catalog.pangeo.io/browse/master/ocean/CESM_POP/.

AUTHOR CONTRIBUTIONS

AS and RA: study conception and design and draft manuscript preparation. AS: training and testing of statistical models and analysis and interpretation of results. All authors reviewed the results and approved the final version of the manuscript.

FUNDING

The authors acknowledge support from NSF Award OCE 1553594 and NASA award NNX16AJ35G (SWOT Science Team).

ACKNOWLEDGMENTS

The work was started in 2018 and an early proof-of-concept was reported in AS's PhD dissertation (Sinha, 2019).

REFERENCES

Arbic, B. K., Richman, J. G., Shriver, J. F., Timko, P. G., Metzger, E. J., and Wallcraft, A. J. (2012). Global modeling of internal tides: within an eddying ocean general circulation model. *Oceanography* 25, 20–29. doi: 10.5670/oceanog.2012.38

Bar-Sinai, Y., Hoyer, S., Hickey, J., and Brenner, M. P. (2018). Data-driven discretization: a method for systematic coarse graining of partial differential equations. *arXiv*.

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer-Verlag.

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

Bonjean, F., and Lagerloef, G. S. E. (2002). Diagnostic model and analysis of the surface currents in the tropical Pacific Ocean. *J. Phys. Oceanogr*. 32, 2938–2954. doi: 10.1175/1520-0485(2002)032<2938:DMAAOT>2.0.CO;2

Chapman, C., and Charantonis, A. A. (2017). Reconstruction of subsurface velocities from satellite observations using iterative self-organizing maps. *IEEE Geosci. Rem. Sens. Lett*. 14, 617–620. doi: 10.1109/LGRS.2017.2665603

Chollet, F. (2015). *Keras*. Available online at: https://keras.io

Ducet, N., Le Traon, P. Y., and Reverdin, G. (2000). Global high-resolution mapping of ocean circulation from topex/poseidon and ERS-1 and -2. *J. Geophys. Res. Oceans* 105, 19477–19498. doi: 10.1029/2000JC900063

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett*. 45, 5742–5751. doi: 10.1029/2018GL078202

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep Learning, Vol. 1*. Cambridge, MA: MIT Press.

Gregor, L., Kok, S., and Monteiro, P. M. S. (2017). Empirical methods for the estimation of southern ocean CO_2: support vector and random forest regression. *Biogeosciences* 14, 5551–5569. doi: 10.5194/bg-14-5551-2017

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

Kingma, D. P., and Ba, J. (2017). Adam: A method for stochastic optimization.

Knox, J. A., and Ohmann, P. R. (2006). Iterative solutions of the gradient wind equation. *Comput. Geosci*. 32, 656–662. doi: 10.1016/j.cageo.2005.09.009

Lagerloef, G. S. E., Mitchum, G. T., Lukas, R. B., and Niiler, P. P. (1999). Tropical pacific near-surface currents estimated from altimeter, wind, and drifter data. *J. Geophys. Res. Oceans* 104, 23313–23326. doi: 10.1029/1999JC900197

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521:436. doi: 10.1038/nature14539

Lguensat, R., Sun, M., Fablet, R., Mason, E., Tandeo, P., and Chen, G. (2017). Eddynet: a deep neural network for pixel-wise classification of oceanic eddies. *CoRR* abs/1711.03954. doi: 10.1109/IGARSS.2018.8518411

Ling, J., Kurzawski, A., and Templeton, J. (2016). Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *J. Fluid Mech*. 807, 155–166. doi: 10.1017/jfm.2016.615

Menemenlis, D., Campin, J. M., Heimbach, P., Hill, C., Lee, T., Nguyen, A., et al. (2008). "ECCO2: high resolution global ocean and sea ice data synthesis," in *Mercator Ocean Quarterly Newsletter*, 13–21.

Morrow, R., Blurmstein, D., and Dibarboure, G. (2018). "Fine-scale altimetry and the future SWOT mission," in *New Frontiers in Operational Oceanography*, eds E. Chassignet, A. Pascual, J. Tintoré, and J. Verron (GODAE OceanView) 191–226. doi: 10.17125/gov2018.ch08

Nielsen, M. A. (2015). *Neural Networks and Deep Learning*, Vol. 25. Determination Press USA.

Pathak, J., Hunt, B., Girvan, M., Lu, Z., and Ott, E. (2018). Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. *Phys. Rev. Lett*. 120:024102. doi: 10.1103/PhysRevLett.120.024102

Rocha, C. B., Chereskin, T. K., Gille, S. T., and Menemenlis, D. (2016). Mesoscale to submesoscale wavenumber spectra in drake passage. *J. Phys. Oceanogr*. 46, 601–620. doi: 10.1175/JPO-D-15-0087.1

Sinha, A. (2019). *Temporal variability in ocean mesoscale and submesoscale turbulence* (Ph.D. thesis), Columbia University in the City of New York, New York, NY, United States.

Small, R. J., Bacmeister, J., Bailey, D., Baker, A., Bishop, S., Bryan, F., et al. (2014). A new synoptic scale resolving global climate simulation using the community earth system model. *J. Adv. Model. Earth Syst.* 6, 1065–1094. doi: 10.1002/2014MS000363

Smith, R., Jones, P., Briegleb, B., Bryan, F., Danabasoglu, G., Dennis, J., et al. (2010). *The Parallel Ocean Program (POP) Reference Manual Ocean Component of the Community Climate System Model (CCSM) and Community Earth System Model (CESM)*. Rep. LAUR-01853 141:1–140. Available online at: http://n2t.net/ark:/85065/d70g3j4h

Sudre, J., Maes, C., and Garcon, V. (2013). On the global estimates of geostrophic and ekman surface currents. *Limnol. Oceanogr. Fluids Environ.* 3, 1–20. doi: 10.1215/21573689-2071927

Sudre, J., and Morrow, R. A. (2008). Global surface currents: a high-resolution product for investigating ocean dynamics. *Ocean Dyn.* 58:101. doi: 10.1007/s10236-008-0134-9

Traon, P. Y., and Morrow, R. (2001). *Ocean Currents and Eddies*, Vol. 69, Chapter 3. San Diego, CA: Academic Press.

Uchida, T., Abernathey, R., and Smith, S. (2017). Seasonality of eddy kinetic energy in an eddy permitting global climate model. *Ocean Model.* 118, 41–58. doi: 10.1016/j.ocemod.2017.08.006

Zanna, L., and Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophys. Res. Lett.* 47:e2020GL088376. doi: 10.1029/2020GL088376

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 15 June 2021
doi: 10.3389/fmars.2021.672334

Application of Three Deep Learning Schemes Into Oceanic Eddy Detection

Guangjun Xu[1,2,3], Wenhong Xie[3,4], Changming Dong[2,3,4] and Xiaoqian Gao[5,6]*

[1] School of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang, China, [2] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China, [3] UNIVER-NUIST Joint AI Oceanography Academy, Nanjing University of Information Science and Technology, Nanjing, China, [4] Oceanic Modeling and Observation Laboratory, Nanjing University of Information Science and Technology, Nanjing, China, [5] College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao, China, [6] Key Laboratory of Marine Science and Numerical Modeling (MASNUM), Ministry of Natural Resources, Qingdao, China

Recent years have witnessed the increase in applications of artificial intelligence (AI) into the detection of oceanic features. Oceanic eddies, ubiquitous in the global ocean, are important in the transport of materials and energy. A series of eddy detection schemes based on oceanic dynamics have been developed while the AI-based eddy identification scheme starts to be reported in literature. In the present study, to find out applicable AI-based schemes in eddy detection, three AI-based algorithms are employed in eddy detection, including the pyramid scene parsing network (PSPNet) algorithm, the DeepLabV3+ algorithm and the bilateral segmentation network (BiSeNet) algorithm. To justify the AI-based eddy detection schemes, the results are compared with one dynamic-based eddy detection method. It is found that more eddies are identified using the three AI-based methods. The three methods' results are compared in terms of the numbers, sizes and lifetimes of detected eddies. In terms of eddy numbers, the PSPNet algorithm identifies the largest number of ocean eddies among the three AI-based methods. In terms of eddy sizes, the BiSeNet can find more large-scale eddies than the two other methods, because the Spatial Path is introduced into the algorithm to avoid destroying the eddy edge information. Regarding eddy lifetimes, the DeepLabV3+ cannot track longer lifetimes of ocean eddies.

Keywords: oceanic eddy detection, deep learning, PSPNet, BiSeNet, DeepLabV3+

OPEN ACCESS

Edited by:
Jun Li,
University of Technology Sydney, Australia

Reviewed by:
Peter Strutton,
University of Tasmania, Australia
Taoping Liu,
University of Technology Sydney, Australia

****Correspondence:***
Changming Dong
cmdong@nuist.edu.cn

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *26 February 2021*
Accepted: *21 May 2021*
Published: *15 June 2021*

Citation:
Xu G, Xie W, Dong C and Gao X (2021) Application of Three Deep Learning Schemes Into Oceanic Eddy Detection. Front. Mar. Sci. 8:672334. doi: 10.3389/fmars.2021.672334

INTRODUCTION

Oceanic eddies are ubiquitous in the global ocean. They play an important role in material and energy transport, and global climate changes. On a global scale, oceanic mesoscale eddies contribute significantly to horizontal heat and salt transports (Dong et al., 2014; Moreau et al., 2017; Patel et al., 2019, 2020). Many eddies identification methods based on different kinds of remote sensing data have been developed (Chelton et al., 2007; Chaigneau et al., 2008;

Nencioli et al., 2010; Dong et al., 2011a,b; Faghmous et al., 2012; Chen et al., 2016).

Deep learning schemes have been used in oceanic eddies detection during the past few years. Lguensat et al. (2017) firstly applied a deep learning algorithm based on the encoder-decoder network to oceanic eddies detection in the classic framework of the semantic segmentation. The encoder-decoder network with the simple convolution module and the upsampling module was also used to detect and track oceanic eddies (Franz et al., 2018). Based on synthetic aperture radar images, deep learning was applied to automatically detect oceanic eddies according to the extracted higher-level features and fused multi-scale features (Du et al., 2019). Xu et al. (2019) applied the pyramid scene parsing network (PSPNet) to identify oceanic eddies and find that the PSPNet has great advantage in the detection of small-scale eddies. Duo et al. (2019) proposed an Ocean Eddy Detection Net (OEDNet) based on an object detection network to recognize the eddy field by enhancing the accurate small sample data to obtain the training dataset.

Systematic comparison of the performances of these AI-based methods discussed above is required to justify which one is the most applicable in eddy detection. Such comparison can also shed light on the application of AI algorithms into oceanography. The present study employs three different AI-based algorithms into eddy detection, including Pyramid Scene Parsing Network (PSPNet), DeepLabV3+ and Bilateral Segmentation Network (BiSeNet), and makes the comparisons of their results based on a few eddy parameters, such as eddy number, size and lifetime.

MATERIALS AND METHODS

Pyramid Scene Parsing Network (PSPNet)

The PSPNet (Zhao et al., 2017) incorporates the pyramid pooling module (He et al., 2014) and the reduced convolution (Yu and Koltun, 2016), which fully uses the global scene to capture more details of the context information between the different category

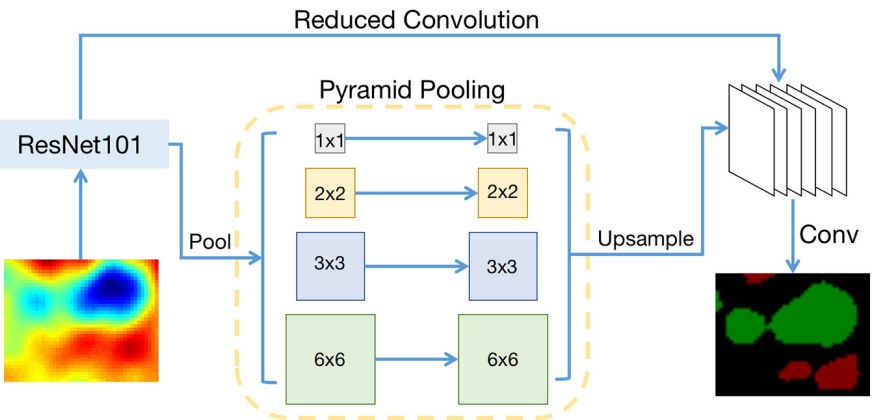


FIGURE 1 | Schematic diagram of PSPNet.

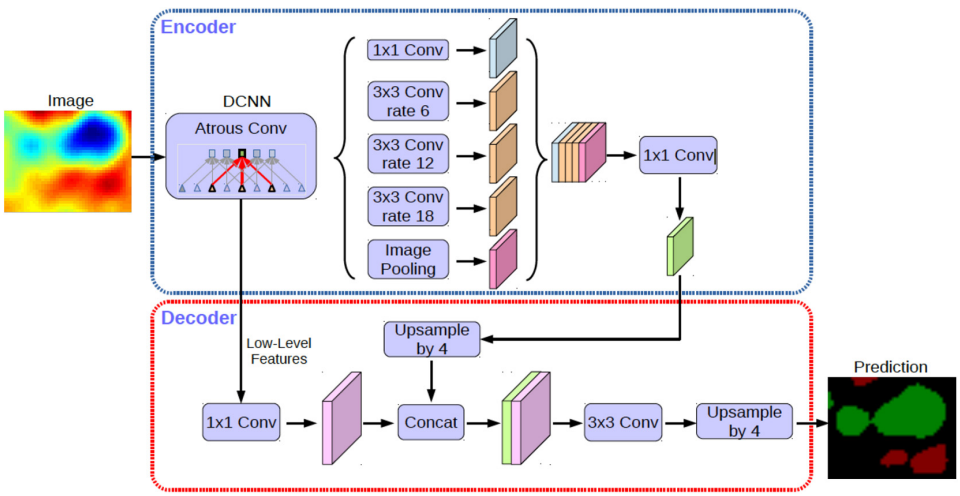


FIGURE 2 | Schematic diagram of DeepLab V3+.

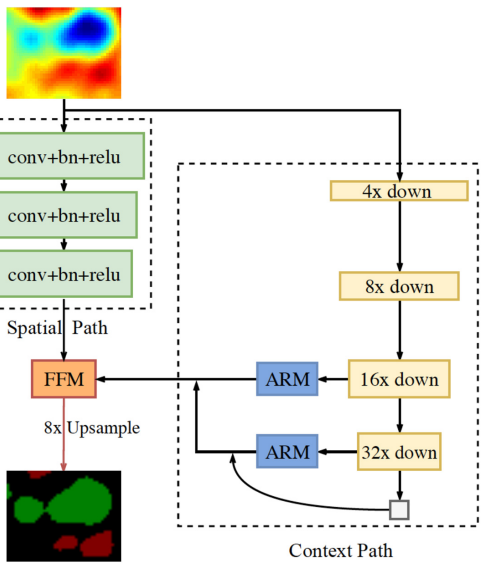


FIGURE 3 | Schematic diagram of BiSeNet.

labels. Sea surface height anomaly (SSHA) data labeled with eddy information are used as the training data. **Figure 1** shows the configuration of the PSPNet. A 101-level ResNet (ResNet101) model with a dilated network strategy (Yu and Koltun, 2016; Chen et al., 2017) is implemented to an input SSHA image to extract the feature map at different levels. The final feature map is reduced to 1/8 of the input image. The pyramid pooling module is then employed to obtain context information. A four-level pyramid fuses the images at the different sizes as the global prior. The prior is connected to the original feature and a convolution layer to generate the final prediction. The PSPNet program, proposed by Zhao et al. (2017), is publicly available at https://github.com/hszhao/PSPNet.

DeepLabV3+

The DeepLabv3+ (Chen et al., 2017) employs the Spatial Pyramid Pooling module (SPP) and the encoder-decoder structure for the semantic segmentation based on the deep-learning network (**Figure 2**). SPP applies atrous convolution with different rates to obtain convolutional features at multiple scales to mine the multi-scale context information. The low-level features of the encoder are able to capture larger spatial information. The spatial information is used to recover the details and spatial dimensions of the target during the decoder stage, and to refine the segmentation results along object boundaries. The DeepLabV3+ program, proposed by Chen et al. (2017), can be obtained from https://github.com/tensorflow/models/tree/master/research/deeplab.

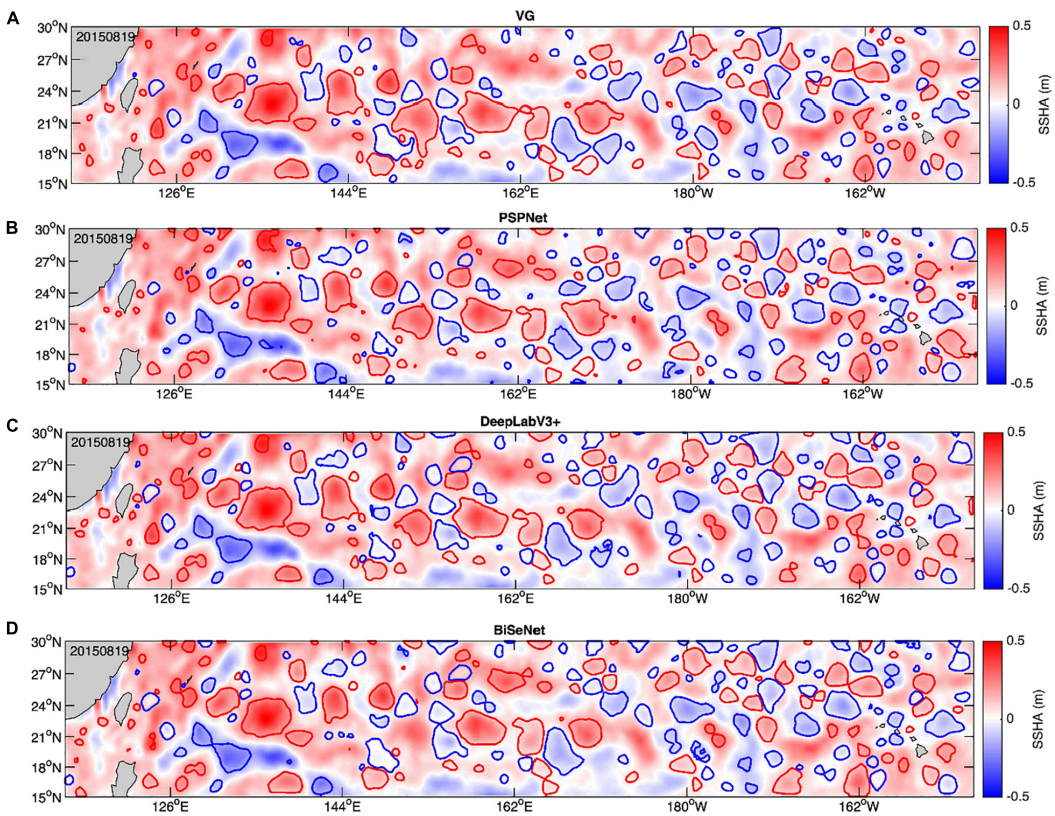


FIGURE 4 | Oceanic eddies identified by VG **(A),** PSPNet **(B),** DeepLabV3+ **(C)**, and BiSeNet **(D)** algorithms in the STCC region on 19 August 2015.

Bilateral Segmentation Network (BiSeNet)

The BiSeNet (Yu et al., 2018) is divided into three parts: Spatial Path, Context Path, and Feature Fusion Module, as shown in **Figure 3**. The Spatial Path encodes the detailed information captured from the image into spatial information. The Context Path is applied to encode the context information and the Feature Fusion Module is used to fuse the spatial information and the context information. According to the characteristics of the different levels, the spatial and the context information are connected in series, and a batch normalization is adopted to balance the feature information scale. The connected features are applied to obtain a weight vector employed to select and combine the feature information, and to get the final result. The BiSeNet program, proposed by Yu et al. (2018), can be download from https://github.com/CoinCheung/BiSeNet.

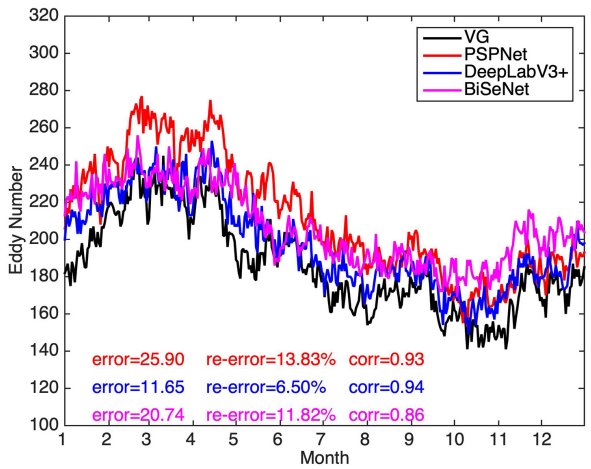


FIGURE 5 | Time series of the daily number of eddies identified by the VG and three AI algorithms in the STCC region during 2015. Black, red, blue, and purple curves represent the results from VG, PSPNet, DeepLabV3+, and BiSeNer algorithm, respectively. "Error" is the difference between the three AI-based results and the VG results, "re-error" is the relative error which is defined as the error divided by the VG results, and "corr" is the correlation coefficient between the results from the three AI-based algorithms and the VG algorithm.

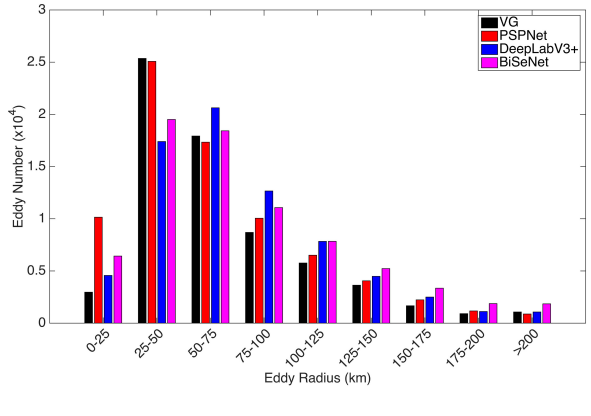


FIGURE 6 | Size distribution of the identified eddies from the four different algorithms in the STCC region during 2015. Black, red, blue, and purple bars represent the results from VG, PSPNet, DeepLabV3+, and BiSeNet algorithm, respectively.

Vector Geometry—Based Eddy Detection Algorithm (VG)

We apply the method based on the geometry of velocity vectors of the flow field (Dong et al., 2009; Nencioli et al., 2010) for mesoscale eddy identification and tracking from geostrophic current, which is obtained from SSHA data. Eddy centers are determined by four criterions as follows (Nencioli et al., 2010): (a) along an east-west section, meridional velocity v has to reverse in sign across the eddy center and its magnitude has to increase away from it; (b) along a north—south section, zonal velocity u has to reverse in sign across the eddy center and its magnitude has to increase away from it: the sense of rotation has to be the same as for v; (c) velocity magnitude has a local minimum at the eddy center; and (d) around the eddy center, the directions of the velocity vectors have to change with a constant sense of rotation and the directions of two neighboring velocity vectors have to lay within the same or two adjacent quadrants. Eddy boundary is determined by the closed contour of the stream function field. Eddy tracks are retrieved by comparing the distribution of eddy centers at successive time steps. The tracking methods for the AI-based algorithms are the same as that for VG. The tracking method is one part of the VG, which can be found in Nencioli et al. (2010).

Data

Eddies were identified from daily SSHA data with a spatial resolution of 1/4° × 1/4°. The data, obtained from Copernicus Marine Environment Monitoring Service (CMEMS)[1], was a

[1]http://marine.copernicus.eu

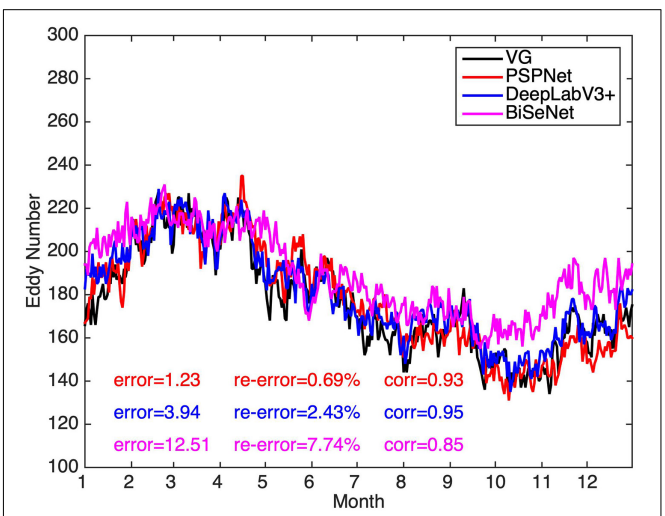


FIGURE 7 | Same as **Figure 5** except for the comparison of the daily number of eddies with radii greater than 25 km.

global product from multiple satellite altimeter along-track data. The SSHA data in the period from 2011 to 2015 were used in this study. The SSHA data were linearly interpolated onto a 1/8 × 1/8° grid to extend the eddy field for a larger number of grid points in order to improve further the performance of the eddy detection scheme (Liu et al., 2012). The SSHA data from 2011 to 2014 was used as the training data containing the labels of eddy information, while the 2015 were used as the validation set. This study focused on the North Pacific Subtropical Countercurrent (STCC, 15° N ∼ 30° N, 115° E ∼ 150° W), extending from east of Luzon Strait to the Hawaii Islands.

RESULTS

The training SSHA data from 2011 to 2014 in the STCC region, which is labeled with cyclonic and anticyclonic eddies, is produced based on the traditional VG algorithm. In order to ensure validity and consistency of the training dataset, the eddy information are cleaned, dealing with invalid and missing data. Finally, the three different AI schemes are employed for deep learning with the training dataset from 2011 to 2014 and for eddy identification with the validation dataset during 2015.

The eddies identified based on the three different algorithms in the STCC region are compared on 19 August (**Figure 4**). Using the traditional VG algorithm, 172 ocean eddies are detected, including 84 cyclonic and 88 anticyclonic eddies. However, the other three deep learning algorithms all identify more eddies: 185 eddies (87 cyclonic and 98 anticyclonic eddies) from the PSPNet algorithm, 184 eddies (87 cyclonic and 97 anticyclonic eddies) from the DeepLabV3+ algorithm and 188 eddies (89 cyclonic and 99 anticyclonic eddies) from the BiSeNet algorithm.

The comparisons of oceanic eddies detected from the AI-based algorithms are made based on a few eddy parameters, such as eddy number, size and lifetime.

Figure 5 compares the daily number of eddies detected by the VG, PSPNet, DeepLabV3+ and BiSeNet algorithms in the STCC region during 2015. A total of 68,010 oceanic eddies are identified by the VG algorithm, including 32,783 cyclonic and 35,227 anticyclonic eddies, which are less than those identified by the three AI algorithms. Among the three AI methods, the PSPNet algorithm detects the largest number of oceanic eddies (a total of 77,462 eddies). The DeepLabV3+ and the BiSeNet algorithm identify 72,264 and 75,579 eddies, respectively.

Compared with the traditional VG method results, the PSPNet,DeepLabV3+ and BiSeNet algorithm on average identify 25.90, 11.65, and 20.74 more eddies per day, respectively. There is a good correlation between the daily eddy numbers from the VG algorithm and the PSPNet (0.93), and the DeepLabV3+ (0.94) algorithm. The correlation of the daily eddy numbers between the BiSeNet and the VG method is less than the other two AI methods, with a correlation coefficient of 0.86. In **Figure 6**, the time series of the daily number of eddies detected by the DeepLabV3+ algorithm is the most consistent with that by the VG algorithm. In addition, based on the VG algorithm results, the differences in the results from the PSPNet and BiSeNet algorithms have seasonal variation characteristics. The PSPNet algorithm detected more eddies in spring and summer, while the BiSeNet algorithm detected a larger number of eddies in winter.

The radii of the oceanic eddies detected by the four different methods are compared in **Figure 6**. All the identified eddies peak at the 25–50 km bin, except for the DeepLabV3+, which has the highest number at the 50–75 km bin. The PSPNet algorithm has an obvious advantage in detecting small-scale eddies with radii

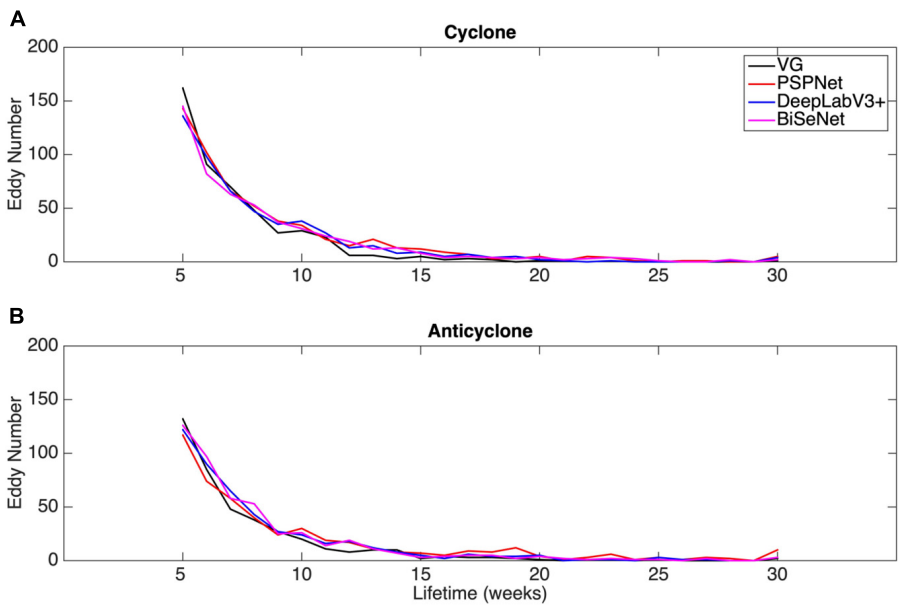


FIGURE 8 | Lifetime distribution of eddies identified by the VG (black), PSPNet (red), DeepLabV3+ (blue), and BiSeNet (purple) algorithms in the STCC region during 2015. **(A)** Cyclones; **(B)** anticyclones.

of less than 25 km. The DeepLabV3+ algorithm identified the highest number of eddies with radius between 50 and 100 km. Furthermore, the BiSeNet algorithm detects more bigger eddies (greater than 100 km) than the other three methods.

The resolution of the altimetry data limits the presentation of small eddies. The facticity of the small-scale eddy needs further confirmation, so the results of detected eddies with radii less than 25 km are removed and the comparison among the algorithms is plotted in **Figure 7**. 64586 eddies are afterward detected by the VG method, while 65,034, 66,023, and 69,153 eddies are, respectively, detected by the PSPNet, DeepLabV3+ and BiSeNet algorithms. The differences between the traditional and AI-based results decrease with a similar pattern during 2015. The BiSeNet algorithm identifies the largest number of oceanic eddies among the three AI-based methods. 12.51 eddies more, on average, are identified per day with a relative error of 7.74%. It is suggested that the BiSeNet algorithm takes an advantage in identifying large-scale eddies, which can be verified in **Figure 6**. However, the majority of the additional eddies detected by the PSPNet algorithm are small-scale ones.

Lifetime is another important eddy characteristic. The lifetime distribution of the eddies identified by the four algorithms are

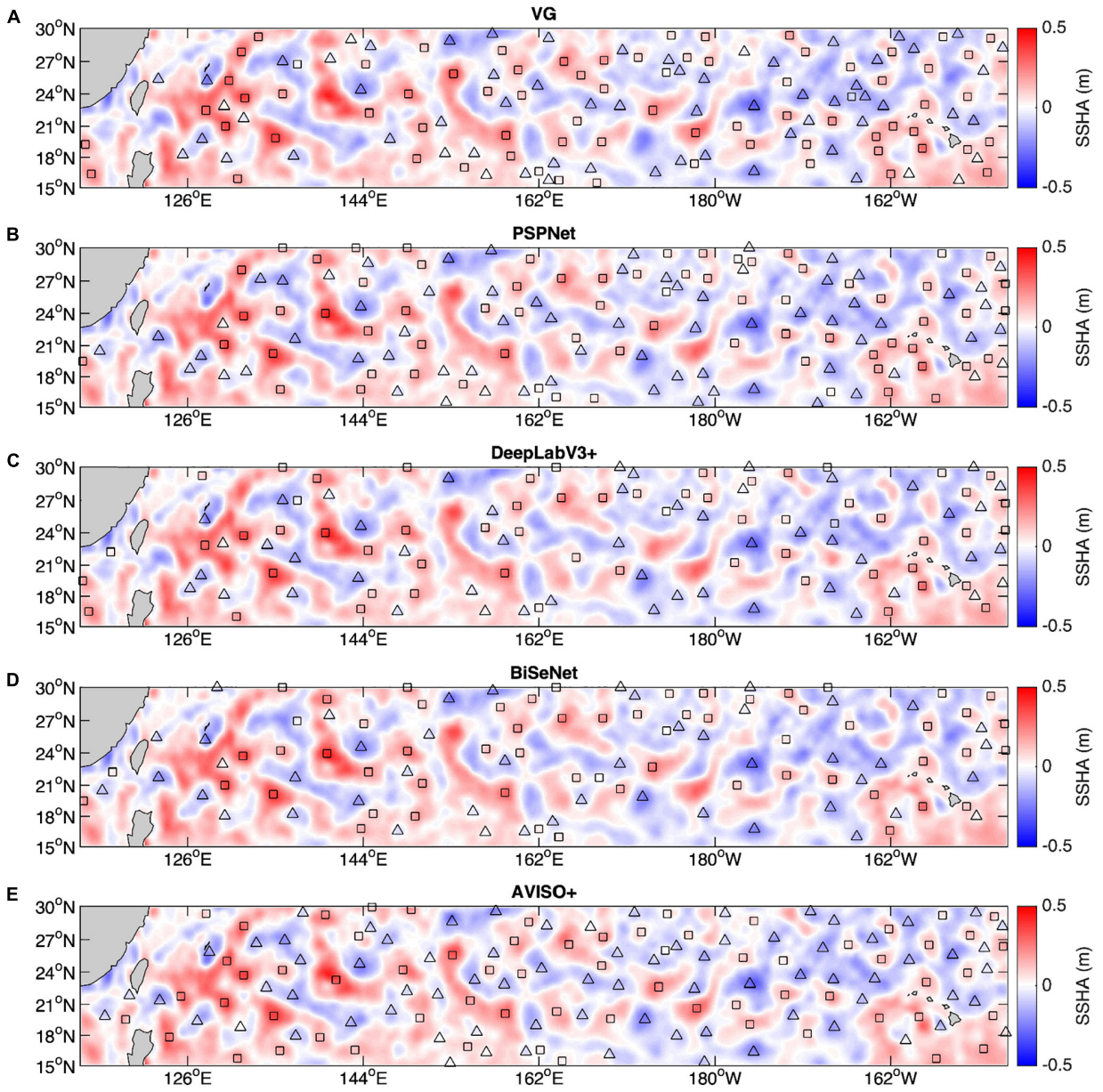


FIGURE 9 | Comparison of the oceanic eddies longer than 4 weeks from the five versions in the STCC region on June 07, 2015. **(A)** The VG version; **(B)** the PSPNet version; **(C)** the DeepLabV3+ version; **(D)** the BiSeNet version; **(E)** the AVISO+ version. Triangles and squares represent the center of the cyclonic and anticyclonic eddies, respectively.

TABLE 1 | Comparisons of eddy properties base on the five results.

		Lifetime (days)	Amplitude (cm)	Radius (km)	Displacement (km)	Translational speed (km day^{-1})
VG	Mean	50.1	7.2	80.9	332.2	6.7
	Maximum	223	40.5	298.7	1453.7	25.0
PSPNet	Mean	63.4	8.2	89.4	343.7	6.8
	Maximum	347	45.4	296.5	3174.0	25.0
DeepLabV3+	Mean	57.3	8.4	92.8	299.2	6.3
	Maximum	365	40.6	284.5	1498.6	25.0
BiSeNet	Mean	57.6	9.3	104.6	408.9	7.8
	Maximum	233	39.3	296.7	1769.4	25.0
AVISO+	Mean	72.2	8.0	99.8	544.5	6.9
	Maximum	365	34.3	267	3355.3	25.0

presented in **Figure 8**. A total of 844 eddies with lifetimes greater than 4 weeks (387 anticyclones and 457 cyclones) are identified by the traditional VG algorithm. Considering eddies which survive more than 4 weeks, the PSPNet, DeepLabV3 + and BiSeNet algorithms detect 875 eddies (400 anticyclonic and 475 cyclonic eddies), 805 eddies (370 anticyclonic and 435 cyclonic eddies), and 819 eddies (383 anticyclonic and 436 cyclonic eddies), respectively. The eddies identified by the three AI algorithms survive longer than that identified by the VG algorithm because the AI algorithms can detect small-sized eddies during the growing and decaying periods. Among the three AI-based methods, eddies detected by the DeepLabV3+ algorithm have the shortest lifetimes.

DISCUSSION

In order to further discuss the AI-based results, the eddies identified by the AI-based algorithms and by the VG algorithm in the STCC region during 2015 are compared with mesoscale eddy trajectory atlas product (Ver 2.0), which is obtained from AVISO+[2]. Since the AVISO+ product provides the eddies with the lifetimes longer than 4 weeks but without the boundary information (only radius), only the center locations of the eddies with the lifetimes longer than 4 weeks and with the radii greater than 25 km in the STCC region on June 07, 2015 are plotted in **Figure 9** for the comparison of the results from the five versions. The PSPNet algorithm detects the most oceanic eddies (176 eddies) of the three versions, followed by 142 ones for the AVISO+ version, 137 ones for the VG version, 128 ones for the BiSeNet version and 125 ones for the DeepLabV3+. During the 1 year period, 637 eddy tracks are totally obtained from the AVISO product, all of which survive longer than 4 weeks. The PSPNet algorithm detects 875 eddy tracks with lifetimes longer than 4 weeks, while 844 tracks by the VG algorithm, 819 tracks by the BiSeNet and 805 tracks by the DeepLabV3+. The number of eddy tracks from the AVISO+ version is the smallest among those from the five versions.

[2]https://www.aviso.altimetry.fr/en/data/products/value-added-products/global-mesoscale-eddy-trajectory-product.html

Several eddy parameters are compared among these five results in **Table 1**. The mean lifetimes of eddies identified by the AI-based algorithms are all shorter than that from the AVISO+ version but longer than that detected by the VG algorithm. The eddies from the AVISO+ version, the PSPNet version and the BiSeNet version survive for up to a year. The mean amplitude (9.3 cm) and the mean radius (104.6 km) of the eddies from the BiSeNet version are greater than those from the other four versions, which suggests that the BiSeNet has an advantage in large-scale eddy. And the maximum eddy radii are all larger than 260 km, even close to 300 km, for the five versions. The mean displacements of the eddies from the AI-based versions and the VG version are both shorter than that from the AVISO+ version. But the eddies from the PSPNet version and the AVISO+ version moves up to more than 3,000 km. Furthermore, the eddies from the three versions propagate with the similar mean speed at 6.3 ∼7.8 km day^{-1}.

CONCLUSION

Compared to the traditional eddy detection method, deep learning-based algorithms are novel. Three different AI-based algorithms were applied to identify ocean eddies, including the PSPNet, DeepLabV3+ and BiSeNet algorithms. SSHA data from 2011 to 2014 in the STCC region, labeled with eddy information detected by the VG algorithm, are employed for training, and SSHA data in 2015 used for validation. Eddies detected by the three AI-based methods are compared with each other and with the results from the traditional method. All three AI-based algorithms extract more eddies than the traditional VG algorithm. The PSPNet algorithm detects the highest number of eddies. The BiSeNet algorithm performs better in large-scale eddy identification than the other two AI-based algorithms, because the Spatial Path is introduced into the algorithm to avoid destroying the eddy edge information. The eddies identified by the AI-based algorithms tend to survive longer than those identified by the VG method. The lifetimes of the eddies extracted by the DeepLabV3+ algorithm are the shortest among the results from the AI-based methods.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

GX and CD contributed to the designing and planning the experiments and writing the manuscript. WX and XG performed the experiments and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the project supported by Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (SML2020SP007), the Guangdong Basic and Applied Basic Research Foundation (2019A1515110840), the Key Program of Marine Economy Development (Six Marine Industries) Special Foundation of Department of Natural Resources of Guangdong Province [GDNRC(2020)049], the National Natural Science Foundation of China (42076162), the Natural Science Foundation of Guangdong Province of China (2020A1515010496), and the Research Startup Foundation of Guangdong Ocean University (R20009).

ACKNOWLEDGMENTS

The authors would like to thank AVISO+ for providing SSHA data and mesoscale eddy trajectory atlas product. The authors would like to thank GitHub for providing AI programs. The authors would also like to thank Zhoutong Technology ILC (UNIVER) for their competent comments and technical guidance.

REFERENCES

Chaigneau, A., Gizolme, A., and Grados, C. (2008). Mesoscale eddies off peru in altimeter records: identification algorithms and eddy spatio-temporal patterns. *Prog. Oceanogr.* 79, 106–119. doi: 10.1016/j.pocean.2008.10.013

Chelton, D. B., Schlax, M. G., Samelson, R. M., and De Szoeke, R. A. (2007). Global observations of large oceanic eddies. *Geophys. Res. Lett.* 34:L15606. doi: 10.1029/2007GL030812

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2016). *Image segmentation with deep convolutional fully connected CRFs. arXiv[Preprint]*. Available online at: https://arxiv.org/abs/1412.7062 (accessed March 23, 2019).

Chen, L. C., Papandreou, G., Schroff, F., and Adam, H. (2017). *Rethinking atrous convolution for semantic image segmentation. arXiv[Preprint]*. Available online at: https://arxiv.org/abs/1706.05587 (accessed March 23, 2019).

Dong, C., Liu, Y., Lumpkin, R., Lankhorst, M., Chen, D., Mcwilliams, J. C., et al. (2011a). A scheme to identify loops from trajectories of oceanic surface drifters: an application in the Kuroshio Extension Region. *J. Atmos. Ocean. Tech.* 28, 1167–1176. doi: 10.1175/JTECH-D-10-05028.1

Dong, C., Mavor, T., Nencioli, F., Jiang, S., Uchiyama, Y., Mcwilliams, J. C., et al. (2009). An oceanic cyclonic eddy on the lee side of Lanai Island, Hawai'i. *J. Geophys. Res.* 114:C10008. doi: 10.1029/2009JC005346

Dong, C., McWilliams, J. C., Liu, Y., and Chen, D. (2014). Global heat and salt transports by eddy movement. *Nat. Commun.* 5:3294. doi: 10.1038/ncomms4294

Dong, C., Nencioli, F., Liu, Y., and McWilliams, J. C. (2011b). An automated approach to detect oceanic eddies from satellite remotely sensed sea surface temperature data. *IEEE Geosci. Remote Sens.* 8, 1055–1059. doi: 10.1109/LGRS.2011.2155029

Du, Y., Song, W., He, Q., Huang, D., Liotta, A., and Su, C. (2019). Deep learning with multi-scale feature fusion in remote sensing for automatic oceanic eddy detection. *Inform. Fusion* 49, 89–99. doi: 10.1016/j.inffus.2018.09.006

Duo, Z., Wang, W., and Wang, H. (2019). Oceanic mesoscale eddy detection method based on deep learning. *Remote Sens.* 11:1921. doi: 10.3390/rs11161921

Faghmous, J. H., Chamber, Y., Boriah, S., Liess, S., and Mesquita, S. (2012). "A novel and scalable spatio-temporal technique for ocean eddy monitoring," in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, (Menlo Park, CA: AAAI).

Franz, K., Roscher, R., Milioto, A., Wenzel, S., and Kusche, J. (2018). "Ocean eddy identification and tracking using neural networks," in *Proceedings of the International Geoscience and Remote Sensing Symposium 2018*, (Valencia).

He, K., Zhang, X., Ren, S., and Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1904–1916.

Lguensat, R., Sun, M., Fablet, R., Mason, E., Tandeo, P., and Chen, G. (2017). EddyNet: a deep neural network for pixel-wise classification of oceanic eddies. *arXiv[Preprint]*. doi: 10.1109/IGARSS.2018.8518411

Liu, Y., Dong, C., Guan, Y., Chen, D., McWilliams, J., and Nencioli, F. (2012). Eddy analysis in the subtropical zonal band of the North Pacific Ocean. *Deep Sea Res. 1 Oceanogr. Res. Pap.* 68, 54–67. doi: 10.1016/j.dsr.2012.06.001

Moreau, S., Penna, A. D., Llort, J., Patel, R., Langlais, C., Boyd, P. W., et al. (2017). Eddy-induced carbon transport across the Antarctic Circumpolar Current. *Glob. Biogeochem. Cycles* 31, 1368–1386. doi: 10.1002/2017GB005669

Nencioli, F., Dong, C., Dickey, T., Washburn, L., and McWilliams, J. C. (2010). A vector geometry-based eddy detection algorithm and its application to a high-resolution numerical model product and high-frequency radar surface velocities in the Southern California Bight. *J. Atmos. Ocean. Tech.* 27, 564–579. doi: 10.1175/2009JTECHO725.1

Patel, R. S., Llort, J., Strutton, P. G., Phillips, H. E., Moreau, S., Pardo, P. C., et al. (2020). The biogeochemical structure of Southern Ocean mesoscale eddies. *J. Geophys. Res. Oceans* 125:e2020JC016115. doi: 10.1029/2020JC016115

Patel, R. S., Phillips, H. E., Strutton, P. G., Lenton, A., and Llort, J. (2019). Meridional heat and salt transport across the subantarctic front by cold-core eddies. *J. Geophys. Res. Oceans* 124, 981–1004. doi: 10.1029/2018JC014655

Xu, G., Cheng, C., Yang, W., Xie, W., Kong, L., Hang, R., et al. (2019). Oceanic eddy identification using an AI scheme. *Remote Sens.* 11:1349. doi: 10.3390/rs11111349

Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., and Sang, N. (2018). *Bisenet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. arXiv[Preprint]*. Available online at: https://arxiv.org/abs/1808.00897 (accessed May 15, 2019).

Yu, F., and Koltun, V. (2016). *Multi-Scale Context Aggregation by Dilated Convolutions. arXiv[Preprint]*. Available online at: https://arxiv.org/abs/1511.07122 (accessed May 15, 2019).

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). "Pyramid scene parsing network," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Geneva).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 17 June 2021
doi: 10.3389/fmars.2021.680079

ConvLSTM-Based Wave Forecasts in the South and East China Seas

*Shuyi Zhou[1], Wenhong Xie[1], Yuxiang Lu[1], Yuanlin Wang[1], Yulong Zhou[1], Nian Hui[1] and Changming Dong[1,2]**

[1] School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing, China, [2] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China

Numerical wave models have been developed for the wave forecast in last two decades; however, it faces challenges in terms of the requirement of large computing resources and improvement of accuracy. Based on a convolutional long short-term memory (ConvLSTM) algorithm, this paper establishes a two-dimensional (2D) significant wave height (SWH) prediction model for the South and East China Seas trained by WaveWatch III (WW3) reanalysis data. We conduct 24-h predictions under normal and extreme conditions, respectively. Under the normal wave condition, for 6-, 12-, and 24-h forecasting, their correlation coefficients are 0.98, 0.93, and 0.83, and the mean absolute percentage errors are 15, 29, and 61%. Under the extreme condition (typhoon), for 6 and 12 h, their correlation coefficients are 0.98 and 0.94, and the mean absolute percentage errors are 19 and 40%, which is better than the model trained by all the data. It is concluded that the ConvLSTM can be applied to the 2D wave forecast with high accuracy and efficiency.

Keywords: ConvLSTM, wave forecasting, significant wave height, typhoon, deep learning

OPEN ACCESS

Edited by:
Zhengguang Zhang,
Ocean University of China, China

Reviewed by:
Yuping Guan,
South China Sea Institute of Oceanology (CAS), China
Yibin Ren,
Institute of Oceanology (CAS), China

***Correspondence:**
Changming Dong
cmdong@nuist.edu.cn

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *13 March 2021*
Accepted: *26 May 2021*
Published: *17 June 2021*

Citation:
Zhou S, Xie W, Lu Y, Wang Y, Zhou Y, Hui N and Dong C (2021) ConvLSTM-Based Wave Forecasts in the South and East China Seas. Front. Mar. Sci. 8:680079. doi: 10.3389/fmars.2021.680079

INTRODUCTION

Ocean surface gravity waves (hereinafter, waves) are strongly non-linear and significantly affect ocean engineering activities, maritime operations, and transportation. Traditional wave forecasting models have been continuously developed and improved. Currently, the most widely used models are the WaveWatch III (WW3) by the US National Centers for Environmental Prediction, the Simulating Waves Nearshore (SWAN) by the Netherlands Delft University, and so on. The traditional numerical wave models are based on the wave action balance equation and adopt gridded discretization instead of differential equations. This inevitably introduces numerical errors and faces problems such as non-convergence and instability in numerical computations. Additionally, numerical models are highly sensitive to the simulated area terrain (especially in shallow nearshore waters) and computational domain boundaries. They also require input information of many other variables, such as wind data (Niu and Feng, 2021). Uncertainties of these external variables lead to additional model errors, which further affect the model's accuracy. Numerical models also consume large amounts of computational resources and need to run for long periods of time, which is often impractical in emergency situations and thus is a significant bottleneck that restricts the development of fast and accurate wave forecasts.

With the rapid development of artificial intelligence (AI), due to its applicability across diverse fields and the ability to consider non-linearities in complex physical mechanisms, AI techniques have been widely applied in the field of marine sciences. These range from the automatic detection

and prediction of mesoscale eddies (Zeng et al., 2015; Xu et al., 2019), El Niño–Southern Oscillation, Arctic sea ice density, and sea surface temperature prediction (Aparna et al., 2018; Kim et al., 2018, 2020; Ham et al., 2019; Zheng et al., 2020). Wave forecasting has also been attempted through AI techniques though this is mostly a single-point wave forecasting. For example, Kaloop et al. (2020) combined wavelets, particle swarm optimization (PSO), and extreme learning machine (ELM) to create a joint wavelet-PSO-ELM (WPSO-ELM) model and found that in the case of lower complexity and fewer input variables, 36-h wave height forecasts for coastal and in the offshore areas have higher prediction accuracies. Londhe and Panchang (2006) used an artificial neural network (ANN) based on existing wave datasets to predict wave heights at six geographically separated buoy locations. This article uses ANN technology to reproduce ocean surface wave observed by buoys for 24 h. It is found that the method has a good forecast for the next 6 h, and the correlation between the observation and forecast for the next 12 h can reach 67%. Emmanouila et al. (2020) improved the numerical prediction of SWH by using a Bayesian network (BN).

Recently, the long short-term memory (LSTM) network has been applied to wave forecasting applications. The LSTM was proposed by Hochreiter and Schmidhuber (1997), which has many advantages over other networks. For example, it can selectively choose to remember or forget long-term information through a series of gates, which is very useful in the study of waves that evolve rapidly in space and time. Its usage has seen application where Lu et al. (2019) combined the LSTM network and multiple linear regression to establish an M-LSTM hybrid forecast model that limits a single predictor, thereby optimizing wave height forecasts. Fan et al. (2020), by contrast, coupled LSTM and SWAN for single-point forecasting and found that this model has better forecasting performances than models such as ELM and SVM. Additionally, in their combined SWAN-LSTM model, forecast accuracy was increased by 65% compared to using SWAN alone.

From the above literature review, we can see that the application of AI in ocean wave forecasting is still largely limited to single-point forecasting. However, a wave field is two-dimensional (2D), and few AI predictions of 2D wave fields have been reported. This paper intends to use the convolutional LSTM (ConvLSTM) algorithm recently proposed by Shi et al. (2015) to perform AI forecasting of the 2D wave field, thus adding to the available literature on its efficacy. ConvLSTM has been successfully applied to 2D precipitation nowcasting (Shi et al., 2015). It shows good spatiotemporal correlation and is always better than the fully connected LSTM (FC-LSTM) network and thus solves the problem of spatial information loss and improves the accuracy of 2D predictions. Presently, ConvLSTM has also been applied to human behavior recognition (Majd and Safabakhsh, 2019), dynamic gesture recognition (Peng et al., 2020), and stock prediction (Lee and Kim, 2020). Its application to the short-term prediction of waves is limited to a study conducted by Choi et al. (2020) that estimated wave height from raw images provided by buoys. This paper adds to the literature by conducting the short-term prediction of the 2D wave field by applying the ConvLSTM network in the South and East China Seas. The remainder of this paper is structured as follows: Section "Materials and Methods" describes the materials and methodology, including the ConvLSTM network and model evaluation methods, and model training and verification materials used in this study. Section "Results" presents the results, mainly discussing the forecast results under different sea conditions using ConvLSTM, and Section "Discussion" concludes with the discussion.

MATERIALS AND METHODS

Materials

Significant Wave Height Reanalysis Product

In this study, significant wave height (SWH) data are obtained from the WW3 third-generation numerical wave model reanalysis dataset produced by the National Oceanic Atmospheric Administration (NOAA)[1]. This reanalysis product is used to train and validate the ConvLSTM network. Usage of this product is justified as researchers have extensively validated the dataset and found that it is in good agreement with observations (Mondon and Warner, 2009; Zheng and Li, 2015; Triasdian et al., 2019). The study area is defined as the coastal waters in the northwestern Pacific Ocean enclosed by 105° E to 126° E and 4° N to 43° N. The study period is selected from 2011 to 2019. The temporal resolution of the data is hourly, and the spatial resolution is 1/2° 1/2°.

Selected Typhoons

Typhoons (those systems that reached a maximum Beaufort wind force of 12–13, and a central wind speed of 32.7–41.4 m/s) that entered the study area enclosed by 105°E to 126°E and 4°N to 43°N over the period 2011–2019 were selected to generate a typhoon-induced SWH dataset. Typhoon data were acquired from the Central Meteorological Observatory[2]. The dataset contains a total of 64 typhoons, of which 51 are used in a training set and the remainder used as test sets (**Table 1**).

Methodology

Convolutional Long Short-Term Memory Network

The LSTM is a special type of recurrent neural network (RNN). The basic idea of the LSTM is to control the input and output of information in the cell by introducing three gates: input, output, and forget gates. These are used to control the flow of information between the cells. Respectively, the input gate determines the value to be updated, the output gate mainly controls the information transmission to the next cell, and the forget gate selectively forgets the information in the information transfer. The LSTM has two states, cell state (c_t) and hidden state (h_t), which are related to c_{t-1} and h_{t-1} of the previous cell (Hochreiter and Schmidhuber, 1997). These structural features enable the LSTM to learn long-term temporal information and avoid long-term dependence problems. Since there are

[1]https://coastwatch.pfeg.noaa.gov/erddap/griddap/NWW3_Global_Best.html
[2]http://typhoon.nmc.cn/web.html

TABLE 1 | Partitioning of typhoons into training and validation sets.

Partitioning	Number	Typhoons
Training	51	Songda (2011), Muifa (2011), Nanmadol (2011), Nesat (2011), Nalgae (2011), Vicente (2012), Saola (2012), Damrey (2012), Haikui (2012), Kai-tak (2012), Tembin (2012), Bolaven (2012), Jelawat (2012), Son-tinh (2012), Bopha (2012), Soulik (2013), Utor (2013), Trami (2013), Usagi (2013), Wutip (2013), Fitow (2013), Nari (2013), Krosa (2013), Haiyan (2013), Neoguri (2014), Rammasun (2014), Matmo (2014), Kalmaegi (2014), Noul (2015), Chan-hom (2015), Linfa (2015), Soudelor (2015), Goni (2015), Dujuan (2015), Mujigae (2015), Nepartak (2016), Nida (2016), Meranti (2016), Malakas (2016), Megi (2016), Chaba (2016), Sarika (2016), Haima (2016), Nock-ten (2016), Nesat (2017), Hato (2017), Pakhar (2017), Talim (2017), Doksuri (2017), Khanun (2017), Damrey (2017)
Test	13	Tembin (2017), Prapiroon (2018), Soulik (2018), Mangkhut (2018), Trami (2018), Kong-rey (2018), Lekima (2019), Lingling (2019), Tapah (2019), Mitag (2019), Nakri (2019), Kammuri (2019), Phanfone (2019)

many ways to propagate gradients in the LSTM, vanishing and exploding gradient problems can be better avoided. The following are the information state transfer formulae in one cell of the LSTM:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \quad (2)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \circ c_t + b_o) \quad (4)$$

$$h_t = o_t \circ tanh(c_t) \quad (5)$$

where i_t represents the input gate, f_t represents the forget gate, o_t represents the output gate, c_t represents the state of the current moment, c_{t-1} represents the state of the previous moment, h_t represents the final output, W represents the weight coefficient for a given gate, b represents the corresponding bias coefficient for a given gate, $\circ$ is the Hadamard product, and σ is the sigmoid function.

Presently, the LSTM is widely used in time series forecasting. However, when the LSTM is applied to 2D data, if it is expanded into full connected layer processing, it not only consumes substantial computing resources but also it is difficult to capture the spatial correlation and spatial characteristics of the 2D space field (Shi et al., 2015). To overcome these deficiencies, Shi et al. (2015) replaced the FC-LSTM layers with a convolutional structure, leading to the development of the ConvLSTM network. The primary difference between the LSTM and the ConvLSTM is the replacement of matrix multiplication by a convolutional operation:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \quad (6)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hj} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \quad (7)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ tanh(W_{xc} * X_t + W_{hc} * h_{t-1} + b_c) \quad (8)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * h_{t-1} + W_{co} \circ c_t + b_o) \quad (9)$$

$$h_t = o_t \circ tanh(c_t) \quad (10)$$

where $*$ is convolution operator.

Convolution operation can extract the spatial characteristics of the data, while the LSTM can extract the temporal variability of the data. Therefore, the ConvLSTM has the ability to well depict both a variable's spatiotemporal characteristics and is hence highly suitable for regional ocean wave predictions.

Architecture of the ConvLSTM Model for Wave Forecasts

In this paper, a regional wave prediction model is established based on the ConvLSTM network (**Figure 1**). The SWH data of three continuous time steps are taken as the input data. The SWH data at a certain time in the future are output through three ConvLSTM layers and finally through a convolution layer for a total of four layers. For example, SWH at times 13:00, 14:00, and 15:00 on January 1, 2018 is given (input) to the model and SWH at time 16:00 is predicted (output). To improve the model's ability to capture non-linearities, the recursive linear unit (ReLU) is employed as the activation function in each layer, with the hard sigmoid used as the activation function in the loop step. The convolutional kernels of each of four layers are set to $5 * 5$, $3 * 3$, $3 * 3$, and $5 * 5$, respectively, to capture different characteristics at different spatial scales. The root-mean-square error (RMSE) is used as the loss function during model training, the number of epochs is set to 100, and all other remaining parameters remain constant throughout all training exercises.

Data Pre-processing

To improve the training dataset quality, the WW3 SWH reanalysis product is linearly interpolated from a resolution of $1/2° * 1/2°$ to $1/4° * 1/4°$. The input data are wave field data of three consecutive time steps. According to the different prediction time steps (e.g., 1, 2, and 3 h), the corresponding training dataset and verification set are generated using the data from 2011 to 2018, in which the data volume ratio of training set and verification set is 4:1. For predictions of SWH, the data for the year of 2019 are the test set, which are excluded from the model training to ensure relative independence between the training and test datasets.

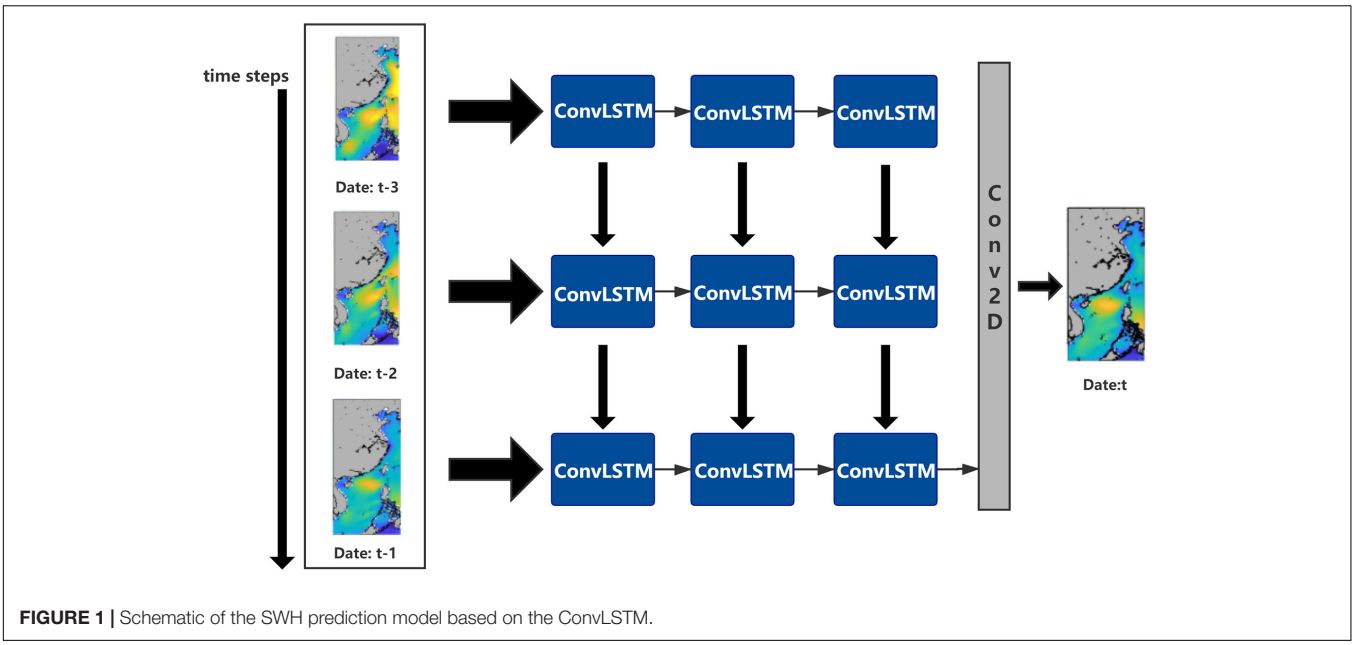


FIGURE 1 | Schematic of the SWH prediction model based on the ConvLSTM.

Evaluation Functions

In order to better evaluate the accuracy of model forecasts, this article defines the following evaluation functions. Difference error (DR), mean absolute error (MAE), RMSE, spatially averaged RMSE (SARMSE), and spatially averaged mean absolute percentage error (SAMAPE) are used to evaluate the deviation between predicted values and WW3 reanalysis data. In addition, the spatially averaged correlation coefficient (SACC) is used to measure the linear correlation between the predicted values and WW3 values. The expressions of the above variables are as follows:

$$DR(i,j) = h_p(i,j) - h_m(i,j) \tag{11}$$

$$MAE = \frac{1}{I \cdot J}\sum_{i=1}^{I}\sum_{j=1}^{J}\left|h_p(i,j) - h_m(i,j)\right| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{I \cdot J}\sum_{i=1}^{I}\sum_{j=1}^{J}(h_p(i,j) - h_m(i,j))^2} \tag{13}$$

$$SARMSE = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{I \cdot J}\sum_{i=1}^{I}\sum_{j=1}^{J}(h_p(i,j,k) - h_m(i,j,k))^2} \tag{14}$$

$$SAMAPE = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{I \cdot J}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left|h_p(i,j,k) - h_m(i,j,k)\right|}{h_m(i,j,k)} \times 100\% \tag{15}$$

$$SACC = \frac{\frac{1}{n}\sum_{k=1}^{n}(h_p(i,j,k) - \overline{h_p(i,j,k)})(h_m(i,j,k) - \overline{h_m(i,j,k)})}{\sqrt{\frac{1}{n}\sum_{k=1}^{n}(h_p(i,j,k) - \overline{h_p(i,j,k)})^2}\sqrt{\frac{1}{n}\sum_{k=1}^{n}(h_m(i,j,k) - \overline{h_m(i,j,k)})^2}} \tag{16}$$

where i and j denote the coordinates of space lattice points, k denotes cases, n represents the total number of cases, I denotes the total number of latitudinal lattice points, and J denotes the total number of meridional lattice points. $DR(i,j)$ is the error value of a certain point in space, $h_p(i,j)$ is the SWH value predicted based on the ConvLSTM model, and $h_m(i,j)$ is the WW3 SWH value corresponding to a certain point in space. $h_p(i,j,k)$ represents the ConvLSTM model-predicted SWH at a certain point in the case space, $\overline{h_p(i,j,k)}$ represents the mean SWH predicted by the ConvLSTM model at a certain point in the case space, and $h_m(i,j,k)$ represents the SWH of WW3 at a certain point in the case space. $\overline{h_m(i,j,k)}$ represents the mean SWH of WW3.

RESULTS

In this study, the ConvLSTM is applied to wave forecasting in the northwestern Pacific Ocean under both normal and extreme conditions. This section is divided into two subsections. In Section "Wave Forecast Under Normal Conditions," the test set results are presented and discussed for the normal condition. In Section "Wave Forecast Under Extreme Conditions," the wave forecast is presented for extreme condition (i.e., typhoon cases).

Wave Forecast Under Normal Conditions

First, we discuss the performance of the ConvLSTM algorithm in predicting SWH under normal conditions. **Figures 2**, **3** show

examples of the ConvLSTM wave forecasts, which all use SWH data at 5:00, 6:00, and 7:00 (UTC) on October 3, 2019, as inputs. **Figures 2A,D,G** correspond to the spatial distribution of the forecasted SWH after 1, 3, and 6 h. **Figures 2B,E,H** show the spatial distribution of the WW3 data that corresponds to the forecast moments. The errors between the forecast and the WW3 data are displayed in **Figures 2C,F,I**. In the 1-h forecast (**Figures 2A,B**), the model results show good consistency with the baseline in terms of both the magnitude of the wave height and the spatial distribution. In general, the 1-h forecast error fluctuates within −0.1 to 0.1 m (**Figure 2C**). Relatively high values are observed in the open sea and Bohai Sea. Low SWH covers the rest of the domain. A slightly larger error can be seen only in the Yellow Sea. We can identify that when forecast time span is increased to 3 h, a high degree of accuracy is obtainable in the Yellow and East China Seas (**Figures 2D,E**). The large deviations of the forecasts from the WW3 correspond to the high SWH area in the Bohai Sea (**Figures 2G,H**), with a large forecast error also noted in **Figure 2I**. The 6-h wave forecasts underestimate the SWH noticeably in the Bohai and South China Seas, while the forecasts overestimate the SWH in the Yellow and the East China Seas. **Figure 3** shows the predicted and the WW3 SWH at each sample point in space at the three moments from **Figure 2** from low to high latitudes. It can be found that the predicted spatial distribution of SWH at these three moments is basically consistent with that from the WW3. From **Figure 3A**, the 1-h forecast has the best performance, a good consistency is observed between forecast and the WW3 data with MAE, RMSE, and the correlation at 0.03 m, 0.04 m, and 0.997, respectively. The forecast values are smaller than the WW3 baseline. A similar pattern can also be seen in **Figure 3B**. This deviation causes the larger MAE and RMSE of the 3-h forecast compared with the 1-h forecast, and the correlation decreases to 0.991. At the 6-h window, the least accurate forecast was observed that had a correlation coefficient of 0.962.

To further illustrate this model's performance, it is necessary to discuss the impact of the forecast time span on forecast errors. The forecast time span is the time scale of wave forecasting. **Figure 4** shows the error variations with the forecast time span in 2019, which is based on the training and validation sets for the period from 2011 to 2014. It can be seen that the SARMSE increases as the forecast time span increases, while this trend is reversed for SACC. The SARMSE for the forecast time span of less than 6 h is less than 0.2 m, with the SACC close to 1.0, the forecast accuracy is still high, so in this case the spatial characteristics of the wave field from the training set at the three moments can better respond to the changing wave field trends of the next 6 h that ensure a relatively high forecast accuracy. When the forecast time span increases to 12 h, the SARMSE gradually increases to 0.29 m. However, when the forecast time span exceeds 12 h and increases to 24 h, the model still keeps the SACC around 0.8, but the SARMSE is close to 0.6 m. The adaptability of the training samples to forecast larger time spans decreases, which may be due to a great number of unknown factors encumbering the training set from accurately and adequately responding to a rapidly evolving wave field. The above results indicate that the model still requires further experiments, testing and evaluation to improve forecast beyond 12 h. The selection of training samples can be adjusted to improve the model performance, which will be discussed in Section "Discussion." Therefore, it can be concluded that the larger the forecast time span, the larger SARMSE and the smaller SACC.

In the preceding paragraph, the SARMSE is presented for the discussion of the model performance. To better represent the model errors, the changes of SAMAPE with the forecast time span are listed in **Table 2**, which shows that the changes of SAMAPE are similar to those of the SARMSE. The SAMAPE increases and the forecast accuracy decreases as the forecast time span increases (**Table 2**). When the forecast time span is 3 h, the SAMAPE is only 11.1%, and when the forecast time span increases to 24 h, the SAMAPE is 62.8%.

Wave Forecast Under Extreme Conditions

Typhoons, generated over tropical and subtropical oceans, produce intense surface wind speed that can force wind waves to grow to approximately 10 m in height. Along the coast of China, typhoons mainly affect the South China Sea in May, June, and October to December, and usually affect southeastern coastlines and even the coastal areas in the East China Sea from July to September (Lu and Qian, 2012). Due to the large discrepancy between normal and extreme (typhoon-induced) wave states, in conjunction with the low frequency of typhoons in proportion to the full dataset, the ConvLSTM trained by the data under the normal conditions may fail to learn all typhoon characteristics and so typhoon-induced SWH forecasts may be flawed. In this section, we propose that the ConvLSTM can be trained by the typhoon-induced wave data for the wave forecast under the extreme conditions.

To verify this hypothesis, we have defined Model 1 and Model 2, where Model 1 is trained by the normal-condition wave data and Model 2 is trained by typhoon-induced waves. For the experiments, their forecast time spans are held at a 3-h constant. **Figure 5** shows the forecast effect of using Model 1 and Model 2 using snapshots of Typhoons Lekima that occurred in the East China Sea at 18:00 on August 8, 2019 and Tembin that occurred in the South China Sea at 21:00 on December 24, 2017. Comparing **Figures 5A,D**, with **Figures 5B,E**, we can find that although Model 1 is able to capture the primary features of the wave field near the typhoon center, the forecast results at the center are generally smaller than the WW3, large SWH generated near the center is not adequately reproduced. Far from the typhoon center where the system has a reduced impact on the waves, Model 1 can more adequately capture SWH patterns than at the typhoon center. By contrast, Model 2 is better able to capture the spatial patterns of SWH during typhoons as observed from **Figures 5B,E**. This, however, comes at a minor cost of slightly overestimating SWH at the typhoon center. That is, the area around the highest values in **Figures 5B,E** are slightly larger than that in **Figures 5C,F**. There are insufficient examples of typhoon-induced SWH in the Model 1 dataset due

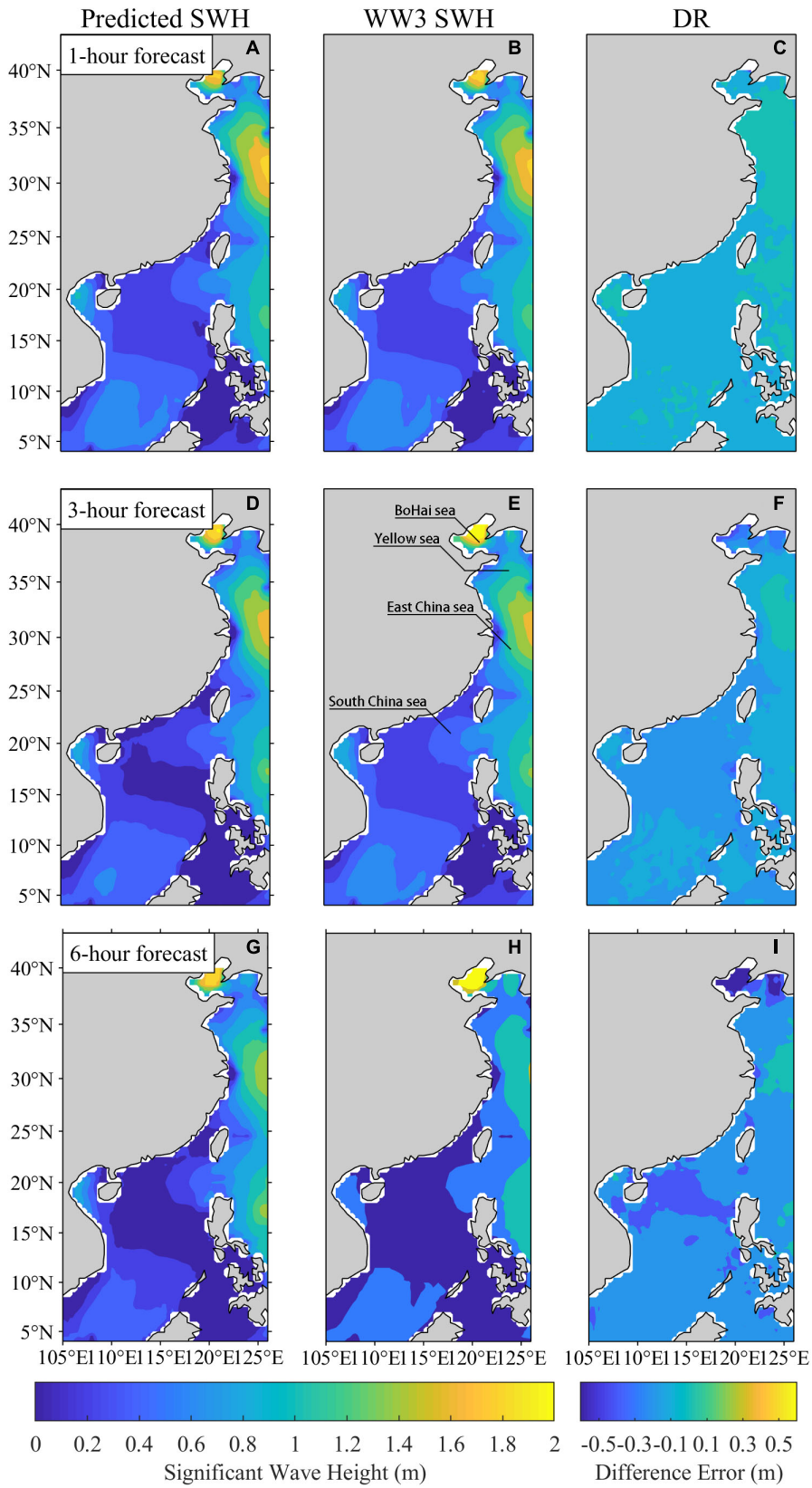


FIGURE 2 | Comparison of SWH from the WW3 and the ConvLSTM algorithm prediction. **(A,D,G)** are 1-, 3-, and 6-h predictions, respectively, based on the SWH data at 5:00, 6:00, and 7:00 on October 3, 2019. **(B,E,H)** The WW3 wave fields at the corresponding forecast moments. **(C,F,I)** are the difference error between the WW3 data and the predictions.

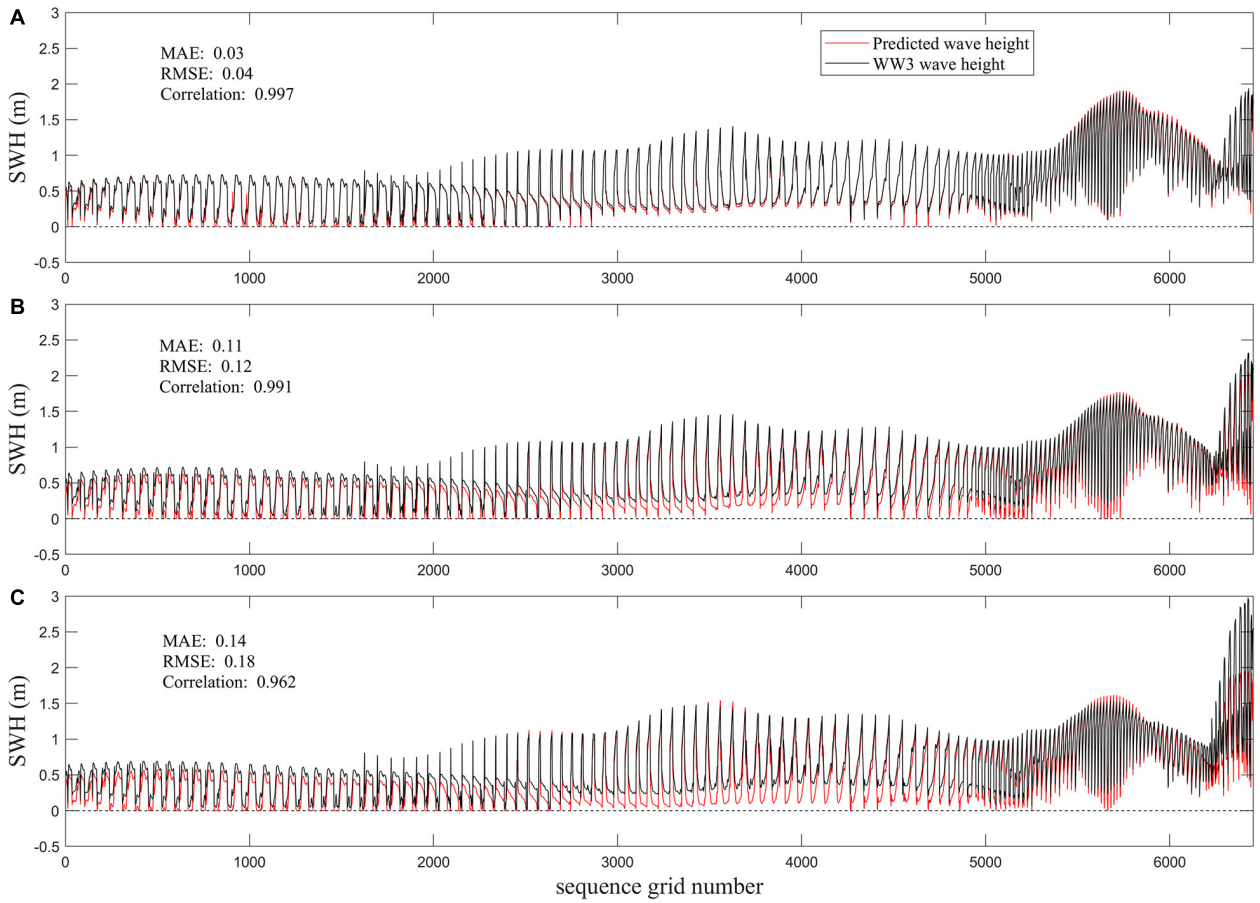


FIGURE 3 | Predicted (red line) vs. the WW3 (black line) SWH throughout the study area for **(A)** 1-, **(B)** 3-, and **(C)** 6-h forecast moments.

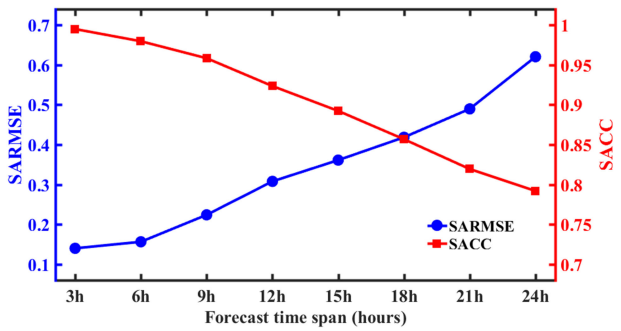


FIGURE 4 | Variations in the SARMSE (blue) and SACC (red) of each training result with forecast time span.

to the low frequency and short duration of typhoons and thus in all training datasets, there were insufficient examples of typhoon-induced SWH for Model 1 to learn from. Consequently, typhoon characteristics are difficult to be extracted from Model 1. Generally, both in the center of the typhoon and surrounding areas, wave forecasts were greatly improved in Model 2 as compared to Model 1.

To further illustrate the difference in the wave forecasts with Model 1 and Model 2 under extreme state, changes of the forecast time span of Model 1 and Model 2 are examined. The error statistic results (SACC, SARMSE, and SAMAPE) are shown in **Figures 6A–C**, respectively. In **Figure 6A**, SACC of both Model 1 and Model 2 decreases almost linearly with the increase of the forecast time span, but the SACC of Model 2 is always higher than that of Model 1. In **Figures 6B,C**, both SARMSE and SAMAPE of Model 1 and Model 2 show an almost linear increasing trend. The SARMSE of Model 2 within 6 h is less than 0.5 m and the SAMAPE is less than 20%, while the SARMSE from Model 1 is at about 1 m and the SAMAPE at approximately 40%. When the forecast time span is 24 h, the SAMAPE of Model 2 becomes slightly higher than that of Model 1.

In summary, for within the 24-h forecast under typhoon conditions, Model 2 shows better forecast capability than that of Model 1. Compared with Model 1, the prediction error

TABLE 2 | Changes of SAMAPE with forecast time span.

Lead time	3 h	6 h	9 h	12 h	15 h	18 h	21 h	24 h
SAMAPE	12.8%	15.0%	16.9%	29.0%	27.4%	29.0%	46.5%	61.4%

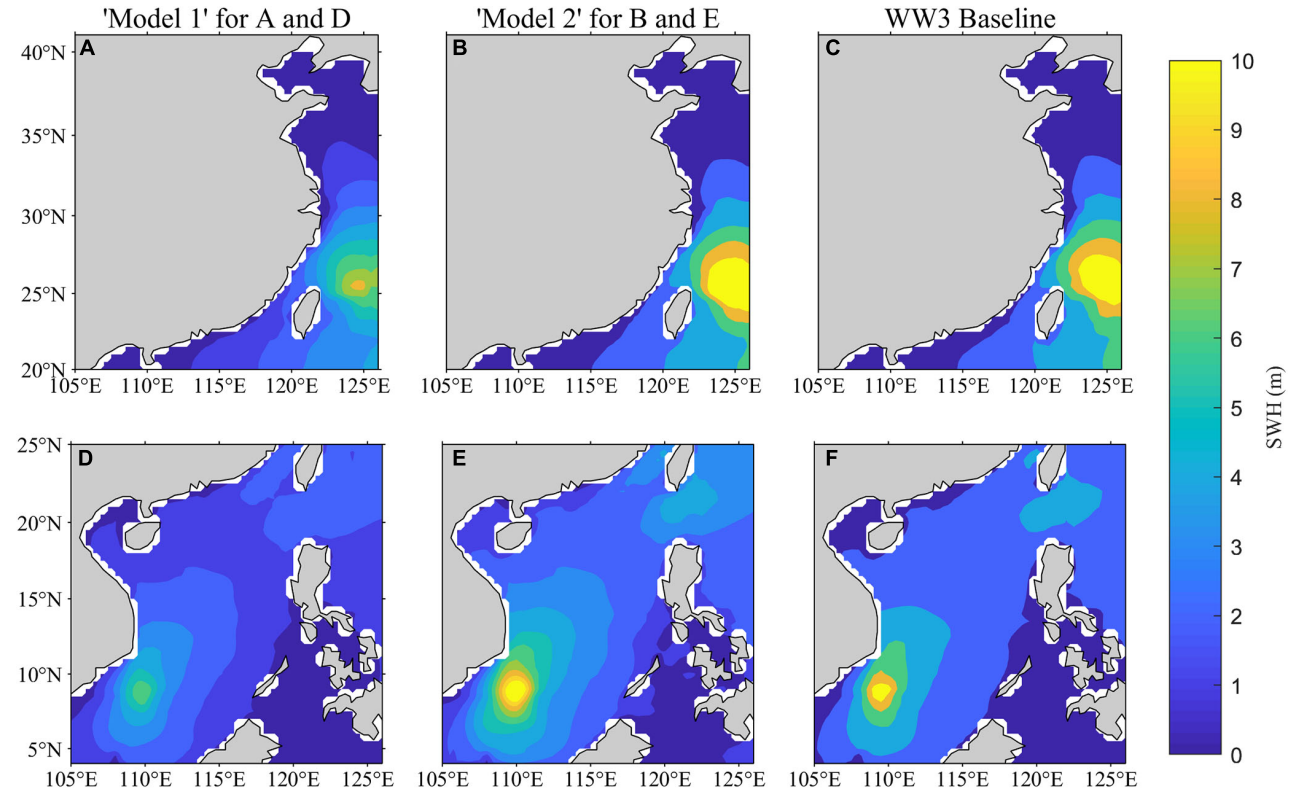


FIGURE 5 | The SWH prediction results of Model 1 and Model 2 under typhoon conditions in the East China Sea and South China Sea. **(A–C)** SWH of Typhoon "Lekima" over the East China Sea at 18:00 on August 8, 2019. **(D–F)** SWH of Typhoon "Tembin" over the South China Sea at 21:00 on December 24, 2017.

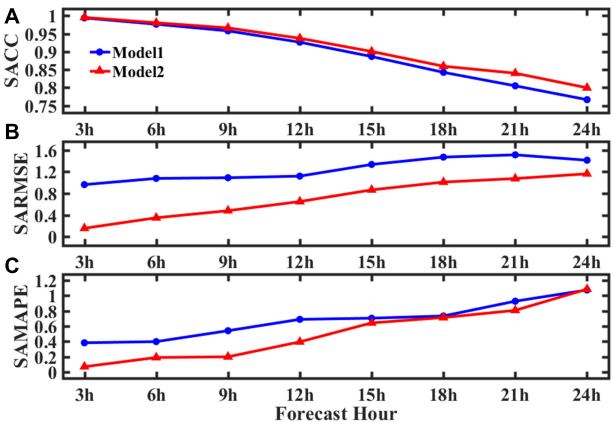


FIGURE 6 | Error statistics of Models 1 and 2 relative to the WW3 baseline. **(A)** SACC, **(B)** SARMSE, and **(C)** SAMAPE.

of Model 2 is smaller and the prediction in the typhoon-affected area is more accurate. This is because the training set of Model 1 is mainly from the wave data under normal conditions including few typhoon-induced wave data. As a result, Model 1 cannot accurately capture wave characteristics under typhoon conditions.

DISCUSSION

In this paper, an intelligent forecast model for waves in the South China Sea and East China Sea based on the ConvLSTM algorithm is established. The model relies on wave hindcasts as input to forecast wave spatial distributions. It can be seen from the results in Section "Results" that the prediction of the SWH based on the ConvLSTM is feasible under normal and extreme conditions. The 1–12-h results are acceptable, with relatively larger errors in 24-h forecasts.

Figure 7 discusses the model's performance from the input data time span and the training sets time span, respectively. We first discuss the input data time span that represents the input data quantity. For example, "2 h" (**Figure 7A**) means the continuous input of 2-h wave field data for forecasting. **Figure 7A** shows the forecast error results against the input data time span. When the input data time span is 2 h, the SACC is low (about 0.97), and the SARMSE is large (about 0.22). Since the data volume of 2-h wave field data is small, the model reproduces few wave features during the training process, which leads to low forecast accuracy. As the input data time span increases, the SACC improves, and the SARMSE gradually decreases, stabilizes, but increases at 5-h mark. The reason is that if the amount of input data is too much, it causes data redundancy, which cannot further improve

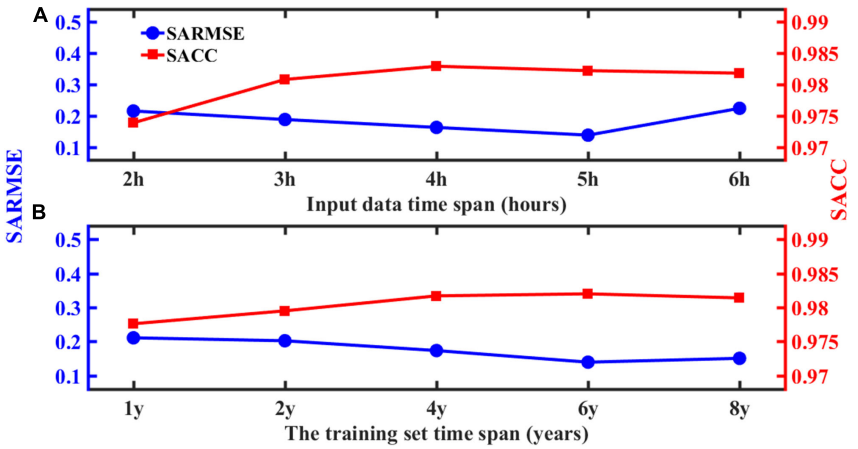


FIGURE 7 | (A) Changes in SARMSE and SACC with the time span of the input data. **(B)** Changes in SARMSE and SACC with the time span of the training set data.

forecast accuracy. Therefore, it can be seen that 3 h is the optimal selection for the input data time span in terms of the forecast accuracy and computing time cost. This selection of the input data time span guarantees the forecast accuracy, in addition to saving computational resources and minimizing the amount of time required for computations to ensure the forecast timeliness.

Because of the difference in the characteristics of the different training set time span, it is necessary to find the optimal training set time span. We set the forecast time span to 6 h and then select the same number of data samples under different time spans as training sets. **Figure 7B** shows the error results under different training set time spans. Generally, with the increase in the training set time span, the SARMSE has a downward trend, and the SACC has an upward trend. When the training set time span is 1y, the SARMSE is 0.21 m, and the SACC is about 0.97. As the training set time span increases, the SARMSE gradually decreases and stabilizes at about 0.15 m. When the training set time span exceeds 4 years, the model performance has not greatly been improved. So, this study chooses 4 years as the training set time span. Therefore, combining the input data and the training set time span, the SWH prediction based on the ConvLSTM does not require a very large training set, nor does it require long-term data as input. So, this method not only improves the forecast efficiency but also greatly saves computational resources. This model is trained on the GeForce RTX 2080 Ti, the training of a single model takes about 2 h, and it only takes less than 20 s to complete the prediction of the test set in 2019. Therefore, the SWH prediction based on the ConvLSTM is feasible and efficient.

The present study only uses SWH data to predict the SWH; however, the generation of waves is closely related to the overlying wind field, and SWH can also be predicted by wind and other variables (Wang et al., 2021), but this relationship is specific to wind waves as may be caused by typhoons. Swell contamination of the wave field can weaken the wind–wave relationship (Niu and Feng, 2021) and thus it is necessary to introduce other input variables. Consequently, in future research, additional physical phenomena such as ocean currents can be added to improve wave forecasts. Additionally, due to the coarse nature of the input data, no wave information on any coastline was available and thus restricts its operational usage for coastal communities. The usage of a high-resolution wave model can be used in place of the reanalysis dataset that can provide coastal wave information and should be pursued in future research.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://coastwatch.pfeg.noaa.gov/erddap/griddap/NWW3_Global_Best.html.

AUTHOR CONTRIBUTIONS

SZ and WX designed the experiment. SZ, YL, YW, YZ, and NH contributed to the data analysis and the writing of the first draft of the manuscript. CD is the creator and person in charge of the project, directing experimental design, data analysis, thesis writing, and revision. All authors read and agreed to the final text.

FUNDING

This study was supported by the project supported by the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (SML2020SP007), the Key Program of Marine Economy Development (Six Marine Industries) Special Foundation of Department of Natural Resources of Guangdong Province [GDNRC(2020)049], the National College Students' Platform for Innovation and Entrepreneurship

Training Program (202010300017Z), the Jiangsu Province College Students' Platform for Innovation and Entrepreneurship Training Program (202010300017Z), and the NUIST Students' Platform for Innovation and Entrepreneurship Training Program (202010300017Z).

ACKNOWLEDGMENTS

The SWH data were collected from NOAA, and the relevant typhoon data were obtained from the Central Meteorological Observatory.

REFERENCES

Aparna, S. G., D'Souza, S., and Arjun, N. B. (2018). Prediction of daily sea surface temperature using artificial neural networks. *Int. J. Remote Sens.* 39, 4214–4231. doi: 10.1080/01431161.2018.1454623

Choi, H., Park, M., Son, G., Jeong, J., Park, J., Mo, K., et al. (2020). Real-time significant wave height estimation from raw ocean images based on 2D and 3D deep neural networks. *Ocean Eng.* 201:107129. doi: 10.1016/j.oceaneng.2020.107129

Emmanouila, S., Aguilarc, S. G., Nane, G. F., and Schoutenc, J. J. (2020). Statistical models for improving significant wave height predictions in offshore operations. *Ocean Eng.* 206:107249. doi: 10.1016/j.oceaneng.2020.107249

Fan, S., Xiao, N., and Dong, S. (2020). A novel model to predict significant wave height based on long short-term memory network. *Ocean Eng.* 205:107298. doi: 10.1016/j.oceaneng.2020.107298

Ham, Y. G., Kim, J. H., and Luo, J. (2019). Deep learning for multi-year ENSO forecasts. *Nature* 573, 568–572. doi: 10.1038/s41586-019-1559-7

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural. Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Kaloop, M. R., Kumar, D., Zarzoura, F., Roy, B., and Hu, J. W. (2020). A wavelet–particle swarm optimization–extreme learning machine hybrid modeling for significant wave height prediction. *Ocean Eng.* 213:107777. doi: 10.1016/j.oceaneng.2020.107777

Kim, J., Kim, K., Cho, J., Kang, Y. Q., Yoon, H. J., and Lee, Y. W. (2018). Satellite-based prediction of arctic sea ice concentration using a deep neural network with multi-model ensemble. *Remote Sens.* 11:19. doi: 10.3390/rs11010019

Kim, Y. J., Kim, H. C., Han, D., Lee, S., and Im, J. (2020). Prediction of monthly Arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks. *Cryosphere* 14, 1083–1104. doi: 10.5194/tc-14-1083-2020

Lee, S. W., and Kim, H. Y. (2020). Stock market forecasting with super-high dimensional time-series data using ConvLSTM, trend sampling, and specialized data augmentation. *Expert. Syst. Appl.* 161:113704. doi: 10.1016/j.eswa.2020.113704

Londhe, S. N., and Panchang, V. J. (2006). One-day wave forecasts based on artificial neural networks. *J. Atmos. Ocean. Technol.* 23, 1593–1603. doi: 10.1175/JTECH1932.1

Lu, B., and Qian, W. H. (2012). Seasonal lock of rapidly intensifying typhoons over the South China offshore in early fall. *Chin. J. Geophys.* 55, 1523–1531. doi: 10.6038/j.issn.0001-5733.2012.05.009

Lu, P., Liang, S., Zou, G., Zheng, Z., and Zou, P. (2019). M-LSTM, a hybrid prediction model for wave heights. *J. Nonlinear. Convex.* A 20, 775–786.

Majd, M., and Safabakhsh, R. (2019). Correlational convolutional LSTM for human action recognition. *Neurocomputing* 396, 224–229. doi: 10.1016/j.neucom.2018.10.095

Mondon, E., and Warner, P. (2009). Synthesis of a validated nearshore operational wave database using the archived NOAA Wave watch III ocean model data and swan nearshore model. *J. Coastal Res.* 2009, 1015–1019. doi: 10.2307/25737940

Niu, Q., and Feng, Y. (2021). Relationships between the typhoon-induced wind and wave in the northern South China Sea. *Geophys. Res. Lett.* 48:e2020GL091665. doi: 10.1029/2020GL091665

Peng, Y., Tao, H., Li, W., Yuan, H., and Li, T. (2020). dynamic gesture recognition based on feature fusion network and variant convLSTM. *IET Image. Process.* 14, 2480–2486. doi: 10.1049/iet-ipr.2019.1248

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., and Woo, W. C. (2015). "Convolutional lstm network: a machine learning approach for precipitation nowcasting," In Proceedings of the 29th Annual Conference *in Neural Information Processing sSystems*, Montreal, MTL, 802–810.

Triasdian, B., Indartono, Y. S., Ningsih, N. S., and Novitasari, D. (2019). Device selection of the potential wave energy site in indonesian seas. *IOP Conf. Ser. Earth Environ. Sci.* 291:012040. doi: 10.1088/1755-1315/291/1/012040

Wang, H., Yang, J., Zhu, J., Ren, L., Liu, Y., Li, W., et al. (2021). Estimation of significant wave heights from ASCAT scatterometer data via deep learning network. *Remote Sens.* 135:195. doi: 10.3390/rs13020195

Xu, G., Cheng, C., Yang, W., Xie, W., Kong, L., Hang, R., et al. (2019). Oceanic eddy identification using an AI scheme. *Remote Sens.* 11:1349. doi: 10.3390/rs11111349

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

Zheng, C., and Li, C. (2015). Variation of the wave energy and significant wave height in the China Sea and adjacent waters. *Renew. Sust. Energ. Rev.* 43, 381–387. doi: 10.1016/j.rser.2014.11.001

Zheng, G., Li, X., Zhang, R., and Liu, B. (2020). Purely satellite data-driven deep learning forecast of complicated tropical instability waves. *Sci. Adv.* 6:1482. doi: 10.1126/sciadv.aba1482

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

BRIEF RESEARCH REPORT
published: 20 July 2021
doi: 10.3389/fmars.2021.671791

A Glider Simulation Model Based on Optimized Support Vector Regression for Efficient Coordinated Observation

*Fangjie Yu[1,2], Zhiyuan Zhuang[1], Jie Yang[1,2] and Ge Chen[1,2]**

[1] College of Information Science and Engineering, Ocean University of China, Qingdao, China, [2] Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

Multi-gliders have been widely deployed as an array in nowadays ocean observation for fine and long-term ocean research, especially in deep-sea exploration. However, the strong, variable ocean currents and the delayed information feedback of gliders are remaining huge challenges for the deployment of glider arrays which may cause that the observed data cannot meet the study needs and bring a prohibitive cost. In this paper, we develop a Glider Simulation Model (GSM) based on the support vector regression with the particle swarm optimization (PSO)-SVR algorithm to integrate the information feedback from gliders and ocean current data for rapid modeling to effectively predict the gliders' trajectories. Based on the real-time predictive information of the trajectories, each glider can select future movement strategies. We utilize the in-suit datasets obtained by sea-wing gliders in ocean observation to train and test the simulation model. The results show that GSM has an effective and stable performance. The information obtained from the modeling approaches can be utilized for the optimization of the deployment of the glider arrays.

Keywords: glider, simulation model, PSO-SVR, rapid modeling, trajectory prediction

OPEN ACCESS

Edited by:
Jun Li,
University of Technology Sydney, Australia

Reviewed by:
Shinya Kouketsu,
Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan
Xiping Jia,
Guangdong Polytechnic Normal University, China

***Correspondence:**
Ge Chen
gechen@ouc.edu.cn

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *24 February 2021*
Accepted: *14 June 2021*
Published: *20 July 2021*

Citation:
Yu F, Zhuang Z, Yang J and Chen G (2021) A Glider Simulation Model Based on Optimized Support Vector Regression for Efficient Coordinated Observation. Front. Mar. Sci. 8:671791. doi: 10.3389/fmars.2021.671791

INTRODUCTION

Underwater gliders are characterized as a type of persistent (can operate continuously for weeks to months even to years), long operation range (up to thousand kilometers), small power consumption, and unmanned marine vehicle which are propelled by buoyancy (Rudnick et al., 2004). These characteristics make them conduct long-term ocean studies and understand the inner working mechanisms of the ocean in a controllable way which are better than existing traditional instruments for ocean observation (Testor et al., 2019). In recent years, with the widespread attention aroused by researchers on ocean observation, especially on deep ocean, gliders are widely deployed in ocean studies (Liblik et al., 2016).

Furthermore, in the major ocean observation, multiple gliders are deployed as a glider array to develop surveys to optimize and refine the ocean sampling. Shu et al. (2019) deploy 12 gliders as an observation network in the northern South China Sea (NSCS) for the research on anticyclonic eddies. The observations reveal not only the fine subsurface structure of temperature and salinity but also the time evolution of the three-dimensional structure of the eddies. Li et al. (2019) not only equip the glider observation array with CTDs to reveal the three-dimensional structure of the anticyclonic eddies in NSCS but also mount the biogeochemical sensors on the gliders which

can show the influence of anticyclonic eddies on the biogeochemical structure. On this basis, a compound three-dimensional structure of the eddy is constructed.

Although glider arrays present a compelling advantage in ocean observation, there are challenging difficulties on deployment. The movements of gliders are closely related to dynamic ocean currents (Thompson et al., 2010); the high nonlinearity of disturbances may make the actual observation trajectories of gliders not consistent with the planned running trajectories, which can cause the redundant measurement or the incomplete measurement during the experiments (Zhang et al., 2007). Meanwhile, the position information of gliders will only be transmitted when the gliders come out of the water; the asynchrony of information synchronization prevents the accuracy of the simulation on the trajectory control of glider arrays. Furthermore, conducting a coordinated observation with multiple gliders in large-scale ocean experiments is expensive.

Several prior studies have focused on these problems. Rao and Williams (2009) considered the influence of ocean currents and proposed a path planning method based on rapidly exploring random trees (RRT). However, this path planning method is based on the assumption that we have accurate knowledge of future currents, which is nondeterministic in practice. Thompson et al. (2010) presented a method with historical ocean current predictions to address the glider path planning and control in the uncertain, time-varying ocean currents. Besides, to solve the problem that currents may drive glider arrays into clumps without feedback information (Leonard et al., 2007). Paley et al. (2008) designed a control system named Glider Coordinated Control System (GCCS) which, through the real-time feedback control law, coordinates the gliders and optimizes the glider measurement formations. The study of Leonard et al. (2010) confirmed the capacity of the GCCS applied to ocean sampling. Considering avoiding the collisions between gliders, a navigation system based on Networked Decentralized Model Predictive Control was developed to coordinate the gliders into the desirable formation (Fonti et al., 2011). This method has high autonomy and reliability but only considers the horizontal plane motion. Xiong et al. (2020) discussed a rapidly exploring random tree star (RRT*) method which is a variant algorithm of standard RRT to meet the need of collision avoidance and achieve continuous sampling effectively. Xue et al. (2018) also proposed a strategy of the control of hybrid gliders that merge the coordinate control model based on artificial potential fields with a motion optimization method. In a recent study, Clark et al. (2020) decoupled a vehicle motion model with a predictive ocean current model forming an adaptive planner that glider observation arrays can maintain stability.

However, most of the studies are pointing at the nonlinear mechanical model of the underwater glider and the uncertain ocean current predictive model, as the actual conditions of gliders and ocean currents can quickly diverge in practice and the existing studies are not feasible to reconfigure the glider arrays according to actual conditions in different sea states. Thus, in this paper, we design a glider simulation model (GSM), considering multiple factors, which can rapidly feedback information from gliders and provide real-time predictive information of coordination of gliders. Based on the predictive information, each glider can determine future motion strategies. In this model, we introduce support vector regression based on PSO-SVR as the supporting algorithm. Although the information feedback from gliders contains non-inear fluctuations, noise, and outliers, the PSO-SVR algorithm can well grasp the movement trend. As far as we know, few studies have employed PSO-SVR to optimize glider control and formation deployment. We also introduce three observation datasets obtained by sea-wing gliders into the simulation experiments; the results demonstrate that the GSM has great stability and effectiveness on sea-wing gliders.

The rest of the paper is made up as follows: the summary of the dataset is shown in section "Data." In section "Methods," we give a comprehensive introduction to the methods involved in the model. Section "Experiments" presents the results of the experiments using different algorithms, and a full comparison of the performance is evaluated. In the last section, we draw the conclusion and increase the future works needed to be done to improve our simulation model.

DATA

Glider Data

Sea-wing gliders are selected to verify the validity of the model (**Figure 1C**). Three sets of in-suit observation data collected by sea-wing gliders are employed to train and test the simulation model. These datasets come from two ocean observation experiments where the investigation areas are located in the northern South China Sea.

G-001 dataset: The experiment started from April 28, 2015, to June 1, 2015, employing one glider (g-001) for the cross-sectional observation of eddies, and the trajectory of the glider is shown in **Figure 1A**.

G-002 dataset: The experiment started from July 3, 2016, to July 16, 2016, and lasted for 17 days. During the experiment, a sea-wing glider (g-002) with multiple sensors is deployed for the inner observation of eddies, and the trajectory is shown in **Figure 1B**; the black line represents the trajectory of g-002.

G-003 dataset: The experiment started from July 3, 2016, to July 16, 2016, with a sea-wing glider (g-003) carrying one sensor for cross-eddy observation. We can see the trajectory in **Figure 1B**; the green line is the trajectory of g-003.

The triangles indicate the starting points of observations. The backgrounds are the sea-level anomaly in the observation areas where the higher the anomaly, the greater the possibility of the existence of eddies. The arrows indicate the ocean current field, and the size of arrows represents the strength of the geostrophic currents.

AVISO Ocean Current Dataset

The ocean current dataset used in this study was distributed by AVISO and derived from the absolute dynamic topography (ADT), which can be downloaded from the website[1].

[1]https://www.aviso.altimetry.fr/

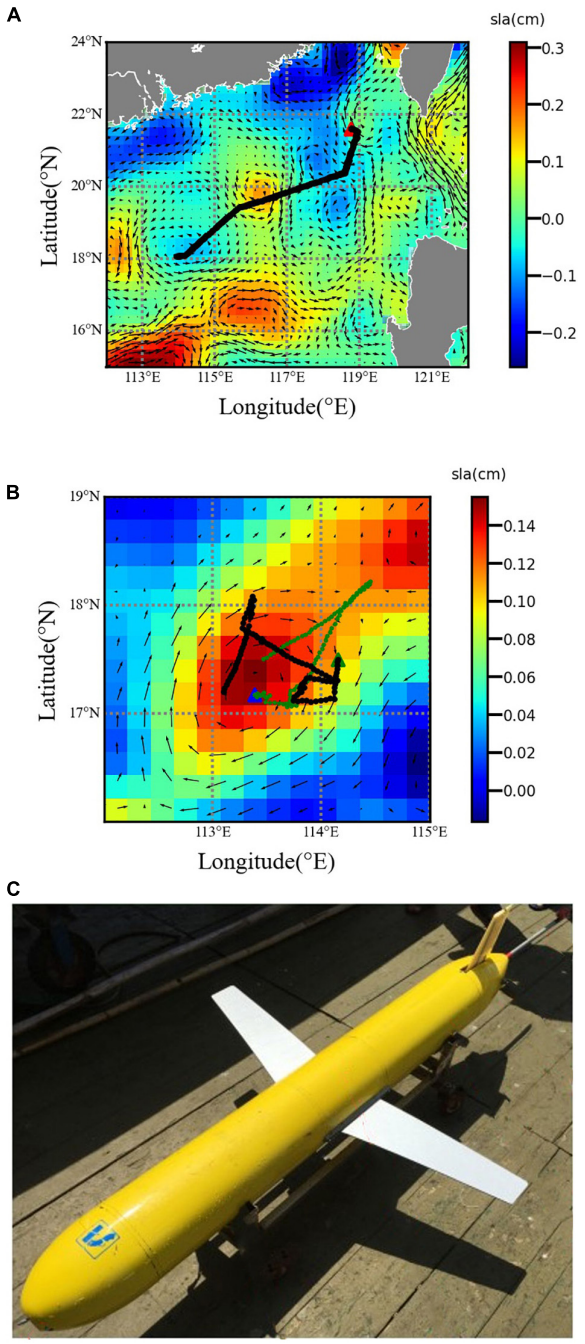


FIGURE 1 | The trajectories of gliders in the ocean observation. **(A)** The observation started from April 28, 2015 to June 1, 2016. **(B)** The observation started from July 3, 2016 to July 16, 2016. The triangles indicate the starting points of observations. The backgrounds are the sea-level anomaly in the observation areas, and the arrows indicate the ocean current field. **(C)** Picture of the sea-wing glider.

METHODS

In this section, we present an overview of the method that we adopt to train and test the simulation model.

SVR for Data Regression

SVR is an extended algorithm of the support vector machine (SVM) which is a classic and powerful machine-learning algorithm to solve the nonlinear regression problem (Brereton and Lloyd, 2010). SVR calculates the loss function based on the structural risk minimization, allowing a deviation of ε between the output of the model and the real value, which differs from the traditional regression model based on the error between the output of the model and the real output. It can avoid the disadvantages caused by pursuing experience risk minimization.

In this subsection, we concisely describe the basic theory of the SVR algorithm. We can learn more detailed theories from Smola and Scholkopf (2004). In our study, the whole diving and climbing process of the glider is described as the dataset $S = \{(x_i, y_i)\}$. $x_i = \{P_i, D_i, V_i\}$ denotes the parameters of the ith diving process. P_i denotes the position in which the glider dives into the water which is observed by the GPS equipped on the glider. D_i is the target diving depth, and V_i represents the velocity at the diving point from the AVISO. y_i represents the real GPS coordinate of the point that the glider comes out of the water during the ith climbing process. The objective of the algorithm is to build a regression model to predict the location that the glider comes out of the water in the following form:

$$\begin{aligned} f(x_i) &= w^T x_i + b, \\ y_i - w^T x_i - b &\leq \varepsilon, \\ w^T x_i + b - y_i &\leq \varepsilon, \\ \varepsilon &\geq 0 \end{aligned} \tag{1}$$

where w and b are, respectively, the slope and offset coefficient generated in the process of mapping x_i to $f(x_i)$, which is optimized during the training of the model. $f(x_i)$ is the predicted location that the glider comes out of the water, whileε denotes the deviation between $f(x_i)$ and y_i. The loss is calculated when the error between the predicted value and the real value is greater than ε. So, the regression problem can be formalized as minimizing the convex optimization problem as follows:

$$min \ \frac{1}{2}||w||^2 + C\sum_{i=1}^{m} \beth\left(f(x_i) - y_i\right) \tag{2}$$

In this equation, C is the regularization parameter, and $\beth$ is the ε –insensitivate loss function to assess the accuracy of the model. Generally speaking, it is difficult to define the appropriate ε to ensure that $f(x)$, all pairs (x_i, y_i), and the slack variables ξ_i and ξ_i^* are introduced. The regression problem can be reformulated as the following constraints:

$$\begin{aligned} min &\tfrac{1}{2}||w||^2 + C\sum_{i=1}^{m}\left(\xi_i + \xi_i^*\right), \\ s.t.\ &y_i - w^T x_i - b \leq \varepsilon + \xi_i, \\ &w^T x_i + b - y_i \leq \varepsilon + \xi_i^*, \\ &\xi_i, \xi_i^* \geq 0, \ i = 1, 2, ..., N \end{aligned} \tag{3}$$

By applying the Lagrange multiplier method and KKT condition, the problem can be expressed as the following equations:

$$\begin{aligned} f(x) &= \sum_{i=1}^{m} (\widehat{\alpha}_i - \alpha_i) K\left(x_i^T, x_i\right) + b, \\ s.t. &\sum_{i=1}^{m} (\widehat{\alpha}_i - \alpha_i) = 0, \\ &0 \le \alpha_i, \widehat{\alpha}_i \le C \end{aligned} \tag{4}$$

where α_i and $\widehat{\alpha}_i$ are the Lagrange coefficients, and $K\left(x_i^T, x_i\right)$ is the kernel function. The Lagrange multiplier method can transform constrained inequalities into unconstrained equations, and the KKT condition can transform the nonlinear regression problem into an approximate linear regression problem. In our study, the kernel function named radial basis function (RBF) was mentioned which can project vectors into any dimensional space. The equation is described by:

$$\begin{aligned} K\left(x_i, x_j\right) &= exp\left(-\frac{\left|\left|x_i - x_j\right|\right|^2}{2\sigma^2}\right), \\ \sigma &\ge 0 \end{aligned} \tag{5}$$

where x_i and x_j are the input vector space, and σ is the bandwidth of RBF.

From the above equations, we can see that two hyperparameters heavily affect the performance of the SVR, namely, the regularization parameter C which represents the tolerance for error and the bandwidth of RBF σ which determines the influence scope of each vector. The appropriate selection of these hyperparameters can significantly improve the performance of SVR. However, a trial-and-error approach to optimize hyperparameters needs much time and is not practical. In our study, we employ the PSO algorithm to optimize parameters.

PSO for Optimizing Parameters

Although SVR is a powerful algorithm for regression problems, its effectiveness is easily affected by the selection of hyperparameters (Kim et al., 2020). Inappropriate parameter selection can easily lead to over-fitting or under-fitting of the prediction results which can affect the coordination of glider arrays. In early studies of the SVR, the grid search method is widely adopted to tune hyperparameters (Jiang et al., 2013). In recent years, many heuristic algorithms have been used to optimize the hyperparameters.

PSO is a technology of evolutionary computing proposed by Kennedy and Eberhart (1995), which has advantages at accelerating the process of tuning the SVR parameters and getting the optimum. The main idea of the algorithm is to analogize the process of optimizing parameters as birds looking for food. Under the constraints of the fitness function, particles keep adjusting speed and position through collaboration and information sharing between each particle to find the final optimum solution.

The particle speed update equation is:

$$\begin{aligned} v_i = w \times v_i + c_1 &\times rand() \times \left(pbest_i - x_i\right) + c_2 \\ &\times rand() \times (gbest - x_i) \end{aligned} \tag{6}$$

where $i = 1, ..., n \in R^n$ denotes the number of particles, v_i represents the speed of the particle, $pbest_i$ is the optimal parameter under current training steps, *gbest* indicates the global optimal parameter under current training steps, x_i is the particle position, *rand*() represents the random number between (0,1), w is the inertia factor which affects the capability of global optimization and local optimization of the algorithm, c_1 indicates the level that the particle updating is affected by the local optimal particle, and c_2 is the level that the particle updating is affected by the global optimal particle.

The particle position update equation is:

$$x_i = x_i + v_i \tag{7}$$

In our study, to find the best hyperparameters (C,σ), the process is shown as follows:

First, we set PSO parameters, the number of particles is 30, the dimension is 2, the iteration steps are 50, and the space ranges of the solutions are from 0.1 to 5. Second, randomly generating the initial value of hyperparameters includes the prime velocities and the initial positions of the swarm of particles in search space. Third, we utilize the parameters in each particle to construct the SVR model and calculate the values of the fitness function, through the values of fitness function to update $pbest_i$ and *gbest*. Last, if the experiment meets the iteration steps or the value of fitness function reaches the extent predefined, we stop updating and choose *gbest* as the best hyperparameters (C,σ); otherwise, we return to Third.

EXPERIMENTS

In this section, we conduct experiments to evaluate the accuracy and efficiency of the simulation model. We use the glider datasets to train and test the simulation model where 75% of each dataset are employed as the training samples for the training and parameter optimization. The rest are applied as the testing samples. We also select several other algorithms as the baseline to compare them with our algorithm. The key idea of our experiment is to utilize the information feedback from gliders for rapid modeling to generate the trajectories of gliders and minimize the error between real trajectory and predicted trajectory.

Data Standardization

In this subsection, to eliminate adverse effects caused by singular sample data, we standardized the data before conducting experiments. The training data $S = \{(x_i, y_i)\}$ are standardized in the normalization function of min–max normalization:

$$S_{train} = \frac{S - min(S)}{max(S) - min(S)} \tag{8}$$

where S_{train} represents the standardized value.

The Experiment Results and Discussion

The Results of Parameter Optimization

In this subsection, we demonstrate the hyperparameters optimized by the PSO algorithm. The optimization of parameters uses an iterative approach, and the iterative results can be learned from **Figure 2** where the x-axis represents the number of iteration steps, the y-axis represents the fitness values, the yellow lines indicate the fitness values in the longitudinal direction, and the blue lines represent the fitness values in the latitudinal direction. The process of optimization is the coordinated adjustment of particles to find the optimal fitness value. We can see that with the increase of iteration steps, the fitness values keep decreasing until achieving stability. The stability of fitness values means that all particles achieve an individual best fitness and the PSO-SVR model achieves the optimal hyperparameters (C, σ); we can learn the optimization results in **Figure 2**.

The Comparative Experiment Results

To verify the effectiveness and stability of our model, we conduct comparison experiments with four baselines, random forest (RF), K-nearest neighbor (KNN), AdaBoost (Ada), and the basic SVR, respectively.

Figure 3 shows the test regression results of three glider datasets by using RF, KNN, Ada, SVR, and PSO-SVR, respectively. **Figures 3A–C** demonstrate the real trajectories and predicted trajectories of g-001, g-002, and g-003, respectively. Since the difference between the results regressed by several methods is not obvious enough and the trajectories of gliders are not regular, we have to add the comparison experiments of the errors between predicted values and real values to validate the results. **Figures 3D–F** introduce the error comparison results. The errors are computed by the function as follows:

$$error = \sqrt{\left(pred_{lon} - real_{lon}\right)^2 + \left(pred_{lat} - real_{lat}\right)^2} \quad (9)$$

Cooperating the figures of trajectory comparison in conjunction with the figures of error comparison, we can learn that the trajectory predicted by the simulation model proposed by us has the best fit with the real trajectory, and what is completely different from other methods is that the proposed model keeps a stable and slight error which confirms the robustness of our model.

Based on the simulation information of PSO-SVR and other comparative experiments, we can further evaluate the effectiveness and stability of the proposed simulation model by the mean square error (MSE), the relative error (RE), and the coefficient of determination (R2).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - f\left(x_i\right)\right)^2 \quad (10)$$

$$RE = \frac{\left|\left|y_i - f(x_i)\right|\right|}{y_i} \quad (11)$$

$$R2 = \frac{\sum_i^N \left(f\left(x_i\right) - \overline{y_i}\right)^2}{\sum_i^N \left(y_i - \overline{y_i}\right)^2} \quad (12)$$

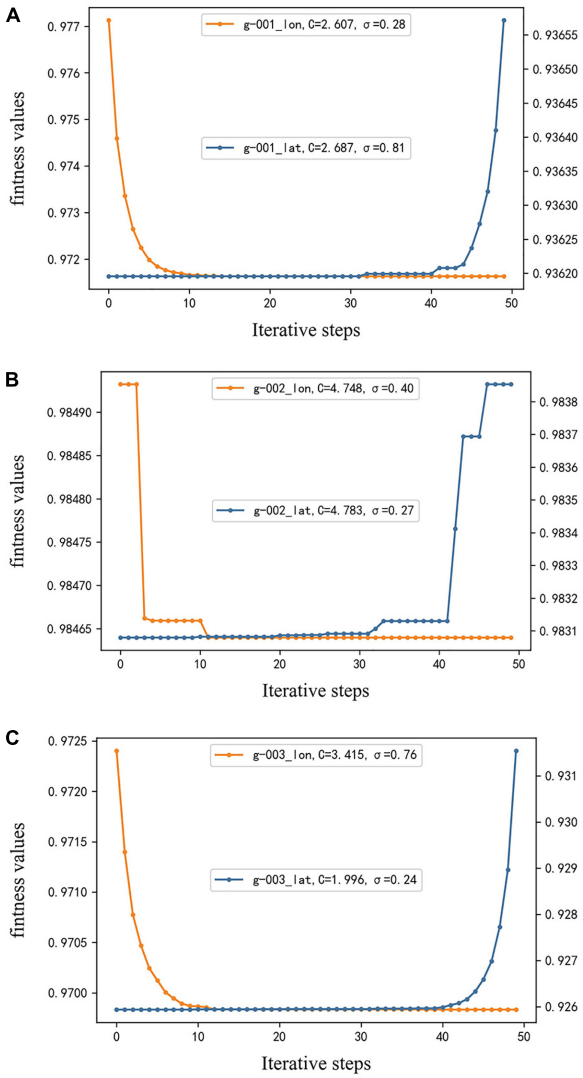


FIGURE 2 | The relationship lines between the iterative steps and the fitness values of parameter optimization of g-001, g-002, and g-003. **(A)** The optimal results of parameters of g-001. **(B)** Represents the parameter optimization of g-002. **(C)** The optimal results of parameters of g-003. The yellow line represents the fitness values in the longitudinal direction, and the blue line demonstrates the fitness values in the latitudinal direction; the legends demonstrate the best hyperparameters.

Here, y_i denotes the real position that the glider comes out of the water, $f\left(x_i\right)$ is the location of the glider predicted by the simulation model. The quantity $\overline{y_i}$ is the mean value of y_i over the whole moving process. N is the number of test samples.

The MSE, RE, and R2 over the testing samples of three glider datasets are shown in **Figure 4**. **Figure 4A** represents the performance of the simulation model applied in g-001 where MSE = 0.00039, RE = 0.0005, and R2 = 0.9998; **Figure 4B** shows the assessment factors of g-002 where MSE = 0.00095, RE = 0.0013, and R2 = 0.993; the results of g-003 can be found in **Figure 4C**, the value of MSE is 0.0007, RE is 0.0012, and R2 is 0.994. Obviously, in contrast to other methods, the PSO-SVR

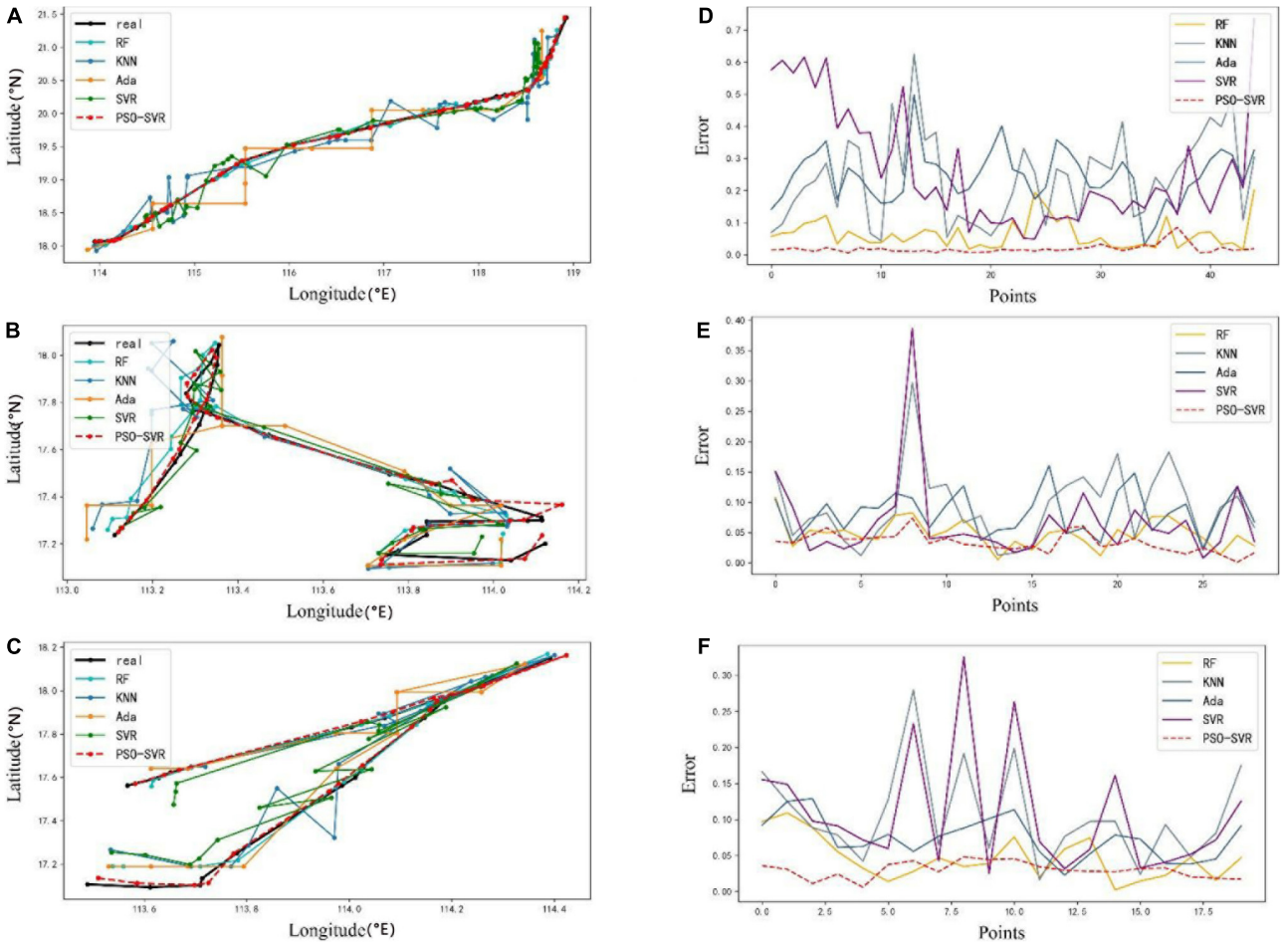


FIGURE 3 | The comparisons of the trajectories of three gliders reconstructed by several algorithms with the original trajectories and the error comparisons between the proposed model with the comparison models. **(A)** The comparison of the regression results of g-001. **(D)** The comparison of the error of g-001. **(B)** The comparison of the regression results of g-002. **(E)** The comparison of the error of g-002. **(C)** The comparison of the regression results of g-003. **(F)** The comparison of the error of g-003.

yields better regression results with the lowest MSE and RE and the highest R2, and these assessment indicators demonstrate that the simulation model can accurately predict the trajectories of gliders and meet the key idea of our experiments.

We can find out that the simulation model tested on the same type of gliders has a different level of performance. The mode tested with g-001 has the best performance, the performance on g-003 is second, and g-002 has the worst performance. From the information obtained from **Figure 1**, the strength of the geostrophic currents of the g-001 trajectory coverage is much lighter than that of g-002 and g-003. Besides, we can learn that the movement direction of g-003 is consistent with the direction of the flow field, while the direction of g-002 is opposite to the flow field which makes the simulation model with g-003 have better performance. Further speaking, both g-001 and g-003 conduct the experiment for crossing the eddy, but g-001 only conducts cross-sectional observation, and the level at which g-001 is affected by the ocean dynamic process is much weaker than that of g-003, which makes the model test with g-001 have better performance than that with g-003.

Besides, the error curves of the PSO-SVR algorithm in **Figures 3D–F** have many bumps. Analyzed in conjunction with **Figure 1**, the bump in the g-001 simulation experiment appears when g-001 crossed through the eddy. The bumps in the g-002 simulation experiment mostly appear at the intersection of trajectories and the period of reverse movement. In the simulation experiment of g-003, the larger error values are distributed when the glider sails to the boundary of the eddy.

From the above discussion, we suppose that the strength and direction of the flow field play an indispensable role in the accuracy of the simulation model. The strong currents, to a certain extent, reduce the performance of the simulation model. The eddies also have a great influence on the effectiveness of the simulation model. When the gliders cross over the eddies or move to the boundary of the eddies, the performance of the model will decrease. Meanwhile, comparing g-002 with g-003, the

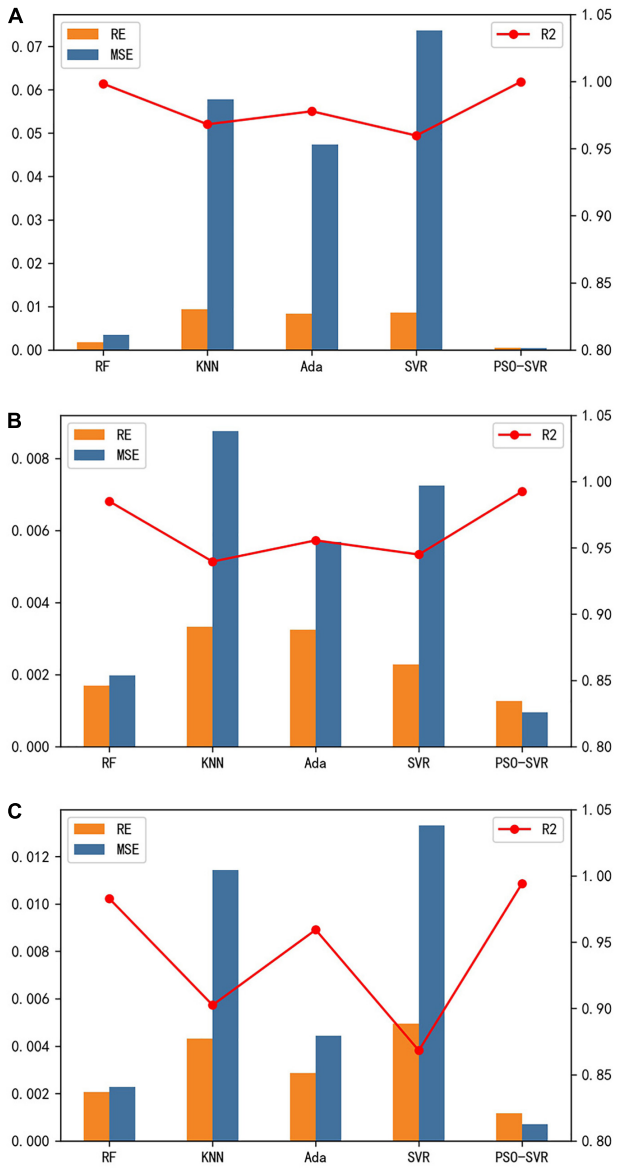


FIGURE 4 | The performances of the model constructed by RF, KNN, Ada, SVR, and PSO-SVR, respectively. **(A)** The performances of the model constructed with the dataset of g-001. **(B)** The performances of the model constructed with the dataset of g-002. **(C)** The performance of the model constructed with the dataset of g-003. The orange histogram represents the relative errors (RE), and the blue histogram shows the mean square error (MSE). The coefficient of determination (R2) of the models is represented by the red line.

difference between them is small which may be caused by their different weights.

CONCLUSION

The tremendous potential value of the ocean observation with the glider arrays is well known. However, the strong, variable ocean currents, the lack of real-time feedback information of gliders, and the high cost of cooperative observation with glider arrays make it necessary to design a simulation model to provide strategies for the deployment and adjustment of glider arrays. The main target of our simulation model is to assimilate the data collected by gliders into numerical models and model rapidly to predict the future trajectories of glider arrays in advance which can direct the gliders more efficiently to adjust measurement formation. In this model, we collected three main influence factors, including the strength and direction of ocean currents, the current coordinate of the glider, and the diving depth of the glider, to predict the trajectory of the glider after working for a certain period of time. In addition, in the case that the accuracy of the datasets is high enough, other more valuable factors can be introduced. Moreover, to improve the performance of our model, we use the PSO algorithm to dynamically optimize the hyperparameters of the SVR algorithm and then cooperate PSO with SVR as the supporting algorithm of the simulation model.

We conducted the experiments to demonstrate the effectiveness of the model by applying the model to sea-wing gliders. The results show that the simulation model has good performance on sea-wing underwater gliders. However, we need more datasets to evaluate the performance of the model applied on different makes of gliders.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The datasets analyzed in this article are not publicly available. Requests to access these datasets should be directed to yufangjie@ouc.edu.cn.

AUTHOR CONTRIBUTIONS

FY and GC conducted the analysis on the datasets of the gliders and led the design of the simulation model. ZZ performed the run of the model. FY and ZZ wrote the first draft. JY validated and refined the algorithm. All authors reviewed, edited the final manuscript, and conceived the research question.

FUNDING

This research was jointly supported by the National Key Research and Development Program of China under contract nos. 2016YFC1402608 and 2019YFD0901001 and the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao), grant no. 2018SDKJ0102-7.

ACKNOWLEDGMENTS

The study benefited from the glider datasets collected by Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China.

REFERENCES

Brereton, R. G., and Lloyd, G. R. (2010). Support vector machines for classification and regression. *Analyst* 135, 230–267.

Clark, E. B., Branch, A., Chien, S., Mirza, F., Farrara, J., Chao, Y., et al. (2020). Station-keeping underwater gliders using a predictive ocean circulation model and applications to SWOT calibration and validation. *IEEE J. Ocean. Eng.* 45, 371–384. doi: 10.1109/joe.2018.2886092

Fonti, A., Freddi, A., Longhi, S., and Monteriù, A. (2011). Cooperative and decentralized navigation of autonomous underwater gliders using predictive control. *IFAC Proc. Vol.* 44, 12813–12818. doi: 10.3182/20110828-6-it-1002.02980

Jiang, M., Jiang, S., Zhu, L., Wang, Y., Huang, W., and Zhang, H. (2013). Study on parameter optimization for support vector regression in solving the inverse ECG problem. *Comput. Math. Methods Med.* 2013:158056.

Kennedy, J., and Eberhart, R. (1995). "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks* (Perth, WA: Institute of Electrical and Electronics Engineers).

Kim, D., Lee, S., and Lee, J. (2020). Data-driven prediction of vessel propulsion power using support vector regression with onboard measurement and ocean data. *Sensors* 20:1588. doi: 10.3390/s20061588

Leonard, N. E., Paley, D. A., Davis, R. E., Fratantoni, D. M., Lekien, F., and Zhang, F. (2010). Coordinated control of an underwater glider fleet in an adaptive ocean sampling field experiment in monterey bay. *J. Field Robot.* 27, 718–740. doi: 10.1002/rob.20366

Leonard, N. E., Paley, D. A., Lekien, F., Sepulchre, R., Fratantoni, D. M., and Davis, R. E. (2007). Collective motion, sensor networks, and ocean sampling. *Proc. IEEE* 95, 48–74. doi: 10.1109/jproc.2006.887295

Li, S., Wang, S., Zhang, F., and Wang, Y. (2019). Constructing the three-dimensional structure of an anticyclonic eddy in the south china sea using multiple underwater gliders. *J. Atmos. Ocean. Technol.* 36, 2449–2470. doi: 10.1175/jtech-d-19-0006.1

Liblik, T., Karstensen, J., Testorc, P., Aleniusd, P., Hayese, D., Ruizf, S., et al. (2016). Potential for an underwater glider component as part of the global ocean observing system. *Methods Oceanogr.* 17, 50–82. doi: 10.1016/j.mio.2016.05.001

Paley, D. A., Zhang, F., and Leonard, N. E. (2008). Cooperative control for ocean sampling: the glider coordinated control system. *IEEE Trans. Control. Syst. Technol.* 16, 735–744. doi: 10.1109/tcst.2007.912238

Rao, D., and Williams, S. B. (2009). Large-scale path planning for Underwater Gliders in ocean currents.

Rudnick, D. L., Davis, R. E., Eriksen, C. C., Fratantoni, D. M., and Perry, M. J. (2004). Underwater gliders for ocean research. *Mar. Technol. Soc. J.* 38, 73–84.

Shu, Y., Chen, J., Li, S., Wang, Q., Yu, J., and Wang, D. (2019). Field-observation for an anticyclonic mesoscale eddy consisted of twelve gliders and sixty-two expendable probes in the northern South China Sea during summer 2017. *Sci. China Earth Sci.* 62, 451–458. doi: 10.1007/s11430-018-9239-0

Smola, A. J., and Scholkopf, B. (2004). A tutorial on support vector regression. *Stat. Comput.* 14, 199–222. doi: 10.1023/b:stco.0000035301.49549.88

Testor, P., de Young, B., Rudnick, D. L., Glenn, S., Hayes, D., Lee, C. M., et al. (2019). OceanGliders: a component of the integrated GOOS. *Front. Mar. Sci* 6:422. doi: 10.3389/fmars.2019.00422

Thompson, D. R., Chien, S., Chao, Y., Li, P., Cahill, B., Levin, J., et al. (2010). "Spatiotemporal path planning in strong, dynamic, uncertain currents," in *Proceedings of the IEEE International Conference on Robotics and Automation ICRA* (Anchorage, AK: Institute of Electrical and Electronics Engineers), 4778–4783.

Xiong, C., Zhou, H., Lu, D., Zeng, Z., Lian, L., and Yu, C. (2020). Rapidly-exploring adaptive sampling tree*: a sample-based path-planning algorithm for unmanned marine vehicles information gathering in variable ocean environments. *Sensors* 20:2515. doi: 10.3390/s20092515

Xue, D. Y., Wu, Z. L., Wang, Y.-H., and Wang, S.-X. (2018). Coordinate control, motion optimization and sea experiment of a fleet of petrel-II gliders. *Chin. J. Mech. Eng.* 31:17.

Zhang, F., Fratantoni, D. M., Paley, D. A., Lund, J. M., and Leonard, N. E. (2007). Control of coordinated patterns for ocean sampling. *Int. J. Control* 80, 1186–1199. doi: 10.1080/00207170701222947

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fmars.2021.717184

Multi-Year ENSO Forecasts Using Parallel Convolutional Neural Networks With Heterogeneous Architecture

Min Ye[1], Jie Nie[1], Anan Liu[2], Zhigang Wang[1], Lei Huang[1], Hao Tian[3], Dehai Song[4] and Zhiqiang Wei[1]*

[1] College of Information Science and Engineering, Ocean University of China, Qingdao, China, [2] College of Electrical and Information Engineering, Tianjin University, Tianjin, China, [3] College of Mathematical Sciences, Ocean University of China, Qingdao, China, [4] Key Laboratory of Physical Oceanography, Ocean University of China, Qingdao, China

The El Niño-Southern Oscillation (ENSO) is one of the main drivers of the interannual climate variability of Earth and can cause a wide range of climate anomalies, so multi year ENSO forecasts are a paramount scientific issue. However, most existing works rely on the conventional iterative mechanism and, thus, fail to provide reliable long-term predictions due to error accumulation. Although methods based on deep learning (DL) apply the parallel modeling scheme for different lead times instead of a single iteration model, they leverage the same DL model for prediction, which can not fully mine the variability of different lead times, resulting in a decrease of prediction accuracy. To solve this problem, we propose a novel parallel deep convolutional neural network (CNN) with a heterogeneous architecture. In this study, by adaptively selecting network architectures for different lead times, we realize variability modeling of different tasks (lead times) and thereby improve the reliability of long-term predictions. Furthermore, we propose a relationship between different prediction lead times and neural network architecture from a unique perspective, namely, the receptive field originally proposed in computer vision. According to the spatio-temporal correlated area and sampling scale of lead times, the size of the convolution kernel and the mesh size of sampling are adjusted as the lead time increases. The Coupled Model Intercomparison Project phase 5 (CMIP5) from 1861 to 2004 and the Simple Ocean Data Assimilation (SODA) from 1871 to 1973 were used for model training, and the GODAS from 1982 to 2017 were used for testing the forecast skill of the model. Experimental results demonstrate that the proposed method outperforms the other well-known methods, especially for long-term predictions.

Keywords: ENSO, long-term prediction, deep learning, convolutional neural networks, heterogeneous architecture

OPEN ACCESS

Edited by:
Andrei Herdean, University of Technology Sydney, Australia

Reviewed by:
Jingzhi Su, Chinese Academy of Meteorological Sciences, China
Qin-Yan Liu, South China Sea Institute of Oceanology, Chinese Academy of Sciences, China

***Correspondence:**
Jie Nie
niejie@ouc.edu.cn

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *30 May 2021*
Accepted: *19 July 2021*
Published: *19 August 2021*

Citation:
Ye M, Nie J, Liu A, Wang Z, Huang L, Tian H, Song D and Wei Z (2021) Multi-Year ENSO Forecasts Using Parallel Convolutional Neural Networks With Heterogeneous Architecture. Front. Mar. Sci. 8:717184. doi: 10.3389/fmars.2021.717184

1. INTRODUCTION

El Niño-Southern Oscillation (ENSO) is a short-term interannual climate change consisting of three phases: neutral, El Niño, and La Niña (Yang et al., 2018). It is a significant feature of ocean-atmosphere interaction over the equatorial Pacific Ocean and, thus, directly or indirectly leads to global climate abnormalities and regional meteorological disasters such as flooding in the

summer and low temperatures in the winter (Forootan et al., 2016). Generally, an El Niño event occurs when the average sea surface temperature (SST) anomaly in the Nino3.4 region (5°N–5°S, 170°W–120°W) is above the threshold of 0.5°C for at least five consecutive overlapping 3-month periods (Cane et al., 1986). Based on this occurrence regularity, ENSO forecasts expected to be an essential method for minimizing the negative effects of climate change has been researched in the last few years. However, due to the spring prediction barrier, the complexity of the internal oscillation mechanism, and the chaos of climate variability, developing an efficient multi year ENSO forecasts method with high accuracy and low complexity is incredibly challenging (Santoso et al., 2017).

In essence, ENSO forecasts are an example of a spatio-temporal sequence prediction problem, whereby past ENSO data (e.g., SST grid maps) are applied to predict future information. Generally, people use related indexes to predict the development tends of ENSO. The commonly used ENSO indexes include the Nino 3,4 index, SST index, and Oceanic Nino index (Yan et al., 2020). By studying the change law of the index, people try to reveal the underlying complex ENSO change characteristics. The existing approaches for making ENSO forecasts can roughly be classified into two categories such as numerical weather prediction methods and statistical prediction methods. With the aid of physics and human experience, numerical models can predict ENSO events by establishing various equations (Goddard et al., 2001). For instance, Zebiak and Cane (1987) proposed the first dynamic model for making seasonal ENSO forecasts, which was followed by various improved models, such as the intermediate coupled model, the hybrid coupled model, and the coupled general circulation model (Wang et al., 2017). However, these models can only well predict ENSO no more than 6 months in advance (Duan and Wei, 2013). To date, the conventional numerical prediction models have been comprehensively developed and are moving toward high resolution and multi-physical processes. Nevertheless, due to the cognitive limitation of empirical models and uncertainties in the optimization of key parameters, the increasingly accumulating prediction errors hinder the application of numerical models to long-term prediction (Duan and Wei, 2013). Moreover, when the horizontal resolution increases by one order of magnitude, the calculation complexity increases by three orders of magnitude (Masumoto et al., 2004).

Speaking of statistical methods, they can be further divided into two subclasses, namely, the conventional statistics method and the deep learning (DL) method. The former subclass includes the Holt-Winters method and the autoregressive integrated moving average (ARIMA) method. More specifically, So and Chung (2014) used the Holt-Winters method to predict the Niño region 3 SST index, and the final root mean square errors from January 1933 to December 2012 were 0.303 by 1-month-ahead and 1.309 by 12-month-ahead. Given that the long-term prediction was poor, they introduced an improved Holt-Winters model to deal with periodically stationary time series. Rosmiati et al. (2021) used the ARIMA method to predict ENSO regional SST Niño 3.4 and found that the ARIMA model stage is well suited for predicting short-term ENSO events. However, the occurrence and evolution of ENSO are nonlinear and the methods mentioned above are incapable of extracting inherent characteristics from nonlinear problems with a huge quantity of raw data. Therefore, conventional statistical methods are not ideal for complex pattern recognition and knowledge discovery.

In contrast, the DL method, which does a good job constructing an end-to-end mapping model for high-dimensional data with complex associations, is considered to have significant potential to make ENSO forecasts. To be more specific, Shukla et al. (2011) used an adaptive neural network (ANN) model to predict the rainfall index by selecting the Niño 1 + 2, 3, 3.4, and 4 indexes as predictors. The results show that the ANN model outperforms all the linear regression models. Unfortunately, the ANN model is unable to capture and process the sequence information in the input data, so researchers must resort to a recurrent neural network model to improve prediction accuracy. Mcdermott and Wikle (2017) used the Bayesian spatial-temporal recurrent neural network model to predict the Oceanic Nino Index with a lead time of 6 months. In an other study, Zhang et al. (2017) leveraged long short-term memory (LSTM) to predict the variation of SSTs in the Bohai Sea, and Broni-Bedaiko et al. (2019) applied seven input predictors to obtain stable predictions with a lead time of 11 months. Noticeably, LSTM, as an improved recurrent neural network model, offers long-term memory with several control gates and is, thus, suitable for dealing with sequence modeling problems. However, the single-point prediction method based on LSTM ignores the spatial correlation of SST and breaks the continuity of the temperature distribution.

To mitigate the drawbacks of the LSTM model, Xingjian et al. (2015) proposed a convolution LSTM (ConvLSTM) architecture to implement the precipitation prediction, where convolution layers are added based on LSTM to capture spatial features. He et al. (2019) proposed a DLENSO model based on ConvLSTM to forecast ENSO events, and the simulation results indicate that it outperforms the conventional LSTM model. Gupta et al. (2020) proved that using a ConvLSTM network to predict the Niño3.4 index overcomes the spring predictability barrier. For ENSO forecasts, ConvLSTM not only obtains the sequential correlation of a single point over time series but also exploits the spatial correlation of the region at a certain time through the convolution operation (Mu et al., 2019).

Thus, it can be easily found that the above-mentioned DL methods are capable of mining the correlation of high-dimensional data to carry out complex modeling and significantly improve the prediction accuracy of ENSO events. However, the drawback of these DL-based methods is that they only build a single depth model for predicting the next lead time and try to predict longer lead times by iteration mechanism, which still exposes them to the error-accumulation problem. Thus, none of these models can provide reliable long-term forecasts (i.e., over 1 year in advance).

In contrast with the above-mentioned DL methods, which make long-term predictions by iterating a single prediction model, Ham et al. (2019) proposed a parallel modeling scheme for long-term prediction. For each lead time, an independent deep neural network is built, and the parallel predictions of

these models are constructed into a prediction sequence, which improves the reliability of long-term predictions. The reasons for this are as follows: First, long-term prediction is decomposed into multiple parallel independent prediction tasks, so the iterative process is replaced by parallel predictions to avoid error accumulation. Next, different models establish different scale patterns, which leads to multiscale modeling. Therefore, this method extends the ENSO prediction time horizon from 1 to 1.5 years.

There is no doubt that the parallel model can provide reliable long-term predictions. Nevertheless, this work only leveraged the same DL model for making predictions with longer lead times, which does not fully mine the variability of different tasks, resulting in a shortage of reliable predictions. To overcome this drawback, we propose herein an adaptive DL model selection method that, depending on the prediction lead times, adaptively chooses between two key structures of the DL model such as the size of the convolution kernels and the mesh size of the input. The idea of an adaptive selection scheme is inspired by the receptive field in computer vision (Gál et al., 2004). The receptive field is defined as the area of a pixel on the feature map of a convolutional neural network (CNN) that is mapped onto the original image (**Figure 1**). In this study, we regard the spatio-temporal correlated area, which varies with the prediction lead time as the receptive field, and then map it to the size of convolution kernels and mesh size in the input layer of the deep network model. By constructing the relationship between prediction time, the size of convolution kernels, and the mesh size of the input layer we obtain an adaptive selection of network structure with different lead times and thereby improve the modeling ability with different lead times. By enhancing the reliability of a single model, we thus improve the reliability of parallel model predictions and the accuracy of the results.

The main contributions of this study can, thus, be summarized as follows:

1. We propose a novel framework that uses parallel deep CNNs with a heterogeneous architecture to forecast ENSO. In this study, by replacing the traditional iteration process with parallel predictions, the proposed scheme avoids error accumulation. Meanwhile, with the help of the adaptive selection of heterogeneous neural network architecture, the modeling can be adjusted to accommodate different tasks (lead times) and, thus, improve the reliability of long-term predictions.
2. We investigate and formulate the relationship between different prediction tasks (lead times) and neural network architectures from a unique perspective, namely, the receptive field from computer vision. We discover that with lead time lengthened, the correlated spatial area should be appropriately increased. From the view of computer vision, the prediction spatial-temporal correlated area is corresponding to the receptive field range, which is determined by the convolution kernels size and sampling mesh size of the input in a deep neural network. According to this, we design an adaptive network structure selection method with respect to lead time lengthened and applied it in the long-term ENSO prediction.
3. We validate the effectiveness of the proposed method by comparing its results with those of current efficient methods. The experimental results demonstrate the advantages of the proposed approach.

2. DATA AND METHOD

2.1. Data

Since ENSO is related to sea temperature, the monthly SST and heat content (HC) (vertically averaged oceanic temperature in the upper 300 m) anomaly maps are used as two input predictors. The heat exchange in the ocean is carried out by vertical vortex motion, convective mixing, and vertical ocean current convey, so SST and HC are correlated to each other at different moments. The anomaly maps extend over 0°–360°E, 55°S–60°N, and the spatial resolution is 5° latitudes by 5° longitudes.

The dataset used in this study was originally published by Ham et al. (2019). Since the observation data of global oceanic temperature distribution is only available as of 1,871 (Giese and Ray, 2011), only 146 are available for each calendar month. To overcome the limitation, Ham et al. (2019) introduced the Coupled Model Intercomparison Project phase 5 (CMIP5) output to increase the number of samples. The CMIP5 output is the simulation data from the numerical climate model that contains the variables SST and heat content. Therefore, the dataset consists of three-part data (as shown in **Table 1**): the historical numerical simulations produced by the 21 CMIP5 models (Taylor et al., 2012), the reanalysis data from the Simple Ocean Data Assimilation (SODA) (Giese and Ray, 2011), and the Global Ocean Data Assimilation System (GODAS) (Behringer and Xue, 2004).

2.2. Receptive Field of CNN

2.2.1. The Size of Receptive Field

Convolutional neural networks have the powerful ability to hierarchically capture spatial structural information (LeCun et al., 1998). They can map the effect of neighborhood points on the center point through a convolution filter and, multilevel convolution processing, they can map the evolution of dynamic complexity between points. The convolution process of CNNs extracts local characteristics from the global maps and calculates dot products between values in the convolution filter and those in the input layer. The jth feature map in the ith convolution layer at grid point (x, y) is calculated by using

$$v_{i,j}^{x,y} = \tanh\left(\sum_{m=1}^{M_{i-1}}\sum_{p=1}^{P_i}\sum_{q=1}^{Q_i} w_{i,j}^{p,q} v_{(i-1),m}^{\left(x+p-\frac{P_i}{2}, y+q-\frac{Q_i}{2}\right)} + b_{i,j}\right), \quad (1)$$

where the hyperbolic tangent serves as the activation function, P_i and Q_i are horizontal and vertical size of the convolution filter for convolution layer i, M_{i-1} is the number of feature maps in layer $i-1$, $w_{i,j}^{p,q}$ is the weight at grid point (p, q), $v_{(i-1),m}^{(x+p-P_i/2, y+q-Q_i/2)}$ is the value of feature map m for convolution layer $i-1$ at grid

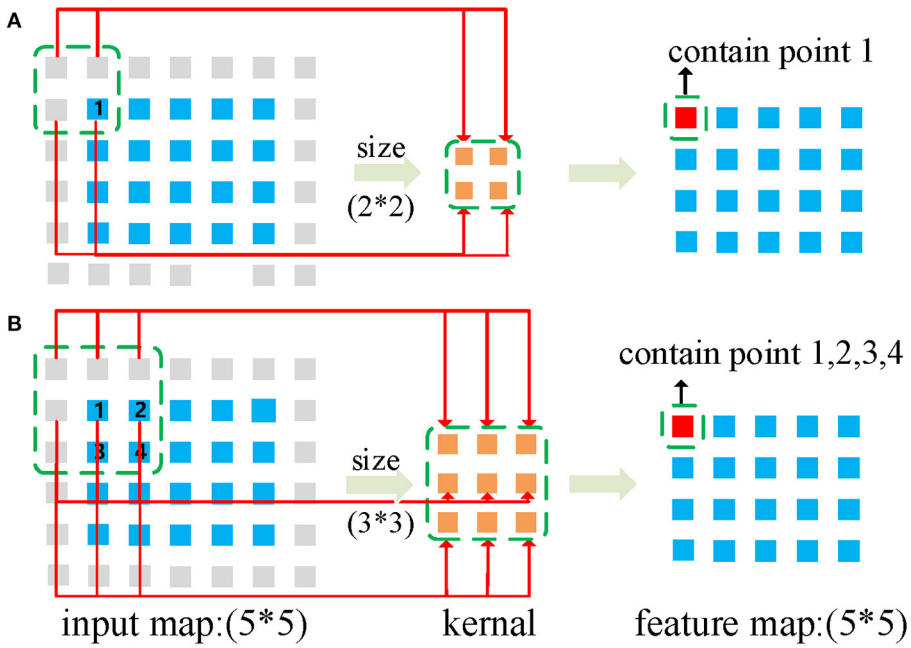


FIGURE 1 | Receptive field: same input and different convolutional kernel. The gray squares represent the padding. **(A,B)** illustrate how the input map converts the output (feature map) through a convolution operation with 2×2 and 3×3 kernels, respectively.

TABLE 1 | The dataset for training and testing the model.

Data	Type	Period
CMIP5	Historical run	1861–2004
SODA	Reanalysis	1871–1973
GODAS	Reanalysis	1982–2017

point $(x + p - P_i/2,\ y + q - Q_i/2)$, and $b_{i,j}$ is the bias of feature map j in convolution layer i.

In computer vision, the receptive field represents the spatial mapping of each point after multilevel convolution operations; that is, the size of the receptive field represents the ability to acquire the spatial correlation range, as shown in **Figure 1**. **Figures 1A,B** illustrate how the input map converts the output (feature map) through a convolution operation with 2 × 2 and 3 × 3 kernels, respectively, so the receptive field of 1 unit in output is a region containing 4 units in the input map in **Figure 1A**, whereas it contains 9 units in **Figure 1B**. Given a deep CNN, the receptive field term usually considers the final output unit in relation to the network input. For example, as shown in **Figure 2**, the receptive field of the unit in the final output map denotes the area in the input map, which consists of every unit that has propagated its value through the whole chain to the given end unit.

It is easy to calculate the size of the receptive field of an output unit in a basic CNN with a fixed structure, since only the convolution and pooling layers can affect its size. The convolution layer uses the kernel to execute the convolution operation to multiply the receptive distance by the kernel size. The pooling layer downsamples following a certain stride to multiply the receptive distance by stride size. Thus, the receptive field size can be formulated as

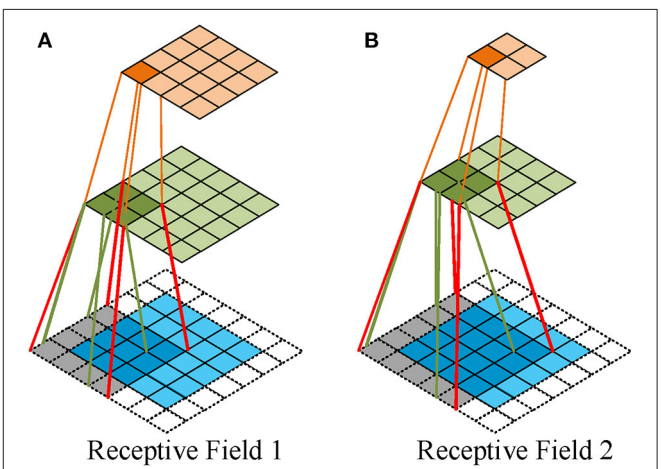


FIGURE 2 | Receptive field a deep CNN. **(A)** The first convolving: a 3 × 3 kernel over a 5 × 5 input padded with a 1 × 1 border of zeros using unit strides. The second convolving: a 2 × 2 kernel using unit strides. **(B)** The first convolving: a 4 × 4 kernel over a 5 × 5 input padded with a 1 × 1 border of zeros using unit strides. The second convolving: a 2 × 2 kernel using 2×2 strides.

$$RF_{i+1} = RF_i + (K_{\text{size}\,i+1} - 1)S_i, \tag{2}$$

where RF_{i+1} and RF_i are the receptive field of layers $(i+1)$ and (i), respectively, with $RF_0 = 1$, S_i is the product of all layer strides before layer $(i+1)$ (i.e., $S_i = \prod_{i=1}^{i} \text{Stride}_i$), and $K_{\text{size}\,i+1}$ is the size of convolution kernels at layer $(i+1)$. Thus, the receptive field area of the last feature map on the input image is (RF_n, RF_n), where n is the last convolution layer in the network.

Note that Equation (2) is calculated based on the assumption that the shape of input images and convolution kernels are square, whereas the input anomaly maps and convolution in the model are rectangular, so Equation (2) must be rewritten as

$$RF_i = \left(RF_i^x, RF_i^y\right)$$

$$RF_{i+1} = RF_i + (P_{i+1} - 1) \times S_i$$

$$RF_{i+1} = RF_i + (Q_{i+1} - 1) \times S_i \tag{3}$$

where the receptive field area of the last feature map on the input image is (RF_n^x, RF_n^y).

2.2.2. Relationship Between Receptive Field and Lead Time

In the simulated prediction, each grid point in the input is spatio-temporally correlated. Specifically, owing to the complicated interactions of the global climate system, the one grid point is highly correlated to the near points and distant points, and hence the slight changes of grid point will lead to the variation of other points. Therefore, in this study, the spatio-temporal correlated area can be regarded as the receptive field in CNN.

All related works applied only a single DL model for multi year ENSO forecasts, which use the same size of convolution kernels and the same step size for different lead times. Therefore, according to Equation (2), the receptive field is the same for different lead times. However, with the increase of prediction lead times, the spatio-temporal correlated area of each point differs. Sønderby et al. (2020) used 64 km as the basic unit to predict the probability of precipitation in the next 8 h. With an input patch covering $1,024 \times 1,024$ km and an average precipitation displacement of 1 km per min, they set the target patch to cover 64×64 km centered on the input patch. They mentioned that accurate predictions of the target tensor for a longer lead time require a larger spatio-temporal context around the target. Therefore, in the prediction problem, the spatio-temporal correlated area (receptive field) of the target should expand with the increase of lead time. As mentioned, the receptive field size is determined by the stride size and the size of the convolution kernels. To this end, we now design a scheme by adjusting the two parameters to appropriately expand the receptive field size.

First, the input mesh size can be selected by changing the stride size through multiple pooling processes. The mesh size of the input means the receptive field resolution; that is, each input grid point represents the number of original grid points. The increase in receptive field resolution allows not only the small range simulations and noises to be smoothed but also the receptive field to be expanded. We assume that the receptive field resolution should increase by a multiple of four as the prediction lead time increases. In prediction modeling, the predicted lead times are divided into three stages for ENSO forecasts: short, medium, and long term, which can be stated as

$$T_{\text{term}} = \begin{cases} \text{Short term (ST)}, & \frac{|lt|-1}{5} \leq 1 \\ \text{Medium term (MT)}, & 1 < \frac{|lt|-1}{5} \leq 3 \\ \text{Long term (LT)}, & 3 < \frac{|lt|-1}{5}, \end{cases} \tag{4}$$

where lt is the lead time of ENSO forecasts, which ranges from 1 to 20 months. Thus, the receptive field resolution for these three terms is 1 : 4 : 16.

The mesh size of the original image is changed and, with the increased lead time, the receptive field is enlarged at the first assumption. However, in computer vision, a larger receptive field does not mean better performance. Luo et al. (2016) also refer to the concept of an effective receptive field. Increasing the mesh size causes the receptive field to expand fourfold, so it is slightly larger.

Second, in the process of convolution feature extraction, we change the kernel size to adjust the size of the expanding receptive field. We assume that the receptive field enlarges at the rate of $2\alpha^{N-1}$ $(0 < \alpha \leqslant 1)$ with the increase of lead term; N represents T_{term} and its value is 1, 2, 3 for each of these three terms. In conclusion, we formulate the relationship between the prediction lead time, input mesh size, and kernel size to increase the receptive field and realize multi-task modeling of spatio-temporal differences.

2.3. Architecture Details

Given that the same single model for multi year ENSO forecast is often unreliable and more difficult to train (Chevillon, 2007; Shi and Yeung, 2018), a more promising strategy must be proposed for long-term prediction. Thus, considering the differences in the spatio-temporal dependencies of different tasks, we propose a framework that uses parallel deep CNNs with heterogeneous architectures to forecast ENSO. **Figure 3** illustrates the schematic structure of the framework, which is called the MS-CNN model and consists of four parts: the input mesh size selection (IMSS) module, the multiterm swappable convolution (MSC) module, the fully connected (FC) module, and the parallel prediction result splicing (PPRS) module.

The IMSS module automatically selects different sampling mechanisms according to the three prediction terms. The sampling mechanism changes the mesh size of the input by selecting the pooling strategy to increase the receptive field. The hypothesis that the resolution of the receptive field should increase with prediction time is realized by the subsampling mechanism. SST and HC anomaly maps for three consecutive months serve as original input predictors.

In the MSC module, we set up a heterogeneous network architecture to select one model of these three-term periods adaptively based on the CNN network model (Ham et al., 2019). This module serves mainly to adjust the expanding receptive field of the first module by setting the size of the convolution kernels to appropriately expand the receptive field. In the proposed model, before the feature map enters the fully connected layer of the

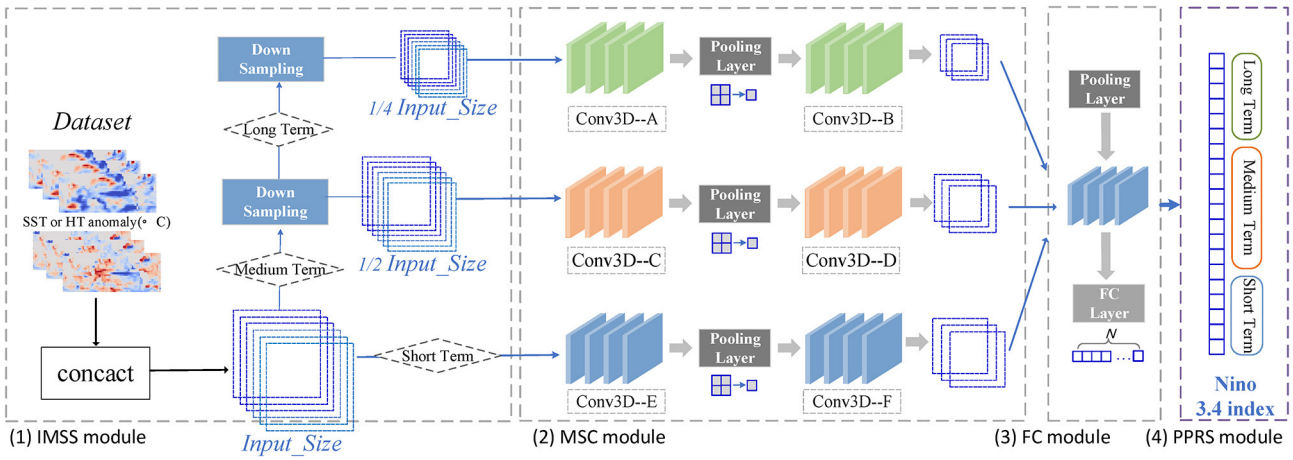


FIGURE 3 | The architecture of the MS_CNN model. (1) Input mesh size selection (IMSS) module aims to increase the input mesh size as the prediction lead time lengthened. (2) Multi-term swappable convolution (MSC) module is mainly to adjust the appropriate receptive field expansion. (3) Fully connected (FC) module transforms the feature map of CNN to vector, and then output the prediction of Nino 3.4 index. (4) Parallel prediction result splicing (PPRS) module splices the Nino 3.4 index of each lead time at the parallel network.

FC module, the process after downsampling is convolution → pool→ convolution → pool → convolution. We assume that the three convolutions have sizes (P_1, Q_1), (P_2, Q_2), (P_3, Q_3), so the receptive field of these three terms is

$$RF = 4^s((4P_3 + 2P_2 + P_1 - 3), (4Q_3 + 2Q_2 + Q_1 - 3)), \quad (5)$$

where s represents the different downsampling scales of the three terms, with $s = 0$ for ST, $s = 1$ for MT, and $s = 2$ for LT.

To satisfy the above assumption that the receptive field grows at the rate of $2\alpha^{N-1}$ $(0 < \alpha \leqslant 1)$ with respect to the increase in the prediction lead time, the receptive field should grow appropriately for different lead terms. Therefore, we formulate the relationships between input mesh size (downsampling strategy), convolution kernel size, and the time period in the network to control the size of the receptive field under these three terms and improve the accuracy of prediction. The formula is

$$(C_1, C_2) = (P_3P_2P_1,\ Q_3Q_2Q_1) \prod_{m=1}^{s} 2^{(m-1)}, \quad (6)$$

where s is the same as in Equation (5), C_1 and C_2 are the same constants for the three terms in the model. We control the receptive field by adjusting C_1 and C_2 and use the large convolution kernels for the short term and the small convolution kernels for the long term to make the product (6) a constant.

The FC module transforms the CNN feature map to a vector and then outputs the prediction of the Niño 3.4 index. The PPRS module is used to splice Niño 3.4 index of each lead time at the parallel network, and the final result is the prediction of the whole lead time series. In this study, by replacing a traditional iteration with the same model with parallel prediction, the proposed scheme avoids error accumulation. Thus, by designing the heterogeneous architecture model, the evolution of different lead times can be fully explored.

TABLE 2 | Correlation skill (Corr) of lead times with different layers.

Layers	2	3	4	5
6-month	0.693	**0.742**	0.697	0.692
12-month	0.522	**0.540**	0.514	0.512
18-month	0.307	**0.328**	0.327	0.305

Bold values is the best value of the four layer parameters.

TABLE 3 | Root Mean Square Error (RMSE) of lead times with different layers.

Layers	2	3	4	5
6-month	0.631	**0.582**	0.606	0.625
12-month	0.754	**0.742**	0.753	0.757
18-month	0.852	**0.848**	0.853	0.859

Bold values is the best value of the four layer parameters.

3. EXPERIMENTS

We ran the numerical experiments on a single NVIDIA RTX2080ti-11G, and the proposed model was implemented by using Python 3.6 with Pytorch.

3.1. Experiment Settings

a) *Input and output data:* Sea surface temperature and HC anomaly maps are used as inputs. The output of the model is the Nino 3.4 index, which is the area-averaged SST anomaly over the Nino3.4 region, as a predictand to be predicted up to 20-months-ahead. That is, only SST is used to calculate the actual result of the Nino3.4 index. HC is just an additional input predictor. More specifically, the model uses SST and HC for three consecutive months as input, so the input size is $6 \times 72 \times 24$ (an anomaly map size is 72×24). A single network output is the predicted

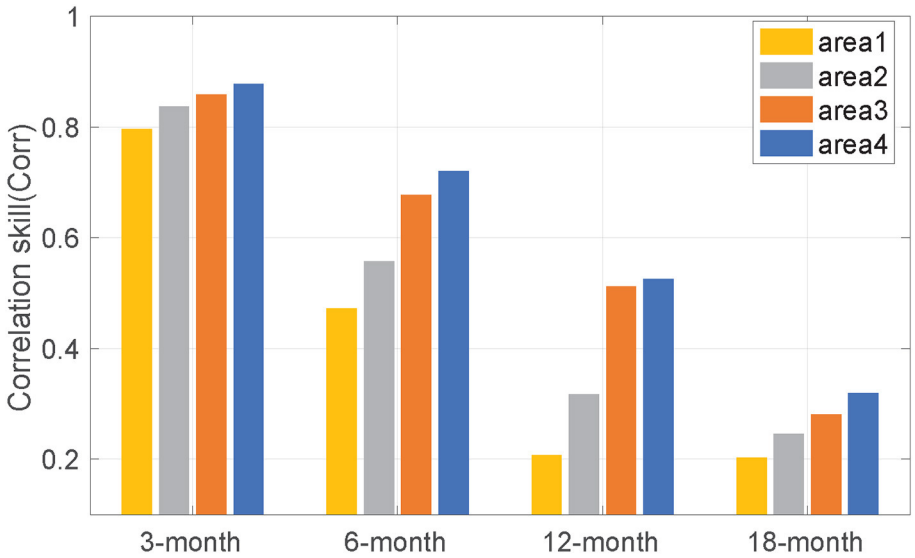


FIGURE 4 | The correlation skill (Corr) between predicted values and actual values with different input sizes. The different input size represents the different regional ranges: area1 (yellow) represents the Nino3.4 region (5°S–5°N, 170°W–120°W), area2 (gray) represents the region of (15°S–15°N, 170°E–100°W), area3 (orange) represents the region of (30°S–30°N, 125°E–55°W), and area4 (blue) represents the region of (55°S–60°N, 0°E–360°E).

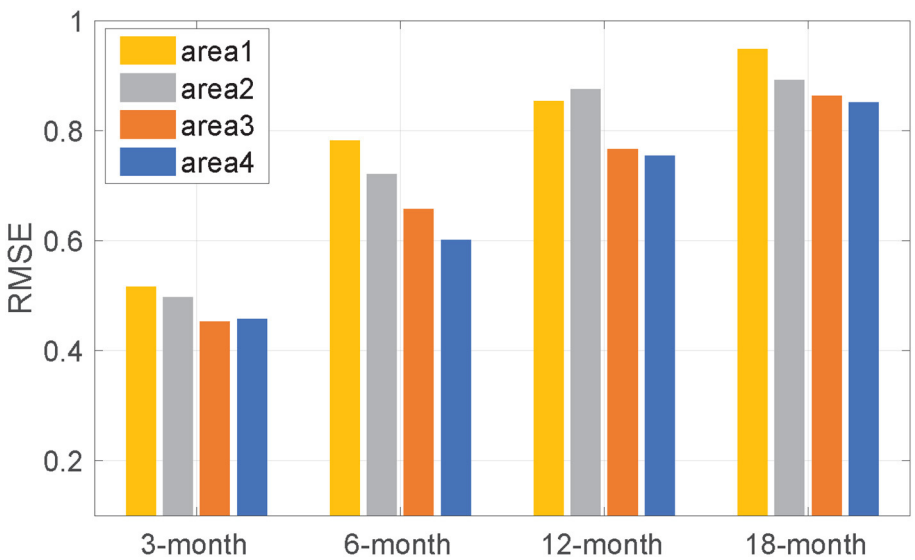


FIGURE 5 | Root Mean Square Error between predicted values and actual values with different input sizes. Refer to **Figure 4** for the information of the area1, 2, 3, 4.

value of the Nino3.4 index for the corresponding lead month, so the size is 1 × 1. The final output of the parallel networks is a 20 × 1 vector of Nino3.4 index from 1- to 20-month-lead. The transfer learning technique is used to optimally train the model. The SODA of 1871–1973 is used for the training period to minimize the systematic errors caused by the CMIP5 samples (from 1861 to 2004). Then, we use the GODAS of 1982–2017 for testing the forecast skill of the model.

b) *Evaluation metrics:* To evaluate the performance of the MS-CNN model, we adopt two commonly used metrics such as Temporal Anomaly Correlation Coefficient Skill (Corr) and Root Mean Square Error (RMSE) of the prediction lead months. Corr measures the linear correlation between the predicted value and the actual value, whereas the RMSE measures the difference. The formulas for calculating the Corr and RMSE are

$$Corr_l = \sum_{m=1}^{12} \frac{\sum_{t=s}^{e} \left(Y_{t,m} - \bar{Y}_m\right)\left(P_{t.m.l} - \bar{P}_{m,l}\right)}{\sqrt{\sum_{t=s}^{e} \left(Y_{t,m} - \bar{Y}_m\right)^2 \sum_{t=s}^{e} \left(P_{t.m.l} - \bar{P}_{m,l}\right)^2}} \quad (7)$$

$$RMSE_l = \sum_{m=1}^{12} \sqrt{\frac{\sum_{t=s}^{e} \left(Y_{t,m} - P_{y \cdot m.l}\right)^2}{|e-s|}} \quad (8)$$

where P and Y are the predicted value and observed value, respectively, $\bar{P}_{m,l}$ and $\bar{Y}_m$ are the mean values of P and Y, m = 1–12 is the calendar month, the label t is the forecast target year, and s = 1982 and e = 2017 are the starting and ending years, respectively.

3.2. Comparison of Different MS-CNN Structures

To further analyze the relationships between the different network structures and simulation results and to prove that the proposed model is reasonable, we compare it with three aspects of the MS-CNN model: (1) the number of layers, which is a fundamental setting for deep neural-network structure; (2) the size of the region on original maps, which is the geographic scope of the original area; and (3) the relationship between the three terms, the size of the input mesh, and the size of the convolution kernel, which represent the change in receptive field size as the lead time lengthens. Experiments (1) and (2) were implemented on the same parallel model based on CNNs to select the optimal basic parameters for the subsequent Experiment (3).

Tables 2, **3** show the comparison of Corr and RMSE between different layers. The results show that, (1) at each layer, both metrics deteriorate as the lead time increases, and (2) more layers do not always guarantee better performance. It is observed that the RMSE of the results decreases at the outset and then increases over the next few lead months. Previous experiments (Mu et al., 2019) show that implementing more layers does not improve performance, because of insufficient training data for a large number of parameters. In this experiment, three layers can be considered as an appropriate choice under the two metrics.

Next, **Figures 4**, **5** show the performance of different original anomaly maps with a three-layer network. In these figures, areas 1, 2, 3, and 4 represent the Nino3.4 region(5°S–5°N, 170°W–100°W), (15°S–15°N, 170°E–100°W), (30°S–30°N, 125°E–55°W), (55°S–60°N, and 0°E–360°E), respectively. The existing DL models for ENSO forecasting use mostly the Nino3.4 area as original input, but Park mentioned that SST anomalies outside the equatorial Pacific Ocean can lead to an ENSO event with a time-lag longer than 1 year (Park et al., 2018). Therefore, we explore how the region size of original anomaly maps affects multi year forecasts. The results show that (1) at each region, both metrics deteriorate as the lead time increases, and (2) larger original anomaly maps result in more accurate forecasting. These results show that the SST patterns are related to each other over a vast area of the ocean, which is consistent with the opinion of Park. We, thus, choose the entire anomaly maps (55°S–60°N, 0°E–360°E) as the original maps.

With the fixed layer and original map area, we next explore the relationship between lead time, input mesh size, and the size of the convolution kernels to determine how the receptive field changes to improve the reliability of ENSO predictions as the lead time increases. **Figures 6**, **7** show that single-scale means that the original map does not downsample for any of these three lead terms. In the same way as with the MSC module, the sizes of the convolution kernels for the three lead terms are

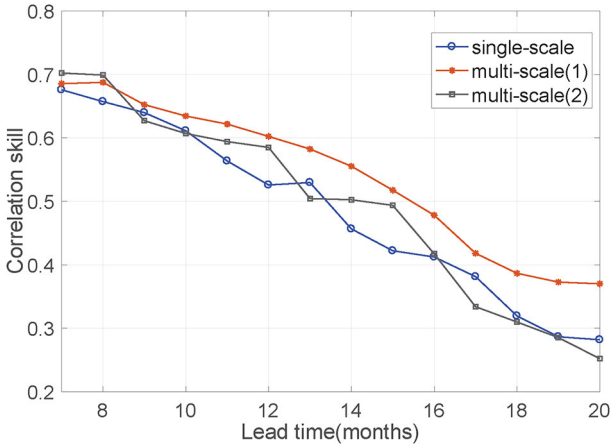


FIGURE 6 | The Corr between predicted values and actual values with different relationships among lead months, input mesh size, and convolution kernel size. Their relationship determines the ratio of the receptive field size in the three prediction lead terms. Single-scale means that input mesh size and convolution kernel size are the same with the increase of lead months. The Conv3d are A/C/E (8,4), B/D/F (4,2) at MSC module. Thus, the receptive field size about these three terms has a ratio of 1:1:1. Multi scale(1) means their relationship is consistent with Equation (6). Noticeably, these three terms correspond to the different input mesh size by downsampling the original map and the ratio of input mesh size is 1:4:16. The Conv3d are different at MSC module, is A(8,4)/B(4,2),C/D (4,2), E/F (2,1). So the receptive field size about these three terms has an approximate ratio of 1:2.5:4. Multi scale(2) have the same input mesh size as multi scale(1) in these three lead terms. But multi scale(2) uses the same Conv3d (A/C/E (8,4), B/D/F (4,2)) at MSC module, so the relationship is not consistent with Equation (6). The size of the receptive field about these three terms has a ratio of 1:4:16.

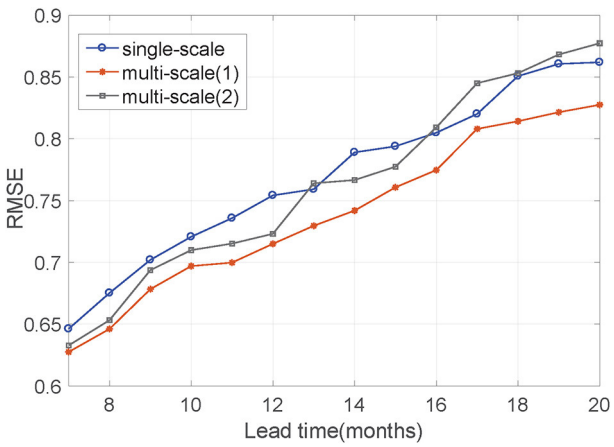


FIGURE 7 | Root Mean Square Error between predicted values and actual values with different relationships among lead months, input mesh size, and convolution kernel size. Their relationship determines the ratio of the receptive field size in the three prediction stages. Refer to **Figure 6** for the information on the single-scale, multi scale (1), and multi scale (2).

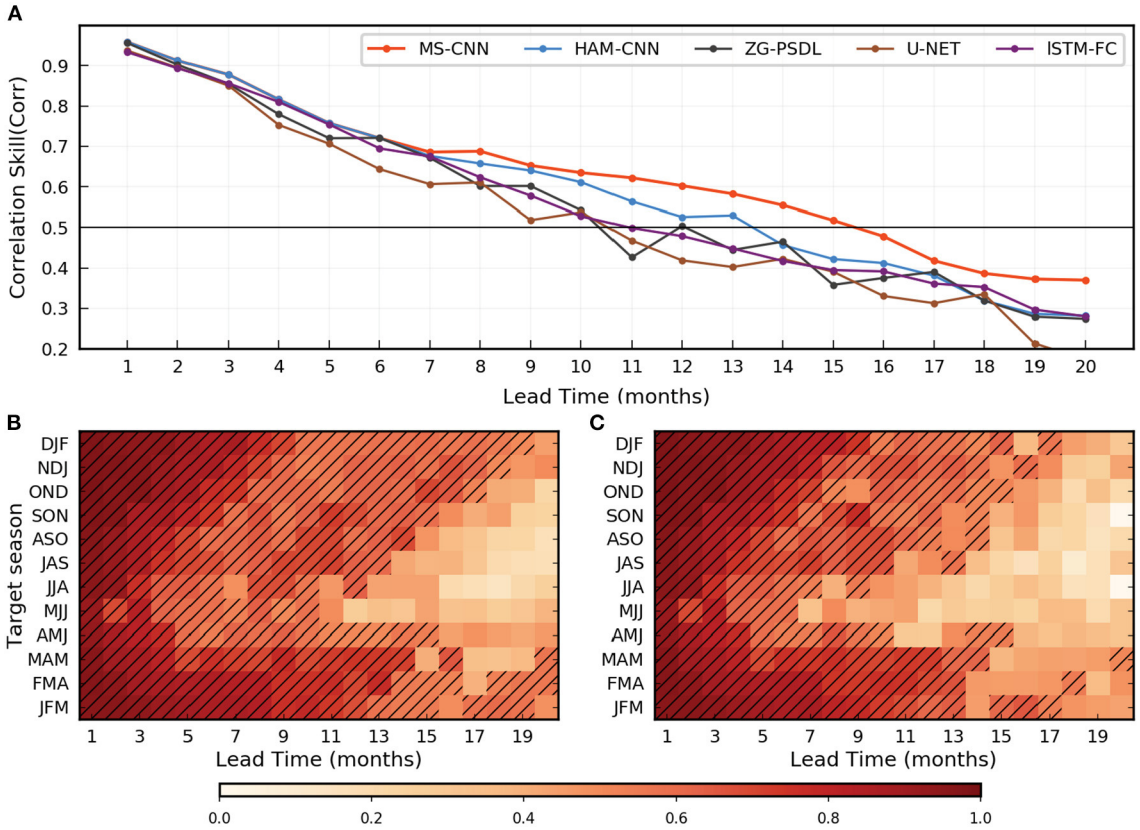


FIGURE 8 | Multi year El Nino-Southern Oscillation (ENSO) forecasts Corr in the MS-CNN model. **(A)** The all-season correlation skill of the 3-month-moving-averaged Nino3.4 index as a function of the forecast lead month in the model (red), HAM-CNN model (blue), ZG-PSDL(black), and U-net model(brown). **(B,C)** The correlation skill of the Nino3.4 index targeted to each calendar month in the MS-CNN model **(B)** and the HAM-CNN model **(C)**. Hatching highlights the forecasts with correlation skills exceeding 0.5.

all the same [A/C/E (8,4) and B/D/F (4,2)]. Thus, the receptive field size about these three terms has a ratio of 1:1:1. Multis cale (1) means that lead times are divided into three terms, and these three terms correspond to the different mesh size of the input map by downsampling. The relationship between the three terms, the mesh size, and the size of the convolution kernel is consistent with Equation (6): $C1 = 64$ and $C2 = 16$, the Conv3d is A(8,4), B/C/D (4,2), E/F (2,1), so the size of the receptive field about these three terms has the approximate ratio of 1:2.5:4. Multiscale (2) is similar to multiscale (1), but the relationship is not consistent with Equation (6). Instead, it uses the same Conv3d [A/C/E (8,4), B/D/F (4,2)] in the MSC module, so the size of the receptive field about these three terms has a ratio of 1:4:16. We can observe from **Figures 6**, **7**, both Corr and RMSE, the multi scale (1) scheme corresponding to Equations (5) and (6) perform best. The receptive field nearly expands at the rate of $2\alpha^{N-1}$ ($\alpha = 1$) for three terms. Moreover, with increasing lead time, the performance gap between the three terms increases. The single-scale scheme does not build different receptive field models to differentiate spatio-temporal dependencies. Multi scale (2) scheme expands the receptive field indefinitely. This scheme only expands the receptive field resolution but does not change the kernel size to adjust the receptive field. Neither of these two schemes fits the hypothesis, and the ratio does not expand at the rate of 2^{N-1} for these three terms. These results prove that the hypothesis is correct and that the method to appropriately expand the receptive field size is reasonable.

3.3. Performance Comparison of Different Models

We now compare the MS-CNN model with the following SST prediction models and multiscale models of computer vision:

a. *The method of Ham et al. (2019).* A multi year ENSO forecast model based on CNNs and denoted HAM-CNN. The parallel model is formulated separately for each forecast lead month and target season.
b. *The method of Zheng et al. (2020).* A model with four stacked composite layers for forecasting the evolution of the SST. It is denoted ZG-PSDL.
c. *UNET (Ronneberger et al., 2015).* A multiscale model for image segmentation. The features of different scales are extracted by using downsampling, and then the output of upsampling is

combined with feature maps of the corresponding scale to the decoder.

d. *LSTM-FC (Zhao et al., 2019).* An LSTM model with a FC neural network.

The deep neural networks are trained with the same number of epochs for a meaningful comparison: 80 epochs are used for the first training using the CMIP5 output, and 30 epochs are used for the second training using the reanalysis data. The total number of convolution filters is 30. An Adam optimizer is applied during the training process. The size of the mini-batch for each epoch is set to 64, and the learning rate is fixed to 0.005.

Inspired by Ham et al. (2019), **Figure 8** facilitates the comparison. **Figure 8A** shows the all-season Corr of the 3-month moving-averaged Niño3.4 index from 1982 to 2017. We experiment with forecasting 1–20 months in advance, and the forecast accuracy of the Niño3.4 index in the MS-CNN model is systematically superior to that of the other models, especially for lead times longer than 6 months. The all-season Corr of the Niño3.4 index in the MS-CNN model is greater than 0.50 for up to a lead time of 16 months, whereas it is 0.41 for a lead time of 16 months in the HAM-CNN model. The HAM-CNN model provides accurate predictions, but when we introduce multiscale network architecture and the method of adaptively controlling the size of convolution kernels to enlarge the receptive field (that is, the MS-CNN model), the prediction accuracy significantly improves for medium to long lead times.

Zhenggang's model and UNET use a multiscale model, but both models have the same convolution kernels for different lead times, which prevents them from controlling the size of receptive fields. Moreover, both models are fused multiscale feature maps, which introduce noise so that the results fluctuate badly. Therefore, we conclude that the MS-CNN model provides a skillful and stable forecast of ENSO phenomena for up to a lead time of 18 months.

The MS-CNN model also provides a greater Corr than the HAM-CNN model for the Niño3.4 index for almost each target season (as shown in **Figures 8B,C**). The improvement in Corr is clear at longer lead times and is robust for target seasons between the late boreal spring and autumn. For example, the forecasts for the May-June-July season have a correlation skill exceeding 0.50 up to lead time of 10 months for the MS-CNN model, but only up to a lead time of 6 months for the HAM-CNN model. In addition, for the HAM-CNN model, the Corr fluctuates significantly for medium lead time (11–16 months), but the result of the model is relatively stable and decreases for longer lead times. This narrows the gap in the ability of the proposed model to forecast different target seasons. The conclusion imposed is that the MS-CNN model is robust and essentially independent of the spring predictability barriers.

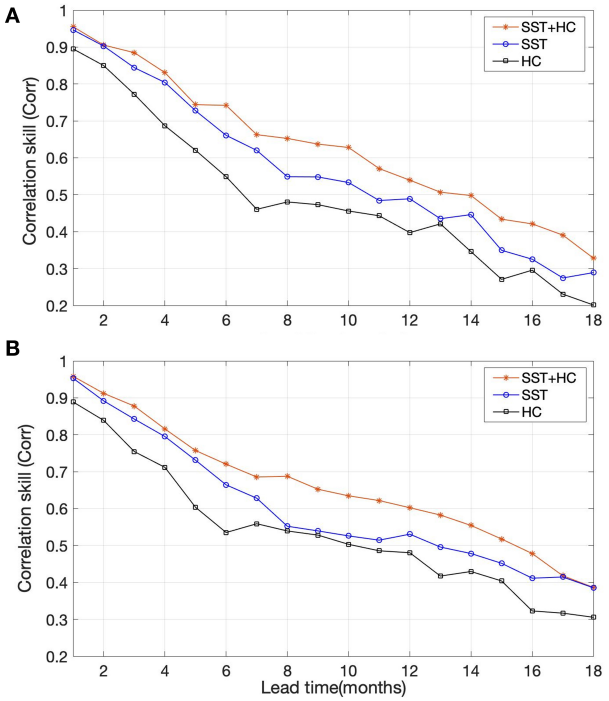


FIGURE 9 | The Corr between predicted values and actual values with different input predictors based on **(A)** CNN model, **(B)** MS-CNN model. The red line shows SST and heat content (HC) were all used as input predictors. The blue line shows only SST was used as input predictors. The black line shows only HC was used as input predictors.

3.4. Results of the MS-CNN Model

Considering the heat exchange in the ocean, we analyze the influence of SST and HC on the prediction results. **Figure 9** shows the performance of different input predictors with CNN and the MS-CNN model. The results show that: (1) the Corr of only SST is higher than only HC, which confirms that SST is the most important factor. The reason is that the Nino3.4 index is obtained by calculating the area-averaged SST anomaly over the Nino3.4 region, so the input of SST is the key factor that decides the fluctuation of results; (2) SST+HC scheme both perform best on CNN and MS-CNN model. These results indicate although the predicted Nino3.4 index is only related to SST, the HC as an input predictor can significantly improve the accuracy of long-term prediction. The reason is that internal ocean motion, such as vertical vortex motion and convective mixing, produces heat exchange. To be more specific, the ocean under the surface can transfer heat to the ocean surface for a while, and vice versa. Therefore, using SST and HC for several previous months as two input predictors, the model can extract some laws and features of ocean internal motion for better long-term prediction.

We regarded the spatio-temporal correlated area as the receptive field in the study. Given the complicated interactions of the global climate, each grid point is highly correlated to near points and distant points, so slight changes of a grid point will cause other points to vary. For short-term forecasting, a small correlated area is enough to predict the change of the grid point. But for long-term forecasting, a larger correlated area should be used. First, we smooth out the fluctuations and noise by changing the mesh size and then adjust the receptive field by changing the convolution kernel size. The correlation

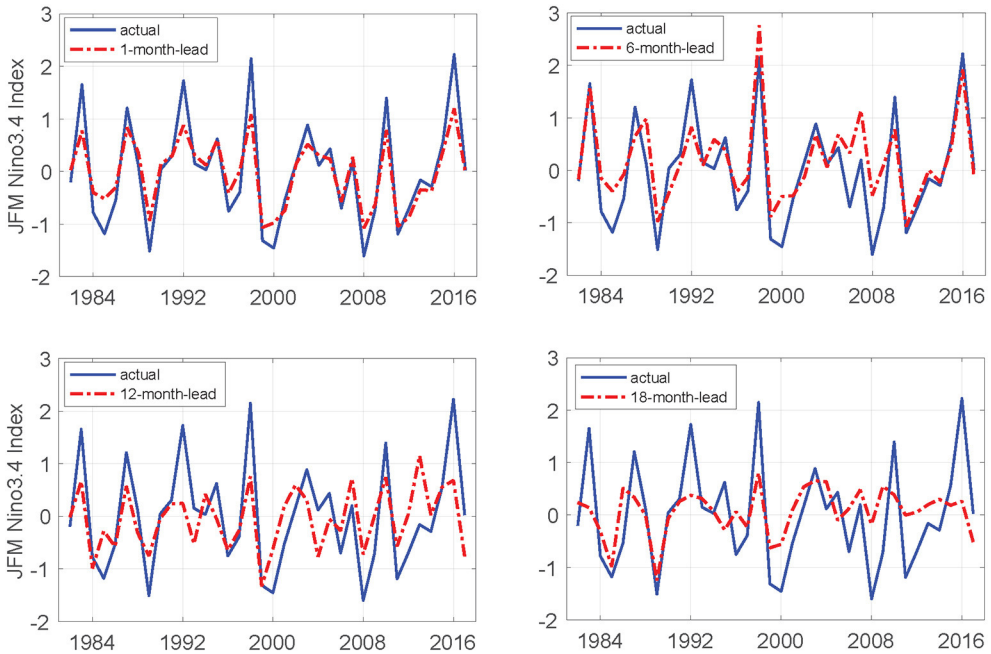


FIGURE 10 | Predicted and actual values of the January-February-March season Nino3.4 indexes for four different lead months.

can be extracted from the convolution kernel of CNN. We make the physical assumption that as the prediction lead time increases, the space temporal correlated area of each grid point should increase appropriately, that is, the receptive field of a pixel in CNN last layer should increase appropriately. More specifically, we assume that the receptive field enlarges at the rate of $2\alpha^{N-1}$ $(0 < \alpha \leqslant 1)$ with respect to the increase in the leading term. **Figure 6** shows that the assumption is reasonable and its performance is best in these three schemes. The traditional method did not consider the dynamic variation that the spatio-temporal correlated area of grid point will change as the prediction lead time increases. The single-scale scheme, which is the traditional method of using only a single model for all lead times, does not build different receptive field models to differentiate spatio-temporal dependencies. The multi scale (2) scheme, which only expands the receptive field resolution by downsampling, does not change the kernel size to adjust the receptive field and lead to the receptive field expand indefinitely. Therefore, the physical assumption that as the prediction lead time increases, the spatio-temporal correlated area of each grid point should increase appropriately is correct.

The Nino3.4 index from 1982 to 2017 was predicted 1–20 months in advance in this experiment. The result from **Figure 8A** shows that the 1-month-lead prediction is the most accurate, and the 20-month-lead prediction is the least accurate. Specifically, the curve produced by the 1-month-lead prediction of Corr is 0.96, and the corresponding RMSE is 0.31; the Corr predictions with a 16-month lead are all greater than 0.50; even for the 20-month-lead prediction, the worst Corr approaches 0.40, and the corresponding RMSE is 0.83. As shown in **Figure 8B**, the MS-CNN model generally provides accurate predictions for autumn and winter (e.g., for December, January, and February, Corr is greater than 0.50 for 19 consecutive months). Although the results for spring and summer are improved compared with other models, the prediction results remain poor in the medium and long term.

The Niño3.4 index for the January-February-March season for 1-, 6-, 12-, and 18-month-lead predictions demonstrates that the proposed MS-CNN model predicts the evolution and the ENSO amplitude by the 12-month-lead time (**Figure 10**). The Corr values are 0.96, 0.87, 0.68, and 0.60, respectively. However, at times with strong EI Niño or La Niña phenomena, the prediction errors become relatively large with increasing lead time. Thus, the proposed method is better than previous methods but does not work well for longer lead times in a more demanding environment.

4. CONCLUSION

Improving the predictability of ENSO is the key to better seasonal predictions across the globe. At present, the DL-based methods for multi year ENSO forecasts only focus on the application of deep neural networks, but none consider the distribution characteristics of oceanic spatio-temporal data. Using SST and HC for several previous months as two input predictors, the model can extract some laws and features of ocean internal motion for better long-term prediction. It can be concluded that it is feasible to improve the reliability of the long-term prediction by adding more related predictors as input.

The spatio-temporal correlated area was regarded as the receptive field in this study. We make the physical assumption that the space-temporal correlated area of each grid point should

increase appropriately with the lead time increases, that is, the receptive field of a pixel in CNN last layer should increase appropriately. Given the differences in the spatio-temporal dependencies of different lead times for forecasting ENSO, we build a framework that we call the "MS-CNN model" by using parallel deep CNN with a heterogeneous architecture for multi year ENSO forecasts. First, by replacing the traditional iteration process with parallel predictions, the proposed scheme avoids error accumulation. Second, by using an adaptive selection of heterogeneous neural network architectures, we adaptively expand the receptive field with increasing prediction lead time. The results of using the model on a real-world dataset show that the hypothesis (i.e., the receptive field expands appropriately as the prediction lead time increases) is reasonable. The reanalysis data for 1982–2017 were predicted 1–20 months in advance. Furthermore, the proposed method provides more accurate predictions than other models, especially for a longer lead times: the Corr is nearly 0.60 for 13-month-ahead and 0.40 for 18-month-ahead.

Although the method proposed for multi year ENSO forecasts significantly improves upon the performance of the conventional methods, some shortcomings remain. For instance, when strong ENSO events happen, the prediction of the Niño3.4 index is relatively inaccurate, and the errors increase with increasing prediction lead time. In addition, the predictions are more accurate than those of other models only for spring, as opposed to all seasons. Thus, in future research, we will focus on the following three themes: (1) introducing more related measurements, such as westerly wind bursts and warm water volume, to model the spatio-temporal process more comprehensively and improve ENSO forecasting for the spring season; (2) integrating the CONVLSTM with the proposed method to determine whether it is generalizable and whether it can further improve the accuracy of multi year ENSO forecasts; (3) learning more knowledge in the marine field and improving the reliability of forecasting by combining the DL methods with numerical simulation models.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analysed in this study. This data can be found here: Ham et al. (2019), https://doi.org/10.5281/zenodo.3244463.

AUTHOR CONTRIBUTIONS

MY performed methodology, data processing, modeling, and writing of the original draft. ZWe performed validation and review. LH, HT, and DS performed validation and investigation. AL and ZWa optimized the model framework. JN contributed to conceptualization, writing review and editing, and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by the National Natural Science Foundation of China (62072418), the Fundamental Research Funds for the Central Universities (202042008), and the Major Scientific and Technological Innovation Project (2019JZZY020705).

REFERENCES

Behringer, D., and Xue, Y. (2004). "Evaluation of the global ocean data assimilation system at ncep: the Pacific ocean," in *24th Conference on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*. Seattle, WA.

Broni-Bedaiko, C., Katsriku, F. A., Unemi, T., Atsumi, M., Abdulai, J.-D., Shinomiya, N., and Owusu, E. (2019). El ni no-southern oscillation forecasting using complex networks analysis of lstm neural networks. *Arti. Life Rob.* 24, 445–451. doi: 10.1007/s10015-019-00540-2

Cane, M. A., Zebiak, S. E., and Dolan, S. C. (1986). Experimental forecasts of el nino. *Nature* 321, 827–832. doi: 10.1038/321827a0

Chevillon, G. (2007). Direct multi-step estimation and forecasting. *J. Econ. Surv.* 21, 746–785. doi: 10.1111/j.1467-6419.2007.00518.x

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

Forootan, E., Awange, J. L., Schumacher, M., Anyah, R., van Dijk, A. I., Kusche, J., et al. (2016). Quantifying the impacts of enso and iod on rain gauge and remotely sensed precipitation products over australia. *Remote Sens Environ.* 172, 50–66. doi: 10.1016/j.rse.2015.10.027

Gál, V., Hámori, J., Roska, T., Bálya, D., Borostyánkői, Z., Brendel, M., et al. (2004). Receptive field atlas and related cnn models. *Int. J. Bifurcat. Chaos* 14, 551–584. doi: 10.1142/S0218127404009545

Giese, B. S., and Ray, S. (2011). El ni no variability in simple ocean data assimilation (soda), 1871-2008. *J. Geophys. Res. Oceans* 116:C02024. doi: 10.1029/2010JC006695

Goddard, L., Mason, S. J., Zebiak, S. E., Ropelewski, C. F., Basher, R., and Cane, M. A. (2001). Current approaches to seasonal to interannual climate predictions. *Int. J. Climatol.* 21, 1111–1152. doi: 10.1002/joc.636

Gupta, M., Kodamana, H., and Sandeep, S. (2020). Prediction of enso beyond spring predictability barrier using deep convolutional lstm networks. *IEEE Geosci. Remote Sens. Lett.* 1–5. doi: 10.1109/LGRS.2020.3032353

Ham, Y.-G., Kim, J.-H., and Luo, J.-J. (2019). Deep learning for multi-year enso forecasts. *Nature* 573, 568–572. doi: 10.1038/s41586-019-1559-7

He, D., Lin, P., Liu, H., Ding, L., and Jiang, J. (2019). "Dlenso: a deep learning enso forecasting model," in *Pacific Rim International Conference on Artificial Intelligence* (Cuvu: Springer), 12–23.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Luo, W., Li, Y., Urtasun, R., and Zemel, R. (2016). "Understanding the effective receptive field in deep convolutional neural networks," in *Proceedings of the 30th International Conference on Neural Information Processing Systems* (Barcelona), 4905–4913.

Masumoto, Y., Sasaki, H., Kagimoto, T., Komori, N., Ishida, A., Sasai, Y., et al. (2004). A fifty-year eddy-resolving simulation of the world ocean: preliminary outcomes of OFES (OGCM for the Earth Simulator). *J. Earth Simul.* 1, 35–56.

Mcdermott, P. L., and Wikle, C. K. (2017). An ensemble quadratic echo state network for non-linear spatio-temporal forecasting. *Stat* 6, 315–330. doi: 10.1002/sta4.160

Mu, B., Peng, C., Yuan, S., and Chen, L. (2019). "Enso forecasting over multiple time horizons using convlstm network and rolling mechanism," in *2019*

International Joint Conference on Neural Networks (IJCNN) (Budapest: IEEE), 1–8.

Park, J.-H., Kug, J.-S., Li, T., and Behera, S. K. (2018). Predicting el ni no beyond 1-year lead: effect of the western hemisphere warm pool. *Sci. Rep*. 8, 1–8. doi: 10.1038/s41598-018-33191-7

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Cham: Springer), 234–241.

Rosmiati, R., Liliasari, S., Tjasyono, B., and Ramalis, T. R. (2021). Development of arima technique in determining the ocean climate prediction skills for pre-service teacher. *J. Phys*. 1731:012072. doi: 10.1088/1742-6596/1731/1/012072

Santoso, A., Mcphaden, M. J., and Cai, W. (2017). The defining characteristics of enso extremes and the strong 2015/2016 el ni no. *Rev. Geophys*. 55, 1079–1129. doi: 10.1002/2017RG000560

Shi, X., and Yeung, D.-Y. (2018). Machine learning for spatio-temporal sequence forecasting: a survey. *arXiv preprint arXiv*:1808.06865.

Shukla, R. P., Tripathi, K. C., Pandey, A. C., and Das, I. (2011). Prediction of indian summer monsoon rainfall using nio indices: a neural network approach. *Atmos. Res*. 102, 99–109. doi: 10.1016/j.atmosres.2011.06.013

So, M. K., and Chung, R. S. (2014). Dynamic seasonality in time series. *Comput. Stat. Data Anal*. 70, 212–226. doi: 10.1016/j.csda.2013.09.010

Sønderby, C. K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans T., et al. (2020). Metnet: a neural weather model for precipitation forecasting. *arXiv preprint* arXiv:2003.12140. Available online at: https://arxiv.org/abs/2003.12140

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of cmip5 and the experiment design. *Bull. Am. Meteorol. Soc*. 93, 485–498. doi: 10.1175/BAMS-D-11-00094.1

Wang, Y., Jiang, J., Zhang, H., Dong, X., Wang, L., Ranjan, R., and Zomaya, A. Y. (2017). A scalable parallel algorithm for atmospheric general circulation models on a multi-core cluster. *Future Generation Comput. Syst*. 72, 1–10. doi: 10.1016/j.future.2017.02.008

Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-C. (2015). "Convolutional lstm network: a machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*, Vol. 28, 802–810.

Yan, J., Mu, L., Wang, L., Ranjan, R., and Zomaya, A. Y. (2020). Temporal convolutional networks for the advance prediction of enso. *Sci. Rep*. 10, 8055. doi: 10.1038/s41598-020-65070-5

Yang, S., Li, Z., Yu, J.-Y., Hu, X., Dong, W., and He, S. (2018). El ni no-southern oscillation and its impact in the changing climate. *Natl. Sci. Rev*. 5, 840–857. doi: 10.1093/nsr/nwy046

Zebiak, S. E., and Cane, M. A. (1987). A model el ni n-southern oscillation. *Mon. Weather Rev*. 115, 2262–2278. doi: 10.1175/1520-0493(1987)115<2262:AMENO>2.0.CO;2

Zhang, Q., Wang, H., Dong, J., Zhong, G., and Sun, X. (2017). Prediction of sea surface temperature using long short-term memory. *IEEE Geosci. Remote Sens. Lett*. 14, 1745–1749. doi: 10.1109/LGRS.2017.2733548

Zhao, J., Deng, F., Cai, Y., and Chen, J. (2019). Long short-term memory - fully connected (lstm-fc) neural network for pm 2.5 concentration prediction. *Chemosphere* 220, 486–492. doi: 10.1016/j.chemosphere.2018.12.128

Zheng, G., Li, X., Zhang, R. H., and Liu, B. (2020). Purely satellite data-driven deep learning forecast of complicated tropical instability waves. *Sci. Adv*. 6:eaba1482. doi: 10.1126/sciadv.aba1482

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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

frontiers in Marine Science

ORIGINAL RESEARCH
published: 17 September 2021
doi: 10.3389/fmars.2021.672280

Short- to Medium-Term Sea Surface Height Prediction in the Bohai Sea Using an Optimized Simple Recurrent Unit Deep Network

Pengfei Ning[1,2], *Cuicui Zhang*[1], *Xuefeng Zhang*[1]* *and Xiaoyi Jiang*[2]

[1] School of Marine Science and Technology, Tianjin University, Tianjin, China, [2] National Marine Data and Information Service, Ministry of Natural Resources, Tianjin, China

OPEN ACCESS

Edited by:
Zhengguang Zhang,
Ocean University of China, China

Reviewed by:
Sheng Li,
National Institute of Information and Communications Technology, Japan
Changming Dong,
Nanjing University of Information Science and Technology, China
Qing Xu,
Ocean University of China, China

***Correspondence:**
Xuefeng Zhang
xuefeng.zhang@tju.edu.cn

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: 25 February 2021
Accepted: 23 August 2021
Published: 17 September 2021

Citation:
Ning P, Zhang C, Zhang X and Jiang X (2021) Short- to Medium-Term Sea Surface Height Prediction in the Bohai Sea Using an Optimized Simple Recurrent Unit Deep Network. Front. Mar. Sci. 8:672280. doi: 10.3389/fmars.2021.672280

Global warming has intensified the rise in sea levels and has caused severe ecological disasters in shallow coastal waters such as the Northeastern China's Bohai Sea. The prediction of the sea surface height anomaly (SSHA) has great significance in the context of monitoring changes in sea levels. However, the non-linearity of SSHA due to the occurrence of dynamic physical phenomena poses a challenge to current methods(e.g., ROMS, MITgcm) that aim to provide accurate predictions of SSHA. In this study, we have developed an optimized Simple Recurrent Unit (SRU) deep network for the short- to medium-term prediction of the SSHA using Archiving Validation and International of Satellites Oceanographic (AVISO) data. Thanks to the parallel structure of the SRU, the computational complexity of the deep network can be reduced to a considerable extent and this makes the short- to medium-term prediction more efficient. To avoid over-fitting and a vanishing gradient, a skip-connection strategy has been utilized for model optimization, and this improves significantly the accuracy of prediction. Detailed experiments were carried out in the Bohai Sea to evaluate the proposed model and it was demonstrated that the proposed framework (i) outperformed significantly the current deep learning methods such as the BP (Backpropagation), the RNN (Recurrent Neural Network), the LSTM (Long Short-term Memory), and the GRU (Gated Recurrent Unit) algorithms for 1, 5, 20, and 300-day prediction; (ii) can predict the short-term trend in the SSHA (for the next day or 2 days) in real time; and (iii) achieves medium-term prediction in seconds for the next 5–20 days and shows great potential for applications requiring medium- to long-term predictions. To the best of our knowledge, this is the first paper that investigates the effectiveness of the SRU deep learning model for short- to medium-term SSHA predictions.

Keywords: short-to medium-term, sea surface height, recurrent neural network, simple recurrent unit, deeplearning model

1. INTRODUCTION

1.1. Research Background

As a result of global warming, the sea-level has rose considerably in the last few decades (Fasullo et al., 2020). Studies have shown that in recent decades the global sea-level has changed from a relatively low average rate (0.4 mm/a) of increase in the past two thousand years to a much higher rate (3.6 mm/a) (Kittel et al., 2021). However, the magnitude of the rise in the global sea-level is not consistent in space, and the changes in sea-level have significant regional characteristics from the decadal to the multigenerational scale. The global average rate of rise in sea-level for the period 1993–2019 was 3.24 ± 0.3 mm/a; in the same period, the rate of rise of sea-level along China's coast was 3.9 mm/a (Kappelle, 2020). Sea level rise along China's coast is more serious (Jeon et al., 2021). In the last 10 years, the average sea-level of China's coastal areas has been at a high level in the past 40 years, being about 100 mm higher than the average sea-level in 1980–1989. Due to the shallow water of the Bohai Sea (average 18 m), the rate of rise of the sea-level in this area for 1980–2019 was 3.7 mm/a. It is estimated that, in the next 30 years, the sea-level of the Bohai Sea will rise by 55–180 mm (Tang et al., 2021).

Rises in sea-level have a serious impact on the economic and social development of the coastal areas of China. A recent analysis of the Global Navigation Satellite System (GNSS) data over a 10 year period showed that although, in the main, the sea-level in the north coast of the Yangtze River estuary has risen, the sea-level in the west coast of Bohai Bay has decreased significantly. The west bank of Bohai Bay has numerous rivers due to the low lying and flat terrain and strong land subsidence, hence this area has become one of the most sensitive areas to experience regional sea-level changes (Wang et al., 2017; Cui et al., 2018; Feng et al., 2019).

The factors affecting sea-level change include melting glacial ice, precipitation, evaporation, runoff, and other seawater exchange processes, as well as changes in seawater density due to changes in temperature or salinity (Jeon et al., 2021). These factors are classified as spatial effects (Tang et al., 2021). It was reported (Guo et al., 2015) that the seasonal variation and spatial distribution of SSHA were related to the changes in the volume of seawater and the effect of differences in temperature, which are due to changes in the onset of the monsoon and the seawater temperature, respectively. Also, based on satellite altimetry data, the spatial and seasonal variations of the SSHA in the coastal waters to the east of China have been analyzed (Yan et al., 2020).

Given the increase in sea-level, how best to accurately predict the changes in sea-level and then make risk management decisions and adopt effective prevention strategies based on the science has become an important issue of widespread concern to the scientific community and society. However, the stochastic dynamics and instabilities of changes in sea-level brings many challenges with respect to the prediction of the SSHA and may limit our ability to respond in a timely manner to the effects of global change (Yu et al., 2019a; Wang et al., 2020).

1.2. Related Works

The traditional research methods for prediction of the SSHA mainly comprise physics-based and statistical-based methods. Based on the sea-level, temperature, salinity, wind field, and other relevant environmental parameters, the physical processes that control ocean movement and circulation may be described; then a method of converting the physics-based model to obtain a model that describes the changes in sea-level is called a physics-based method. In general, for long-term prediction, physics-based methods tend to have excellent characteristics, while statistical-based methods are more suitable for short-term prediction in specific practical applications. Physics-based models have very strict requirements for high-performance computing systems: first, the solution of the model requires very large computing power, and the real-time requirements for prediction are very high, thus the model must complete the calculation within the specified time. Second, with the continuous improvement of forecasting accuracy, the required computing power will show a geometric growth. Therefore, the effective predictable period of the current prediction system based on physical model is about one week and the accuracy of ∼0.9 (Qiao et al., 2016, 2019). Statistical-based methods and deep learning networks using a data-driven approach to SSHA forecasting, and have the ability to do so with less computational and time costs. The current method achieves average predicting accuracy of coming 24, 48, 72, 96, and 120 h by 0.9, 0.85, 0.79, 0.74, 0.68, respectively (Shao et al., 2015; Song et al., 2020). Statistical-based methods focus on building a prediction model using the relationships between historical data. The interaction of factors that affect the changes in sea-level make the time-series data exhibit variation and non-linear characteristics, which lead to large deviations between the prediction results, based on the traditional time-series, the statistical model, and other conventional methods, and the actual results; moreover it is difficult to obtain satisfactory results. With the development of artificial intelligence (AI)-based algorithms, deep learning methods have shown great power in being able to undertake prediction tasks. The deep learning method has obvious advantages in being able to process a large number of samples and non-linear data, and can describe the highly complex relationships existing between the input data and the output data. Well-known deep learning networks for spatio-temporal learning and predictions include the Recurrent Neural Network (RNN) (Mahata et al., 2019), the Long Short-Term Memory (LSTM) (Graves, 2012), and the Gated Recurrent Unit (GRU) (Li et al., 2021), etc. Recently, the prediction systems based on these neural networks have been applied for the prediction of the SSHA (Shao et al., 2015; Song et al., 2020).

1.2.1. Recurrent Neural Networks (RNN)

Traditional neural networks usually assume that all inputs (or outputs) are independent of each other. However, in the process of practical operation, there is a dependency between the current state of each node and the previous steps, and this is the basic assumption of expanded RNN. The signal feedback structure of the recurrent neural networks (RNN) (Mahata et al., 2019)

adopts the output state of the network at the time of K associated with the historical signal before the time of K, in order for it to have dynamic characteristics and a memory capability. However, RNNs are challenged by the vanishing gradient problem, where the gradient decreases over time (Mahata et al., 2019). Moreover, the RNN may also suffer from the gradient explosion problem. Although many techniques have been developed to address these issue, it remains difficult to obtain long-term memory.

1.2.2. Long Short-Term Memory (LSTM)

The phenomena of the vanishing gradient and the exploding gradient of the RNN may not mean that learning cannot be implemented, but even if it can, the process will be very slow. One solution is to establish the weight of linear self-connections with a value close to 1 in the self-connection part, which is called the leaky unit. At present, the most effective way to adjust the weight of the linear self-linking is through gates, thus allowing the weight of linear self-linking to change and make adjustments at each step. LSTM (Graves, 2012) is an implementation of gated RNNs. The LSTM is a new cyclic network architecture training method featuring a gradient based learning algorithm. The approach can learn how to span a time interval of more than 1,000 steps to overcome the problem of error return. The LSTM has successfully solved the defects of the original cyclic neural network and has become the most popular RNN (Yu et al., 2019b).

1.2.3. Gated Recurrent Unit (GRU)

Due to its complex internal structure, the training of the LSTM network is the very time-consuming and the LSTM exhibits a poor real-time capability. With the rapid growth in demand for speech-to-text applications, computing resources are currently not even keeping up with its needs. To solve this problem, Gated Recurrent Unit (GRU) network model (Li et al., 2021) was proposed on the basis of the original LSTM model. The forget gate and the input gate are combined into a single update gate, and the cell state, the hidden state and other changes are also mixed. The GRU neural network has been successfully applied to sequential or temporal data. The GRU has a simpler structure than the LSTM; nevertheless, its performance is comparable with the LSTM. The GRU even outperforms LSTM but has a lower complexity and faster convergence (Zhou et al., 2021). However, the GRU has a serial structure, which makes parallel computation hard to implement.

1.2.4. Simple Recurrent Unit (SRU)

The common feature of LSTM and GRU is that the calculation of the gate of each time step depends on the output of the previous time step, which leads to a high serial dependence of the network. Also, it is difficult to speed up the calculation by parallel calculation. To solve this problem, the Simple Recurrent Unit (SRU) network was proposed (Lei et al., 2018). The main design feature of the SRU is that the gate calculation depends only on the current input cycle. In this way, only the point-by-point matrix multiplication of the model depends on the previous time step. Thus, the network can be configured in parallel. In addition, the SRU also reduces the number of gates, and the design only features the forget gate and the reset gate. In this way, the calculation efficiency of SRU neural network is higher than that of LSTM and GRU (Jiang et al., 2018).

1.3. The Contribution of This Paper

This paper has developed an SRU model for the short- to medium-term prediction of the SSHA. The parallel computing afforded by the SRU makes it more efficient than the traditional RNN, LSTM, and GRU methods. To solve the problems of over-fitting and vanishing gradient, a skip-connection strategy, which can significantly improve the prediction accuracy, was utilized for model optimization. An outline of the workflow is as follows: We analyzed and normalized the historical SSHA data from the Archiving Validation and International of Satellites Oceanographic (AVISO) center, and then, using the statistical relationships for the historical data, the SRU model for training of the historical data was established. In the training process, we introduced the SRU optimization strategy to solve the gradient explosion phenomenon and the disappearance. Finally, the well-trained model was applied for the short- to medium-term prediction of the SSHA in the Bohai Sea. The model is compared with traditional models (e.g., BP, RNN, LSTM, GRU, and SRU) with regards to prediction accuracy and efficiency.

The organization of this paper is as follows: section 2 introduces the methodology, which includes the principle of the SRU model and its construction. Section 3 analyses the results for 1, 5, 20, and 300-day prediction by comparing the results for the proposed model with the existing BP, RNN, LSTM, and GRU models. We summarize the results of this paper in section 4 and provide future research directions.

2. METHODOLOGY

2.1. The Principle of SRU

The basic structure of SRU is shown in **Figure 1**, where the SRU infrastructure consists of a single forget gate. Assumed input X_t and time t, we need to calculate linear transformation $\hat{x}_t$ and forget gate f_t. This calculation only depends on X_t, therefore parallel processing can be carried out.The forget gate contains internal state C_t, and we need to calculate output state h_t, where G represents the activation function, f represents the forget gate, r denotes the reset gate, h is the output state, C represents the internal state, and x represents the input. The SRU is defined in Equation (1) where it can be shown that computations are dependent only on X_t, allowing computations to be parallelized.

$$\begin{aligned} \hat{x}_t &= WX_t \\ f_t &= \sigma\left(W_f X_t + b_f\right) \\ r_t &= \sigma\left(W_r X_t + b_r\right) \\ C_t &= f_t \odot C_{t-1} + \left(1 - f_t\right) \odot \hat{x}_t \\ h_t &= r_t \odot g\left(C_t\right) + \left(1 - r_t\right) \odot X_t \end{aligned} \tag{1}$$

SRU algorithm is mainly developed to remove the time dependence of the most computing operations, and carry out parallel processing. **Figure 2** compares the structure of SRU with that of the traditional RNN. On the left is the traditional RNN/LSTM/GRU structure. The gray calculation part of each

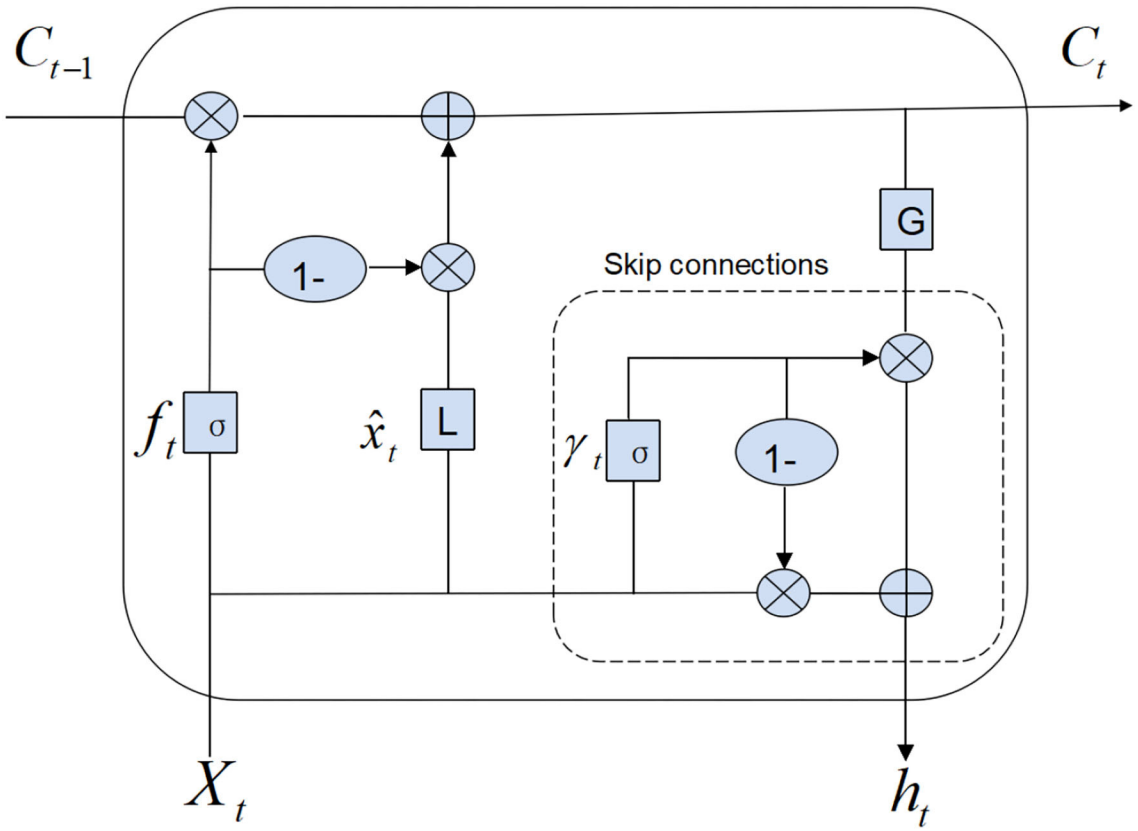


FIGURE 1 | The basic structure of SRU. From the above diagram, we can see that the SRU infrastructure consists of the input X_t and time t, a single forget gate f_t, and reset gate r_t. The forget gate contains internal state C_t, and we need to calculate output state h_t, where G represents the activation function. In SRU, we just need to calculate linear transformation $\hat{x}_t$ and forget gate f_t, which only depends on X_t. Therefore, parallel processing can be carried out.

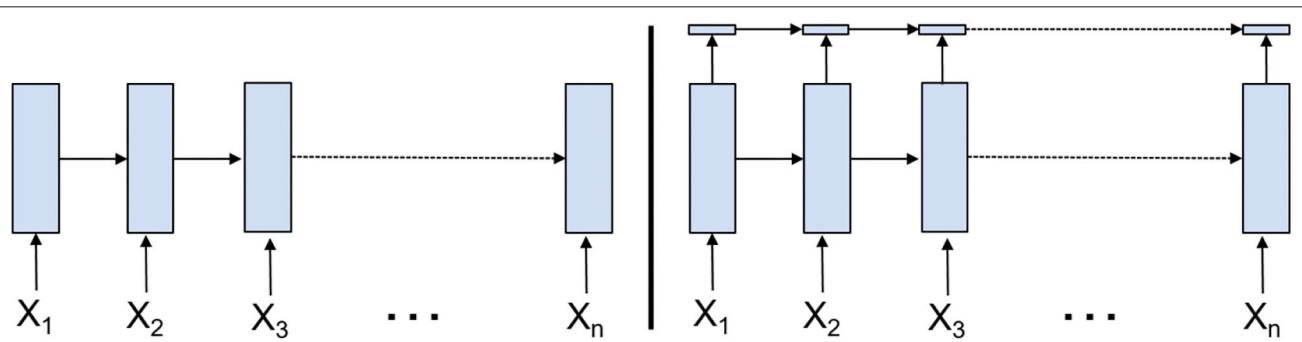


FIGURE 2 | The structure of RNN **(left)** vs. that of SRU cells **(right)**. Compared with the structure of the traditional RNN, SRU eliminates the time dependence of most computing operations and carries out parallel processing.

time has to wait until the last time. It is mainly reflected in the matrix multiplication of the hidden layer and the weight of each door in the last time, which is very time-consuming. While SRU can be parallel in calculating gate, that is to say, matrix operation with large amount of calculation at each time can be parallel. The rest of the dependence is the calculation of C_t and H_t. These calculations are all element-wise multiplication and addition, which is very fast.

SRU is much faster than the current loop implementation, and the loop unit simplifies the state calculation. For SRU, CUDA level optimization method can be used to integrate all elements into a single core function call, which shows the same parallelism similar to CNN, attention model and feedforward network. In particular, although the internal state C_t still using the previous state C_{t-1}, but in the loop step, it is no longer dependent on h_{t-1}. Finally, all matrix multiplication

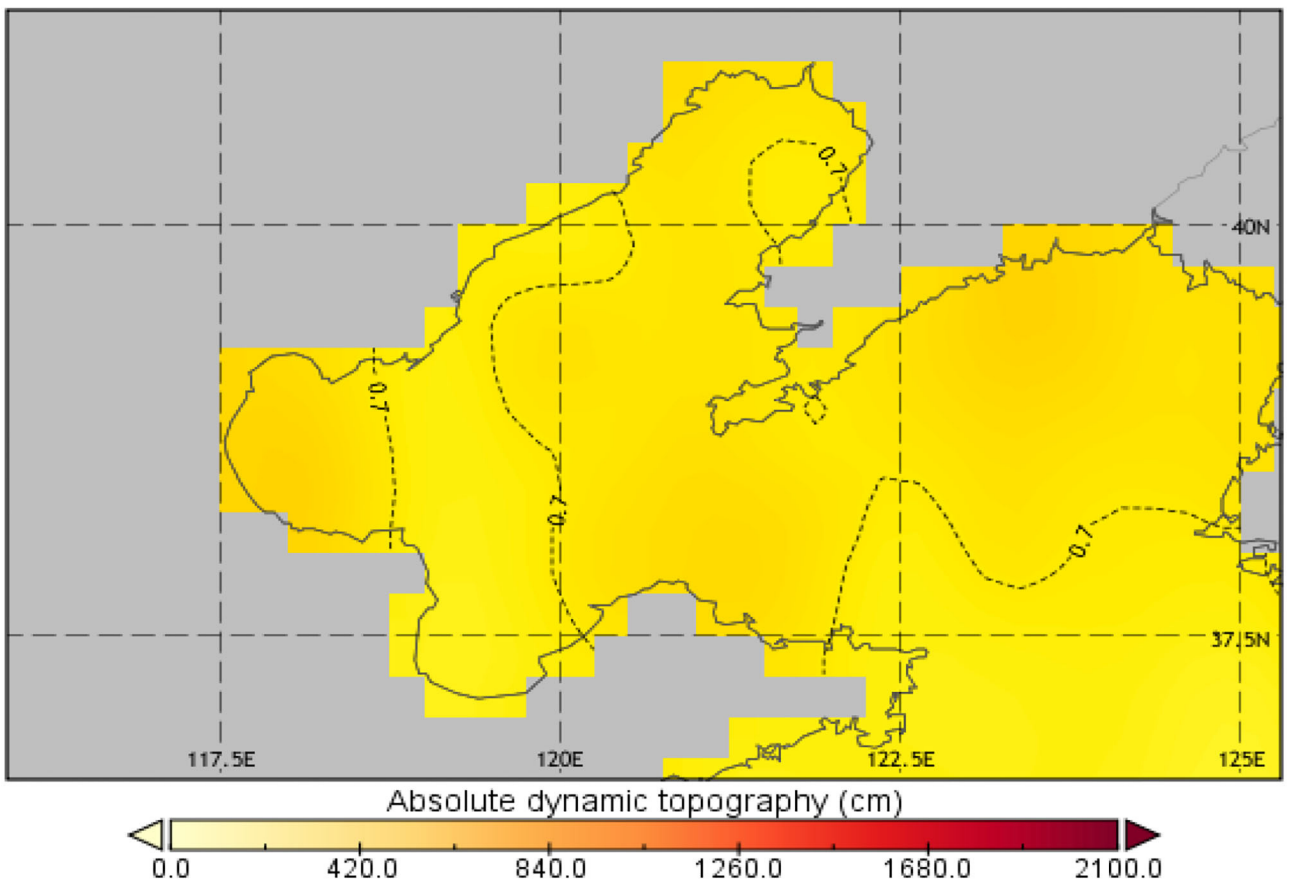


FIGURE 3 | The Research area and the Absolute dynamic topography of the Bohai Sea. As can be seen from the contour lines (dashed lines) in the graph, there is little diversity across the region.

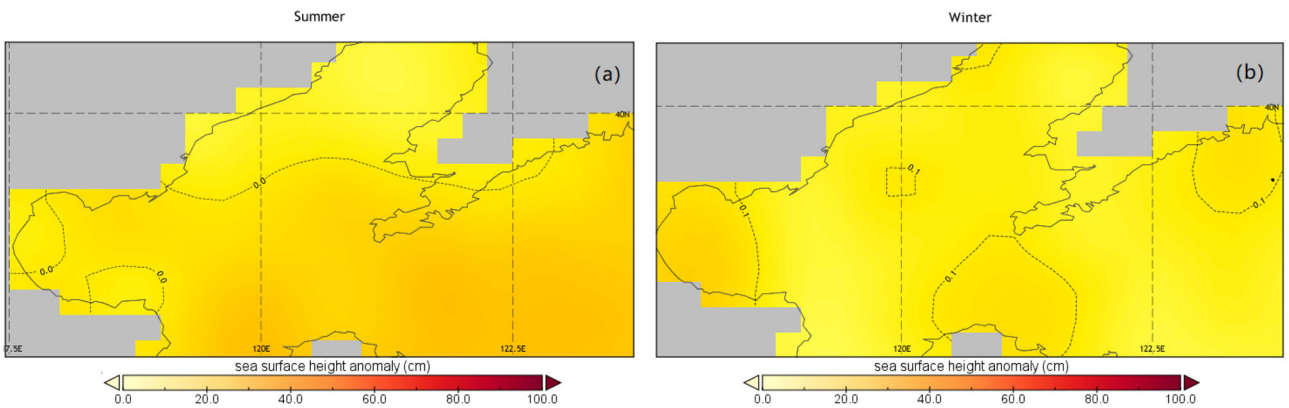


FIGURE 4 | The climatology of SSHA distribution in the Bohai Sea. The graph shows that the SSHA in the Bohai Sea is slightly higher in summer **(A)** than in winter **(B)**, but the overall difference in the Bohai Sea is smaller at the same time of the year.

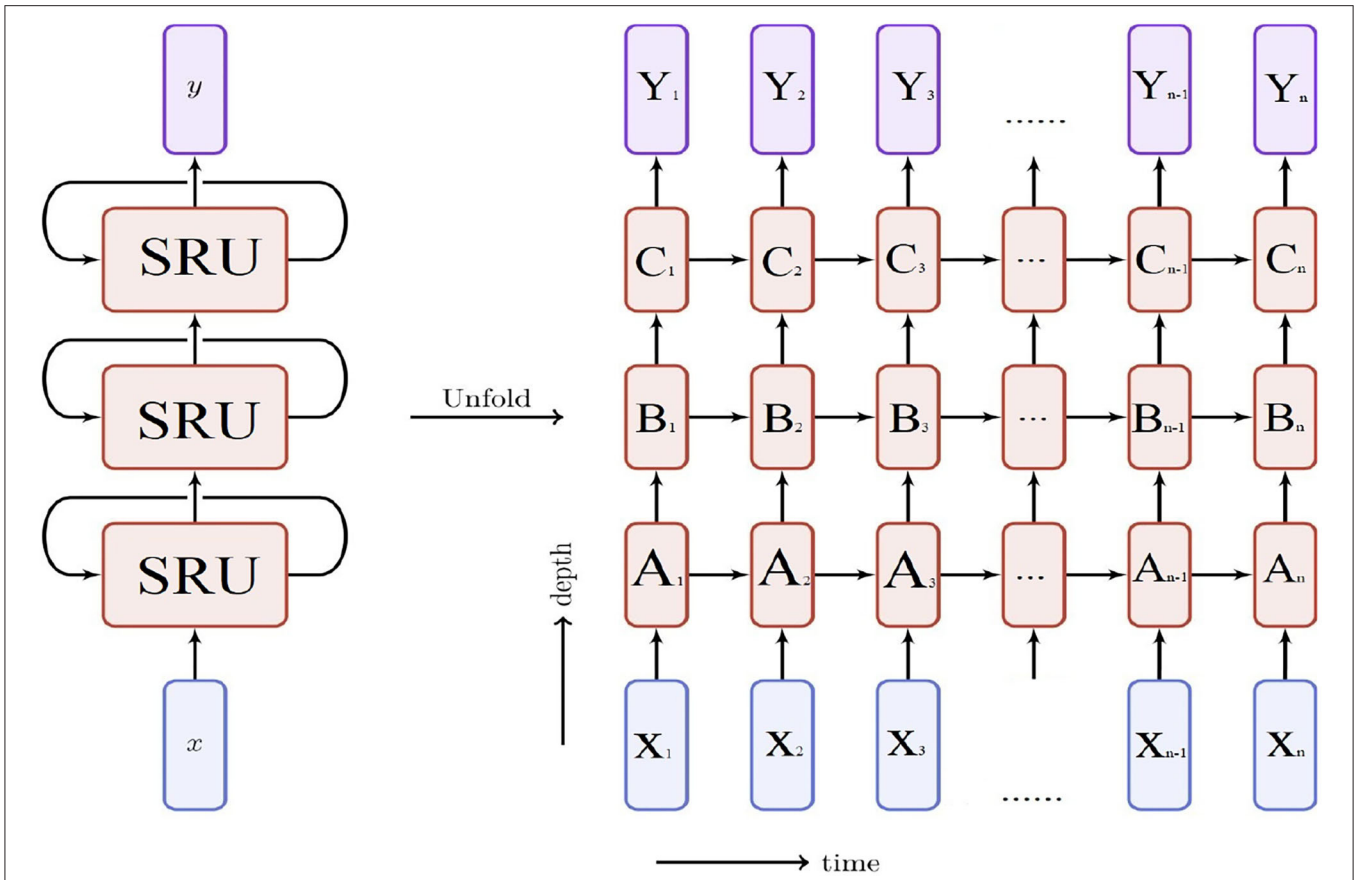


FIGURE 5 | The structure of our model. From the above diagram, we can see that the X_t is the input, Y_t is the output. The SRU network ensures most of the operations are all element-wise multiplication and addition, these calculations can be put into parallel processing. With A, B, and C three-layer SRU modules that can be accelerated in parallel across the time dimension.

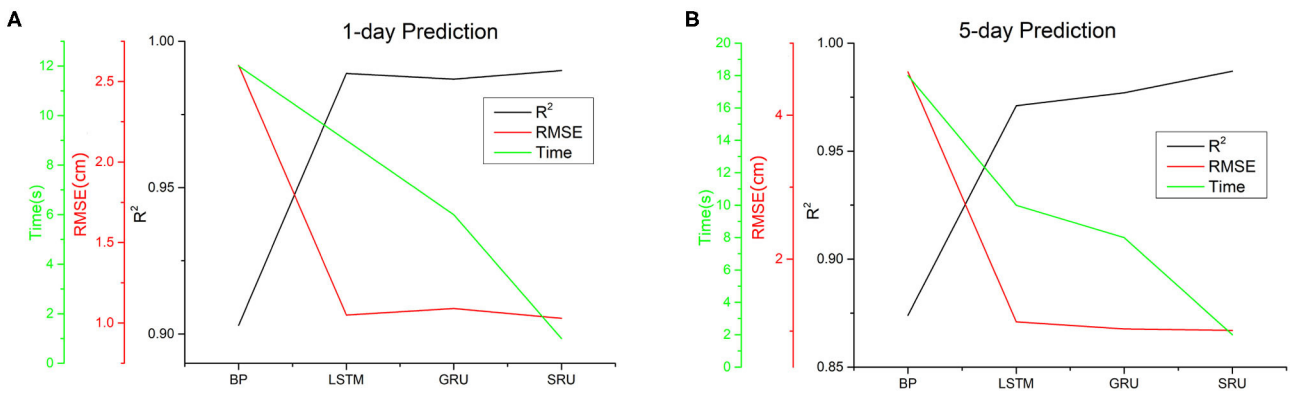


FIGURE 6 | The experimental results of prediction,1-day prediction in **(A)** and 5-day prediction in **(B)**. LSTM and its variants GRU and SRU methods have the advantage of gating mechanism, which leads to the very high goodness of fit values.

operations in SRU can be easily parallelized. That is, SRU network ensures most of the operations are put into parallel processing. Only the steps with small amount of operations are serialized.

2.2. The Data Source and Model Construction

In this paper, the sources of data were the altimetry data of TOPEX/Poseidon, Jason 1, ERS-1, and ERS-2 downloaded

from the Archiving Validation and International of Satellites Oceanographic (AVISO) website (http://marine.copernicus.eu/). This data consist of daily product on a 1/4° × 1/4° Mercator grid ranging from 1993 to 2018. The study area is the Bohai Sea (37–41° N, 117–122° E). The SSHA data ranges from January 1 1993 to December 31 2018 and is divided into training and testing sets on a 4:1 ratio. **Figure 3** shows the Research area and the Absolute dynamic topography of the Bohai Sea. **Figure 4** shows the climatology of SSHA distribution in the Bohai Sea. The graph shows that the overall difference in the Bohai Sea is smaller at the same time of the year. For the Bohai Sea, this study first calculate the daily mean SSHA value over the whole Bohai Sea region, and then used this one-dimensional time series data for training and testing. The length of the data sequence was 10,000, and 80% of the data were selected as the training dataset and the remaining 20% were used as the test set.

Before implementation of the training model, it was necessary to perform data normalization in the pre-processing step. This plays an important role in establishing a robust training model, which can (i) ensure a positive impact during updating of the parameters; (ii) reduce the training time; and (iii) help avoid the side effects caused by the variable distributions and ranges of input data, and make the calculation of the measurement parameters accurate as a result of training.

In this paper, we construct a SRU framework for SSHA prediction with a three-layer SRU modules. Its structure is shown in **Figure 5**, where x, y denote the input and output data and A, B, C denote the first, second, and third level of SRU module, respectively. This framework eliminates the dependence on $T-1$ time step, so it can be accelerated in parallel across time dimension (or input position). During the model training, several

TABLE 1 | The experimental results of 1 and 5-day prediction.

Model	1-day			5-day		
	R^2	RMSE (cm)	Time (s)	R^2	RMSE (cm)	Time (s)
BP	0.903	2.60	12	0.874	4.6	18
LSTM	0.989	1.05	9	0.971	1.13	10
GRU	0.987	1.09	6	0.977	1.03	8
SRU	0.99	1.03	1	0.987	1.01	2

TABLE 2 | The experimental results of 20-day prediction.

Model	R^2	RMSE (cm)	Time (s)
BP	0.743	6.86	24
LSTM	0.968	1.17	18
GRU	0.972	1.12	12
SRU	0.974	1.09	9
SRU*	0.994	0.89	4

**GPU-based parallel SRU algorithm*

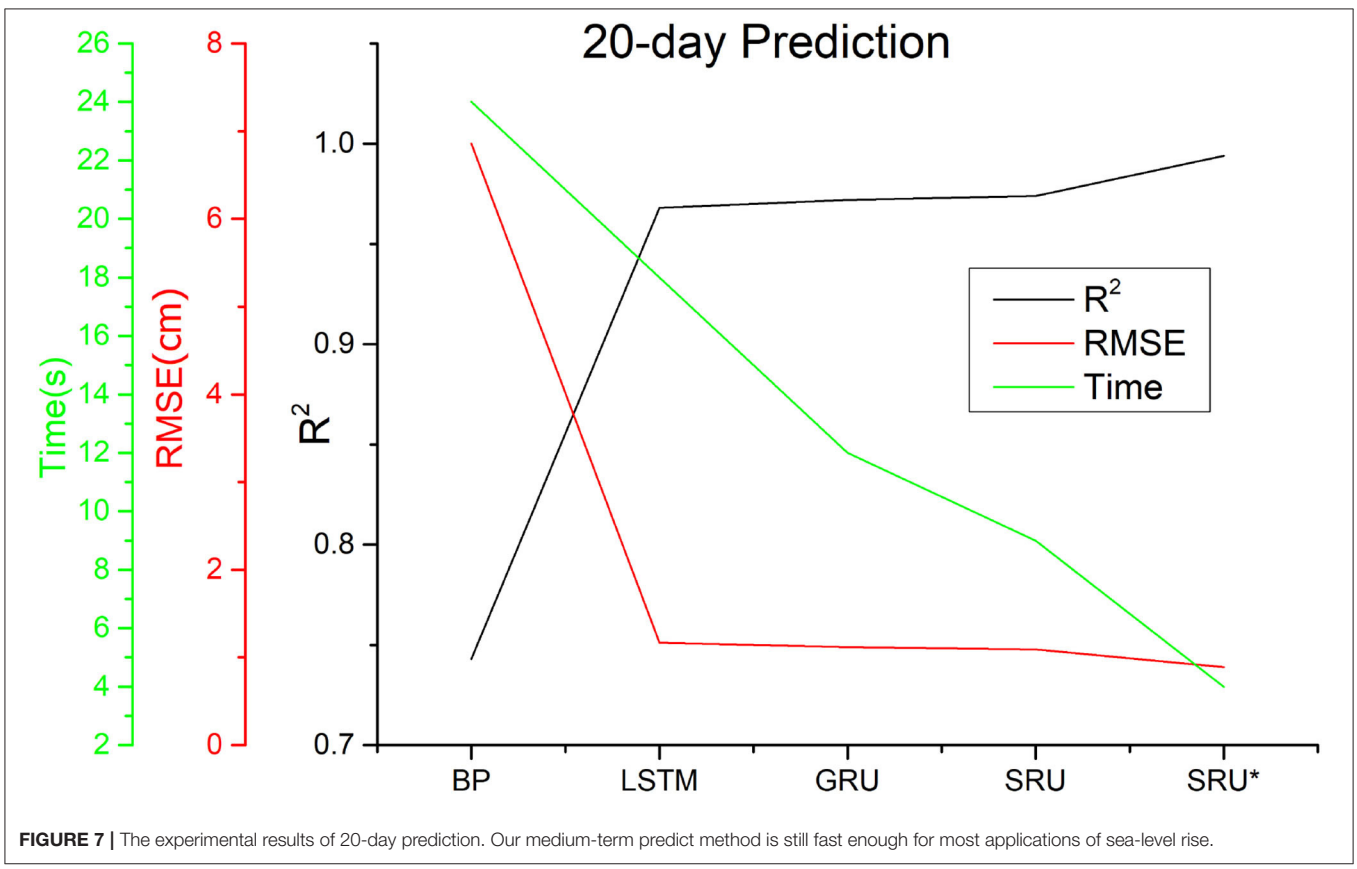


FIGURE 7 | The experimental results of 20-day prediction. Our medium-term predict method is still fast enough for most applications of sea-level rise.

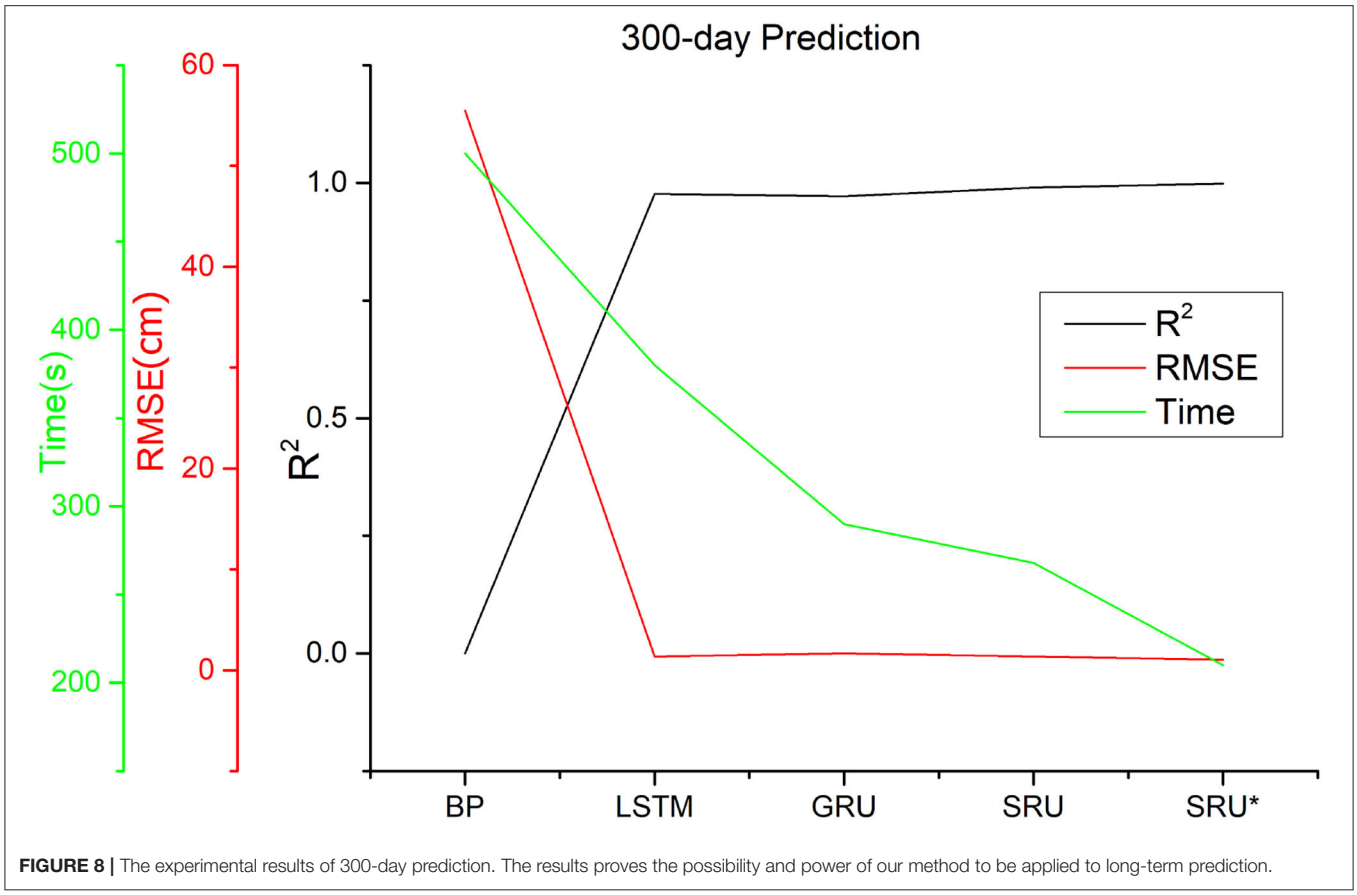


FIGURE 8 | The experimental results of 300-day prediction. The results proves the possibility and power of our method to be applied to long-term prediction.

TABLE 3 | The experimental results of 300-day prediction.

Model	R^2	RMSE (cm)	Time (s)
BP	0	55.48	500
LSTM	0.984	1.37	380
GRU	0.971	1.67	290
SRU	0.982	1.35	268
SRU*	0.991	1.05	210

**GPU-based parallel SRU algorithm*

parameters need to be initialized. The learning rate (Lr) was initialized to 0.001, batch size was initialized to 100, epoch was initialized to 50, and dropout was initialized to 0.5; each epoch ran 1,000 steps, and each step ran all the input data in the network for one round completely. The parameters of our method were optimized by the Adam algorithm to minimize the loss on the training dataset.

Well-known data normalization method include the maximum-minimum normalization and the average standard deviation normalization. Here, we use the first one, which has been widely used in deep learning systems. It can scale the original data into 0–1 range through dividing the original data by the difference of the maximum and minimum value. The formula is shown in Equation (2), where x_i^{scaler} denotes the value after normalization:

$$x_i^{scaler} = \frac{x_i - min\{x_i\}}{max\{x_i\} - min\{x_i\}}. \tag{2}$$

To evaluate the performance of our method, we define two error indicators to evaluate the performance of our method: R-Squared (R^2) and Root Mean Square Error (RMSE). R-Squared (R^2) is an important statistical measure of fit which indicates how much variation of a dependent variable is explained by the independent variable(s) in a regression model. It ranges from 0 to 1, where the best fit closes to 1. (R^2) can be calculated by Equation (4), where the mean square error (MSE) is defined in Equation (3) as follows:

$$MSE = \frac{1}{N}\Sigma_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{3}$$

$$R^2 = 1 - \frac{\Sigma_{i=1}^{N}\frac{(\hat{y}^{(i)} - y^{(i)})^2}{N}}{\Sigma_{i=1}^{N}\frac{(y^{(i)} - \bar{y})^2}{N}} = 1 - \frac{MSE(\hat{y}, y)}{Var(y)} \tag{4}$$

RMSE is the standard deviation of the residuals (prediction errors), which is a measure of how spread out these residuals are. It identifies how concentrated the data is near the best fit. RMSE is commonly used in prediction analysis to verify experimental results. In the AVISO dataset used in this paper, the SSHA data

are in centimeters and the unit of RMSE is the same as SSHA. It is defined in Equation (5) as follows:

$$RMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}[\hat{y}(i) - y(i)]} \quad (5)$$

3. PREDICTION RESULTS

To evaluate the effectiveness and efficiency of the proposed method, the performance of the proposed method based on the SRU deep network was compared with that of several existing models, including BP (He et al., 2018), LSTM (Graves, 2012), GRU (Li et al., 2021), and the original SRU (Lei et al., 2018) models without optimization and parallel computing. The performance of the original SRU network without parallel computing was considered mainly because in most scenarios parallel computing is supported in marine survey platforms and equipment. For certain applications, however, and especially for small survey vessels or on small islands, high performance equipment is not readily available and hence parallel computing is not supported. Thus, for these situations, it is necessary to ensure that the SRU framework can be applied for prediction of the SSHA. The short- to medium-term prediction was first performed for the 1, 5, and 20-day prediction. Also, to demonstrate the capability of the method for long-term prediction, a 300-day experiment was performed. The training and testing environment was equipped with 2 × 10 Intel Platinum9200 Xeon scalable processors (128 GB memory) and 16 Tesla V100 SXM2 GPU (total 512 GB).

3.1. Short-Term Prediction Results

The results for the 1 and 5-day predictions are presented in **Figure 6** and **Table 1** as follows, where SRU denotes the original SRU without optimization and parallel computing. This experiment was performed to show the effectiveness and efficiency of the SRU compared to the existing BP and LSTM family algorithms (LSTM, GRU) for short-term prediction. Given that the 1 and 5-day predications can be performed with the SRU very quickly, it was desirable to include parallel computing in the SRU implementation.

From the experimental results, it can be seen that the LSTM and the GRU and SRU variants outperform the BP method significantly in terms of both accuracy and efficiency. The respective R^2 values are all >0.9, and the RMSE values are around 1, which is much lower than 2.6, the value for the BP method. Compared with the commonly used BP neural network, the LSTM family of methods have the advantage of the gating mechanism, which leads to very high goodness-of-fit values. The forget gate in the network can filter the unimportant information automatically and leave the important information for prediction assessment. This mechanism enables the LSTM method to avoid the problem of the vanishing gradient during learning, and to select and memorize important information to make the prediction more accurate.

By comparing the LSTM, GRU, and SRU methods, it can be seen that the SRU gave much better performance than the LSTM and GRU due to its simpler structure. The accuracy of the algorithm is comparable with and even outperforms that of the LSTM and GRU, while it is much faster than the LSTM and GRU methods. Further, the time consumption factor has been reduced to 1 s for 1-day prediction and 2 s for 5-day prediction, which makes the short-term SSHA prediction essentially equivalent to real-time prediction. Overall, this experiment demonstrates good performance for the SRU method in terms of short-term prediction of the SSHA.

3.2. Medium-Term Prediction Results

The medium-term prediction, which plays a significant role in monitoring sea-level changes for decision making, is the most appropriate task for demonstrating the performance of the proposed method. Here, the revised SRU model (SRU*), which features skip-connection optimization and parallel computation, was compared with the BP, LSTM, GRU, and the original SRU method (without optimization and parallel computation). The experimental results, presented in **Figure 7** and **Table 2**, show that the proposed method achieves much more satisfactory results than the current published state-of-the-art methods. The R^2 value is much lower. This means that the skip-connection optimization can help alleviate the over-fitting and vanishing gradient problem in the SRU to reduce fitting errors to a large extent. Above all, the time consumption for the SRU* is much lower than that for the existing methods (BP, LSTM, GRU, and original SRU), hence demonstrating the effectiveness of parallel computing in SRU. Moreover, in this experiment, we also tried to build higher-layer SRU models. When the SRU deep network was increased from three to six layers, it was found that the accuracy increased (R^2 increased by 9.2% and RMSE decreased by 30%), while the training time changed little. In comparison with the results for short-term prediction, medium-term prediction cannot be performed in real time; however, the method is still fast enough for most applications that require medium-term prediction of sea-level rise.

3.3. Long-Term Prediction Results

To demonstrate the full extent of the prediction and fitting capabilities of the neural network model, the long-term 300-day prediction was examined. The training time for the three-layer SRU was about 210 s, and the prediction accuracy of the six-layer SRU was found to be better than that of the three-layer SRU; however, the increase in the prediction time was not significant because the jump connection structure greatly increases the calculation speed. In addition, comparing the R^2 and RMSE values for the SRU with those of the other neural network models, indicated that the run time for the whole SRU program was shorter than that of the LSTM and the GRU, and the actual prediction outcome was better. The main reason for the improvements is that the SRU's high network and light cycle structure make the SRU outperform the other neural networks in terms of the accuracy and speed of calculation, confirming that the SRU prediction time-series model is superior to that of the other RNN neural networks.

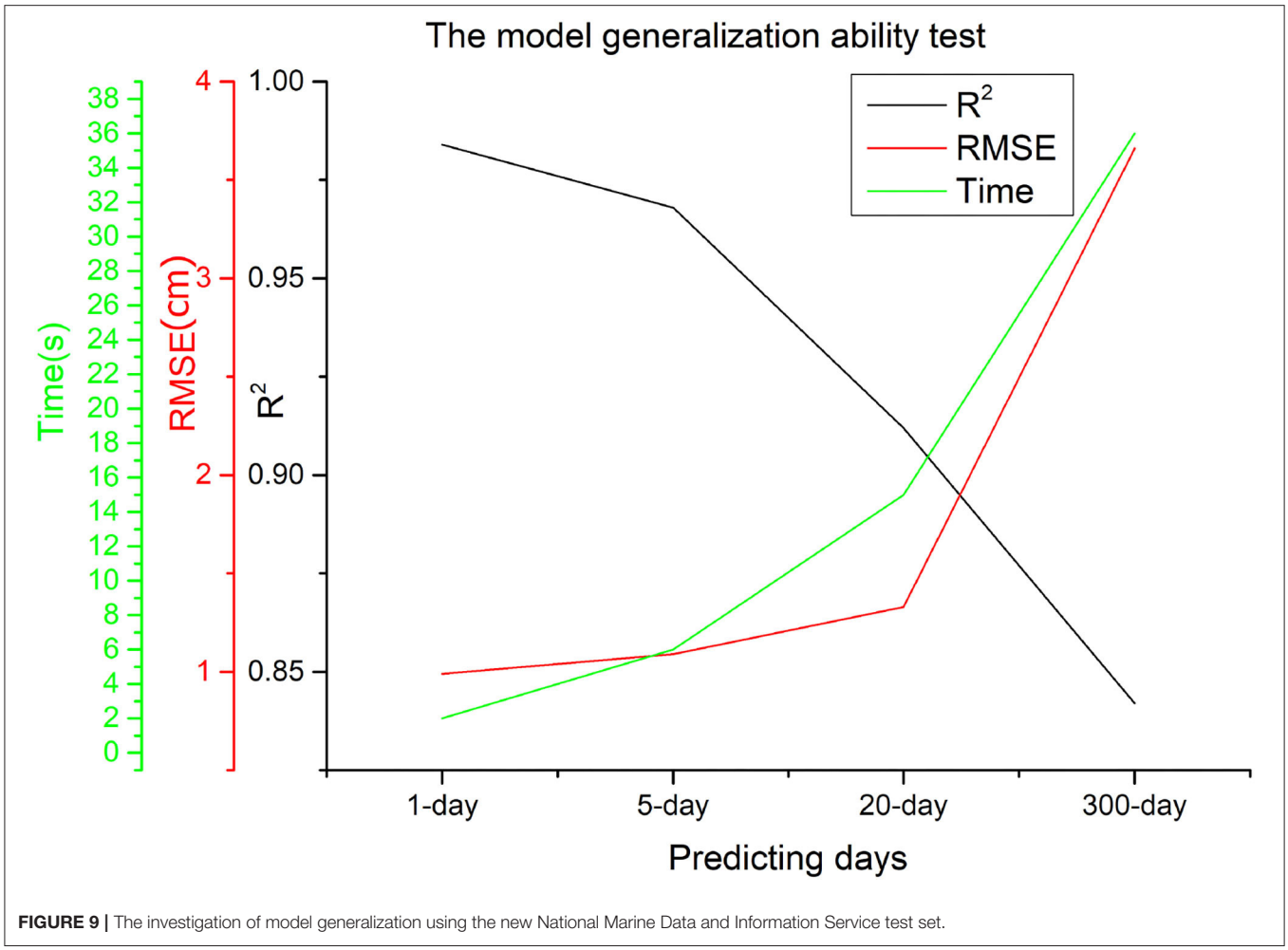


FIGURE 9 | The investigation of model generalization using the new National Marine Data and Information Service test set.

To demonstrate the capability for long-term SSHA prediction, the 300-day prediction experiment was performed. The results are shown in **Figure 8** and **Table 3**, it was observed that the method needed about 210 s to train a three-layer SRU model. Although the computational complexity of long-term prediction is somewhat higher than that of short- and medium-term predictions, the experiment demonstrates the possibilities and power of the method when it is applied to long-term prediction. In addition, compared to the BP and other LSTM models, the run time of the present method is much lower while the prediction accuracy is comparable or even better than the aforementioned methods.

The above-mentioned improvements arise mainly from the benefits from the skip connection optimization strategy, which effectively reduces the fitting error. The short- to long-term prediction results all demonstrate the effectiveness and efficiency of the SRU based framework to model the changes in the nonlinear chaotic structure of sea levels. The skip-connection strategy helps the SRU avoid the problem of the vanishing gradient in the empirical prediction, and, in so doing, simultaneously improves the prediction accuracy and reduces the computational complexity.

3.4. The Generalization of the Deep Learning Model

The model generalization of a deep learning model is also a key factor for considering to be used for practical applications. In this section, we investigate the generalization ability of the deep learning model by introducing a new test set: the datasets of National Marine Data and Information Service (NMDIS) reanalysis data (http://mds.nmdis.org.cn/). This dataset belongs to the reanalysis dataset. It is different from the observational AVISO dataset, which is used for training in this paper. The new dataset comprises of daily SSHA data from January 1, 1958 to December 31, 2018. The spatial resolution is $0.5^\circ \times 0.5^\circ$.

Experimental results of the deep learning model (SRU* with parallel computing and skip-connection) on the 1, 5, 20, and 300-day predictions are shown in **Figure 9** and **Table 4**, respectively. It can be seen from the results that for short and medium prediction, the performance of the deep learning model is almost the same with that on the original dataset. Although for the long term prediction, the R^2 of the deep learning model is a little bit lower and its RMSE is relatively higher than that on the original dataset. The performance of the deep learning method for long-term prediction on the new test set is still comparable with that

TABLE 4 | The investigation of model generalization using an additional test set.

Experiments	Model	R^2	RMSE (cm)	Time (s)
1-day	SRU*	0.984	0.99	2
5-day	SRU*	0.968	1.09	6
20-day	SRU*	0.912	1.33	15
300-day	SRU*	0.842	3.66	36

**GPU-based parallel SRU algorithm*

on the original dataset. That is, the deep learning model has very good generalization ability and it can be widely applied to SSHA prediction problems with variant kinds of dataset.

4. SUMMARY AND DISCUSSION

An SRU deep learning network for the short- to medium-term prediction of the sea surface height anomaly (SSHA) has been developed. The parallel structure of the SRU ensures good efficiency of the prediction. To avoid the over-fitting and vanishing gradient problems, the skip-connection strategy was utilized to speed up the convergence in the loss function in the SRU. To evaluate the effectiveness and efficiency of the method, detailed experiments were performed for 1, 5, 20, and 300-day SSHA predictions in the Bohai Sea using AVISO data. The main findings were as follows:

(1) For short-term (1 and 5-day) prediction, the original SRU model achieved the best performance with the lowest time consumption among the LSTM family of methods (LSTM and GRU) and also significantly outperformed the traditional BP method. The SRU method can effectively undertake short-term SSHA prediction in real time.
(2) For medium-term (20-day) prediction, the revised SRU and SRU* frameworks gave the most satisfactory results compared with the existing methods (BP, LSTM, GRU, original SRU) by introducing the skip-connection optimization strategy and parallel computing. When the SRU deep network was increased from three to six layers, the accuracy clearly increased (R^2 increased by 9.2% and RMSE decreased by 30%), while the training time changed little. Although the revised SRU cannot perform medium-term SSHA prediction, it is fast enough for most applications.
(3) For long-term (300-day) prediction, the revised SRU and SRU* framework also had great capabilities and possibilities for use in long-term prediction. Compared to the BP and other LSTM models, the running time of the method is much shorter while the prediction accuracy is comparable or even superior. These improvements arise mainly from the skip-connection optimization strategy, which effectively reduces the fitting error.

This study shows that the prediction of the changes in sea-level from the perspective of non-linearity can not only explain the physical laws underpinning the data pertaining to changes in sea-level, but can also play an essential role in establishing the prediction theory based on non-linear deep learning methods. The sensitivity of the non-linear chaotic system to the initial value shows that for the traditional linear model it is difficult to track and predict such non-linear data. In the critical period of global climate change, the changes in sea-level not only follow the basic laws of physics, but also are affected by the unique and non-repetitive events of the climate, such as the impact of global warming and the ENSO (El Nino southern oscillation) events. There are many factors that affect the sea-level, hence proper understanding of the changes which affect the sea-level is a complex task. Given such circumstances, there is an immediate need to improve the accuracy of the prediction algorithms such as the non-linear deep learning methods; these methods are of great significance in the context of prediction capability and forward-looking decision-making.

For the future work, we will put our efforts in improving the performance of SRU framework by introducing more effective optimization and speeding-up techniques to make our method more powerful to be applied for medium and long-term prediction. Moreover, the current work treats the Bohai Sea as a whole region, and uses its mean value as one-dimensional data for SSHA prediction. In the future, we also want to build two-dimensional prediction model, which can be applied for SSHA prediction over each grid point in the Bohai Sea.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PN and XZ: conceptualization and software. PN, XZ, and XJ: formal analysis. PN, XZ, and CZ: methodology and writing—review and editing. PN: writing—original draft. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Natural Science Foundation of China (No. 41806116) and Municipal Natural Science Foundation of Tianjin (No. 20JCQNJC01230).

ACKNOWLEDGMENTS

This research was supported/partially supported by Tianjin University, National Marine Data and Information Service. We have to express our appreciation to the Jingsheng Zhai for sharing pearls of wisdom with us during the course of this research. We are also immensely grateful to Jinkun Yang and Yulong Liu for their comments on earlier versions of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed professionals. We are also grateful to Xiaoyu He for assistance with python programming technology, and Xi He who polished this paper and in that line improved the manuscript significantly.

REFERENCES

Cui, Q., Xie, W., and Liu, Y. (2018). Effects of sea level rise on economic development and regional disparity in China. *J. Clean. Prod*. 176, 1245–1253. doi: 10.1016/j.jclepro.2017.11.165

Fasullo, J. T., Gent, P. R., and Nerem, R. S. (2020). Sea level rise in the CESM large ensemble: the role of individual climate forcings and consequences for the coming decades. *J. Clim*. 33, 6911–6927. doi: 10.1175/JCLI-D-19-1001.1

Feng, J., Li, D., Wang, T., Liu, Q., Deng, L., and Zhao, L. (2019). Acceleration of the extreme sea level rise along the chinese coast. *Earth Space Sci*. 6, 1942–1956. doi: 10.1029/2019EA000653

Graves, A. (ed.) (2012). "Long short-term memory," in *Studies in Computational Intelligence, Vol. 1* (Berlin; Heidelberg: Springer), 37–45. doi: 10.1007/978-3-642-24797-2_4

Guo, J., Wang, J., Hu, Z., Hwang, C., Chen, C., and Gao, Y. (2015). Temporal-spatial variations of sea level over China seas derived from altimeter data of TOPEX/Poseidon, Jason-1 and Jason-2 from 1993 to 2012. *Chin. J. Geophys*. 58, 3103–3120. doi: 10.6038/cjg20150908

He, L., Chen, J., Zhang, Y., Guo, T., and Li, G. (2018). An EEMD and BP neural network hybrid approach for modeling regional sea level change. *Desal. Water Treat*. 121, 139–146. doi: 10.5004/dwt.2018.22378

Jeon, T., Seo, K.-W., Kim, B.-H., Kim, J.-S., Chen, J., and Wilson, C. R. (2021). Sea level fingerprints and regional sea level change. *Earth Planet. Sci. Lett*. 567:116985. doi: 10.1016/j.epsl.2021.116985

Jiang, C., Chen, S., Chen, Y., Bo, Y., Han, L., Guo, J., et al. (2018). Performance analysis of a deep simple recurrent unit recurrent neural network (SRU-RNN) in MEMS gyroscope de-noising. *Sensors* 18:4471. doi: 10.3390/s18124471

Kappelle, M. (2020). WMO statement on the state of the global climate in 2019. *World Meteorol. Organ*. 1, 1–10. doi: 10.13140/RG.2.2.13705.19046

Kittel, C., Amory, C., Agosta, C., Jourdain, N. C., Hofer, S., Delhasse, A., et al. (2021). Diverging future surface mass balance between the Antarctic ice shelves and grounded ice sheet. *Cryosphere* 15, 1215–1236. doi: 10.5194/tc-15-1215-2021

Lei, T., Zhang, Y., Wang, S. I., Dai, H., and Artzi, Y. (2018). "Simple recurrent units for highly parallelizable recurrence," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Vol. 1* (Brussels:Association for Computational Linguistics), 4470–4481. doi: 10.18653/v1/D18-1477

Li, N., Hu, L., Deng, Z.-L., Su, T., and Liu, J.-W. (2021). Research on GRU neural network satellite traffic prediction based on transfer learning. *Wireless Pers. Commun*. 118, 815–827. doi: 10.1007/s11277-020-08045-z

Mahata, S. K., Das, D., and Bandyopadhyay, S. (2019). MTIL2017: machine translation using recurrent neural network on statistical machine translation. *J. Intell. Syst*. 28, 447–453. doi: 10.1515/jisys-2018-0016

Qiao, F., Wang, G., Khokiattiwong, S., Akhir, M. F., Zhu, W., and Xiao, B. (2019). China published ocean forecasting system for the 21st-century maritime silk road on december 10, 2018. *Acta Oceanol. Sin*. 38, 1–3. doi: 10.1007/s13131-019-1365-y

Qiao, F., Zhao, W., Yin, X., Huang, X., Liu, X., Shu, Q., et al. (2016). "A highly effective global surface wave numerical simulation with ultra-high resolution," in *SC16: International Conference for High Performance Computing, Networking, Storage and Analysis* (Beijing), 46–56. doi: 10.1109/SC.2016.4

Shao, C., Zhang, W., Sun, C., Chai, X., and Wang, Z. (2015). Statistical prediction of the South China Sea surface height anomaly. *Adv. Meteorol*. 2015:907313. doi: 10.1155/2015/907313

Song, T., Jiang, J., Li, W., and Xu, D. (2020). A deep learning method with merged LSTM neural networks for SSHA prediction. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens*. 13, 2853–2860. doi: 10.1109/JSTARS.2020.2998461

Tang, W., Zhan, W., Jin, B., Motagh, M., and Xu, Y. (2021). Spatial variability of relative sea-level rise in Tianjin, China: insight from InSAR, GPS, and tide-gauge observations. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens*. 14, 2621–2633. doi: 10.1109/JSTARS.2021.3054395

Wang, X., Wang, Y., Tan, M., and Cai, F. (2020). Deep-water deposition in response to sea-level fluctuations in the past 30 kyr on the northern margin of the South China Sea. *Deep Sea Res. Part I* 163:103317. doi: 10.1016/j.dsr.2020.103317

Wang, Z., Wu, J., Liu, C., and Gu, G. (2017). "The impact of sea level rise," in *Integrated Assessment Models of Climate Change Economics, Vol. 3* (Singapore: Springer), 43–57. doi: 10.1007/978-981-10-3945-4_3

Yan, D., Wünnemann, B., Gao, S., and Zhang, Y. (2020). Early Holocene tidal flat evolution in a western embayment of East China Sea, in response to sea level rise episodes. *Quat. Sci. Rev*. 250:106642. doi: 10.1016/j.quascirev.2020.106642

Yu, K., Liu, H., Chen, Y., Dong, C., Dong, J., Yan, Y., et al. (2019a). Impacts of the mid-latitude westerlies anomaly on the decadal sea level variability east of China. *Clim. Dyn*. 53, 5985–5998. doi: 10.1007/s00382-019-04909-8

Yu, Y., Si, X., Hu, C., and Zhang, J. (2019b). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput*. 31, 1235–1270. doi: 10.1162/neco_a_01199

Zhou, C., Liu, H., and Jiang, X. (2021). *Advanced Manufacturing and Automation X, Vol. 1*. Singapore: Springer.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Marine Science

ORIGINAL RESEARCH
published: 17 September 2021
doi: 10.3389/fmars.2021.690962

The Synthesis of Unpaired Underwater Images for Monocular Underwater Depth Prediction

Qi Zhao [1†], Ziqiang Zheng [2†], Huimin Zeng [1], Zhibin Yu [1,3], Haiyong Zheng [1] and Bing Zheng [1,3]*

[1] College of Electronic Engineering, Ocean University of China, Qingdao, China, [2] Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR China, [3] Sanya Oceanographic Institution, Ocean University of China, Sanya, China

Underwater depth prediction plays an important role in underwater vision research. Because of the complex underwater environment, it is extremely difficult and expensive to obtain underwater datasets with reliable depth annotation. Thus, underwater depth map estimation with a data-driven manner is still a challenging task. To tackle this problem, we propose an end-to-end system including two different modules for underwater image synthesis and underwater depth map estimation, respectively. The former module aims to translate the hazy in-air RGB-D images to multi-style realistic synthetic underwater images while retaining the objects and the structural information of the input images. Then we construct a semi-real RGB-D underwater dataset using the synthesized underwater images and the original corresponding depth maps. We conduct supervised learning to perform depth estimation through the pseudo paired underwater RGB-D images. Comprehensive experiments have demonstrated that the proposed method can generate multiple realistic underwater images with high fidelity, which can be applied to enhance the performance of monocular underwater image depth estimation. Furthermore, the trained depth estimation model can be applied to real underwater image depth map estimation. We will release our codes and experimental setting in https://github.com/ZHAOQIII/UW_depth.

Keywords: underwater vision, underwater depth map estimation, underwater image translation, generative adversarial network, image-to-image translation

OPEN ACCESS

Edited by:
Jun Li,
University of Technology Sydney, Australia

Reviewed by:
Boguslaw Cyganek,
AGH University of Science and Technology, Poland
Ancheng Lin,
University of Technology Sydney, Australia

****Correspondence:***
Zhibin Yu
yuzhibin@ouc.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *05 April 2021*
Accepted: *23 August 2021*
Published: *17 September 2021*

Citation:
Zhao Q, Zheng Z, Zeng H, Yu Z, Zheng H and Zheng B (2021) The Synthesis of Unpaired Underwater Images for Monocular Underwater Depth Prediction. Front. Mar. Sci. 8:690962. doi: 10.3389/fmars.2021.690962

1. INTRODUCTION

As an important part of underwater robotics and 3D reconstruction, underwater depth prediction is crucial for underwater vision research. However, the quality of collected images is restricted by light refraction and absorption, suspended particles in the water, and color distortion, making it difficult and challenging to obtain reliable underwater depth maps. Due to the influence of strong absorption and scattering, some widely used devices designed to obtain in-air depth maps, such as Kinect units (Dancu et al., 2014), lidar (Churnside et al., 2017), and binocular stereo cameras (Deris et al., 2017), exhibit limited performance in underwater environments (Massot-Campos and Oliver-Codina, 2015; Pérez et al., 2020). As quite a few underwater RGB-D datasets (Akkaynak and Treibitz, 2019; Gomez Chavez et al., 2019; Berman et al., 2020) are currently available, many researchers have sought to adopt image processing methods to estimate the depth from a single monocular underwater image or a consecutive underwater image sequence. To perform single monocular underwater depth prediction, several restoration-based methods have been developed

(e.g., UDCP; Drews et al., 2016; Ueda et al., 2019). The transmission map is regarded as an intermediate step for obtaining depth maps and restoring underwater images. In theory, the physical process is highly dependent on the calibrated intrinsic parameters and the well-described structural information of the scene. However, it is extremely laborious to select and measure these parameters relevant to the physical process (Abas et al., 2019), and limited to some special task.

Recently, deep learning methods have shown great potential in image processing (Li et al., 2018) applications, such as image-to-image translation (Isola et al., 2017; Zhu et al., 2017a; Choi et al., 2018; Wang et al., 2018b; Zheng et al., 2020), image restoration (Peng et al., 2015), and depth estimation (Gupta and Mitra, 2019). Due to the lack of the underwater depth ground truth to formulate full supervision, supervised learning models cannot be directly adopted for underwater depth estimation. Due to the introduction of cycle-consistency loss designed for unpaired image-to-image translation, many researchers aim to translate the in-air images to the desired underwater images and preserve the original depth annotation (Li et al., 2017, 2018; Gupta and Mitra, 2019). With the synthetic underwater images from the original in-air images paired with the corresponding depth annotation, we can obtain the pseudo underwater and depth image pairs. Previous methods such as WaterGAN (Li et al., 2017) and UMGAN (Li et al., 2018) adopted a two-stage optimization framework for underwater depth estimation. The former underwater image synthesis and the downstream vision task (such as depth prediction or underwater image restoration) are optimized separately. The two models have no direct connection at the training stage. UW-Net (Gupta and Mitra, 2019) has addressed this problem and aims to perform underwater image synthesis and underwater depth estimation parallel. However, two competitive tasks with cycle-consistent learning lead to low training efficiency and inaccurate depth estimation outputs. The leakage of texture is another challenge. The depth value of a fish should be about equal. However, the bright color and textures of a fish may lead to an incorrect depth estimation result (**Figures 1B–E**).

To address these problems, we propose a novel joint-training generative adversarial network for both multi-style underwater image synthesis and depth estimation performed in an end-to-end manner. For the former image synthetic task, we aim to transfer the hazy in-air RGB-D images to multi-style underwater images while retaining the objects and the structural information of the in-air images and controlling the underwater style through one conditional input message. To take advantage of multi-task learning (Zhang and Yang, 2017) between underwater image synthetic and depth estimation tasks, we design a joint-training generator to estimate the depth from the synthesized underwater images through full supervision. Overall, our system includes two consecutive generators (responsible for the underwater image synthesis and underwater depth estimation, separately), which are trained simultaneously. To ensure that the generated underwater images retain the objects and the structural information of the in-air images, we consider perceptual loss (Johnson et al., 2016) computed at the selected layers as a structural loss along with the adversarial loss to optimize the whole network. Furthermore, we develop a depth loss to alleviate the texture leakage phenomenon as shown in **Figure 1**. Finally, we evaluate the effectiveness of our proposed method to synthesize underwater images and estimate the depth map of real underwater images, and the comprehensive experimental results demonstrate the superiority of the proposed method. Overall, our main contributions of this paper are summarized as follows:

- We propose a novel joint-training generative adversarial network, which can simultaneously handle the controllable translation from the hazy RGB-D images to the multi-style realistic underwater images by combining one additional label, and the depth prediction from both the synthetic and real underwater images.
- To construct a semi-real underwater RGB-D dataset, we take the hazy in-air RGB-D image pairs and conditional labels as inputs to synthesize multi-style underwater images. During the training process, we introduce perceptual loss to preserve the objects and structural information of the in-air images during the image-to-image translation process.
- To improve the results of underwater depth estimation, we design the depth loss to make better use of high-level and low-level information. We verify the effectiveness of our proposed method on a real underwater dataset.

2. RELATED WORK

2.1. Image-to-Image Translation

In the past several years, a series of image-to-image translation methods based on generative adversarial networks (GANs) (Mirza and Osindero, 2014; Odena et al., 2017) have been proposed. These approaches can mainly be divided into two categories of paired training and unpaired training methods. Pix2pix (Isola et al., 2017) is a typical powerful paired model and first proposes cGAN (Mirza and Osindero, 2014) learns the one-side mapping function from the input images to target images. To achieve the image-to-image translation of unpaired datasets, CycleGAN (Zhu et al., 2017a) translates images into two domains using two generators and two discriminators and proposes the cycle-consistent loss to tackle the mode collapse of unpaired image translation. To address the multimodal problem, methods including BicycleGAN (Zhu et al., 2017b), MUNIT (Huang et al., 2018), DRIT (Lee et al., 2018), StarGAN (Choi et al., 2018), etc. have been proposed. The BicycleGAN (Zhu et al., 2017b) learns to transfer the given input with a low-dimensional latent code to more diverse results. It takes advantage of the bijective consistency between the latent and target spaces to avoid the mode collapse problem. MUNIT (Huang et al., 2018) achieves multidomain translation by assuming two latent representations that present style and content, respectively and combining different representations of content and style. StarGAN (Choi et al., 2018) learns multiple mapping functions between multiple domains. It only uses a single generator and a discriminator to transfers the source images to the target domain. Then to avoid mode collapse, the generator takes the generated images and the original

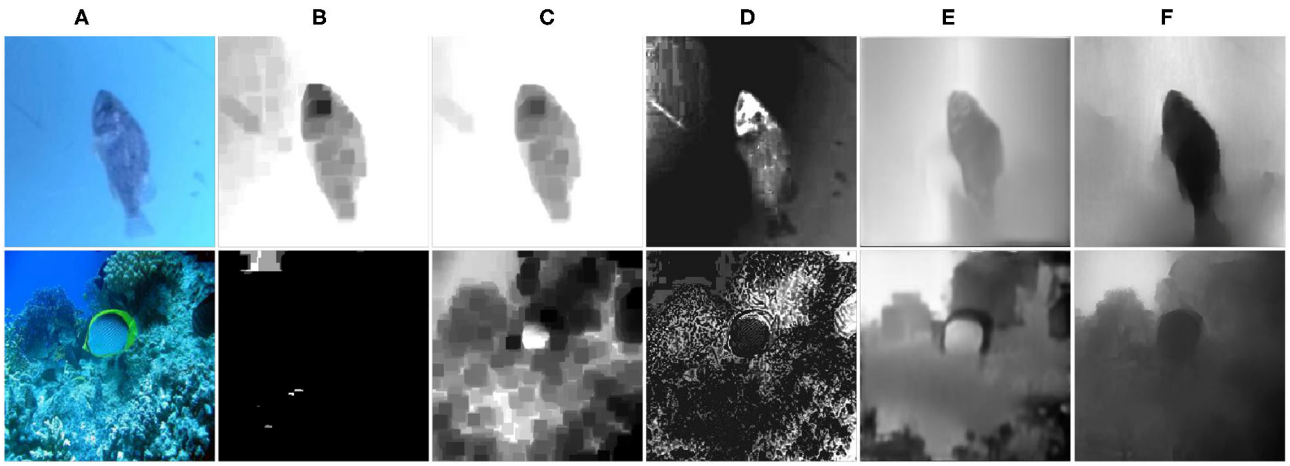


FIGURE 1 | Examples of texture leakage during the underwater depth map estimation process using different methods. **(A)** Real underwater images. **(B)** DCP (He et al., 2010), **(C)** UDCP (Drews et al., 2016), **(D)** Berman et al. (2017), **(E)** UW-Net (Gupta and Mitra, 2019), **(F)** ours.

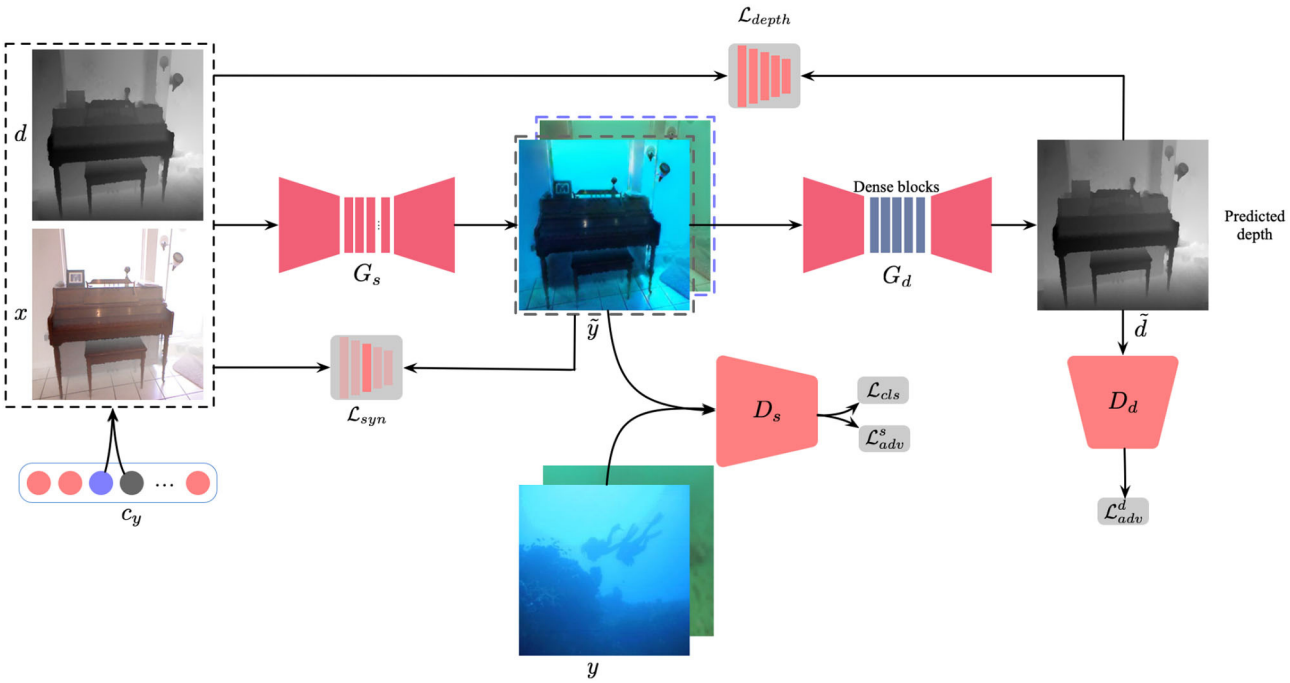


FIGURE 2 | The network framework of our proposed model is designed to synthesize multi-style underwater images and estimate underwater depth maps. The generator G_s and the discriminator D_s are used to synthesize multi-style underwater images, and the generator G_d and discriminator D_d learn to estimate underwater depth map based on the synthesized underwater RGB-D dataset.

labels as input and transfers them to the original domain. The subsequently developed image-to-image translation methods, such as pix2pixHD (Wang et al., 2018b), GauGAN (Park et al., 2019), vid2vid (Wang et al., 2018a), FUNIT (Liu et al., 2019), NICE-GAN (Chen et al., 2020), and StarGAN v2 (Choi et al., 2020) pay more attention to generate higher visual quality, multiple outputs and have been applied in video and small sample studies.

To synthesize underwater images, due to the lack of a large paired underwater image dataset, studies have mainly focused on unsupervised learning. In a pioneering approach of underwater image synthesis, WaterGAN (Li et al., 2017) synthesized the underwater images from the in-air image and the paired depth map for real-time color correction of monocular underwater images. To achieve multidomain translation, UMGAN (Li et al., 2018) proposes an unsupervised method that combines

CycleGAN (Zhu et al., 2017a) and cGAN (Mirza and Osindero, 2014) with an additional style classifier to synthesize multi-style underwater images. UW-Net developed by Gupta and Mitra (2019) learns the mapping functions between unpaired hazy RGB-D images and arbitrary underwater images to synthesize underwater images and estimate the underwater depth map. This method translates the hazy RGB-D image to underwater images while it learns to convert underwater images to the hazy RGB-D images. However, WaterGAN (Li et al., 2017) and UW-Net (Gupta and Mitra, 2019) only provide a solution for single domain underwater image generation. UMGAN (Li et al., 2018) does not consider the transmission map as an extra clue to generate underwater images. Moreover, all of the synthesized underwater images using these methods still lack the characteristics of real underwater images and clear structural information.

2.2. Underwater Depth Map Estimation

Underwater depth map estimation has mainly been studied in the field of traditional image processing. Since, He et al. (2010) first proposed a dark channel prior (DCP) for dehazing, many methods based on DCP (He et al., 2010) have been proposed for underwater depth map estimation in recent years. Drews et al. (2016) proposed a method based on a physical model of light propagation and the statistical priors of the scene to obtain the medium transmission and scene depth in typical underwater scenarios. Peng et al. (2015) proposed a three-step approach consisting of pixel blurriness estimation, rough depth map generation, and depth map refinement for depth map estimation. Berman et al. (2017) took different optical underwater types into account and proposed a more comprehensive physical image formation model to recover the distance maps and object colors. They mainly considered transmission map estimation as an intermediate step to obtain a depth map. Due to the unknown scattering parameters and multiple possible solutions, the results of these methods are most likely to be incorrect (Gupta and Mitra, 2019).

Recently, many deep learning-based methods have been proposed for depth estimation. However, most of these approaches focus on depth estimation from in-air RGB images with full supervision, which are not suitable for underwater depth map estimation due to the lack of the paired RGB-D data. The above mentioned UW-Net developed by Gupta and Mitra (2019) proposed an unsupervised method to learn depth map estimation. It considers an in-air transmission map as a cue to synthesize underwater images and obtains the required depth map from the synthesized underwater images. However, this method cannot estimate the depth map from underwater images of multiple water types. Because two competitive tasks (hazy in-air image reconstruction and depth estimation) are assigned to one generator, the depth prediction results of UW-Net lack sharp outlines. Ye et al. (2019) proposed another unsupervised adaptation networks. They developed a joint learning framework which can handle underwater depth estimation and color correction tasks simultaneously. Unlike their work, in which the two networks (style adaptation network and task network) should be trained separately, our model is more simple and can be trained simultaneously. The depth loss and a fine-tune strategy make our model more efficient in practice for underwater depth map prediction.

3. MATERIALS AND METHODS

3.1. Overall Framework

In this paper, we aim to estimate the depth map from real underwater images. Because there are no paired underwater RGB-D images, we cannot perform supervised learning directly. Therefore, we choose to translate the original in-air images with corresponding depth to underwater images and obtain pseudo-paired images. To perform this task, we design an end-to-end system with two joint-training modules: multi-style underwater image synthesis and underwater depth estimation based on the synthetic paired samples. The former module is trained through unpaired training, while the latter adopts supervised training to achieve precise underwater depth estimation. The overall framework is shown in **Figure 2** and consists of two generators, namely, G_s: $x \rightarrow \tilde{y}$ and G_d: $\tilde{y} \rightarrow \tilde{d}$, where x and $\tilde{y}$ are the original in-air image and the synthesized underwater image with specific underwater style. $\tilde{d}$ is the estimated depth output. For discrimination, we also design two discriminators D_s and D_d to perform adversarial training to boost the underwater image synthesis and depth estimation, respectively. D_s aims to distinguish between real and fake images and identify the domains from which both the real images and the generated images originate. The discriminator D_d only learns to distinguish between the real and fake depth maps.

3.1.1. Multi-Style Underwater Image Synthesis

As shown in **Figure 2**, we refer to the training of StarGAN (Choi et al., 2018) to generate multi-style underwater images. To synthesize specified underwater style images, we adopt an additional one-hot vector c to represent domain attributes. To make the generator G_s depth-aware and preserve the original depth representation after translation, we concatenate the three inputs, namely, the in-air image (x), the target underwater style (c_y), and the corresponding in-air depth (d) to synthesize an underwater image $\tilde{y} = G_s[\mathcal{C}(x, d, c_y)]$ with the required style (c_y), where $\mathcal{C}$ denotes depthwise concatenation. To guarantee that the synthetic image $\tilde{y}$ has the target underwater style, we include an adversarial domain classifier D_s with two branches (one for domain classification and another for real/fake discrimination). The classification branch with the domain classification loss $\mathcal{L}_{cls}$ aims to recognize the underwater style (c_y) of both the synthesized image $\tilde{y}$ and the real underwater image y. Noted that y does not have the corresponding depth annotation due to the lack of underwater ground truth. The adversarial loss $\mathcal{L}^s_{adv}$ is computed to promote the naturalness of the synthetic images. The generator G_s from CycleGAN (Zhu et al., 2017a) and StarGAN (Choi et al., 2018) is one symmetric encoder-decoder architecture with 6 residual blocks.

3.1.2. Underwater Depth Estimation

In the training stage, we perform underwater estimation on the above-mentioned synthetic underwater images $\tilde{y}$ by adopting

a generator G_d with dense-block architectures. The output of generator G_s ($\tilde{y}$) is the input of generator G_d used to estimate its depth map $G_d(\tilde{y})$. Considering that we have the depth annotation d of the in-air images, we can obtain pseudo pairs to compute the $\mathcal{L}_{depth}$ between d and $\tilde{d}$. The discriminator D_d is also designed and has only one discrimination output. Furthermore, the adversarial loss $\mathcal{L}_{adv}^{d}$ in the depth space is conducted. For underwater depth map estimation, we use DenseNet (Jégou et al., 2017) as the generator. In UW-Net (Gupta and Mitra, 2019), the authors proved the importance of using hazy above-water images and compared the results of underwater depth maps estimation with different generator networks, including ResNet (He et al., 2016), Unet (Ronneberger et al., 2015), DenseNet (Jégou et al., 2017), and so on. In their work, DenseNet is proved to be the best choice.

3.2. Loss Functions

3.2.1. Multi-Style Underwater Image Synthesis

3.2.1.1. Adversarial Loss

Regular GANs use sigmoid activation output and the cross-entropy loss function (Goodfellow et al., 2014), which may cause a vanishing gradient during the learning process. To stabilize the training process and generate underwater images with higher quality, we adapt the least-squares loss (Mao et al., 2017) in our method. $\mathcal{L}_{adv}^{s}$ can be expressed as follows:

$$\begin{aligned}\mathcal{L}_{adv}^{s} = \min_{G}\max_{D}\{&\mathbb{E}x, y \sim P_{dta}(x,y)[(D_s(y)-1)^2] \\ &+ \mathbb{E}_{x\sim P_{data}(x)}[(D_s(\tilde{y})^2]\}, \\ &where \quad \tilde{y} = G_s(\mathcal{C}(x,d,c_y))),\end{aligned} \tag{1}$$

where G_s targets the transfer of a hazy in-air RGB-D image x by concatenating an underwater condition label c_y to synthesize image $G_s[\mathcal{C}(x,d,c_y)]$. The discriminator D_s attempts to distinguish the real underwater image y and the synthesized underwater image $\tilde{y}$.

3.2.1.2. Domain Classification Loss

For the given hazy in-air image x and an underwater domain style c_y, G_s translates x into an underwater image $\tilde{y}$, which can be properly classified to the desired target domain by D_s. To achieve this goal, the classification branch of D_s imposes the domain classification. For the real underwater image y, the domain classification loss $\mathcal{L}_{cls}^{r}$ is computed as:

$$\mathcal{L}_{cls}^{r} = \mathbb{E}_{y,c_y}[-\log D_s(c_y|y)]. \tag{2}$$

where the term $D_s(c_y|y)$ denotes a probability distribution over the underwater domain labels (c_y) computed by D_s. By minimizing this objective, D_s learns to classify an underwater image y to its original domain c_y. We assume that the underwater image and domain label pair (y, c_y) is given by the training data. For generator G_s, the loss function for the domain classification of synthetic underwater images is defined as:

$$\mathcal{L}_{cls}^{f} = \mathbb{E}_{\tilde{y},c_y}[-\log D_s(c_y|\tilde{y})]. \tag{3}$$

During the training, G_s tries to synthesize underwater image $\tilde{y}$ that can fool the classification branch of D_s.

3.2.1.3. Feature-Level Loss

Beyond the pixel-level loss, we design feature-level loss functions between the feature representations extracted from a pre-trained VGG19 network. The hybrid feature-level loss can effectively preserve the similarity of the object between the hazy in-air images and the synthesized underwater images. For the multi-style underwater image synthesis, we introduce a perceptual loss, namely, $\mathcal{L}_{syn}$. $\mathcal{L}_{syn}$ is designed to preserve the object content and loosen the restrictions on the color and textile changes after translation. $\mathcal{L}_{syn}$ is expressed as follows:

$$\mathcal{L}_{syn} = [||\Phi^{(i)}(x) - \Phi^{(i)}(G_s(x|c_y))||_1]. \tag{4}$$

where $\Phi^{(i)}$ denotes the parameters at the i-th layer of a pre-trained VGG19 network. Following the work by Kupyn et al. (2019), we compute the 1-norm distance at the same selected $i = 14$ layer of the VGG19 network between the hazy in-air images and the synthesized underwater images.

3.2.1.4. Reconstruction Loss

To perform unpaired training between in-air and underwater images, we include the cycle consistency loss (Zhu et al., 2017a) in our framework. The reconstruction loss $\mathcal{L}_{rec}$ between $\hat{x}$ and x is defined as follows:

$$\begin{aligned}\mathcal{L}_{rec} &= \mathbb{E}_{x,c_y,c_x}[||x-\hat{x}||_1], \\ \hat{x} &= G_s(\mathcal{C}(G_s(\mathcal{C}(x,d,c_y)),d,c_x)),\end{aligned} \tag{5}$$

where c_x and c_y indicate the original hazy in-air domain label and the target underwater domain style, respectively. G_s takes the counterpart $G_s(x|c_y)$, its corresponding depth, and the original domain label c_x as input and tries to reconstruct the original hazy in-air image. We adapt the L1 loss as our reconstruction loss. Note that we use the generator G_s twice, first to translate the hazy in-air RGB-D images into an underwater image in the target domain and then to reconstruct the hazy in-air RGB images from the translated images.

3.2.2. Underwater Depth Estimation

3.2.2.1. Adversarial Loss

For the second underwater depth estimation procedure, the adversarial loss $\mathcal{L}_{adv}^{d}$ is described as:

$$\begin{aligned}\mathcal{L}_{adv}^{d} = \min_{G}\max_{D}\{&\mathbb{E}_{G_s(\tilde{y}),d\sim P_{data}(\tilde{y},d)}[(D_d(d)-1)^2] \\ &+ \mathbb{E}_{\tilde{y}\sim P_{data}(\tilde{y})}[(D_d(\tilde{d}))^2]\}, \\ &where \quad \tilde{d} = G_d(G_s(\mathcal{C}(x,d,c_y))),\end{aligned} \tag{6}$$

where G_d learns the mapping function from the synthesized underwater images $\tilde{y}$ to the in-air depth d as $G_d(\tilde{y}) \rightarrow d$. D_d is responsible to recognize the fake ingredient from the synthesized depth output $\tilde{d}$.

3.2.2.2. Depth Loss

For underwater depth estimation, the pixel-level distance between the estimated value and the ground truth, such as 1-norm and 2-norm, is generally adopted to favor less blurring.

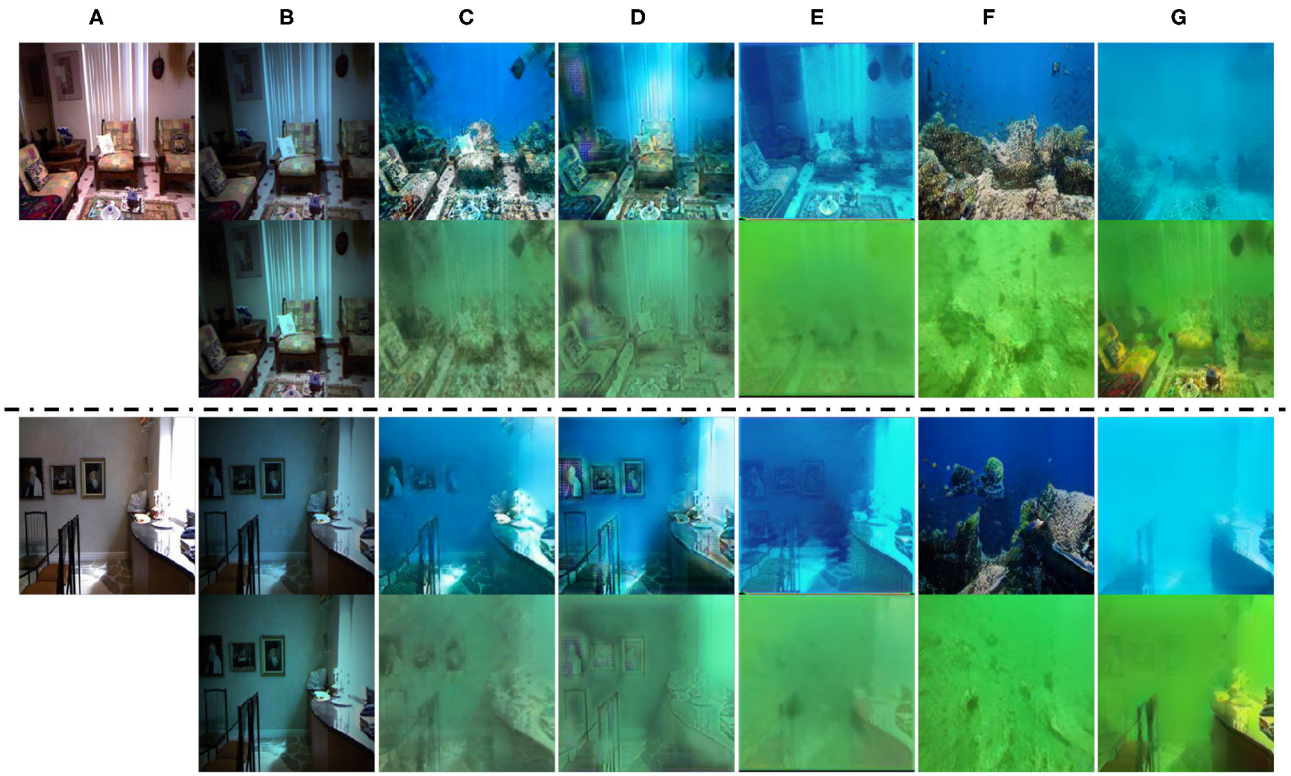


FIGURE 3 | Comparison of the visual quality of synthesized underwater images obtained by different methods. From left to right, **(A)** are original in-air images, **(B–G)** are the results of the WaterGAN (Li et al., 2017), CycleGAN (Zhu et al., 2017a), StarGAN (Choi et al., 2018), UW-Net (Gupta and Mitra, 2019), StarGAN v2 (Choi et al., 2020), and our method.

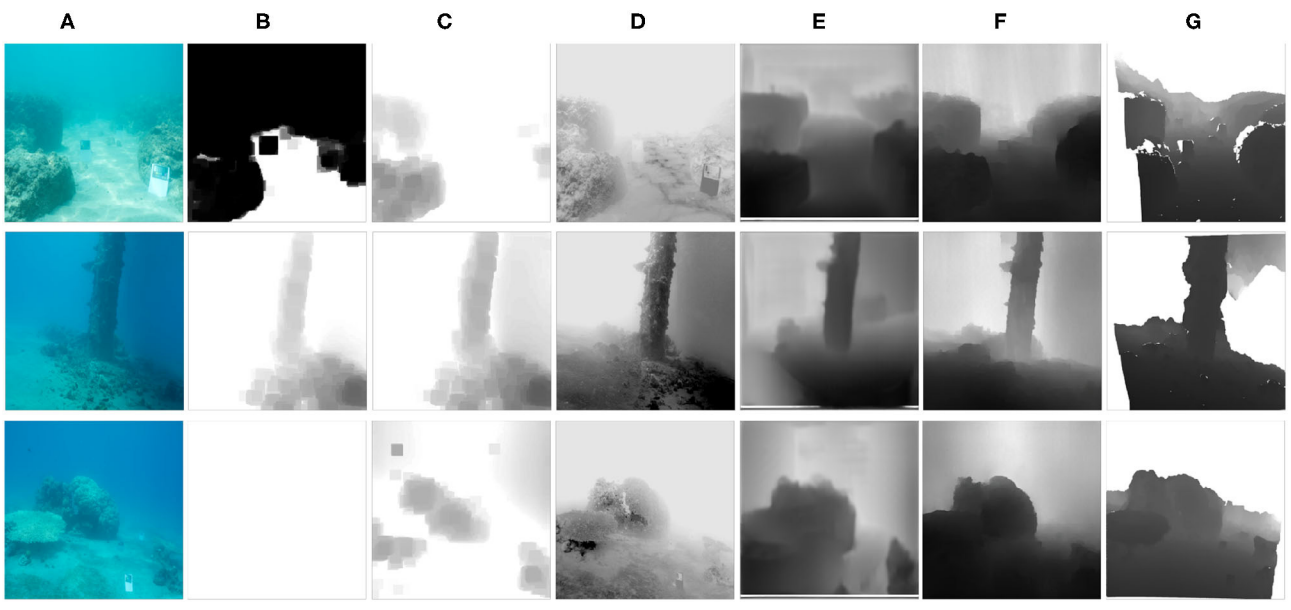


FIGURE 4 | Comparison of our method with other underwater depth estimation methods. From left to right, **(A)** are real underwater images from the dataset of Berman et al. (2017), **(B–F)** are the results of DCP (He et al., 2010), UDCP (Drews et al., 2016), Berman et al. (2017), Gupta and Mitra (2019), and our method, and **(G)** are the ground truths.

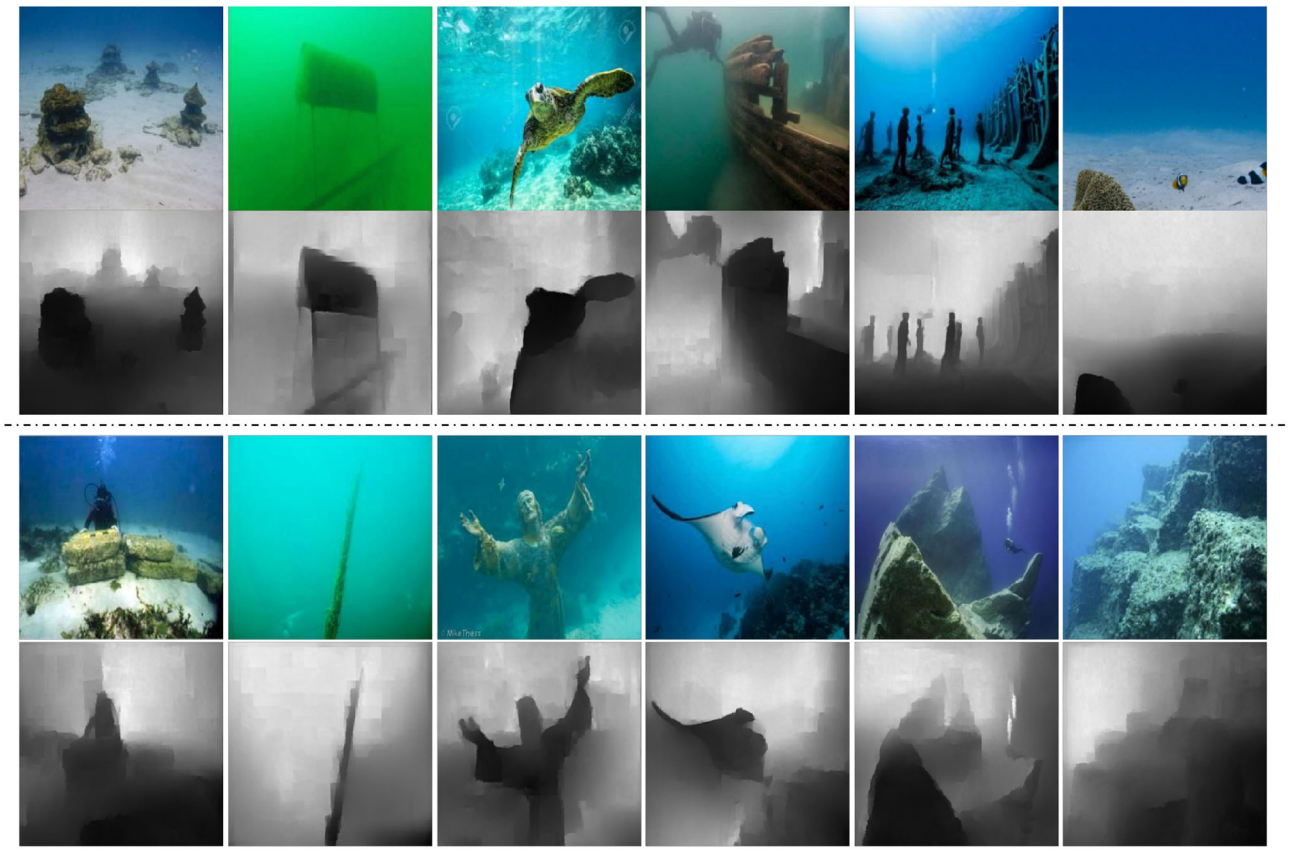


FIGURE 5 | The results of our model for depth map estimation. Every two rows from top to bottom are real underwater images with different illumination and scattering conditions and the results of our model for depth map estimation.

However, we find that only the pixel-level loss between the predicted depth map and the ground truth often leads to poor performance due to the influences of noise, water with various turbidity, etc (Please refer to section 4.3 for more details). To force the model to pay more attention to the objects, we make use of the feature representations extracted from a pre-trained VGG19 network for multi-level information. We also introduce pixel-level distance for low-level details. Finally, to obtain improved results, we combine 1-norm loss and the multi-layer feature constraint between $\tilde{d}$ and d and define the depth loss, namely $\mathcal{L}_{depth}$:

$$\mathcal{L}_{depth} = [||d - G_d(G_s(x|c_y))||_1] + \sum_{i=0}^{N}[||\Phi^{(i)}(d) - \Phi^{(i)}(G_d(G_s(x|c_y)))||_1]. \quad (7)$$

Similarly, $\Phi^{(i)}$ represents the pre-trained parameter of the i-th layer. Here, following the work of Wang et al. (2018b) and Wang C. et al. (2018), we compute the L1 distance at the same selected six layers: $i = 1, 6, 11, 20, 29$.

3.3. Full Objective

Finally, the objective functions can be written, respectively, as:

$$\mathcal{L}_{D_s} = \mathcal{L}^s_{adv} + \alpha\mathcal{L}^r_{cls} \quad (8)$$

$$\mathcal{L}_{G_s} = \mathcal{L}^s_{adv} + \gamma\mathcal{L}_{rec} + \alpha\mathcal{L}^f_{cls} + \lambda\mathcal{L}_{syn} \quad (9)$$

$$\mathcal{L}_{D_d} = \mathcal{L}^d_{adv} \quad (10)$$

$$\mathcal{L}_{G_d} = \mathcal{L}^d_{adv} + \eta\mathcal{L}_{depth} \quad (11)$$

where α, γ, λ, and η are the hyperparameters that control the effect of each loss in the final objective function. We set $\alpha = 5$, $\gamma = 10$, $\lambda = 0.1$, $\eta = 50$ in all of our experiments, and we optimize the objective function with the Adam optimizer (Kingma and Ba, 2015). To choose appropriate weights, we design ablation studies for each hyperparameter except for γ. We follow StarGAN (Choi et al., 2018) to set $\gamma = 10$. For the choice of the rest of hyperparameters, please refer to section 4.3 for more details.

TABLE 1 | Quantitative comparison of our method and other methods for underwater image synthesis.

	WaterGAN (FT)	CycleGAN (FT)	StarGAN (FT)	UW-Net (FT)	StarGAN v2 (FT)	Our (FT)
SI-MSE	0.5994	0.3514	0.4597	0.3594	0.5454	**0.2709**
ρ	0.5031	0.6024	0.5339	0.5795	0.4561	**0.6917**

We evaluate all models for underwater depth map estimation using the generated RGB-D datasets. FT represents a fine-tuned (FT) underwater model on the dataset of Berman et al. (2017). Higher ρ-values and lower SI-MSE (Eigen et al., 2014) values represent a better result. The bold values indicate the best result among different methods.

TABLE 2 | Quantitative comparison of our method and other methods on the dataset of Berman et al. (2017).

	DCP	UDCP	Berman et al.	UW-Net(FT)	Ours(FT)
SI-MSE	1.3618	0.6966	0.6755	0.3708	**0.1771**
ρ	0.2968	0.4894	0.6448	0.6451	**0.7796**

FT represents a fine-tuned (FT) underwater model. Higher ρ-values and lower SI-MSE (Eigen et al., 2014) values represent a better result. The bold values indicate the best result among different methods.

TABLE 3 | Comparison of Floating Point Operations (FLOPs) and total number of parameters among different generators with a size of 256 × 256.

Methods	FLOPs	Params
StarGAN (Choi et al., 2018)	52.32	8.417
CycleGAN (Zhu et al., 2017a)	56.83	11.38
StarGANv2 (Choi et al., 2020)	198.0	33.89
WaterGAN (Li et al., 2017)	132.7	24.18
Ours (G_s)	52.93	8.426
Ours (G_d)	12.98	1.348

4. RESULTS

4.1. Datasets and Implementation Details

In our experiments, we translate the hazy in-air images to two underwater domains (*green and blue*). We also choose the hazy in-air D-Hazy dataset (Ancuti et al., 2016) as the input images; this dataset contains the indoor scenes. For the two underwater domains, we adapt the real underwater images from the SUN (Xiao et al., 2010), URPC,[1] EUVP (Islam et al., 2020), UIEB (Li et al., 2019), and Fish datasets.[2] We collect 1,031 blue and 1,004 green underwater images from these datasets and the Google website, respectively. The D-Hazy dataset (Ancuti et al., 2016) includes 1,449 images. We randomly choose 1,300 images as the in-air images x to train the model. The remaining 149 images of the dataset are selected for evaluation. We use random-crop to obtain 128 × 128 patches for training. For the evaluation stage, we take complete images of 256 × 256. The entire network is trained on one Nvidia GeForce GTX 1070 using the Pytorch framework. To avoid the mode collapse problem, we apply spectral normalization (Miyato et al., 2018) in both the discriminators and the generators. Because of the introduction of spectral normalization (Miyato et al., 2018), we use a two-timescale update rule (TTUR) based on BigGAN (Brock et al., 2019) and SAGAN (Zhang et al., 2018). The Adam algorithm is applied with a learning rate of 0.0002 for the discriminators while 0.00005 for the generators. Because of the limited computing resources, we set the batch size to 10 and perform 100,000 training iterations in our experiments.

4.2. Comparison Methods

Our method achieves underwater depth map estimation using multi-style synthesized underwater images. In this section, we first evaluate the performance of WaterGAN (Li et al., 2017), CycleGAN (Zhu et al., 2017a), StarGAN (Choi et al., 2018), UW-Net (Gupta and Mitra, 2019), StarGAN v2 (Choi et al., 2020), and our method on multiple synthetic underwater images. Additionally, to evaluate the effectiveness of underwater depth map estimation, we compare the results obtained using DCP (He et al., 2010), UDCP (Drews et al., 2016), Berman et al. (2017), Gupta and Mitra (2019), and our method.

4.2.1. Qualitative Evaluation

To evaluate the effectiveness of the proposed method, we perform underwater image synthesis on the NYUv2 (Silberman et al., 2012) and D-Hazy (Ancuti et al., 2016) datasets. **Figure 3** shows a visual comparison of the synthesized underwater images generated by different methods. WaterGAN (Li et al., 2017) takes advantage of in-air RGB-D images to synthesize underwater images. As shown in **Figure 3B**, the results are somewhat single-hued and lack water characteristics. Although WaterGAN supports multi-style image generation, the two styles (blue and green) obtained by WaterGAN in **Figure 3B** are difficult to distinguish. The results of CycleGAN (Zhu et al., 2017a) retain most of the contents and structures of the original images. Compared to WaterGAN, they are similar to the natural underwater scenes shown in **Figure 3C**. By contrast, the outputs of CycleGAN (Zhu et al., 2017a) include serious distortions of the details of the image with incorrect depth information. StarGAN (Choi et al., 2018) can simultaneously translate in-air images into multiple underwater styles. However, the results lack the characteristics of real underwater images, such as depth information, and clear structural information of the objects. Besides, many artifacts are observed in **Figure 3D**. UW-Net (Gupta and Mitra, 2019) also takes hazy in-air RGB-D images as input, the results are presented in **Figure 3E** and show fuzzy structures for the objects. The results of StarGAN

[1]http://www.cnurpc.org/
[2]http://www.fishdb.co.uk/

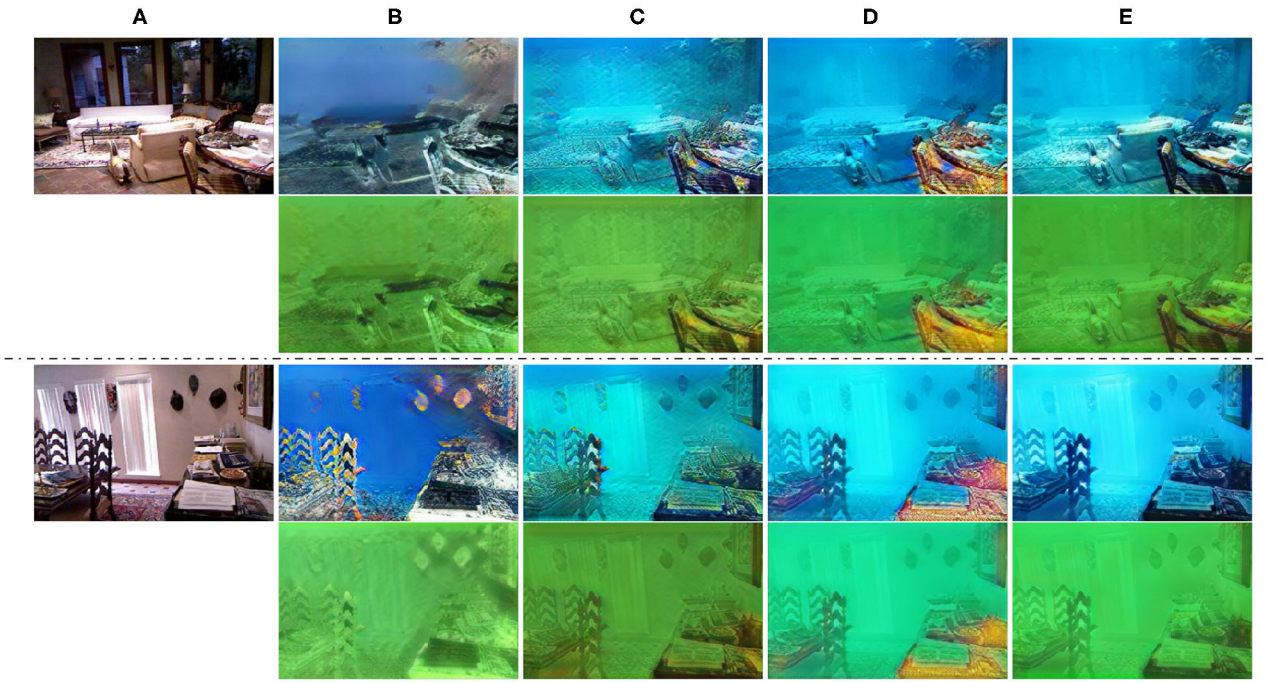


FIGURE 6 | Sample results of our method for synthesizing underwater images using different losses. $\mathcal{L}_{ssim}$, $\mathcal{L}_{msssim}$ and $\mathcal{L}_{syn}$, respectively represent SSIM loss, MS-SSIM loss, and perceptual loss. **(A)** Are in-air images, **(B)** are the results without any structural loss (Baseline), **(C–E)** are the results with $\mathcal{L}_{ssim}$, $\mathcal{L}_{msssim}$, and $\mathcal{L}_{syn}$, respectively.

TABLE 4 | Comparison of our method for the synthesis of underwater images with different combinations.

	Baseline	w/ $\mathcal{L}_{ssim}$	w/ $\mathcal{L}_{msssim}$	w/ $\mathcal{L}_{D_d}$	w/ $\mathcal{L}_{syn}$
SI-MSE	0.3538	0.2308	0.3331	0.2864	**0.1771**
ρ	0.6986	0.7547	0.7111	0.7355	**0.7796**

ResNet (He et al., 2016) represents a basic network for the synthesis of underwater images (Baseline). Our synthesized underwater images are mainly used to estimate depth maps. We show the results of depth maps estimation using ResNet (He et al., 2016) and ResNet (He et al., 2016) with extra losses. The bold values indicate the best result among different methods.

TABLE 5 | Comparison of weights used in the objective function of our model, including α and λ.

SI-MSE/ρ	$\alpha = 1$	$\alpha = 3$	$\alpha = 5$	$\alpha = 7$
$\lambda = 0.05$	0.2586/0.7438	0.2676/0.7502	0.2325/0.7593	0.2957/0.7402
$\lambda = 0.1$	0.2291/0.7513	0.2020/**0.7844**	**0.1771**/0.7796	0.2321/0.7717
$\lambda = 0.2$	0.2955/0.7331	0.2164/0.7688	0.2548/0.7524	0.2535/0.7331
$\lambda = 0.4$	0.2966/0.7236	0.2882/0.7306	0.2929/0.7499	0.2577/0.7577

We separately set α = 1, 3, 5, 7, and λ = 0.05, 0.1, 0.2, 0.4. We discover that α = 5 and λ = 0.1 perform better. The bold values indicate the best result among different methods.

v2 (Choi et al., 2020) are shown in **Figure 3F**. There is no denying that StarGAN v2 (Choi et al., 2020) possesses a powerful style network to extract style codes from reference images. However, the underwater images provided by StarGAN v2 fail to help the depth estimation tasks. As shown in **Figure 3F**, StarGAN v2 removed some objects and structural information during the image synthetic process, which makes the synthetic underwater images and their corresponding in-air depth maps unmatched. The quantitative results in section 4.2.2 further confirm this point.

Our model is optimized to synthesize underwater images with multiple styles based on the unpaired datasets. The results of our method (**Figure 3G**), in which the structural information is well preserved, are better than those obtained from other methods in terms of visual quality.

For underwater depth map estimation, **Figure 4** shows the results of our method and other methods developed by He et al. (2010) (DCP), (Drews et al., 2016) (UDCP), Berman et al. (2017), and Gupta and Mitra (2019) based on the underwater images obtained by Berman et al. (2017). In **Figures 4B–D**, these methods fail to capture relative depth of the scene with respect to the camera. Moreover, these methods mainly obtain the transmission maps of the scene and have excessive texture leakage in the results. Gupta and Mitra (2019) used an unsupervised method to estimate the depth map, obtaining the results shown in **Figure 4E**, and this method appears to be better than the other methods, whose results are presented in **Figures 4B–D**. However, this method still suffers from excessive texture leakage and only estimates the depth map for single-domain underwater images. Our results have a much more reasonable appearance with a

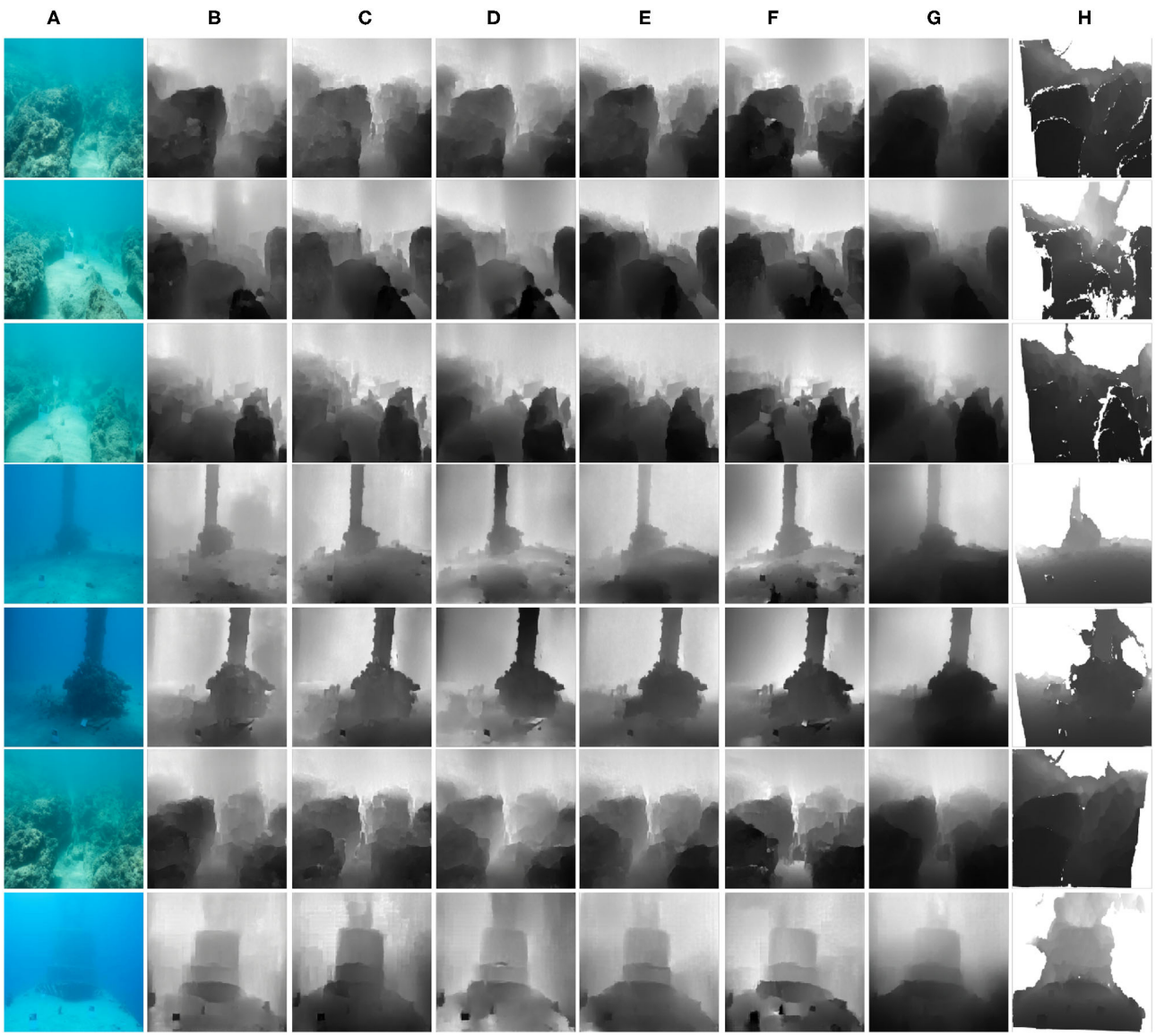


FIGURE 7 | Effectiveness evaluation of the $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_{ssim}$, $\mathcal{L}_{msssim}$, and $\mathcal{L}_{depth}$. From left to right, respectively, **(A)** are real underwater images, **(B–H)** are the results of depth map estimation with L1 loss, L2 loss, $\mathcal{L}_{ssim}$, $\mathcal{L}_{msssim}$, $\mathcal{L}_{pan}$, $\mathcal{L}_{depth}$, and their corresponding ground truths.

linear depth variation. On the other hand, we observe that our network successfully captures the depth information from multi-style underwater images. More results for real underwater images with different underwater characteristics are seen in **Figure 5**. Furthermore, the UW-Net (Gupta and Mitra, 2019) and our method synthesize underwater images using the underwater dataset provided by Berman et al. (2017) to fine-tune the models of the depth map estimation. We fine-tune our model for 10,000 iterations on Berman et al.'s dataset for better depth map estimation.

4.2.2. Quantitative Evaluation

The dataset of Berman et al. (2017) consists of 114 paired underwater RGB-D images from Katzaa, Michmoret, Nachsholim, and Satil. We use 71 images belonging to the three regions Katzaa, Nachsholim, and Satil. Because the Michmoret region has very few natural objects and is of the same scene. Following UW-Net (Gupta and

TABLE 6 | Results with different η values.

	$\eta = 40$	$\eta = 50$	$\eta = 60$	$\eta = 70$
SI-MSE	0.2657	**0.1771**	0.2620	0.2405
ρ	0.7266	**0.7796**	0.7315	0.7635

Higher ρ and lower SI-MSE (Eigen et al., 2014) values are better. The bold values indicate the best result among different methods.

TABLE 7 | Quantitative comparison of our method with different losses on the dataset of Berman et al. (2017).

	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_{ssim}$	$\mathcal{L}_{msssim}$	$\mathcal{L}_{pan}$	L_{depth}
SI-MSE	0.3103	0.2896	0.3983	0.2598	0.2856	**0.1771**
ine ρ	0.7279	0.7419	0.6515	0.7655	0.7397	**0.7796**

Higher ρ values and lower SI-MSE (Eigen et al., 2014) values indicate better results. The bold values indicate the best result among different methods.

Mitra, 2019), we use two metrics for comparison, namely, log scale-invariant mean squared error (SI-MSE) (Eigen et al., 2014) and the Pearson correlation coefficient (ρ). Considering the fact that the depth map provided by the stereo camera is not complete (e.g., the ground truth of the white regions in **Figure 7H** are not provided), we only calculate the pixels with a defined depth-value in the ground truth (GT).

The underwater image synthesis assists to estimate depth maps from real underwater images. Thus, how much the synthetic underwater images can be used to boost the performance of underwater image-based depth prediction is the key evaluation index. We evaluate performance on depth prediction tasks with a series of the state-of-the-art methods, which consist of WaterGAN (Li et al., 2017), CycleGAN (Zhu et al., 2017a), StarGAN (Choi et al., 2018), UW-Net (Gupta and Mitra, 2019), and StarGAN v2 (Choi et al., 2020). We aim to calculate the depth map estimation results on a semi-real underwater RGB-D dataset. UW-Net suggests that fine-tuning the models with a few unlabeled images from the target underwater environment could further boost the depth prediction performance. During the fine-tuning process, we only use the RGB underwater images without considering the depth ground truth of the data from Berman et al. to show the ability that our model can adapt itself to a new environment well. To make it fair, we fine-tune all models to generate a similar underwater style of the dataset of Berman et al..

Although our model already provides a solution for a depth estimation task, we choose a typical independent supervised image-to-image model, pix2pix (Isola et al., 2017), to fairly evaluate the potential of synthetic underwater images on the application of depth prediction. We use identical pix2pix models to learn the mapping function between the generate underwater images of different underwater image synthetic methods and their corresponding in-air depth maps. Finally, we test and evaluate all models on the dataset of Berman et al.. **Table 1** shows the results, and our model obtains higher ρ values and lower SI-MSE.

For the underwater depth estimation task, **Table 2** shows the quantitative results. Our method obtains the least scale-invariant error (SI-MSE) (Eigen et al., 2014) and the highest Pearson correlation coefficient (ρ).

We also investigate the parameters and Floating Point Operations (Tan and Le, 2019) (FLOPs) among different generators in **Table 3**. In the case of CycleGAN, we only count the FLOPs and parameters of a single generator. We can find that the proposed method can achieve better performance with fewer network parameters and computational cost. Benefiting from the dense blocks, the G_d of our model has fewer parameters and FLOPs than G_s. Please note that G_s is only used in training stage. In testing phase, we only need G_d to estimate the depth map.

4.3. Ablation Study

4.3.1. Loss Selection of Underwater Image Synthesis

To preserve clear structural information, we consider the perceptual loss $\mathcal{L}_{syn}$, structural similarity index (SSIM) $\mathcal{L}_{ssim}$, and multiscale structural similarity index (MS-SSIM) $\mathcal{L}_{msssim}$ as the structural loss. We evaluate the efficiency of each loss, including $\mathcal{L}_{syn}$, $\mathcal{L}_{ssim}$, and $\mathcal{L}_{msssim}$, and based on the visual effect of the synthesized underwater images and the results of depth map estimation, we choose the perceptual loss. To verify the effectiveness of the extra losses in our network, we design ablation experiments and perform a comparison on D-Hazy (Ancuti et al., 2016) which consists of 1,449 images. **Figure 6** shows that each loss affects the quality of the generated underwater images. It is observed from **Figure 6B**, that the generated underwater images using ResNet without any extra loss have more color blocks and artifacts. Additionally, during the training, it is extremely unstable and tends to produce color inversions and serious distortions situations. In **Figures 6C,D**, many artifacts are still retained for ResNet with $\mathcal{L}_{ssim}$ or $\mathcal{L}_{msssim}$. **Table 4** shows the results of depth map estimation based on different synthetic underwater image datasets, which are generated by ResNet and ResNet with extra losses, separately. Using $\mathcal{L}_{syn}$, we obtain the best results of underwater depth map estimation. Based on the experiments mentioned above, we introduce a perceptual loss $\mathcal{L}_{syn}$ to preserve the details and restrain the artifacts in **Figure 6E**. To minimize the negative effects of the synthesized images, we design experiments to determine the proper weight of α and λ. In **Table 5**, we show the results of different weights, including α and λ. We note that both UW-Net and our model can be fine-tuned on the dataset of Berman et al. to obtain better results of underwater depth map estimation. Fine-tuning processing provides a flexible approach for adjusting our model and the estimation of depth maps from unexplored underwater regions within a relatively short period.

4.3.2. The Design of Underwater Depth Map Estimation

With the support of synthetic paired RGB-D data, we consider L1 loss, L2 loss, L_{ssim} loss, or L_{msssim} loss to learn the mapping functions for supervised depth map prediction. During the training, we observe the all above-mentioned losses are not enough to generate more correct depth maps. The results in **Figures 7B–E** show that depth prediction based on the above-mentioned losses are easily affected by the shape, noise, etc. As mentioned in section 3.2.2, we design depth loss L_{depth} to make better use of low-level and high-level feature information and avoid the risk of texture leakage. We take advantage of a pre-trained VGG19 network to extract feature maps between the generated depth maps and the ground truths. We assume the feature maps between the generated depth map and its

FIGURE 8 | Sample results for the synthesis of underwater images. **(A)** Show in-air images. **(B–E)** Represent blue style, green style, white style and yellow style, respectively.

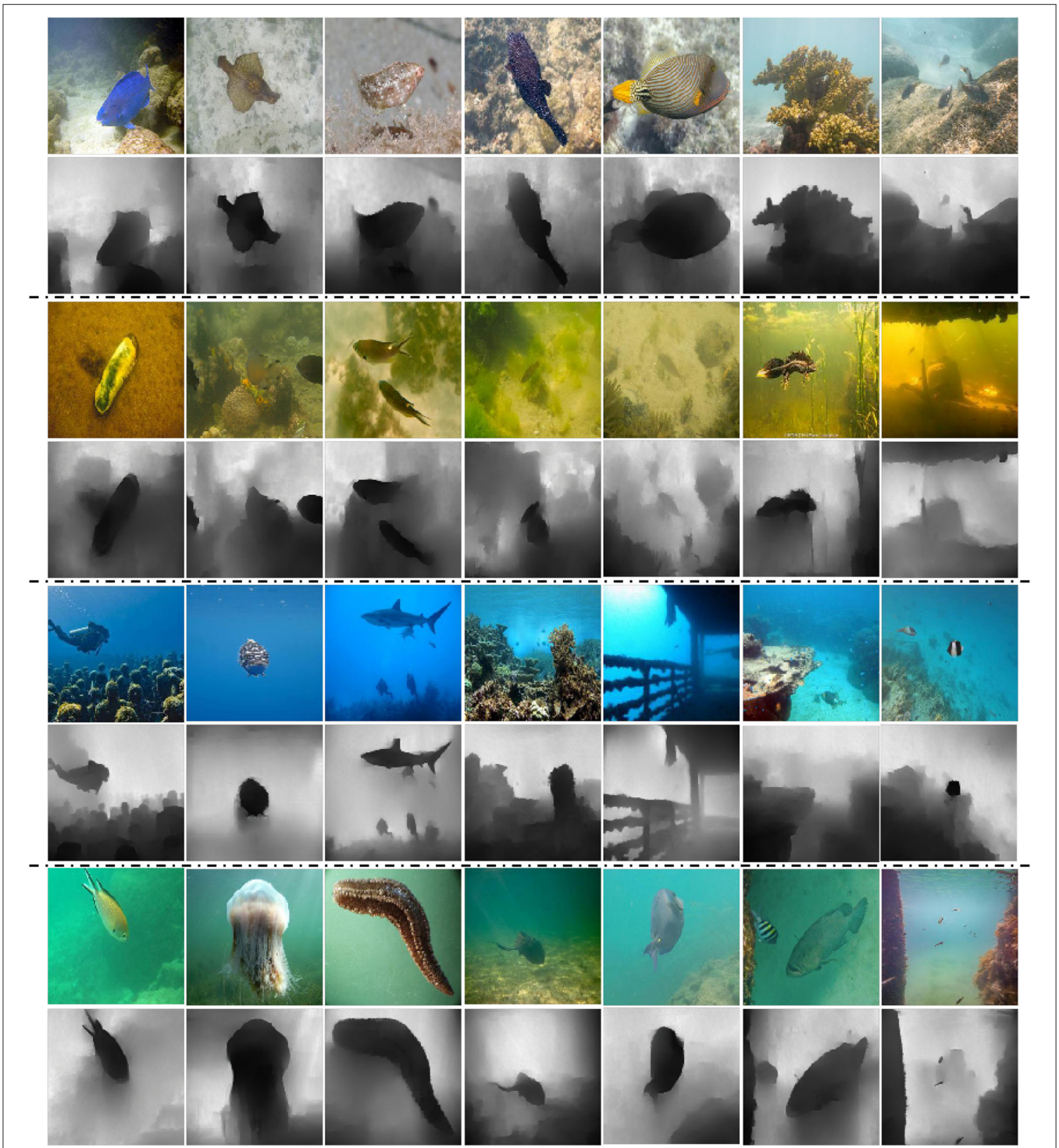


FIGURE 9 | Multi-style underwater depth map estimation. The rows from top to bottom are real underwater images with four different water types and the results of our model for depth map estimation. Every two rows are real underwater images and their predicted depth maps of our method.

corresponding ground truth in each layer from a pre-trained VGG19 network should be equal. The loss L_{depth} makes our model pay more attention to the objects and the relative distance in the underwater images. Inspired by Wang et al.'s work (Wang C. et al., 2018), we also attempt to extract feature maps from the discriminator D_d, namely $\mathcal{L}_{pan}$, rather than a pre-trained VGG19 network. In **Figure 7F**, we can see that our model with $\mathcal{L}_{pan}$ are often overwhelmed with incorrect boundary prediction

due to the insufficient layers of our discriminator D_d to extract high-level feature maps comparing with $\mathcal{L}_{depth}$. Furthermore, we investigate the optimal parameter setting of η with a greedily searching strategy (**Table 6**), and we discover that $\eta = 50$ is the best choice among all the parameters.

Based on **Figure 7** and **Table 7**, we can easily conclude that the results of depth map estimation using L_{depth} loss are more accurate and continuous. The results show sharper outlines. We can clearly distinguish the relative distance and the objects.

5. DISCUSSIONS AND CONCLUSION

To further explore the potential of our model on depth prediction, we considered the work by Li et al. (2018) and prepared a more complex underwater image dataset including four different styles. In this experiment, we still consider the depth map as a conditional input to synthesize a corresponding underwater image. But we did not utilize the physical parameters (e.g., the water turbidity or any optical parameters) for the unpaired image-to-image translation. Instead, we roughly divide the images with different water turbidity into four groups and follow the manner of StarGAN (Choi et al., 2018) to perform conditional image translation. Some synthetic examples of four different styles are shown in **Figure 8**. Due to the lack of ground truth of the depth map, we cannot quantitatively evaluate the effectiveness of our model for multi-style underwater depth map estimation. Instead, we prepared several qualitative evaluation results, as shown in **Figure 9**. Intuitively, we find that the depth estimation of a side-view underwater image is better than that from a vertical view. This result is caused by the lack of vertical view in-air images from the in-air D-Hazy dataset required to produce sufficient synthetic underwater vertical view images. We plan to improve the performance on this point by data augmentation in the future.

In this paper, we proposed an end-to-end system that can synthesize multi-style underwater images using one-hot encoding and estimate underwater depth maps. The system can convert the in-air RGB-D images into more realistic underwater images with multiple watercolor styles. Then we use the synthesized underwater RGB images to construct a semi-real underwater RGB-D dataset. With the synthetic underwater RGB-D dataset, our model can learn to estimate underwater depth maps using supervised learning. Finally, we compare our method with existing state-of-the-art methods to synthesize underwater images and estimate underwater depth maps, and we verify that our method outperforms these methods both qualitatively and quantitatively. Furthermore, our model can be fine-tuned on the untrained datasets to synthesize a similar underwater style. It effectively makes our model to be applied for depth map estimation on new underwater datasets.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

QZ performed the experiments and wrote the manuscript. ZZ and HZe revised the manuscript. ZY provided the ideas and revised the article. HZh and BZ provided advices and GPU devices for parallel computing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the finance science and technology project of 630 Hainan province of China under Grant Number ZDKJ202017 and National Natural Science Foundation of China under Grant Numbers 62171419 and 61701463.

REFERENCES

Abas, P. E., and De Silva, L. C. (2019). Review of underwater image restoration algorithms. *IET Image Process.* 13, 1587–1596. doi: 10.1049/iet-ipr.2019.0117

Akkaynak, D., and Treibitz, T. (2019). "Sea-thru: a method for removing water from underwater images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Long Beach, CA: Computer Vision Foundation; IEEE), 1682–1691. doi: 10.1109/CVPR.2019.00178

Ancuti, C., Ancuti, C. O., and De Vleeschouwer, C. (2016). "D-hazy: a dataset to evaluate quantitatively dehazing algorithms," in *IEEE International Conference on Image Processing* (IEEE), 2226–2230. doi: 10.1109/ICIP.2016.7532754

Berman, D., Levy, D., Avidan, S., and Treibitz, T. (2020). Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 2822–2837. doi: 10.1109/TPAMI.2020.2977624

Berman, D., Treibitz, T., and Avidan, S. (2017). "Diving into haze-lines: color restoration of underwater images," in *Proceedings of the British Machine Vision Conference* (BMVA Press).

Brock, A., Donahue, J., and Simonyan, K. (2019). "Large scale GAN training for high fidelity natural image synthesis," in *Proceedings of the 7th International Conference on Learning Representations, ICLR 2019* (New Orleans, LA).

Chen, R., Huang, W., Huang, B., Sun, F., and Fang, B. (2020). "Reusing discriminators for encoding: towards unsupervised image-to-image translation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Seattle, WA: IEEE), 8168–8177. doi: 10.1109/CVPR42600.2020.00819

Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., and Choo, J. (2018). "Stargan: unified generative adversarial networks for multi-domain image-to-image translation," in *IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE Computer Society), 8789–8797. doi: 10.1109/CVPR.2018.00916

Choi, Y., Uh, Y., Yoo, J., and Ha, J.-W. (2020). "Stargan v2: diverse image synthesis for multiple domains," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 8188–8197. doi: 10.1109/CVPR42600.2020.00821

Churnside, J. H., Marchbanks, R. D., Lembke, C., and Beckler, J. (2017). Optical backscattering measured by airborne lidar and underwater glider. *Remote Sens.* 9:379. doi: 10.3390/rs9040379

Dancu, A., Fourgeaud, M., Franjcic, Z., and Avetisyan, R. (2014). "Underwater reconstruction using depth sensors," in *Special Interest Group Graph. Interact. Techn* (Association for Computing Machinery), 1–4. doi: 10.1145/2669024.2669042

Deris, A., Trigonis, I., Aravanis, A., and Stathopoulou, E. (2017). Depth cameras on UAVs: a first approach. *Int. Arch. Photogr. Remote Sens. Spat. Inform. Sci.* 42:231. doi: 10.5194/isprs-archives-XLII-2-W3-231-2017

Drews, P. L., Nascimento, E. R., Botelho, S. S., and Campos, M. F. M. (2016). Underwater depth estimation and image restoration based on single images. *IEEE Comput. Graph. Appl.* 36, 24–35. doi: 10.1109/MCG.2016.26

Eigen, D., Puhrsch, C., and Fergus, R. (2014). "Depth map prediction from a single image using a multi-scale deep network," in *Advances in Neural Information Processing Systems* (Montreal, QC), 2366–2374.

Gomez Chavez, A., Ranieri, A., Chiarella, D., Zereik, E., Babić, A., and Birk, A. (2019). Caddy underwater stereo-vision dataset for human-robot interaction (HRI) in the context of diver activities. *J. Mar. Sci. Eng.* 7:16. doi: 10.3390/jmse7010016

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

Gupta, H., and Mitra, K. (2019). "Unsupervised single image underwater depth estimation," in *IEEE International Conference on Image Processing* (IEEE), 624–628. doi: 10.1109/ICIP.2019.8804200

He, K., Sun, J., and Tang, X. (2010). "Single image haze removal using dark channel prior", *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)* (Miami, FL: IEEE Computer Society), 1956–1963.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 770–778. doi: 10.1109/CVPR.2016.90

Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J. (2018). "Multimodal unsupervised image-to-image translation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 172–189. doi: 10.1007/978-3-030-01219-9_11

Islam, M. J., Xia, Y., and Sattar, J. (2020). Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* 5, 3227–3234. doi: 10.1109/LRA.2020.2974710

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI: IEEE Computer Society), 1125–1134. doi: 10.1109/CVPR.2017.632

Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., and Bengio, Y. (2017). "The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 11–19. doi: 10.1109/CVPRW.2017.156

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, 694–711. doi: 10.1007/978-3-319-46475-6_43

Kingma, D. P., and Ba, J. (2015). "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations*, eds. Y. Bengio and Y. LeCun (San Diego, CA).

Kupyn, O., Martyniuk, T., Wu, J., and Wang, Z. (2019). "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE International Conference on Computer Vision* (Seoul: IEEE), 8878–8887. doi: 10.1109/ICCV.2019.00897

Lee, H.-Y., Tseng, H.-Y., Huang, J.-B., Singh, M., and Yang, M.-H. (2018). "Diverse image-to-image translation via disentangled representations," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 35–51. doi: 10.1007/978-3-030-01246-5_3

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389. doi: 10.1109/TIP.2019.2955241

Li, J., Skinner, K. A., Eustice, R., and Johnson-Roberson, M. (2017). WaterGAN: unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot. Autom. Lett.* 3, 387–394. doi: 10.1109/LRA.2017.2730363

Li, N., Zheng, Z., Zhang, S., Yu, Z., Zheng, H., and Zheng, B. (2018). The synthesis of unpaired underwater images using a multistyle generative adversarial network. *IEEE Access* 6, 54241–54257. doi: 10.1109/ACCESS.2018.2870854

Liu, M.-Y., Huang, X., Mallya, A., Karras, T., Aila, T., Lehtinen, J., et al. (2019). "Few-shot unsupervised image-to-image translation," in *Proceedings of the IEEE International Conference on Computer Vision* (Seoul: IEEE), 10551–10560. doi: 10.1109/ICCV.2019.01065

Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., and Paul Smolley, S. (2017). "Least squares generative adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition* (Venice: IEEE Computer Society), 2813–2821. doi: 10.1109/ICCV.2017.304

Massot-Campos, M., and Oliver-Codina, G. (2015). Optical sensors and methods for underwater 3d reconstruction. *Sensors* 15, 31525–31557. doi: 10.3390/s151229864

Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*.

Odena, A., Olah, C., and Shlens, J. (2017). "Conditional image synthesis with auxiliary classifier GANs," in *International Conference on Machine Learning*, eds D. Precup and Y. W. Teh (Sydney, NSW: PMLR), 2642–2651.

Park, T., Liu, M.-Y., Wang, T.-C., and Zhu, J.-Y. (2019). "Semantic image synthesis with spatially-adaptive normalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Long Beach, CA: Computer Vision Foundation; IEEE), 2337–2346. doi: 10.1109/CVPR.2019.00244

Peng, Y.-T., Zhao, X., and Cosman, P. C. (2015). "Single underwater image enhancement using depth estimation based on blurriness," in *IEEE International Conference on Image Processing* (IEEE), 4952–4956. doi: 10.1109/ICIP.2015.7351749

Pérez, J., Bryson, M., Williams, S. B., and Sanz, P. J. (2020). Recovering depth from still images for underwater dehazing using deep learning. *Sensors* 20:4580. doi: 10.3390/s20164580

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

Silberman, N., Hoiem, D., Kohli, P., and Fergus, R. (2012). "Indoor segmentation and support inference from RGBD images," in *European Conference on Computer Vision* (Springer), 746–760. doi: 10.1007/978-3-642-33715-4_54

Tan, M., and Le, Q. (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning* (PMLR), 6105–6114.

Ueda, T., Yamada, K., and Tanaka, Y. (2019). "Underwater image synthesis from rgb-d images and its application to deep underwater image restoration," in *2019 IEEE International Conference on Image Processing (ICIP)* (IEEE), 2115–2119. doi: 10.1109/ICIP.2019.8803195

Wang, C., Xu, C., Wang, C., and Tao, D. (2018). Perceptual adversarial networks for image-to-image transformation. *IEEE Trans. Image Process.* 27, 4066–4079. doi: 10.1109/TIP.2018.2836316

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Liu, G., Tao, A., Kautz, J., et al. (2018a). "Video-to-video synthesis," in *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018*, eds S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Montreal, QC), 1152–1164.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2018b). "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proceedings of the IEEE Conference on Computer Vision and*

Pattern Recognition (Salt Lake City, UT: IEEE Computer Society), 8798–8807. doi: 10.1109/CVPR.2018.00917

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., and Torralba, A. (2010). "Sun database: Large-scale scene recognition from abbey to zoo," in *IEEE Conference on Computer Vision and Pattern Recognition* (San Francisco, CA: IEEE Computer Society), 3485–3492. doi: 10.1109/CVPR.2010.5539970

Ye, X., Li, Z., Sun, B., Wang, Z., Xu, R., Li, H., et al. (2019). Deep joint depth estimation and color correction from monocular underwater images based on unsupervised adaptation networks. *IEEE Trans. Circ. Syst. Video Technol.* 30, 3995–4008. doi: 10.1109/TCSVT.2019.2958950

Zhang, H., Goodfellow, I., Metaxas, D., and Odena, A. (2018). "Self-attention generative adversarial networks," in *Proceedings of the 36th International Conference on Machine Learning*, eds K. Chaudhuri and R. Salakhutdinov (Long Beach, CA: PMLR), 7354–7363.

Zhang, Y., and Yang, Q. (2017). A survey on multi-task learning. *arXiv preprint arXiv:1707.08114*.

Zheng, Z., Wu, Y., Han, X., and Shi, J. (2020). "ForkGAN: seeing into the rainy night," in *Computer Vision-ECCV 2020: 16th European Conference* (Glasgow: Springer), 155-170. doi: 10.1007/978-3-030-58580-8_10

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017a). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition* (Venice: IEEE Computer Society), 2223–2232. doi: 10.1109/ICCV.2017.244

Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., et al. (2017b). "Toward multimodal image-to-image translation," in *Advances in Neural Information Processing Systems*, eds I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett (Long Beach, CA), 465–476.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

Generator Architectures

In our experiments, the generator G_s from CycleGAN (Zhu et al., 2017a) and StarGAN (Choi et al., 2018) can be described as **Figure A1**. Here, Convolution denotes a 7 × 7 Convolution-InstanceNorm-ReLU layer with 64 filters and stride 1. Convolution/down denotes a 4 × 4 Convolution-InstanceNorm-ReLU layer and stride 2. Residual block denotes a residual block that contains two 3 × 3 Convolution-InstanceNorm-ReLU layers with the same number of filters on both layers. Deconvolution denotes a 4 × 4 fractional-strided-Convolution-InstanceNorm-ReLU layer and stride 2.

The generator G_d from Jégou et al. (2017) is based on dense-block (DB), as **Figure A2**. Convolution denotes a 3 × 3 Convolution-BatchNorm-ReLU layer with 32 filters and stride 1. Transition down is a maxpool2d operation with the same number of filters and a 1 × 1 Convolution-BatchNorm-ReLU layer with the same number of filters and stride 1. Transition up denotes a 4 × 4 deconvolution layer with the same number of filters and stride 2. Dense block denotes four 3 × 3 BatchNorm-ReLU-Convolution layers with 12 filters and stride 1. The output channel number of the dense block is the concatenation from the output of four layers and the input. The encoder and the decoder concatenate with skip connection.

Discriminator Architectures

For discriminator networks, we use 70 × 70 PatchGANs (Isola et al., 2017; Zhu et al., 2017a). Similarly, we do not use InstanceNorm or BatchNorm in any layer and use leaky ReLUs with a slope of 0.2. The discriminator D_s has two outputs from the discrimination branch and the classification branch. Differently, the discriminator D_d only has one discrimination output.

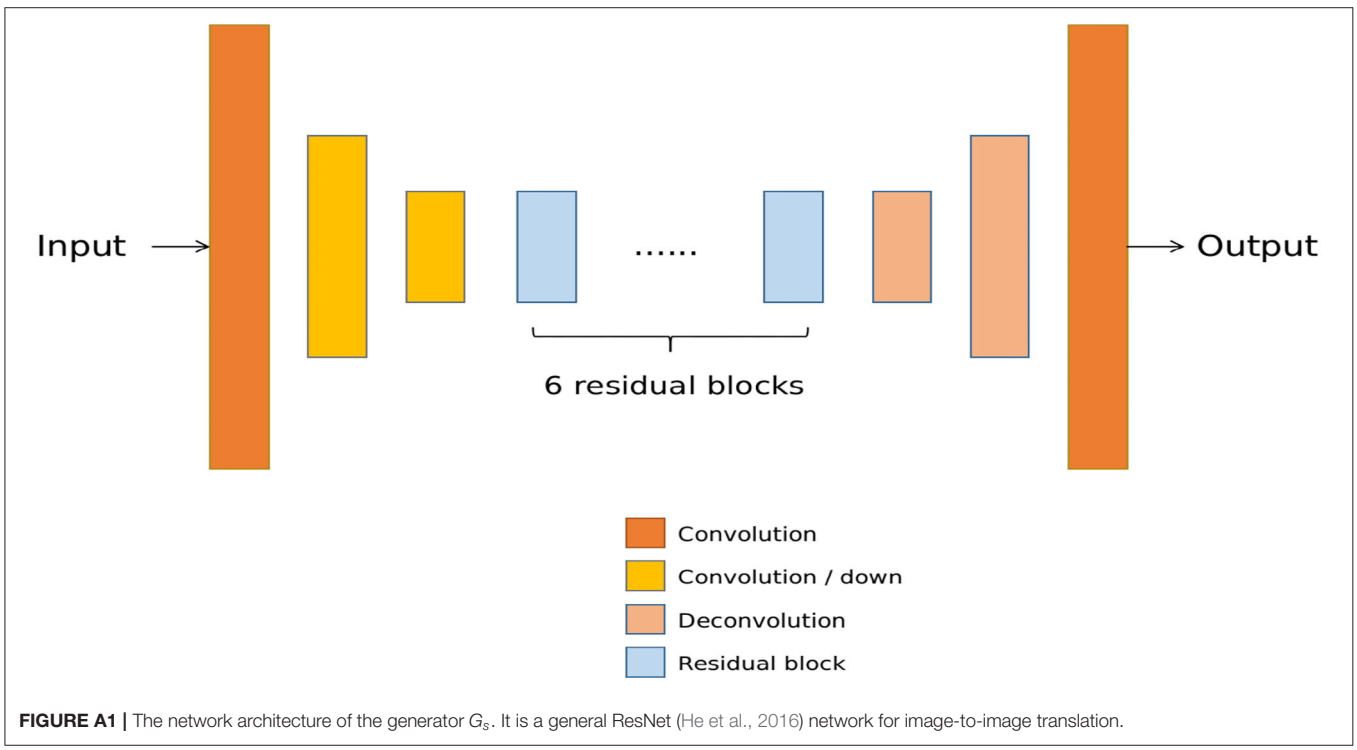


FIGURE A1 | The network architecture of the generator G_s. It is a general ResNet (He et al., 2016) network for image-to-image translation.

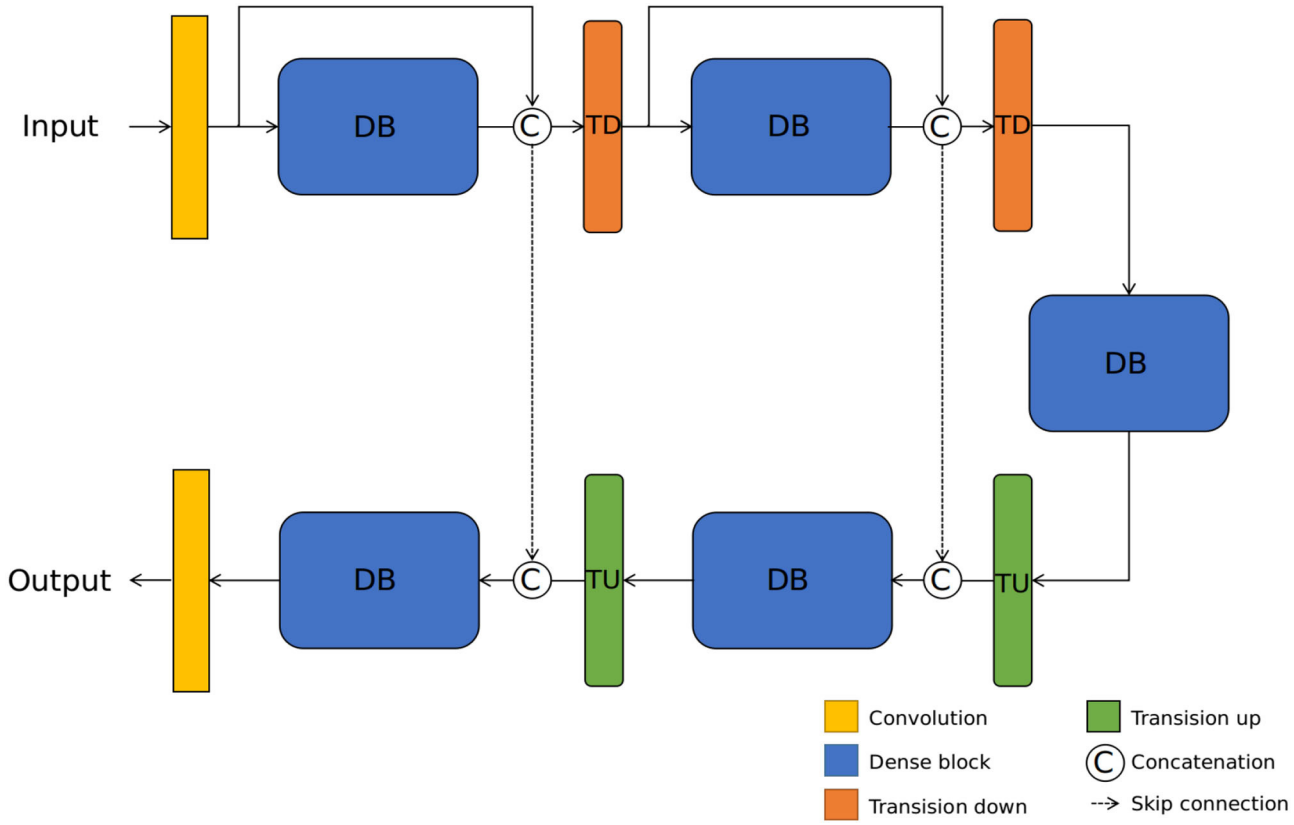


FIGURE A2 | The network architecture of the generator G_d. Following the work of UW-Net (Gupta and Mitra, 2019), we choose DenseNet (Jégou et al., 2017) as the generator G_d.

frontiers in Marine Science

METHODS
published: 24 September 2021
doi: 10.3389/fmars.2021.670683

Data-Driven Mapping With Prediction Neural Network for the Future Wide-Swath Satellite Altimetry

Jiankai Di[1], *Chunyong Ma*[1,2*] *and Ge Chen*[1,2]

[1] College of Information Science and Engineering, Ocean University of China, Qingdao, China, [2] Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

Two-dimensional mapping of sea surface height (SSH) for future wide-swath satellite altimetry (WSA) is a challenge at present. So far, considering the utilization of data-driven methods is a new researching direction for SSH mapping. In general, the data-driven mapping methods rely on the spatial-temporal relationship of the observations. These methods require training in large volumes, and the time cost is high, especially for the WSA observations. This paper proposed the prediction neural networks for mapping (Mapping-PNN) method to improve the training efficiency and maintain stable data and mapping capabilities. By 10-year wide-swath satellite along track observing system simulation experiments (OSSEs) on the HYCOM data, the experiment results indicate that the method introduced in this paper can improve the training efficiency and meet the grid mapping expectations. Compared with other methods, the root mean squared error (RMSE) of the mapping-PNN method can be limited within the range of ∼1.8 cm, and the new method can promote the observation of the ocean phenomena scale with $<$ ∼40 km, which reaches state of the art.

Keywords: two-dimensional mapping, wide-swath satellite altimetry, interpolation method, neural networks, data-driven

OPEN ACCESS

Edited by:
Jun Li,
University of Technology Sydney, Australia

Reviewed by:
Xiping Jia,
Guangdong Polytechnic Normal University, China
Ancheng Lin,
University of Technology Sydney, Australia
Maxime Beauchamp,
IMT Atlantique Bretagne-Pays de la Loire, France

***Correspondence:**
Chunyong Ma
chunyongma@ouc.edu.cn

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *22 February 2021*
Accepted: *25 August 2021*
Published: *24 September 2021*

Citation:
Di J, Ma C and Chen G (2021) Data-Driven Mapping With Prediction Neural Network for the Future Wide-Swath Satellite Altimetry. Front. Mar. Sci. 8:670683. doi: 10.3389/fmars.2021.670683

INTRODUCTION

The 2D SSH mapping is a big challenge for future WSA, which is a major topic of discussion nowadays. The wide-swath satellite missions, such as the surface water and ocean topography (SWOT) mission of the US-France (Gaultier et al., 2016) and the Guanlan satellite mission of China (Chen et al., 2019) will provide 2D altimetric information with a high resolution [15–30 km, depending on sea state (Morrow et al., 2019)]. At present, the optimal interpolation (OI) method (Le Traon et al., 2003) and the dynamic interpolation (DI) method (Ubelmann et al., 2015) are the main classical model-driven two-dimensional data mapping methods (Lguensat et al., 2017) for the altimetric satellite observations [such as the products of the AVISO (Archiving, Verification and Interpretation of data of Satellites Oceanography, CollecteLocalisation Satellites (CLS), AVISO, CNES, 2019)]. The OI method, a static, statistical data mapping approach based on the objective analysis method (Bretherton et al., 1976), with the combined observations of multiple satellites (Morrow and Traon, 2012; Amores et al., 2018; Ballarotta et al., 2019), the SSH grid data products could acquire mesoscale ocean phenomena larger than ∼150 km scales or longer than ∼10 days (Dussurget et al., 2011; Morrow et al., 2019) but cannot observe short-period and small-scale ocean dynamic phenomena (Morrow et al., 2019; Guillou et al., 2020). However, through the OI method with the WSA OSSEs, more details of ocean dynamic phenomena and sub-mesoscale ocean phenomena could be

observed, such as the scale ranges from ∼25 to ∼150 km (Ma et al., 2020). The DI method, which is based on the potential vorticity (PV) conservation theory (Hua and Haidvogel, 1986; Wunsch and Carl, 1996), can be adopted to observe instantaneous nonlinear ocean dynamic phenomena and express them through data reconstruction (Ubelmann et al., 2016), but it may fail to achieve acceptable results when trying to reconstruct the coastal regions or the tropic regions (Roge et al., 2017; Ballarotta et al., 2020). Those classical model-driven methods could be utilized to obtain SSH grid data products for the future WSA.

However, there are some detailed problems in the process of mapping by OI methods, such as eddies missing or the generation of "artifact" eddies (Ma et al., 2020). The DI method has limitations on mapping the sea areas where the PV conservation fails (Roge et al., 2017; Ballarotta et al., 2020). Then, consequently, the "data-driven" mapping methods (Lguensat et al., 2017, 2019b; Zhen et al., 2020) are proposed. Unlike the classical model-driven methods (Lguensat et al., 2017), the data-driven methods rely on the spatial-temporal relationship of the observations (Lguensat et al., 2019b).

Lguensat et al. (2017), who used machine learning (ML) methods in the Mediterranean, South China sea area (Lguensat et al., 2019b), and the Gulf of Mexico (Zhen et al., 2020) through principal component analysis (PCA), K nearest neighbor (KNN), and k-dimensional tree (KD-Tree) technologies, introduced a data-driven mapping method named AnDA (Lguensat et al., 2017). Furthermore, Lopez-Radcenco et al. (2019) extended the AnDA method to the mapping for multi-satellite-combined observations, SWOT observing system simulation experiments (OSSEs)-simulated data, as well as the combination of observations of nadir altimeter satellites and SWOT, and then obtained higher-accuracy SSH grid data products than the result of the OI method (Lopez-Radcenco et al., 2019). By utilizing deep learning (DL) for the DI theory verification in North Atlantic regions, Lguensat et al. (2019a) proved that the combination of the DI method and DL is feasible for data mapping. Compared with the result of the DI method on instantaneous nonlinear data, the accuracy and the error are similar for the data products obtained by "data-driven" methods, machine learning, and deep learning. According to the validation by DL, the DI method is reliable for the instantaneous nonlinear ocean dynamic signals inversion in the active ocean phenomenon regions (Lguensat et al., 2019a).

Shi et al. (2015) proposed the convolutional long short-term memory (ConvLSTM) deep learning method. This approach uses the convolution neural network (CNN) activation method, instead of the rectified linear unit (RELU) or Sigmoid activation functions, to improve the prediction performance in each gate of the classical LSTM network (Hochreiter and Schmidhuber, 1997). The advantage of CNN is that the feature extraction ability of LSTM can be enhanced. Inspired by the ConvLSTM neural network, Lotter et al. (2017) proposed the PredNet method, which grafts the ConvLSTM network to the C gate (the Gates of Controller), further improves the feature extraction ability of the predictive neural network, and makes it more accurate and reliable (Lotter et al., 2017). At the same time, the PredNet method is effective in predicting rapidly changing images and video transport streams (Lotter et al., 2017). Deep learning algorithms have been used for oceanographic applications, such as classification, identification, and prediction. Besides, Lima et al. (2017) used a CNN to identify ocean fronts, which yielded a higher recognition accuracy than the traditional algorithms. Yang et al. (2018) established a sea surface temperature (SST) prediction model based on LSTM networks that have been well tested using coastal SST data of China. By training 20-year AVISO grid data, Ma et al. (2019) used PredNet to conduct DL and implement daily ocean eddy forecast. It was proved that DL could be available on ocean observation prediction.

The purpose of this paper is to find a new 2D mapping method for future WSA observations and to map high-precision, low-error SSH grid data products for future altimetric satellites.

Based on the high-resolution model, this paper puts forward a new "data-driven" mapping method, prediction neural networks for mapping (Mapping-PNN) method by training the OSSEs along-track sampled data of WSA year by year. Additionally, the least recently used access (LRUA) module proposed by Santoro et al. (2016) is adopted, and it is a pure content-based memory write unit that writes memories to either the least used memory location or the most recently used one (Santoro et al., 2016). As for the Mapping-PNN test, it could obtain RMSE result similar as the PredNet method and better than the AnDA method, meeting the expectations. According to the results of experiments, by using the same sampled data volumes of the WSA in the region of Kuroshio and the Kuroshio Extension on OSSEs and testing three data mapping methods, the RMSE result of 2D mapping products can be limited within the range of ∼1.8 cm.

The experiment verification indicates that the Mapping-PNN method is applicable to the 2D mapping for WSA. Compared with the results of the ML method (Lguensat et al., 2017) and the DL method (Lotter et al., 2017), the Mapping-PNN method has the same data-mapping capabilities. It can obtain not only high-precision, high-resolution SSH grid data product with a high rate and low error but also promote the ability of the observation for the scales with $< \sim$40 km.

The remaining sections of this paper are arranged as follows: Material section describes the test materials and data. Methods section introduces the method of this paper. Experiment section illustrates the experiments and results. Conclusion and Discussion section is the conclusion, discussion, and the introduction of the future work.

MATERIALS

Model Data of HYCOM

The data set of hybrid coordinate ocean model (HYCOM) has a resolution of 1/12.5° × 1/12.5°. The Kuroshio and the Kuroshio Extension regions (the sea area researched in this paper at 15°E−39°E, 120°N−144°N region) belong to the middle latitude range. There are seasonal variations of the ocean phenomena in the Kuroshio and the Kuroshio extension regions. Specifically, the seasonal variation of the Kuroshio and the Kuroshio extension is affected by both dynamic factors (SST advection and vertical temperature transport) and thermal

factors (net heat flux at the air-sea interface) (Itoh, 2010; Nagano et al., 2013). The ocean phenomena of the Kuroshio and the Kuroshio extension increase significantly at the end of February in winter and strengthen continuingly in spring (Ji et al., 2018). The absolute geostrophic velocity extreme value of the Kuroshio extension appears slightly later than the Kuroshio (Nagano et al., 2013). The HYCOM data cover the global region, which is conducive to implementing the mapping simulation works in different target sea regions. The HYCOM data, as the representative model data in the application of OSSEs, are updated annually and made available to the public and facilitate scientific research and mutual verification of peer work. Meanwhile, HYCOM is a part of the Global Ocean Data Assimilation Experiment (GODAE) of the United States (Hybrid Coordinate Ocean Model (HYCOM) Data., 2021). The temporal resolution of HYCOM data for training is daily. In general, the HYCOM real-time high-resolution model includes three-dimensional ocean state description, the local coastal model, and the global coupled ocean-atmosphere prediction model with a prescribed ocean boundary.

The Parameters of Guanlan Mission

The 791-Orbit of Guanlan (Chen et al., 2019) was utilized in the experiment. The altitude of 791-Orbit is 791.254 km, which ensures a high observation swath of the WSA and provides the parameters for the OSSEs of the HYCOM. The parameter for generating the input sources of the OSSEs, including the exact repeat cycle, the sub-cycle, the orbit altitude, and the swath width, is required to participate in the calculation process of the along-track sampling simulations. The exact repeat cycle of 791-Orbit is approximately 14 days (4-day sub-cycle) over a swath width of 166.4 km (with a gap width of 27.6 km). A sub-cycle is an integer number of days; after which, the ground track of a satellite repeats itself within a small offset. In other words, a sub-cycle can be viewed as a near-repeat cycle with duration equaling to an integer number of nodal days (Pie and Schutz, 2008). And 4-day-long sub-cycle is chosen for OSSEs in the paper according to the 791-Orbit.

Data of the Simulation Experiment

According to the orbit of the Guanlan satellite and its parameters, the swath width, the nadir gap width, the coordinates of the swath trajectory, and the gap trajectory of the satellite had been calculated. Then, combined with the OSSEs, the sampled data in one cycle were merged to generate global observations, which would be used as the input source data for the subsequent mapping process. In the OSSEs, the data of each orbit on each cycle were calculated by the satellite track analysis algorithms by using reasonably matched satellite parameters.

As shown from **Figure 1A** represents the data sampling simulation in the OSSEs, using the parameter of the 791-Orbit, and the figure shows four ascending-descending tracks in one cycle. **Figure 1B** is an example of a one-cycle data sampling simulation according to the 14-day in one cycle of the 791-Orbit. And **Figure 1C** illustrates one-cycle Global OSSEs results in grids. **Figure 1** is an example of the sampled observations region in the West Pacific Ocean (WPO) to show the results of OSSEs [As shown in **Figures 1B,C**; see details in **Figures 4**, **5** of Experiments section]. In the OSSEs, when the grid points are at the same longitude–latitude, only the data of the last pass through are retained. And the data of the region blocked in blue are the researching sea area in this paper, at 15–39°E, 120–144°N, which is the training, evaluation, and testing volumes for comparison of mapping methods.

METHODS

The Analog Data Assimilation (AnDA)

The observation data volumes can be organized and calculated as follows, according to the AnDA method proposed by Lima et al. (2017).

The following discrete state space (Lguensat et al., 2017):

$$x(t) = \mathcal{M}[x(t-1), \eta(t)], \quad (1)$$

$$y(t) = \mathcal{H}[x(t)] + \mathcal{E}(t), \quad (2)$$

where time $t \in \{0, \ldots, T\}$ refers to the times in which observations are available, assuming the observations are at regular time steps. In Eq. (1), $\mathcal{M}$ characterizes the dynamical model of the true state x(t), while $\eta(t)$ is a random perturbation added to represent model uncertainty. The observation Eq. (2) describes the relationship between the observation y(t) and x(t). And the observation error is considered by the white noise $\varepsilon(t)$. Considering an additive Gaussian noise ε with covariance T in Eq. (2) and the observation operator, the $\mathcal{H} = H$ is quasi-linear (Lguensat et al., 2017).

The Eq. (1) represents the dynamical model governing the evolution of state x through time, while H is a Gaussian-centered noise of covariance Q that models the process error. And Eq. (2) explains the relationship between the observation y(t) and the state to be estimated x(t) through the operator H. The uncertainty of the observation model is represented by the e error, which is considered here to be Gaussian centered and of covariance R. The ε and H are independent, and the Q and R are known.

AnDA relies on the following state-space model, to evaluating filtering, respectively smoothing, posterior likelihood, and the distribution of the state could be estimated x(t) at time t, given past and current observations $y(1, \ldots, t)$, respectively given all the available observation $y(1, \ldots, T)$ (Lguensat et al., 2019b).

$$x(t) = \mathcal{F}(x(t-1)) + \eta(t) \quad (3)$$

$$y(t) = \mathcal{H}(x(t)) + \mathcal{E}(t) \quad (4)$$

The counterpart of a model-driven operator $\mathcal{M}$ of Eq. (1) is the operator $\mathcal{F}$ in Eq. (3), which refers to the analog-forecasting operator. The predicting matrix could be calculated by Kalman smoother to obtain the final result (Lguensat et al., 2019b). The Kalman smoother (KS) algorithm can directly provide the optimal estimation of the state, given the observations and their corresponding errors. The AnDA method relies on the spatial-temporal relationship of the observations and introduces the analog operator through a KNN-search to make the state-space model being applied in practice.

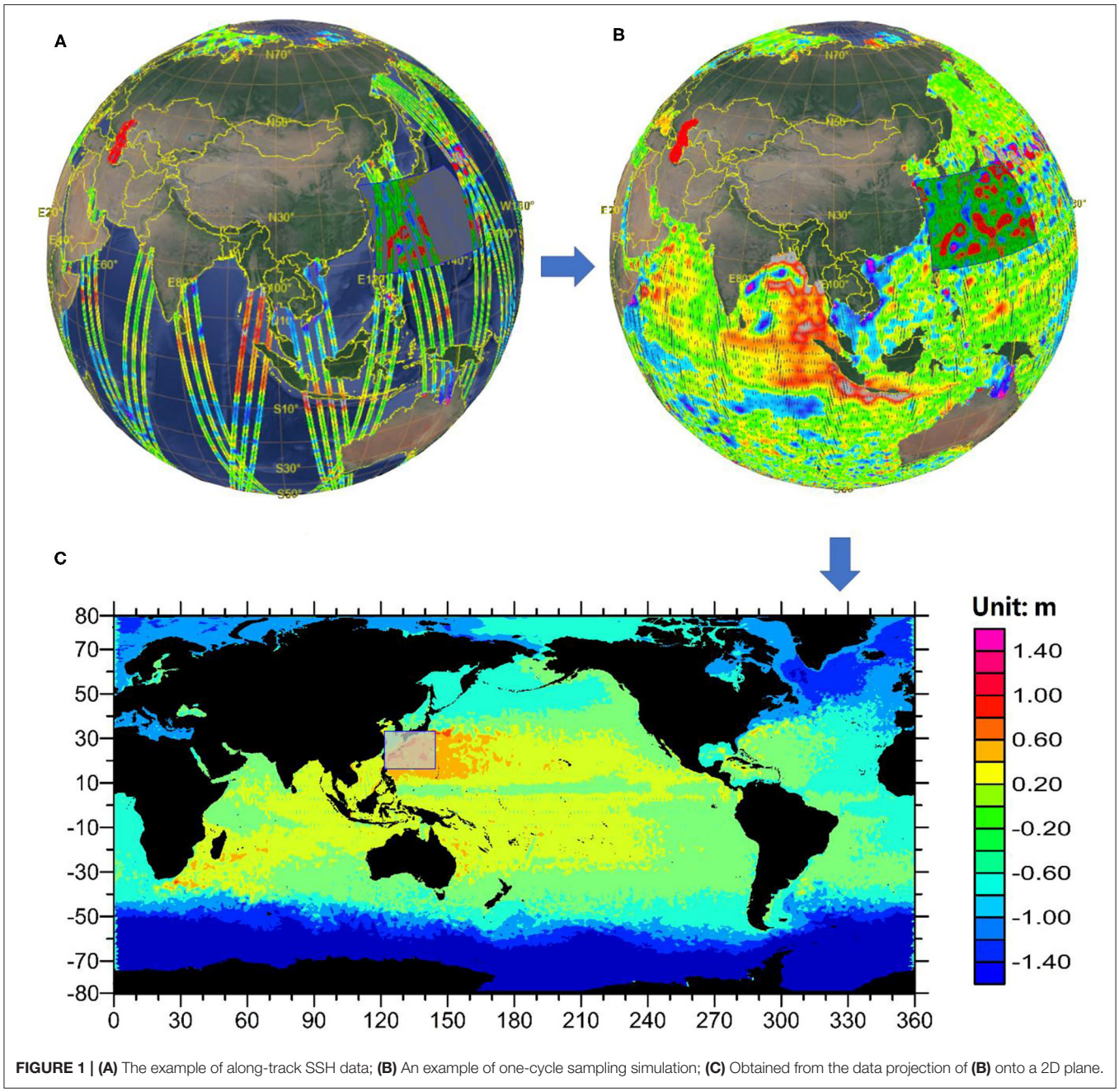


FIGURE 1 | (A) The example of along-track SSH data; **(B)** An example of one-cycle sampling simulation; **(C)** Obtained from the data projection of **(B)** onto a 2D plane.

The proceedings of the AnDA method are as follows: firstly, principal component analysis (PCA) algorithm is performed on satellite along-track sampled data obtained through OSSEs and extract data features to prepare for the later ML. Then, secondly, a KNN search in a catalog of numerical model outputs using a KD-Tree is implemented, and the final mapping result is obtained through Kalman smoothing. Readers can find the AnDA's algorithm sketch block diagram in the reference (Lguensat et al., 2017, 2019b) to learn more details about the AnDA method. This paper compares and discusses the 2D mapping capabilities of the AnDA and the Mapping-PNN methods in the experimental part (The code of python type for AnDA method can be found on the website: https://github.com/ptandeo/AnDA).

The PredNet Method

Shi et al. (2015) proposed the ConvLSTM method. In 2017, Lotter and Kreiman et al. established the PredNet neural network architecture based on ConvLSTM (Lotter et al., 2017). The theory details of ConvLSTM and PredNet are as follows:

ConvLSTM replaced each gate of the LSTM neural network (proposed by Hochreiter and Schmidhuber, 1997) with CNN architecture, which improved the ability of feature

extraction for targets in the original LSTM network. The theory brief introduction and the implementation method are as follows:

ConvLSTM is an extension of full-connection LSTM (FC-LSTM), which has convolutional structures in the input-to-state and state-to-state transitions. The ConvLSTM determines the future state of a certain cell in the grid by the inputs and past states of its local neighbors. The key equations of ConvLSTM (Shi et al., 2015) are shown in Eqs. (5)–(9) below, where '$*$' denotes the convolution operator and '$\circ$' denotes the Hadamard product:

$$i_t = \sigma(W_{xi}*\mathcal{X}_t+W_{hi}*\mathcal{H}_{t-1}+W_{ci}\circ C_{t-1}+b_i) \quad (5)$$

$$f_t = \sigma(W_{xf}*\mathcal{X}_t+W_{hf}*\mathcal{H}_{t-1}+W_{cf}\circ C_{t-1}+b_f) \quad (6)$$

$$C_t = f_t\circ C_{t-1}+i_t\circ tanh\left(W_{xc}*\mathcal{X}_t+W_{hc}*\mathcal{H}_{t-1}+b_c\right) \quad (7)$$

$$o_t = \sigma(W_{xo}*\mathcal{X}_t+W_{ho}*\mathcal{H}_{t-1}+W_{co}\circ C_t+b_o) \quad (8)$$

$$\mathcal{H}_t = o_t\circ tanh(C_t) \quad (9)$$

For a spatiotemporal sequence forecasting problem, the structure consists of an encoding network and a forecasting network (Shi et al., 2015). Compared to classical LSTM, ConvLSTM can model space-time structures by encoding geographic information as tensors, thereby to overcome the limitation of losing spatial information in classic LSTM networks. Readers can find the algorithm sketch block diagram of ConvLSTM in the reference (Shi et al., 2015) to learn more details about the ConvLSTM method (The code of python type for ConvLSTM method can be found on the website: https://github.com/XingguangZhang/ConvLSTM.).

To minimize the weighted sum of the activity of the error units (Ma et al., 2019), based on the ConvLSTM concept, the network structure is enhanced to construct an improved ConvLSTM, which is named the PredNet model, for predicting sequences of images (Lotter et al., 2017). A structure in which the error is fed forward has been added to the network, as shown in **Figure 2**. The network consists of a series of repeatedly stacked blocks, and each of them can be viewed as one layer (Lotter et al., 2017).

The PredNet architecture is illustrated in **Figure 2**. Each module of the network consists of four basic parts: an input convolutional layer (Al), a recurrent representation layer (Rl), a prediction layer (Âl), and an error representation (El). The architecture is rooted in convolutional and recurrent neural networks (RNN) trained with back propagation (BP) (Ma et al., 2019).

$$A_l^t = \begin{cases} x_t & if\ l = 0 \\ MAXPOOL(RELU(CONV(E_{l-1}^t))) & l > 0 \end{cases} \quad (10)$$

$$\hat{A}_l^t = RELU(CONV\left(R_l^t\right)) \quad (11)$$

$$E_l^t = \left[RELU\left(A_l^t-\hat{A}_l^t\right); RELU\left(\hat{A}_l^t-A_l^t\right)\right] \quad (12)$$

$$R_l^t = CONVLSTM(E_l^{t-1},\ R_l^{t-1},\ UPSAMPLE(R_{l+1}^t)) \quad (13)$$

The full set of update rules is listed in Eqs. (10–13).

As follows, **Algorithm 1** is the original PredNet algorithm.

Algorithm 1: Calculation of PredNet States

Require: x_t
$A_0^t \leftarrow x_t$
$E_l^0,\ R_l^0 \leftarrow 0$
for $t \leftarrow 1$ **to** T **do**
  for $l = L$ **to** 0 **do** ► Update R_l^t states
    if $l = L$ **then**:
      $R_L^t = ConvLSTM(E_L^{t-1}, R_L^{t-1})$
    else:
      $R_l^t = ConvLSTM(E_l^{t-1}, R_l^{t-1},\ Up\_Sample(R_{l+1}^t))$
  for $l = 0$ **to** L **do** ► Update $\hat{A}_l^t$,A_l^t, E_l^t states
    if $l = 0$ **then**:
      $\hat{A}_0^t = SatLU(\ RELU\left(\ Conv\left(R_0^t\right)\right))$
    else:
      $\hat{A}_l^t = RELU\left(\ Conv\left(R_l^t\right)\right)$
    $E_l^t = \left[RELU\left(A_l^t - \hat{A}_l^t\right); RELU\left(\hat{A}_l^t - A_t^l\right)\right]$
    if $l < L$ **then**:
      $\hat{A}_{l+1}^t = MaxPool(\ Conv\left(E_t^l\right))$
End

Algorithm 1 was proposed by Lotter et al. (2017). The R_l^t states are computed, and a forward pass is initialized to calculate the predictions, errors, and higher-level targets. The initial prediction is spatially uniform (Lotter et al., 2017).

Readers can learn more details about the PredNet method in the reference (Lotter et al., 2017). This paper also compares the 2D mapping capabilities of the PredNet and the Mapping-PNN methods in the experimental part, and discusses them (The code of python type for the PredNet method can be found on the website: https://coxlab.github.io/prednet).

The Mapping-PNN Method

In the prediction of the target data, the PredNet needs a lot of learning database for the existing data set, which will increase the learning period. To improve the training efficiency of the original version of PredNet, the Mapping-PNN is proposed in this paper, and the external storage gate is added in ConvLSTM to enhance the memory and storage capacity of the learning performance of the ConvLSTM layer, accelerate the learning efficiency of the original PredNet, and save the waiting period of the observations on OSSEs. The external storage gate is utilized in the minimum training period. And the using of the least recently used access (LRUA) module helps to accelerate the read-write process of the memories. The AdamOptimizer (Kingma and Ba, 2015) is adopted to minimize the loss. The specific method is as follows.

The typical external storage gate includes duality of read and write units, as well as the external memory. The controller, neuron C_{convlstm}, is a ConvLSTM network, which receives the current input and controls the read units and write units to interact with the external memories, respectively. Memory encoding and retrieval in external memories are rapid, with

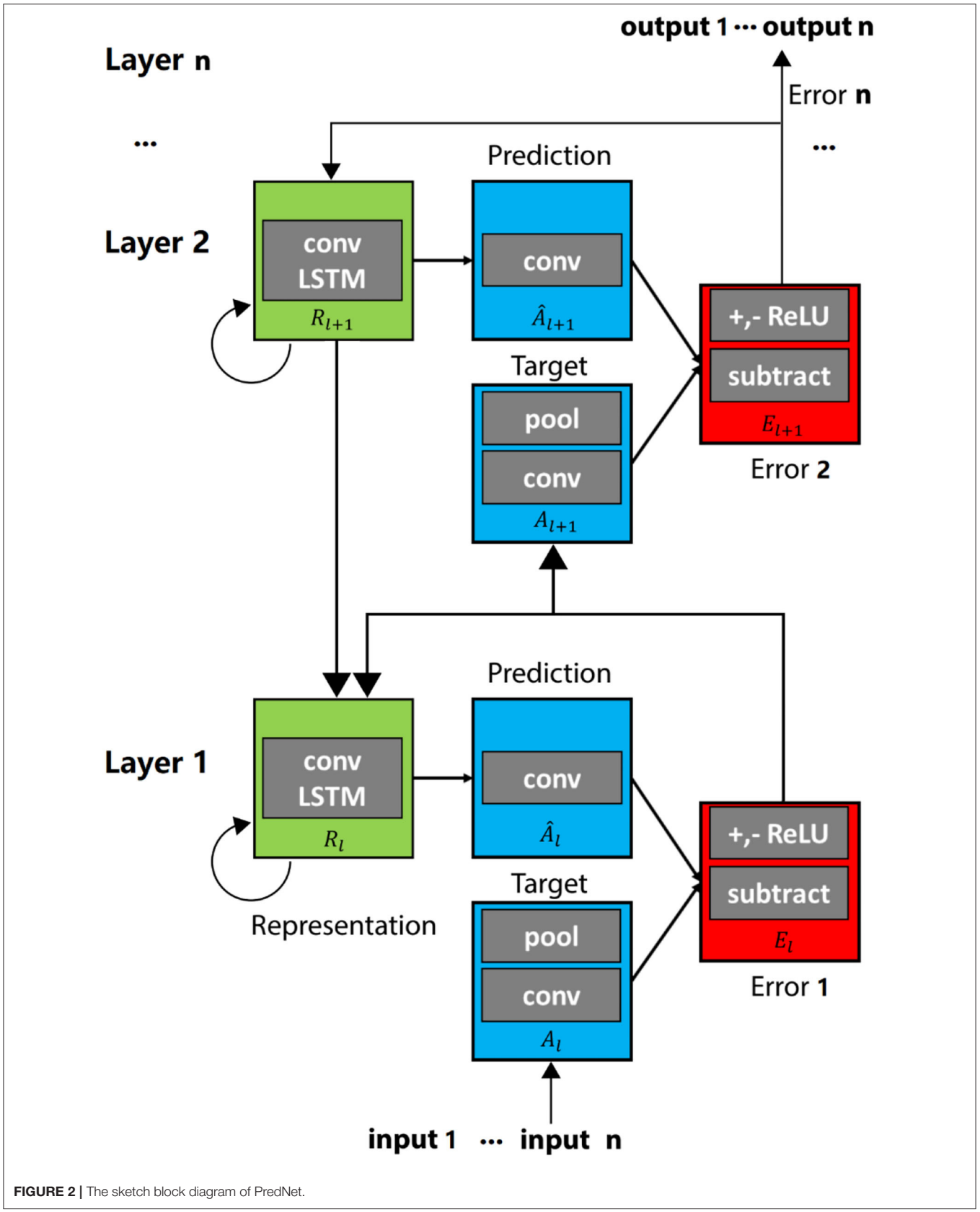


FIGURE 2 | The sketch block diagram of PredNet.

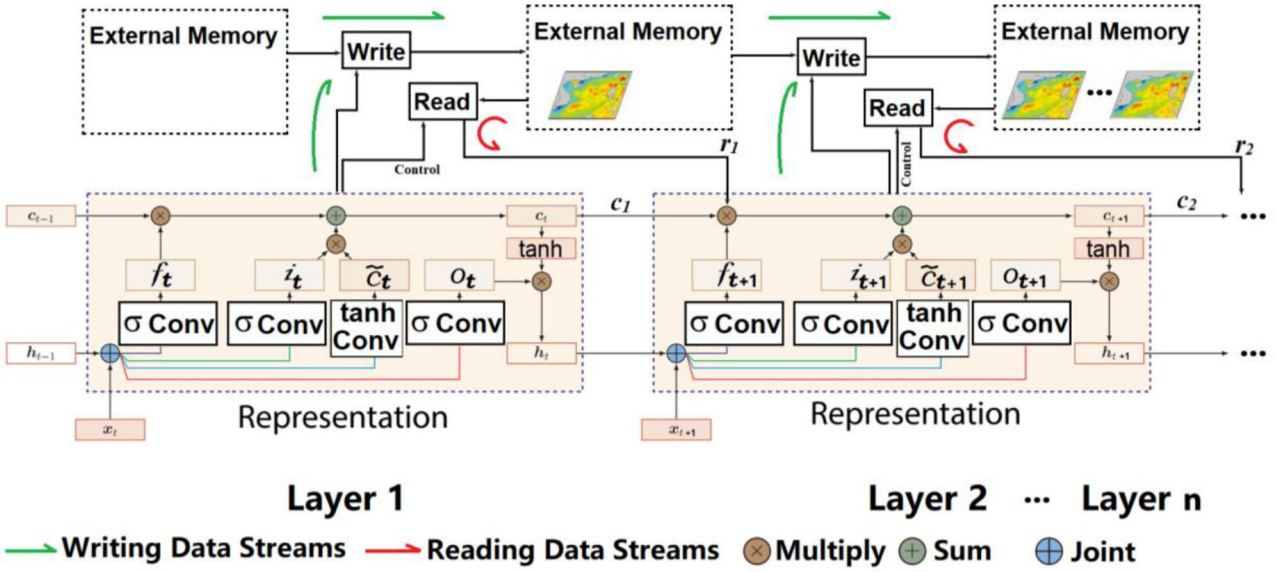


FIGURE 3 | The sketch block diagram of Mapping-PNN.

feature representations being placed into or taken out of memory potentially every time step. Additionally, it can be used for long-term storage by slowly updating the weights and for short-term storage by external memories. Thus, when the model learns the type of representations, it will be placed into memories, and the representations will be used to implement mapping.

As shown in **Figure 3**, the initialized state of the Mapping-PNN network is represented by *init_state*. The cell state of the initialized controller, neuron C_{convlstm}, is represented by c_{k} (k = 1, 2, …, n, nϵR, and n equals to the number of the memory). Given the input SSH observations, the controller receives the memory $r_{\text{t}-1}$ and cell state $c_{\text{t}-1}$ provided by the previous state *prev_state*, and then produces k_{convlstm} used to retrieve a particular memory. Besides, the light-green arrow line represents the writing data streams, while the red arrow line illustrates the reading data streams.

In terms of a new sequence, it is written to a rarely used location with the recently encoded information preserved or to the last used location, which can be used for updating with newer or possibly more relevant information. Then, the whole procedure of the algorithm can be described as follows [including each component of the controller gates $\mathbf{C_t}$, which is transformed from Eq. (7)]:

$$w_t^u \leftarrow \gamma^* w_{t-1}^u + w_t^r + w_t^w \tag{14}$$

$$w_t^{lu}(i,j) = \begin{cases} 1 & if\ w_t^u \le m(w_t^u,\ n) \\ 0 & if\ w_t^u > m(w_t^u,\ n) \end{cases} \tag{15}$$

$$w_t^w \leftarrow RELU(Conv(\sigma(\alpha) \bullet w_{t-1}^r + (1 - \sigma(\alpha)) \bullet w_{t-1}^{lu}))) \tag{16}$$

$$M_t(i,j) \leftarrow M_{t-1}(i,j) + w_t^w(i,j) \bullet k_t \tag{17}$$

where w^{u} denotes the usage weight updated at each time step to keep track of locations most recently read or written; γ denotes the decay parameter; w^{lu} denotes the least-used weight computed using w^{u} for a given time step; the notation $m(v, n)$ is introduced to denote the n^{th} smallest element of the vector v; n is set to equal the number of the writer to memory; w^{w} refers to the written weight computed by the function $RELU(\sigma(.))$, which combines the previous read weights w^{r} and previous least-used weights w^{lu}; α represents a dynamic scalar gate parameter to interpolate between weights. Before writing to memory, the least used memory location is computed from w^{u} and set to zero, and then the memory M_{t} is written by the computed matrix of written weights w^{w}. The parameters will be updated dynamically during back propagation. In addition, $M_{\text{t}}(i, j)$ can be written into the zeroed memory locations or the previously used memory locations; if it is the previously used memory location, the w will simply be erased.

The memory r^{t} is used by the controller as both an input to a classifier and an additional input for the next input sequence. It is calculated by the Eq. (18) for prediction.

$$r_t \leftarrow \sum_i \sum_j w_t^r(i,j) \bullet M_t(i,j) \tag{18}$$

To achieve the learning, the LRUA module proposed by Santoro et al. (2016) is adopted, which is a pure content-based memory write unit that writes memories to either the least used memory location or the most recently used one (Santoro et al., 2016).

Furthermore, w^{r} is ConvLSTM with RELU, following the Eq. (19):

$$w_t^r(i,j) \leftarrow RELU(Conv(K(k\_t, M\_t(i,j)))) \tag{19}$$

where M_{t} refers to the memory matrix at time-step t, and $M_{\text{t}}(i, j)$ refers to a sub-block in this matrix. The block of $M_{\text{t}}(i, j)$ serves

as the memory "slots," with constituting individual memories (Santoro et al., 2016).

$$K(k\_t, M\_t\,(i,j)) \leftarrow LRUA(k\_t, M\_t\,(i,j)) \tag{20}$$

where the read units can amplify or attenuate the precision of the focus by the read weights.

Those read weights w^{r} and corresponding memory M_{t} (i, j) are used to retrieve the memory r_{t}.

The ConvLSTM Layer in **Algorithm 1** (Lotter et al., 2017) was optimized by the introduction of the LRUA method, and **Algorithm 2** was proposed, as shown:

Algorithm 2, the improvement details of the Mapping-PNN, adopted the LRUA method.

Algorithm 2: Mapping-PNN

Input: Given N SSH samples $\{X_1, X_2, \ldots, X_N\}$ belonging to C with Sampled SSH;
$y_t \in Y = \{1, \ldots, C\}$, *for* $t = 1, \ldots, N$;
Output: A Converlutional layer for ssh prediction;

prev_state $\leftarrow$ *init_state* (N)
  $c_0 \leftarrow C_{lstm}\,(N)$
  $r_0 \leftarrow 0_{N\times(unit\_num*memory\_size)}$
  $w_0^r \leftarrow$ *observation_weight* $\left(N, unit\_num, memory\_slots\right)$
  $w_0^u \leftarrow$ *observation_weight* $\left(N, memory\_slots\right)$
  $M_0 \leftarrow \varepsilon_{N\times memory\_slots\times memory\_size}$
  return $\{c_0, r_0, w_0^r, w_0^u, M_0\}$
*_output_* $= [:][:]$
for $t \leftarrow 1$ **to** N **do**:
  $h_t, c_t \leftarrow C_{Convlstm}\left(\left(X_t, y_t\right), prev\_observation\right)$
  for $i \leftarrow 0$ **to** $X_t.length$ **do**:
    for $j \leftarrow 0$ **to** $(X_t.size)/(X_t.length)$ **do**:
      $output, curr\_state \leftarrow$ *MappingPNN* $\left(i,\ j,\ prev\_observation\right)$
      Memory Retrieval:▶ LRUA Method
      $K\left(k_t, M_t\left(i,j\right)\right) \leftarrow$ *LRUA* $\left(k_t, M_t\left(i,j\right)\right)$(Santoro et al., 2016)
      $w_t^r\left(i,j\right) \leftarrow$ *RELU* $\left(\ Conv\left(K\left(k_t, M_t\left(i,j\right)\right)\right)\right)$
      $r_t += w_t^r\left(i,j\right) \bullet M_t\left(i,j\right)$
      Memory Encoding (LRUA):
      $w_t^u \leftarrow \gamma * w_{t-1}^u + w_t^r + w_t^w$
      if $w_t^u \le m\left(w_t^u, n\right)$ **then** $w_t^{lu}\left(i,j\right) = 1$ **else** $w_t^{lu}\left(i,j\right) = 0$
      $w_t^w \leftarrow$ *RELU*$(Conv(\sigma\,(\alpha) \bullet w_{t-1}^r + (1 - \sigma\,(\alpha)) \bullet w_{t-1}^{lu}))$
      $M_t\left(i,j\right) \leftarrow M_{t-1}\left(i,j\right) + w_t^w\left(i,j\right) \bullet k_t$
    return $\{h_t, r_t\}, \{c_t, r_t, w_t^r, w_t^u, M_t\}$
  prev_state $=$ *curr_state*;
  if $i == 0$ **then**:
    $o2o\_w \leftarrow \left(output.length, M_s\right)$, *rand_unif_init* $(minv, maxv)$
    $o2o\_b \leftarrow \left(M_s\right)$, *rand_unif_init* $(minv, maxv)$
  end if:
  $output = RELU\left(output \bullet o2o\_w + o2o\_b\right)$
train_op $=$ *AdamOptimizer.minimize*$(-CrossEntropyCost\left(y_t,\ \_output\_\right))$
End

In the algorithm, the *observation_weight* $(N, unit\_num, memory\_slots)$ function generates a tensor with the zeros set to the *ones*; $\{(N, unit\_num)$, *rand_unif_init* $(minv,\ maxv)\}$ generates a tensor with a uniform distribution, and the value of all elements is set between *minv* and *maxv*.

For the current time-step t, the sampled data $\{X_1,\ X_2,\ \ldots,\ X_N\}$ and the corresponding sample-class y_{t} will be received by the controller C_{Convlstm}. The current state of the network *curr_state* is used by the controller as an additional input for the next time step. According to each sequence of the sample, the algorithm randomly generates the prediction label. If the sampled data X_{t} comes from a new observation, it will be bound to the appropriate y_{t} and stored by the write units in the external memory, which is presented in the subsequent time step. Once a sample from an observed-already data is presented, the controller will retrieve the bound information by the read units from the external memory for SSH prediction. The *cross_entropy_cost*$(\cdot)$ is to measure the loss between the predicted value and the correct prediction label. Then, the adaptive moment estimation, *Adam,optimizer*$(\cdot)$ (Kingma and Ba, 2015), is adopted to minimize the loss. Furthermore, the back-propagated (BP) error signal from the current prediction updates those previous weights and bias, such as the ***o2o_w,o2o_ b***,$\mathbf{w_t^r}$,$\mathbf{w_t^w}$ and $\mathbf{w_t^u}$, followed by the updating of the external memory. Those processes would be repeated until the model converges. Meanwhile, there are some observation gaps or missing data after generation of grid SSH; therefore, an optimal interpolation is needed. The parameters, such as the coefficient matrix, the spatial distances, and searching range, will be designed based on the target region (the coefficient matrix of the observations and the errors, the range of the latitude-longitude, etc.) (Lguensat et al., 2017; Amores et al., 2018; Ma et al., 2020).

EXPERIMENTS

Experimental Setups

The Regions and the Target Date

The AnDA, PredNet, and Mapping-PNN methods all need a large number of training sets during the neural network training. By utilizing 10-year HYCOM data, three data-driven methods were trained, and a target day (at the 15°E ∼ 39°E, 120°N ∼ 144°N region—take a date on March 1, 2017 of HYCOM for example) was considered as the target field of the experimental dataset of the three data-driven methods. The temporal resolution of test datasets is daily. The cycle of the Guanlan 791-Orbit is 14 days, and the target day is the March 1, 2017 on HYCOM.

Furthermore, take the selection of the test data, for example, for AnDA, PredNet, and Mapping-PNN methods. The test data are obtained from January 1, 2017 to March 1, 2017 on HYCOM, which could be divided into "30-day data" (from January 1, 2017 to January 30, 2017) and "60-day data" (from January 1, 2017 to March 1, 2017).

The studied region in this paper is the Kuroshio region (Kuroshio extension as well). With the variation of the seasons in the northern hemisphere, the ocean phenomenon varies significantly in this region. Considering that the early stage of March is in the transition period from winter to spring, "March 1, 2017" is selected as the research target date. At this time, several new unique ocean phenomena of spring are in the generation stage, and unique ocean phenomena of winter are in the transformation stage. The illustration of the comparative analysis and research among the data mapping algorithms in the Kuroshio region and the Kuroshio extension region would be valuable for the ocean science research.

TABLE 1 | Computer hardware resource in this paper.

Hardware resource	Computer configurations
CPU	Intel(R), Core (TM), i7-8700K@3.70 GHz, 16 Cores
Memory	64.0 GB
Operating System	64-bit Windows 10
Hard Disk	SSD 860 EVO250GB(Samsung); Seagate Hub BK SCSI 8TB Disk; and WD20EZRZ-00Z5HB0(WDC) 4TB Disk

The AnDA, PredNet, and Mapping-PNN methods used 2005–2015 OSSEs observations for model training, 2015 ∼2016 original data for evaluation, and 2017 original data for tests. The research on ocean phenomena represented by the Kuroshio and the Kuroshio extension is based on detailed statistics, analysis, and summary from over decades' observations. With consistency, both use the observation data as the research basis.

The evaluation and test methods of AnDA refer to the method introduced by Lguensat et al. (2017). In addition, the methods introduced by Beauchamp et al. (2020) are used as the evaluation methods of PredNet and Mapping-PNN. The validation and evaluation methods are listed in detail to evaluate the mapping performance of comprehensive indicators. For experiment results, please see The Comparison of the Experiment Results of the AnDA, the PredNet, and the 426 Mapping-PNN section.

The Hyper Parameters' Configuration

Some hyper parameters, such as optimizer and learning rate, are as follows: For the AnDA method, the error probability is 0.01. For the PredNet method and the Mapping-PNN method, to minimize the loss, an *Adam,optimizer* (·) is utilized, and the learning rate is 0.01 until the model converges after training 80 times in test No.1 (See **Table 3** in The Comparison of the Experiment Results of the AnDA, the PredNet, and the 426 Mapping-PNN section for details). To avoid the overfitting, the early stopping strategy is utilized during the training. And if the loss of the evaluation data is no longer reduced, the training will be stopped, and the overfitting can be avoided. Meanwhile, 80 training times are the epochs when the early stopping happens.

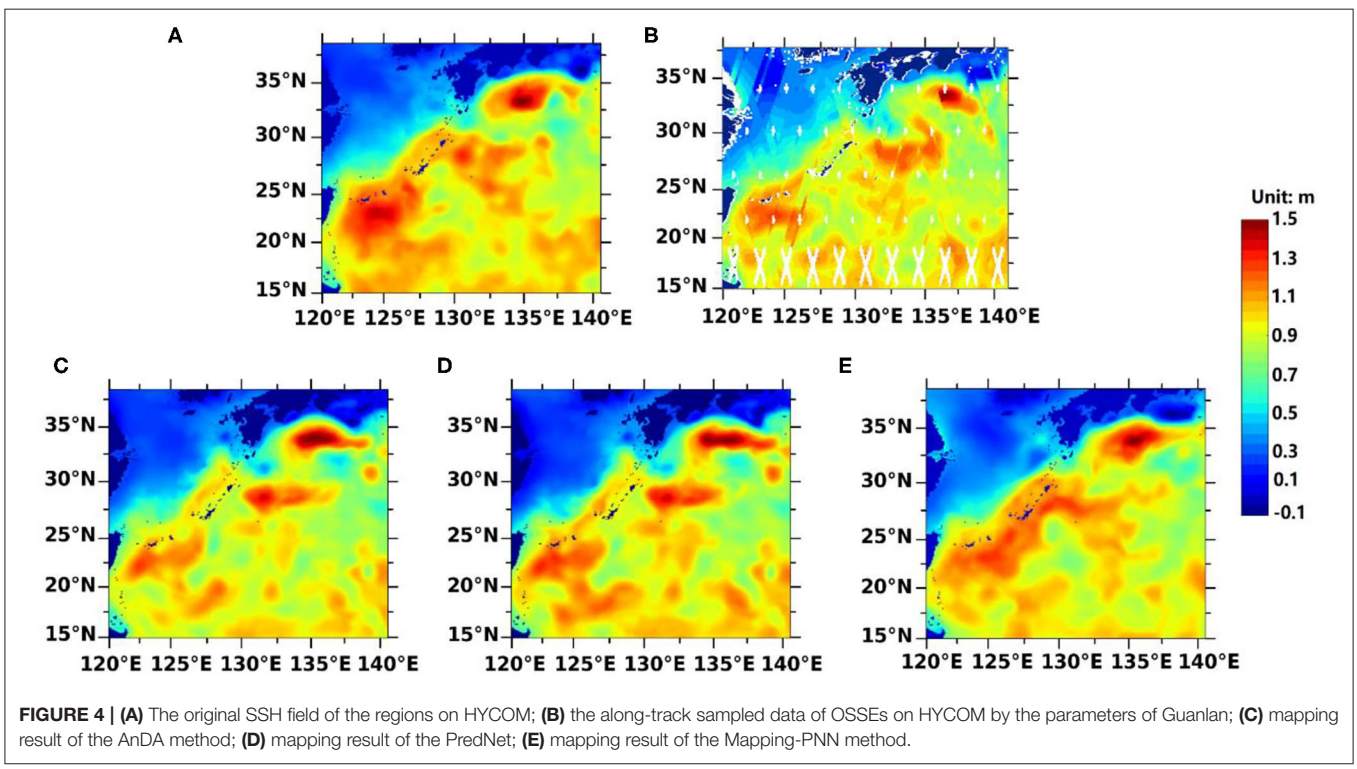


FIGURE 4 | (A) The original SSH field of the regions on HYCOM; **(B)** the along-track sampled data of OSSEs on HYCOM by the parameters of Guanlan; **(C)** mapping result of the AnDA method; **(D)** mapping result of the PredNet; **(E)** mapping result of the Mapping-PNN method.

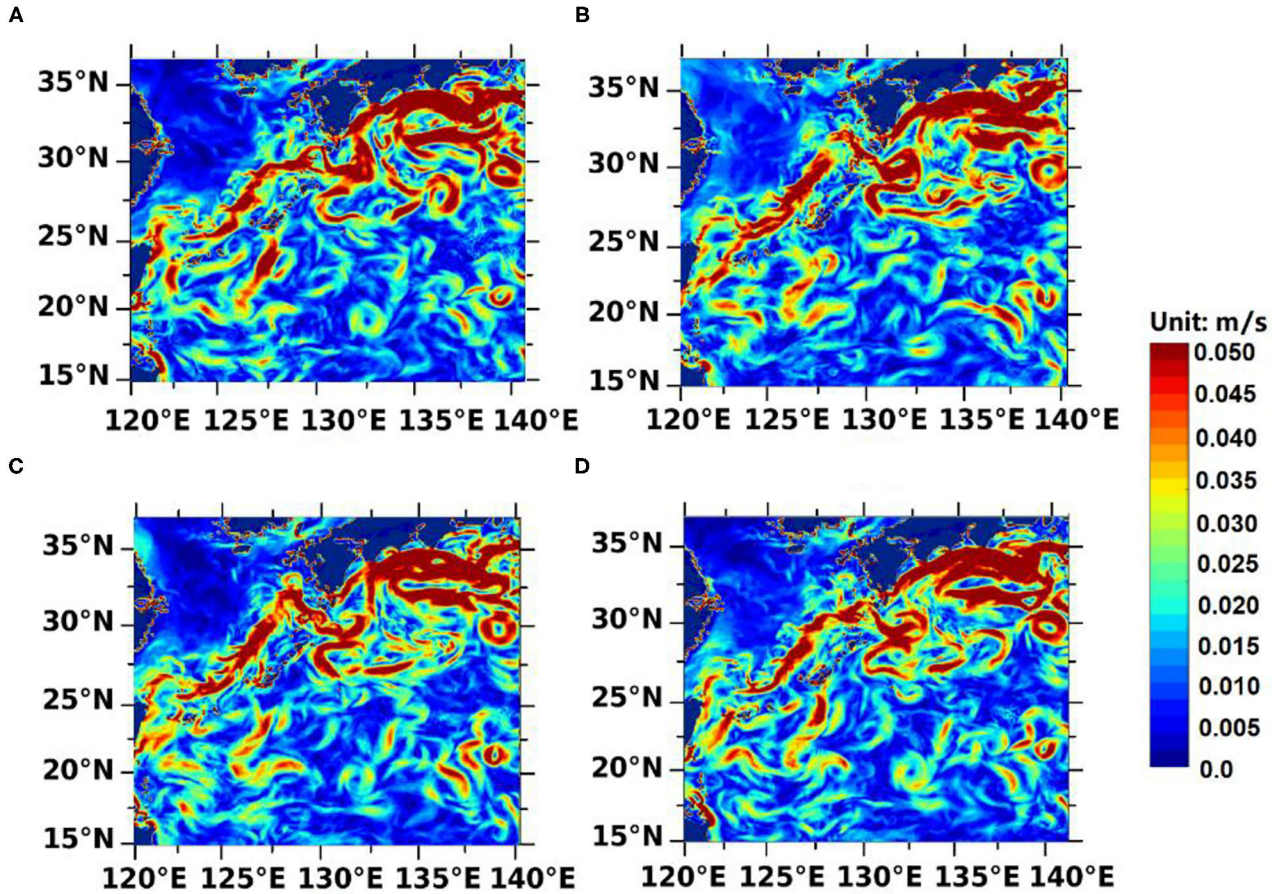


FIGURE 5 | (A) The absolute geostrophic velocity derived from **Figure 4A**, the original SSH; **(B)** the absolute geostrophic velocity from **Figure 4C** (the AnDA method); **(C)** the absolute geostrophic velocity from **Figure 4D** (the PredNet method); **(D)** the absolute geostrophic velocity from **Figure 4E** (the Mapping-PNN method).

Computer Configuration Used in This Paper (Table 1)

Validation Methods

In the analysis of the experimental results, several validation methods have been used to validate the test results of three data-driven methods, such as the method of illustrating the mapping results by the figures, the absolute geostrophic velocities, the RMSE of the mapping results, the Taylor Diagram (Taylor and Karl, 2001; Beauchamp et al., 2020), the power spectral density (PSD) diagram, and the efficiency of the methods.

The Taylor Diagram (Taylor and Karl, 2001), as an effective method, has been widely used to evaluate and verify mapping work. The correlation coefficient (COR), centered root mean square error (Centered RMSE), and standard deviation (STD) can all be expressed on Taylor Diagram and in the form of one point. The Taylor Diagram can centrally express not only the related information of multiple mapping methods but also comprehensively and clearly reflect the data reconstruction capabilities of the methods. The COR indicates the similarity between the reconstruction results and the observations. The centered RMSE represents the error difference between the mapping results and the trues. The ratio of the STD reflects the degree of dispersion between the ability to reconstruct the entire spatial data and the observations. The theoretical expressions of COR, RMSE, and STD are detailed in Eqs. (21), (22), and (23) (Taylor and Karl, 2001; Beauchamp et al., 2020; Zhen et al., 2020), respectively.

$$STD: \sigma^2_{x\_\hat{x}}(t_k) = \frac{1}{N}\sum_{k=1}^{N}\left[(x_k - \hat{x}_k) - \overline{(x_k - \hat{x}_k)}\right]^2 \tag{21}$$

$$COR: COR(t_k) = \frac{COV(x_k, \hat{x}_k)}{\sigma(x_k)\sigma(\hat{x}_k)} \tag{22}$$

$$Centered\ RMSE: RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[(x_k - \overline{x_k}) - (x_r - \overline{x_r})\right]^2} \tag{23}$$

The Comparison of the Experiment Results

Figure 4 illustrates the 2D SSH mapping results from the AnDA, the PredNet, and the Mapping-PNN in **Figures 4C–E**, respectively. Beyond that, **Figure 4A** denotes the original SSH field of HYCOM, and **Figure 4B** illustrates the along-track

sampled data simulated from OSSEs on the date of March 1, 2017 on HYCOM, which utilized the parameter of 791-Orbit of Guanlan.

After the OSSEs process (**Figure 4A**), as the description mentioned, we used the OI as a first guess of the PredNet and Mapping-PNN methods for the gaps and missing data of the target region. By using the methods of AnDA, PredNet and Mapping-PNN, respectively, the 2D SSH grid of the 15°E−39°E, 120°N−144°N region could be obtained. The grid data are shown in **Figures 4C–E**, respectively.

TABLE 2 | The RMSE of the AnDA method, the PredNet method, and the Mapping-PNN method.

	AnDA	PredNet	Mapping-PNN
RMSE (cm)	2.0	1.8	1.8

The absolute geostrophic velocity diagram is shown in **Figure 5**. **Figure 5A** represents the absolute geostrophic velocity inversed from the original SSH field of the region (the Kuroshio and the Kuroshio extension, 15–39°E, 120–144°N region); **Figure 5B** represents the absolute geostrophic velocity of the AnDA method; **Figure 5C** represents the absolute geostrophic velocity of the PredNet method; and **Figure 5D** represents the absolute geostrophic velocity of the Mapping-PNN method.

As indicated in **Figure 5**, the absolute geostrophic velocity inversed from the Mapping-PNN method (**Figure 5D**) is more similar to the true value of HYCOM than the geostrophic velocity inversed from the other two methods. Compared with the AnDA method and the PredNet method, the Mapping-PNN method could obtain more small-scale ocean phenomena.

Table 2 displays the RMSE of the above test. As revealed in **Table 2**, the RMSE level of the 2D SSH grid data by using the methods of the three data-driven methods is basically within the range of <2 cm. The RMSE of the AnDA method, the PredNet method, and the Mapping-PNN method is 2.0, 1.8 and 1.8 cm,

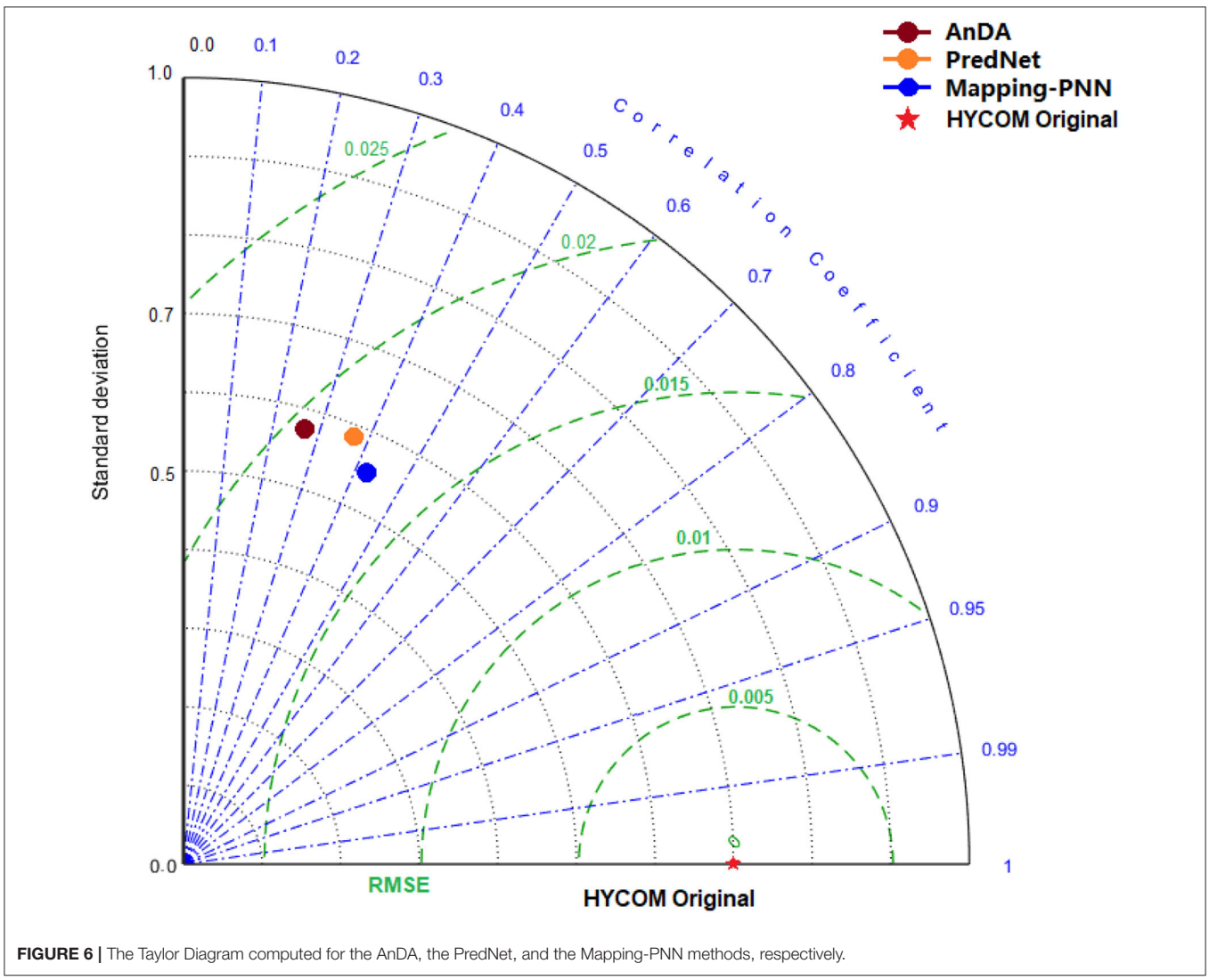


FIGURE 6 | The Taylor Diagram computed for the AnDA, the PredNet, and the Mapping-PNN methods, respectively.

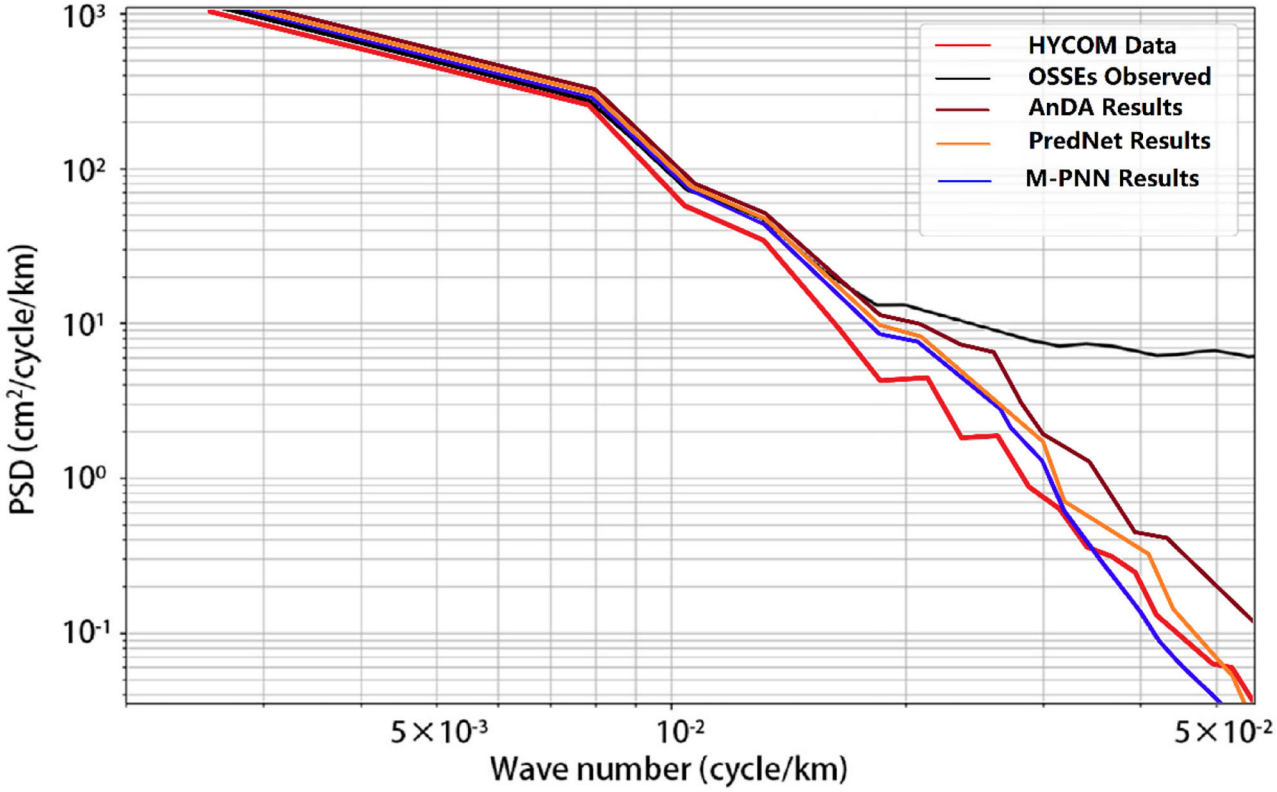


FIGURE 7 | The PSD of the AnDA, the PredNet, and the Mapping-PNN (M-PNN for short).

TABLE 3 | The comparison of efficiency, error probability, accuracy, and learning rate of the three data-driven methods.

No.	Test datasets	Efficiency			Error probability	Learning rate		Accuracy		
		AnDA	PredNet	M-PNN	AnDA	PredNet	M-PNN	AnDA	PredNet	M-PNN
1	30-day	9 days	7 days	4 days	0.01	0.02	0.02	89%	91%	92%
2	60-day	9 days	9 days	6 days	0.01	0.01	0.01	91%	91%	93%

respectively. Therefore, the RMSE of the WSA SSH grid data products can all be limited within the range of ∼1.8 cm.

Figure 6 is the Taylor Diagram corresponding to AnDA, PredNet, and Mapping-PNN methods. Specifically, the red star signifies the location of HYCOM Original value on the diagram. The green dot represents the COR, RMSE, and STD location of the AnDA mapping result on the diagram. The gray dot refers to the error location of the PredNet, and the orange dot refers to the error location of the Mapping-PNN. The RMSE results of the three data-driven methods are in the range of <2 cm, which are the same as the descriptions in **Table 2**. The RMSE value of PredNet and Mapping-PNN is slightly better, which is in the range of 1.8 cm. In addition, the Mapping-PNN method has the same data mapping capabilities and can obtain high-precision, high-resolution SSH grid data product with a high rate and low error. As shown in **Figure 6**, the relative error of the Mapping-PNN method is better than that of the other methods.

Figure 7 is the PSD diagram corresponding to AnDA, PredNet, and Mapping-PNN (M-PNN for short.) methods, respectively. The red line denotes the PSD of the HYCOM true value of the 15–39°E, 120–144°N region. The black line represents the PSD of the observations of the OSSEs, the deep-red line refers to the PSD of the mapping result from the AnDA method, the orange line is that of PredNet, and the blue line in the PSD diagram signifies the Mapping-PNN method. It shows that the Mapping-PNN method is better than the other methods for recognizing scales < ∼40 km (including the sub-mesoscale, the ocean fronts, the internal waves, etc.), but the reconstruction ability of them are almost the same for the scales larger than ∼40 km.

In **Table 3**, the efficiency, accuracy, and learning rate (the error probability for AnDA) of the three data-driven methods [AnDA, PredNet, and Mapping-PNN methods (M-PNN for short)] are described to evaluate their DL ability. The continuous time of the

Mapping-PNN training is better than that of the AnDA method and the PredNet method, which reaches 4 days long in test No. 1 and 6 days long in test No. 2 in the Kuroshio and the Kuroshio extension regions (from January 1, 2017 to March 1, 2017 on HYCOM). The efficiency of Mapping-PNN is higher, and the Mapping-PNN method has similar accuracy and learning rate with the PredNet method and the AnDA method.

CONCLUSION AND DISCUSSION

Focusing on the future scientific research targets of WSA 2D mapping, this paper proposes a new data-driven mapping method called Mapping-PNN. The experiment result, which has been obtained through 10-year WSA satellite along-track OSSEs on the HYCOM data, illustrates that the method of this paper can decrease the mapping RMSE, improve the training efficiency, and meet the grid mapping expectations. Three data-driven methods (the AnDA, the PredNet, and the Mapping-PNN) were used to implement data mapping practice tests in the same region. The future satellite will provide 2D wide-swath altimetric information with an unprecedented high resolution. By comparing the three data mapping methods, this research shows that the data reconstruction ability of the Mapping-PNN method meets the WSA scientific targets.

The Mapping-PNN method proposed in this paper is evaluated with several ways, which are SSH differences, as well as the absolute geostrophic velocities differences, are illustrated with visibility analysis, and the comparisons of errors are presented with the Taylor Diagram and the PSD diagram. More specifically, the OSSEs are implemented in the same region, and the AnDA, the PredNet and the Mapping-PNN methods are used for 2D SSH Mapping. The RMSE of the three data-driven methods is at the range of <2.0 cm. Notably, the RMSE value of PredNet and Mapping-PNN is slightly better, which is in the range of $\sim$1.8 cm. With the same data mapping capabilities, the Mapping-PNN method could obtain high-precision, high-resolution SSH grid data product. The observational dataset is based on a 14-day aggregation; considering to test other aggregation strategies, the experiments with months or seasons will be one of the future works. And, in addition, to improve the efficiency, we will use GPU(s) to implement mapping experiments in the future.

Being different from the classical model-driven method, the data-driven methods rely on the spatial-temporal relationship of the observations so that a data-driven method can capture the ocean phenomena that may not be accounted for in purely numerical models (Lopez-Radcenco et al., 2019). Moreover, the DI method (one of the model-driven methods) is based on the PV conservation theory. When sub-mesoscale ocean phenomena are $< \sim$10 km scale, the kinetic energy is dominated by internal waves, in which the geostrophic balance fails, and the PV conservation theory is no longer applicable. And the PV conservation theory in coastal regions or the tropic regions may also fail, in which the DI method is invalid.

The data-driven method is more suitable for global 2D mapping of ocean phenomena with small scales. The data-driven and model-driven methods can be combined appropriately to obtain a new method, which not only has better mapping error accuracy than the current data-driven method but also has the common advantages of both methods. This will be a challenge for future mapping work and a research direction in the future.

One development direction of the 2D mapping method will be continuing more in-depth research along the direction of the data-driven roadmap, considering the utilization of new methods, such as generative adversarial networks (GANs) (Goodfellow et al., 2014) and enhanced networks for reinforcement learning (Huang et al., 2017), etc.

Furthermore, considering the data-driven mapping method, making error analysis of each layer (especially the hidden layer), replacing the previous BP method of individual neurons with the idea of Capsule (Sabour et al., 2017), to improve the learning rate of the entire network, and avoid the hidden dangers of invariance of CNN (Sabour et al., 2017), then obtaining data reconstruction closer to the real fields will be another research direction for future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

CM, JD, and GC: conceptualization. CM and JD: formal analysis, methodology, investigation, and resources. GC and CM: funding acquisition. JD: software, data curation, writing—original draft preparation, and visualization. CM: writing—review and editing. GC: supervision and project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the Key Research and Development Program of Shandong Province: No. 2019GHZ023; National Natural Science Foundation of China: No. 41906155 and 42030406; the Fundamental Research Funds for the Central Universities: No. 201762005; and the National Key Scientific Instrument and Equipment Development Projects of National Natural Science Foundation of China: No. 41527901.

REFERENCES

Amores, A., Jordà, G., Arsouze, T., and Sommer, J. (2018). Up to what extent can we characterize ocean eddies using present-day gridded altimetric products? *J. Geophy. Res. Oceans* 123, 7220–7236. doi: 10.1029/2018JC014140

Archiving, Verification and Interpretation of data of Satellites Oceanography, CollecteLocalisation Satellites (CLS), AVISO, CNES (2019). Available

online at: https://www.aviso.altimetry.fr/en/home.html (accessed on 2 January 2021).

Ballarotta, M., Ubelmann, C., Pujol, M. I., Taburet, G., and Picot, N. (2019). On the resolutions of ocean altimetry maps. *Ocean Sci.* 15, 1091–1109. doi: 10.5194/os-15-1091-2019

Ballarotta, M., Ubelmann, C., Rog, Marine, Fournier, F., Faugere, Y., Dibarboure, G., Morrow, R., et al. (2020). Dynamic mapping of along-track ocean altimetry: performance from real observations. *J. Atmos. Ocean. Technol.* 37, 1593–1601. doi: 10.1175/JTECH-D-20-0030.1

Beauchamp, M., Fablet, R., Ubelmann, C., Ballarotta, M., and Chapron, B. (2020). Intercomparison of data-driven and learning-based interpolations of along-track nadir and wide-swath SWOT altimetry observations. *Remote Sens.* 12:3806. doi: 10.3390/rs12223806

Bretherton, F. P., Davis, R. E., and Fandry, C. B. (1976). A technique for objective analysis and design of oceanographic experiments applied to mode-73. *Deep-Sea Res.* 23, 559–582. doi: 10.1016/0011-7471(76)90001-2

Chen, G., Tang, J., Zhao, C., Wu, S., and Wu, L. (2019). Concept design of the "Guanlan" science mission: China's novel contribution to space oceanography (Ocean OBS19'). *Front. Mar. Sci.* 6, 1–14. doi: 10.3389/fmars.2019.00194

Dussurget, R., Birol, F., Morrow, R., and Mey, P. D. (2011). Fine resolution altimetry data for a regional application in the Bay of Biscay. *Mar. Geod.* 34, 447–476. doi: 10.1080/01490419.2011.584835

Gaultier, L., Ubelmann, C., and Fu, L. L. (2016). The challenge of using future SWOT data for oceanic field reconstruction. *J. Atmos. Ocean. Technol.* 33, 119–126. doi: 10.1175/JTECH-D-15-0160.1

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., and Ozair, S., et al. (2014). Generative adversarial networks. *Adv. Neural Inf. Process. Syst.* 3, 2672–2680. doi: 10.1145/3422622

Guillou, F. L., Metref, S., Cosme, E., Sommer, J. L., and Verron, J. (2020). Mapping altimetry in the forthcoming SWOT era by back-and-forth nudging a one-layer quasi-geostrophic model. *J. Atmos. Ocean. Technol.* 38, 1–41. doi: 10.1175/JTECH-D-20-0104.1

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hua, B. L., and Haidvogel, D. B. (1986). Numerical simulations of the vertical structure of quasi-geostrophic turbulence. *J. Atmos. Sci.* 43, 2923–2936. doi: 10.1175/1520-0469(1986)043<2923:NSOTVS>2.0.CO;2

Huang, V., Ley, T., Vlachou-Konchylaki, M., and Hu, W. (2017). Enhanced experience replay generation for efficient reinforcement learning. arXiv preprint arXiv: 1705.08245

Hybrid Coordinate Ocean Model (HYCOM) Data. (2021). Available online at: ftp://ftp.hycom.org/datasets/GLBu0.08/expt_91.1/hindcasts/2015/ (accessed on 2 January 2021).

Itoh, S. (2010). Characteristics of mesoscale eddies in the Kuroshio-Oyashio extension region detected from the distribution of the sea surface height anomaly. *J. Phys. Oceanogr.* 40, 1018–1034. doi: 10.1175/2009JPO4265.1

Ji, J., Dong, C., Zhang, B., Liu, Y., and Chen, D. (2018). Oceanic Eddy characteristics and generation mechanisms in the Kuroshio Extension Region. *J. Geophy. Res. Oceans* 123, 8548–8567. doi: 10.1029/2018JC014196

Kingma, D.P., and Ba, J. (2015). "Adam: a method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, San Diego, CA, USA, 7–9 May. arXiv:1412.6980

Le Traon, P. Y., Faugère, Y., Hernandez, F., Dorandeu, J., Mertz, F., and Ablain, M. (2003). Can we merge GEOSTAT follow-on with TOPEX/Poseidon and ERS-2 for an improved description of the ocean circulation? *J. Atmos. Ocean. Technol.* 20, 889–895. doi: 10.1175/1520-0426(2003)020<0889:CWMGFW>2.0.CO;2

Lguensat, R., Sommer, J. L., Metref, S., Cosme, E., and Fablet, R. (2019a). Learning generalized quasi-geostrophic models using deep neural numerical models. arXiv preprint arXiv: 1911.08856.

Lguensat, R., Tandeo, P., Ailliot, P., Pulido, M., and Fablet, R. (2017). The analog data assimilation. *Mon. Weather Rev.* 145, 4093–4107. doi: 10.1175/MWR-D-16-0441.1

Lguensat, R., Viet, P. H., Sun, M., Chen, G., and Fablet, R. (2019b). Data-driven interpolation of sea level anomalies using analog data assimilation. *Remote Sens.* 11:858. doi: 10.3390/rs11070858

Lima, E., Sun, X., Dong, J., Wang, H., Yang, Y., and Liu, L. (2017). Learning and transferring convolutional neural network knowledge to ocean front recognition. *IEEE Geosci. Remote Sens. Lett.* 14, 354–358. doi: 10.1109/LGRS.2016.2643000

Lopez-Radcenco, M., Pascual, A., Gomez-Navarro, L., Aissa-El-Bey, A., Chapron, B., and Fablet, R. (2019). Analog data assimilation of along-track nadir and wide-swath swot altimetry observations in the Western Mediterranean Sea. *IEEE J. Selected Top. Appl. Earth Observ. Remote Sens.* 12, 1–11. doi: 10.1109/JSTARS.2019.2903941

Lotter, W., Kreiman, G., and Cox, D. (2017). Deep predictive coding networks for video prediction and unsupervised learning. arXiv preprint arXiv: 1605.08104.

Ma, C., Guo, X., Zhang, H., Di, J., and Chen, G. (2020). An investigation of the influences of SWOT sampling and errors on ocean eddy observation. *Remote Sens.* 12, 2682–2698. doi: 10.3390/rs12172682

Ma, C., Li, S., Wang, A., Yang, J., and Chen, G. (2019). Altimeter observation-based Eddy nowcasting using an improved ConvLSTM network. *Remote Sens.* 11, 783. doi: 10.3390/rs11070783

Morrow, R., Fu, L. L., Ardhuin, F., Benkiran, M., and Zaron, E. D. (2019). Global observations of fine-scale ocean surface topography with the surface water and ocean topography (SWOT) mission (Ocean OBS19'). *Front. Mar. Sci.* 6, 1–19. doi: 10.3389/fmars.2019.00232

Morrow, R., and Traon, P. Y. L. (2012). Recent advances in observing mesoscale ocean dynamics with satellite altimetry. *Adv. Space Res.* 50, 1062–1076. doi: 10.1016/j.asr.2011.09.033

Nagano, A., Ichikawa, K., Ichikawa, H., Konda, M., and Murakami, K. (2013). Volume transports proceeding to the kuroshio extension region and recirculating in the Shikoku Basin. *Oceanogr. J.* 69, 285–293. doi: 10.1007/s10872-013-0173-9

Pie, N., and Schutz, B. E. (2008). Subcycle analysis for Icesat's repeat groundtrack orbits and application to phasing maneuvers. *J. Astronaut. Sci.* 56, 325–340. doi: 10.1007/BF03256556

Roge, M., Morrow, R., Ubelmann, C., and Dibarboure, G. (2017). Using a dynamical advection to reconstruct a part of the SSH evolution in the context of SWOT, application to the Mediterranean Sea. *Ocean Dyn.* 67, 1–20. doi: 10.1007/s10236-017-1073-0

Sabour, S., Frosst, N., and Hinton, G. E. (2017). Dynamic routing between capsules. arXiv preprint arXiv: 1710.09829.

Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., and Lillicrap, T. (2016). "Meta-learning with memory-augmented neural networks," in *Proceeding of the International Conference on Machine Learning*, 19–24, 1842–1850. doi: 10.5555/3045390.3045585

Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., and Woo, W.C. (2015). "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," in *Proceedings of the Advances in Neural Information Processing Systems*, Montreal, QC, Canada, 7–12, 802–810. doi: 10.5555/2969239.2969329

Taylor, and Karl, E. (2001). Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* 106, 7183–7192. doi: 10.1029/2000JD900719

Ubelmann, C., Cornuelle, B., and Fu, L. L. (2016). Dynamic mapping of along-track ocean altimetry: method and performance from observing system simulation experiments. *J. Atmos. Ocean. Technol.* 33, 1691–1699. doi: 10.1175/JTECH-D-15-0163.1

Ubelmann, C., Klein, P., and Fu, L. L. (2015). Dynamic interpolation of sea surface height and potential applications for future high-resolution altimetry mapping. *J. Atmos. Ocean. Technol.* 32, 177–184. doi: 10.1175/JTECH-D-14-00152.1

Wunsch and Carl. (1996). *The Ocean Circulation Inverse Problem*. Cambridge University Press.

Yang, Y., Dong, J., Sun, X., Lima, E., Mu, Q., and Wang, X. (2018). A CFCC-LSTM model for sea surface temperature prediction. *IEEE Geosci. Remote Sens. Lett.* 15, 207–211. doi: 10.1109/lgrs.2017.2780843

Zhen, Y., Tandeo, P., StéphanieLeroux, Metref, S., and Sommer, J. L. (2020). An adaptive optimal interpolation based on analog forecasting: application to SSH in the Gulf of Mexico. *J. Atmos. Ocean. Technol.* 37, 1–46. doi: 10.1175/JTECH-D-20-0001.1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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

frontiers in Marine Science

ORIGINAL RESEARCH
published: 08 February 2022
doi: 10.3389/fmars.2021.672272

Toward Exploring Topographic Effects on Evolution and Propagation of Ocean Mesoscale Eddies Through Life Cycle Across Izu-Ogasawara Ridge in Northwestern Pacific Ocean

Rui Nian[1], Xue Geng[2], Zhengguang Zhang[3], Minghan Yuan[1], Zhen Fu[1], Hengfu Xu[1], Hua Yang[1], Qi Lai[1], Hui He[1], Chi Wei Su[4], Lina Zang[1], Qiang Yuan[1] and Bo He[1]*

[1] School of Electronic Engineering, Ocean University of China, Qingdao, China, [2] JD.com, Inc., Beijing, China, [3] Physical Oceanography Laboratory, Ocean University of China, Qingdao, China, [4] School of Economics, Qingdao University, Qingdao, China

Ocean mesoscale eddies contribute significantly to water transport on a global scale, constituting the ubiquitous, irregular, discrete, nonlinear components. In this manuscript, we propose to explore whether and how the topographic effect of one meridional ridge, could exert considerable influences on the evolution and propagation of mesoscale eddies through their life cycle, directly from the perspectives of real observation statistics. We systematically investigate the known variability of mesoscale eddy trajectories, derived by multimission satellite altimetry from 1993 to 2018, of a life cycle more than 6 months, over the Izu-Ogasawara Ridge, and quantitatively examined the eddy-ridge interaction by observation statistics and wavelet coherence map, with respect to the intrinsic attributes, namely, the amplitude, the rotation speed, the radius. Due to the spatial-temporal diversity, a series of correlative steps have been particularly designed along time-frequency domain to trace back mesoscale eddy trajectories in a variety of origins, location, lifespan, polarity, either completely or partially passing over the ridge, and to facilitate the standardization in statistics across three phases of their life cycle, i.e., before, during and after the interaction with the ridge. It has been revealed in our experiment that three intrinsic attributes of mesoscale eddies within 25 years, all demonstrated noticeable correlation with the variation of topographic relief over the ridge. We observed that most of the cyclonic eddies obviously tended to begin to decay or even demise, while on the contrary, some of the anticyclonic eddies preferred to intensify slightly, or making no significant difference when encountering the upslope until climbing across the top, basically consistent with the expectation of potential vorticity (PV) conservation. The drifting velocity agreed with the tendency that the direction would be more probably modified toward equatorward or poleward by forcing to meridional component, with zonal component reduced at the beginning. The mesoscale eddies with the passage over the ridge exhibited the relatively high average horizontal scales, amplitude, rotation speed on the whole, compared to those with only partially passage. The developed scheme could integrate more evidences on how mesoscale eddies response to the topographic effects during their time-varying evolution and propagation

OPEN ACCESS

Edited by:
Sabrina Speich,
École Normale Supérieure, France

Reviewed by:
Moacyr Cunha de Araujo Filho,
Federal University of Pernambuco, Brazil
Alex Stegner,
École Polytechnique, France

***Correspondence:**
Rui Nian
nianrui_80@163.com

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *25 February 2021*
Accepted: *13 December 2021*
Published: *08 February 2022*

Citation:
Nian R, Geng X, Zhang Z, Yuan M, Fu Z, Xu H, Yang H, Lai Q, He H, Su CW, Zang L, Yuan Q and He B (2022) Toward Exploring Topographic Effects on Evolution and Propagation of Ocean Mesoscale Eddies Through Life Cycle Across Izu-Ogasawara Ridge in Northwestern Pacific Ocean. Front. Mar. Sci. 8:672272. doi: 10.3389/fmars.2021.672272

process, and help provide opportunities to potentially identify and predict the underlying dynamic patterns and mechanisms that mesoscale eddies engage in ocean dynamics when proceeding toward meridional ridges on a global scale, with the promise of the end-to-end data-driven solution, such as deep learning architecture involved in the future.

Keywords: ocean mesoscale eddies, topographic effects, wavelet analysis, life cycle, satellite altimetry, eddy-ridge interaction, Izu-Ogasawara Ridge

INTRODUCTION

Ocean mesoscale eddies, with typical horizontal scales of less than 100 km and timescales on the order of months, constitute one of the most essential and fundamental components of water transport on a global scale (Sarangi, 2012; Frenger et al., 2013; Zhang et al., 2014, 2018, 2019, 2020; Shu et al., 2018). It is of great scientific importance to quantitatively investigate the known variability of ocean mesoscale eddies through life cycle. On this basis, we could integrate evidences on the spatio-temporal responses to manifold intrinsic and external origins during their evolution and propagation. It could also help provide opportunities to potentially identify and predict the underlying patterns and mechanism that mesoscale eddies involve, in ocean dynamics, or the related biological and chemical processes.

For decades, evidences have been shown that the life cycle evolution and trajectories of mesoscale eddies would be to some extent affected by the topographic effects when encountering with mid-ocean ridges, seamounts, bottom slopes, or over a variety of topography (Sekine, 1989; Jacob et al., 2002; Sutyrin et al., 2011; Torres and Gomez-Valdes, 2017). The eddy-seamount interaction has been examined in a variety of quasi-geostrophic and multi-layer primitive equation settings, as well as in laboratory experiments (Adduce and Cenedese, 2004; Nycander and Lacasce, 2004; Sutyrin et al., 2008, 2011). The evolution of eddies near a topographic obstacle depends on the bathymetric gradients, which can be treated as background potential vorticity (PV) gradient, and on the vorticity of these eddies. This provides several general predictive rules (Kamenkovich et al., 1996; Beismann et al., 1999; Thierry and Morel, 1999; Morrow et al., 2004; Hu et al., 2012; Falcini and Salusti, 2015): (1) When the column PV and the topographic PV are conserved (f-plane), and if the topography is smooth, vortex column squeezing leads to the decay of relative vorticity for cyclonic eddies, while the column stretching leads to the intensification, with strong barotropic component of the eddy, conversely, for anticyclonic eddies. (2) If the topographic gradient is strong but regular, topographic Rossby wave may scatter the lower PV and lead to the vortex compensation at the surface. (3) When the column PV and the planetary PV are conserved (β-plane), cyclones and anticyclones will have different meridional propagation. (4) When PV is not conserved because of the bottom friction, irregular topography will lead to the erosion of the vortex. Apart from the theoretical understanding, the survey from available real observation still remains incomplete.

The northwestern subtropical Pacific Ocean (NWSTP) could be considered to be one of the most convoluted bottom topography area in the ocean (Qiu and Lukas, 1996; Qiu and Miao, 2000; Ohara et al., 2007; Jing et al., 2011; Rudnick et al., 2011; Yuan and Wang, 2011; Li and Wang, 2012; Yang et al., 2013). Trodahl and Isachsen (2018) suggested from the calculation of a nonlinear eddy-permitting ocean model hindcast that the time-mean currents could be strongly guided by bottom topography in the northern North Atlantic and Nordic seas, where baroclinic instability constitutes a consistent source of the mesoscale eddy field but topographic potential vorticity gradients impact unstable growth significantly, and the topographic effects has been systematically observed on finite-amplitude eddy characteristics, including a general suppression of length scales over the continental slopes. Ihara et al. (2002) developed a two-layer primitive ocean model with a submerged ridge that mimics the Izu-Ogasawara Ridge and divides Philippine Basin in the west and the Pacific Basin in the east, suggested that the baroclinic eddies may be generated owing to nonlinearity through the interaction between the basin-wide response to seasonal winds and the localized bottom topography, and believed that the possible link between the eddy generation and the seasonal variation will shed new light on the predictability on the Kuroshio meandering. Ebuchi and Hanawa (2001) investigated trajectories of mesoscale eddies in the kuroshio recirculation region with sea surface height anomaly, and it seemed to be crucially affected by the bottom topography around Izu-Ogasawara Ridge region, where most of eddies pass through the gap between the Hachijojima Island and Ogasawara (Bonin) Islands, and in the region south of Shikoku and east of Kyushu, some of the eddies coalesce with the Kuroshio, which may trigger the path variation of the Kuroshio.

We hope to look into the interaction between the variation of mesoscale eddies in history through real observation and the topographic effects, when proceeding toward the meridional ridge Northwestern Pacific Ocean, Izu-Ogasawara Ridge. Kamenkovich et al. (1996) have examined the influences of one meridional ridge, Walvis Ridge, on the dynamics of Agulhas eddies, in a series of numerical experiments with a two-layer primitive equation model, differing vertical structures of a specified intensity in the upper layer and a prescribed horizontal scale, and found significantly baroclinic eddies could go cross the ridge, but barotropic or near-barotropic ones could not. Eddies with the characteristics and dimensions of Agulhas rings did react to crossing a ridge, exhibiting an intensification in the form of a deeper thermocline and a heightened sea surface amplitude just before reaching the ridge. Beismann et al. (1999) applied the quasi-geostrophic model to investigate the topographic influence of a meridional ridge on the translatory movement of Agulhas

rings, with the vertical ring structure, the initial ring position, and the height of the ridge varied, and found that the general northwestward movement of the model eddies has been modified toward a more equatorward direction by encountering the upslope of the ridge, even forced toward a pure meridional movement for the sufficient topographic heights and strong slopes. The vertical coherence lost during the translation of the eddies, due to the possible rectification of radiated Rossby waves at the topographic slope. Strong, shallow eddies over deep lower layers can cross the ridge without strong modification of their translatory movement. Thierry and Morel (1999) investigated the influence of steep topography on the propagation of a surface intensified vortex in a two-layer quasigeostrophic model, with the planetary beta effect taken into account, and found the steep topography could scatter disturbances created by the upper-layer vortex displacement and maintain the lower-layer motion weak. The numerical experiments showed that both the steep topography and reduced-gravity trajectories remained close up to a large radius, after which a vortex above a strong slope became unstable and was dispersed by a signature of topographic Rossby waves.

We try to further extend the research scope of the topographic effect on the meridional ridge, into the real observation statistics of mesoscale eddy trajectories, directly retrieved from the satellite altimeter and bathymetric digital elevation, with the promise of the end-to-end data-driven solution. It has been noted that most commonly applied time series approaches, such as bootstrapping regression models (Freedman, 1981), rolling window regression (Su et al., 2019b), and auto-regressive distributed lag (ADL) model (Shahbaz et al., 2015), often only consider one single or several independent predictive elements, and tend to ignore the combined effects of the complex structures, while the classic correlation calculation, is only suitable for studying relatively short-term time series and severely limited to stationary series. Wavelet analysis is a mathematically basic tool for analyzing localized intermittent periodicities within non-stationary time series (Daubechies, 1991; Torrence and Compo, 1998; Labat, 2005), which could optimally describe the occurrence of transient events, and well adapt to conditions where the amplitude of the response varies significantly in non-stationary time series (Meyers et al., 1993; Steel and Lange, 2007; Pineda-Sanchez et al., 2013; Li et al., 2015; Rhif et al., 2019; Su et al., 2021a). As the extended usage of wavelet transform, wavelet coherence (Maraun and Kurths, 2004; Reboredo et al., 2017; Su et al., 2019a, 2020, 2021b; Tao et al., 2021) and phase difference (Aguiar-Conraria and Soares, 2011; Nian et al., 2021a,b; Su et al., 2021c,d) can be utilized to recognize whether two time series are quantitatively linked by a certain correlation even causality relationship.

Recently, machine learning (ML), has become one of the most powerful tools in the field of multivariate multi-step time series prediction (Hochreiter and Schmidhuber, 1997; Geurts et al., 2006; Sapankevych and Sankar, 2009; Box et al., 2015; Hu and Zheng, 2020; Nian et al., 2021c). Deep learning could be regarded as one of the hottest topics in the context, and all kinds of most emerging and advanced algorithms have been put forward and made progresses (Hinton and Salakhutdinov, 2006; Krizhevsky et al., 2012; Goodfellow et al., 2014; He et al., 2016; Huang et al., 2017; Wan et al., 2019), such as Recurrent Neural Network (RNN) (Elman, 1990; Lipton et al., 2015; Braakmann-Folgmann et al., 2017; Qin et al., 2017) and Long Short Term Memory (LSTM) (Kalchbrenner et al., 2015; Shi et al., 2015; Greff et al., 2016; Zhang et al., 2017; Shi and Yeung, 2018; Wang et al., 2021; Gangopadhyay et al., 2021; Nian et al., 2021b). We expect to establish a comprehensive predictive model of mesoscale eddy trajectories toward meridional ridges on a global scale in the future, coupling with the topographic effects, via deep learning. So in this manuscript we try to evaluate the possible responses of mesoscale eddies to manifold attributes, particularly the topography effects onto one meridional ridge, from both the real observation statistics and wavelet coherence map, which might help collect evidences to identify the fundamental roles of mesoscale eddy attributes, so as to adaptively select from the relevant indices that dominantly correlative with the time-varying process through life cycle, to further feed into deep learning architecture.

In this manuscript, taking the Izu-Ogasawara Ridge in Northwest Pacific Ocean as an example, we retrieve daily mesoscale eddy attributes and the integrated bathymetric digital elevation in the study area, from AVISO satellite altimeter and ETOPO1, respectively, and systematically investigate the variation of mesoscale eddy trajectories, of a life cycle more than 6 months, to discover the possible expression of the topographic effects with and without passage over the ridge. We employ temporal regularity, spatial normalization, range expansion, to identically align into a standardized representation. We observe the evolution and propagation process of mesoscale eddies in history through life cycle, and quantitatively examine the eddy-ridge interaction by both observation statistics and wavelet coherence map, from multiple perspectives, namely the amplitude, the rotation speed, and the radius. This will also help establish a complete machine learning framework for multivariate time series, including the amplitude, rotation speed, radius, the latitude and longitude at the geographical location, the zonal displacement and meridional displacement, and the variation of the bathymetric topography, which would possibly behave as the potential input vectors when we attempt to determine the comprehensive predictive patterns of mesoscale eddy variation toward meridional ridges.

The remainder of the manuscript is organized as follows: Section "Data and Methods" describes the access of the time series about both topography and mesoscale eddy trajectories referred, the steps of mathematical statistics in our study, the basics in wavelet coherence. Section "Results" quantitatively and systematically investigates the interaction between the variation of mesoscale eddy trajectories in history and the topographic effects, and exhibits the experimental results. Section "Discussion" suggests the potentially mutual relationships among manifold intrinsic and external attributes including topography effects, which would help develop deep learning architecture for the potential prediction of mesoscale eddy trajectories. Finally, the conclusions are drawn in Section "Summary and Conclusion".

DATA AND METHODS

Mesoscale Eddies in the Izu-Ogasawara Ridge

The mesoscale eddy trajectory atlas, retrieved from AVISO satellite altimeter by Chelton et al. (2011), is of 1 day time-resolution, provides the amplitude, the radius, the rotation speed, the life cycle, and the longitude and latitude generated for the mesoscale eddies within the spatial range of the global ocean (excluding 2°S–2°N), and retains only mesoscale eddies with lifetime of more than 4 weeks, which contains 179, 127 cyclones and 173, 245 anticyclones from 1993 to 2018. Our study area refers to 130°E–170°E, 20°N–35°N, as is shown in **Figure 1**, with location the marked in the red rectangle. 74 mesoscale eddy trajectories across the Izu-Ogasawara Ridge from east to west, of a life cycle more than 6 months, have been systematically and statistically investigated, to explore the potential spatio-temporal correlation with the topographic effects, where five mesoscale eddies that are odd and eccentric in trajectories have been discarded. Among the rest 69 mesoscale eddy trajectories, 32 of them completely pass through the ridge from east to west, and 37 of them only partially leap over one side of the ridge and then demise inside it. The records of the amplitude offer the difference between the extreme value of the sea surface height (SSH) and the average value of the SSH within the mesoscale eddy. The radius refers to the circle with area equal to that within the closed contour of SSH in each eddy that has the maximum average geostrophic speed. The rotation speed is defined as the maximum of the average geostrophic speeds around all of the closed contours of SSH inside the eddy, while the life cycle is the number of days from the first to the last detection date. **Table 1** is the statistics of mesoscale eddies in the STCC from 1993 to 2018. **Table 2** displays the statistics of 69 mesoscale eddies in study from 1993 to 2018. **Table 3** lists the seasonal and polarized statistics of all the mesoscale eddies over the principal ridge region (139°E–143°E, 25°N–32°N) from 1993 to 2018.

The topography of Izu-Ogasawara Ridge has been retrieved from ETOPO1, the global integrated bathymetric-topographic digital elevation model (DEM), to discover the potential topographic effects on mesoscale eddies, with the terrestrial topography and ocean water depth provided, at the highest resolution, as a grid-registered, 1 arc-minute grid that spans from pole to pole and from −180° to 180° in longitude. The Izu-Ogasawara Ridge is located in the northwest of the Philippine Sea, which occupies a large part of the Northwest Pacific, and the Philippine Sea Plate forms its seafloor topography, long from north to south, narrow from north to south, at a large water depth, with a variety of islands, ridges, ocean basins, deep trenches, and other topography. The range of Izu-Ogasawara Ridge starts from Nagoya, Honshu, Japan to the southern area of Tokyo, and the open ocean constitutes the Ogasawara-Iwo Jima arc. Along the eastern side of the ridge runs the Izu-Ogasawara Trench, with the maximum depth deeper than 9,000 m, and to the western of the ridge, there is flat Shikoku Basin, the depth of which ranges approximately from 4,000 to 5,000 m. According to the former studies based on the observational data, the vertical scale of the mesoscale eddies could reach about 1,000 m in the region around the ridge (Zhang et al., 2013). On the other hand, the water depth along the trajectories that 69 mesoscale eddies travel within the ridge has been started from 110 m to more by statistics, and this means that the topography of the ridge could indeed directly influence the attributes of mesoscale eddies during the evolution and propagation process.

Mathematical Statistics and Wavelet Analysis

Initialization

As the origin of mesoscale eddies in our study can be traced back in a variety of starting time, formation location, and successive lifespan intervals, we first uniformly reset the arrival time to the east ridge edge as the reference origin time 0, so that the standard timeline could be consistently imposed on all mesoscale eddies. Let the total number of mesoscale eddies be I, with the $i-th$ mesoscale eddy denoted as $M_{\mathrm{i}}(t_{\mathrm{i}})$, $i = \{1, 2, \ldots, I\}$, where t_{i} refers to the survival time that the given mesoscale eddy activity moves forward, and the amplitude, rotation speed, radius of the given mesoscale eddy could then be, respectively, defined as $A_{\mathrm{i}}(t_{\mathrm{i}})$, $S_{\mathrm{i}}(t_{\mathrm{i}})$, and $R_{\mathrm{i}}(t_{\mathrm{i}})$. For each mesoscale eddy, we first transform the time domain in the above three indices to identically align the arrival time as follows:

$$t_i^{'} = t_i - t_{i0} \tag{1}$$

where $t_{\mathrm{i}0}$ is the time point that the mesoscale eddy originally pass through the ridge, and the transformed $t_{\mathrm{i}}^{'}$ reset the arrival time with $t_{\mathrm{i}}^{'} = 0$ when the mesoscale eddy the east ridge edge.

Temporal Regularity

After the initialization stage, we make a temporal regularity for all the mesoscale eddies thoroughly proceeding within the ridge region. Let $\left(x_{\mathrm{i}}(t_i^{'}), y_{\mathrm{i}}(t_i^{'})\right)$ be the longitude and the latitude of the geographical location that each mesoscale eddy moves toward, with r indicating the ridge region. Among all those mesoscale eddies completely passing over the ridge, i.e., for any time $t_{\mathrm{i}}^{'} \geq 0$in a given mesoscale eddy, if $\left(x_{\mathrm{i}}(t_{\mathrm{i}}^{'}), y_{\mathrm{i}}(t_{\mathrm{i}}^{'})\right) \notin r$ exists, we first examine every timespan across the whole ridge region by defining $t_{i\max}^{'} = t_{\mathrm{ir}}^{'}$, with $t_{\mathrm{ir}}^{'}$ representing the time duration when the mesoscale eddy moves to the west edge of the ridge region, and then find out the maximum timespan duration $t_{*\max}^{'} = \underset{\mathrm{i}=1:I}{argmax}(t_{\mathrm{ir}}^{'})$, from all the mesoscale eddies that completely move out of the ridge edge. Furthermore, we define a temporal scaling index between the real and the maximum timespan duration $t_{*\max}^{'}$ across the ridge, with the interval from the arrival to departure time toward the ridge, and then align and extend each mesoscale eddy trajectory to be considered to an equal timespan length via interpolation.

For example, for the amplitude time series of each given mesoscale eddy that completely goes across the ridge, $A_{\mathrm{i}}(t_{\mathrm{i}}^{'})$, $t_{\mathrm{i}}^{'} \in [0, t_{\mathrm{ir}}^{'}]$, the cubic spline interpolation has been applied for

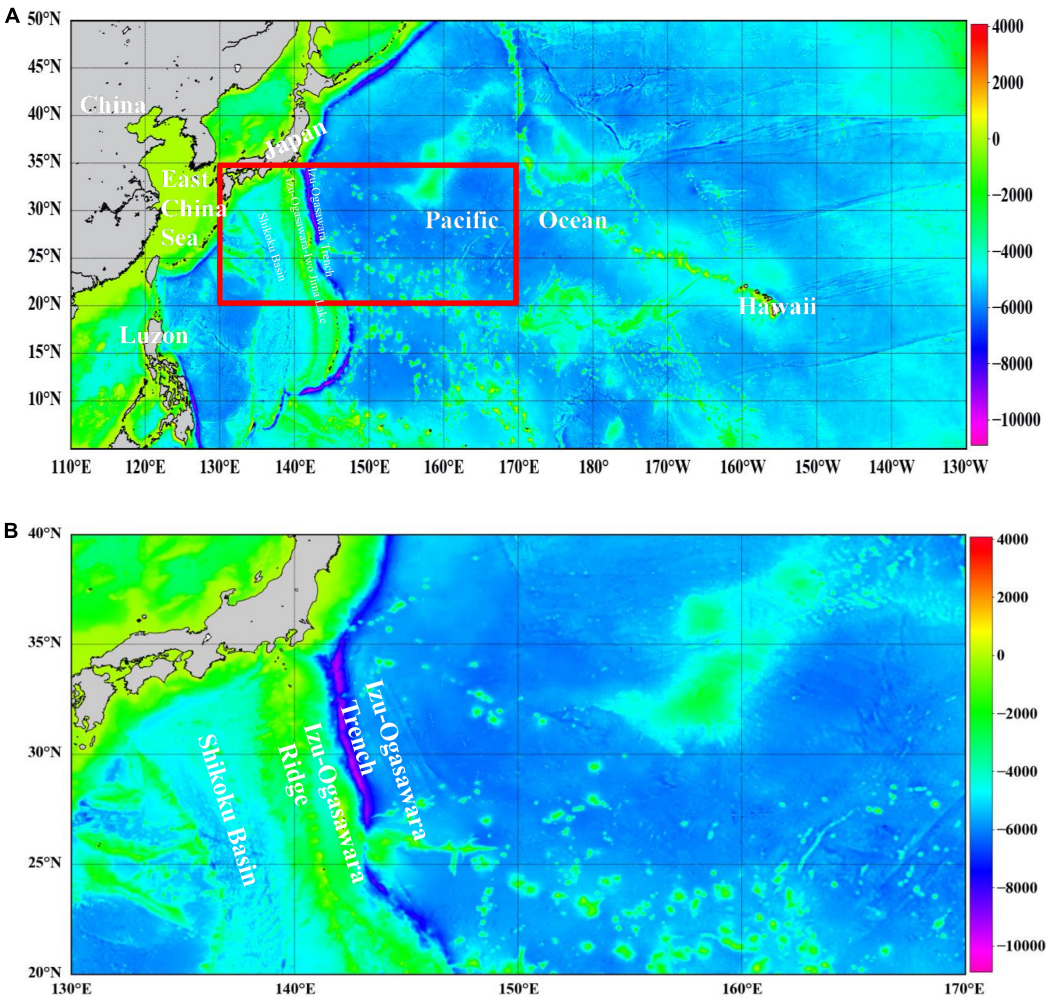


FIGURE 1 | The location and topographic mapping of our study area. **(A)** Location. **(B)** Topographic mapping.

TABLE 1 | Statistics of mesoscale eddies in the STCC from 1993 to 2018.

Region	Character	Number	Average lifespan (week)	Average amplitude (cm)	Average radius (km)	Average speed (cm/s)
The North Pacific Subtropical Countercurrent (STCC)	Cyclonic	3,601	16 ± 18 (4,211)	8.1 ± 5.6 (1.7, 39.0)	91 ± 23 (48,176)	24.98 ± 9.51 (9.08, 92.91)
	Anticyclonic	3,373	17 ± 20 (4,236)	7.3 ± 5.4 (1.9, 23.0)	94 ± 25 (45,179)	22.89 ± 6.11 (8.01, 57.74)

the newly updated one $A_i^{'}(t_i^{'})$ with the assumed form of the cubic polynomial curves to fit for each segment,

$$O_i = \alpha_{ij}\left(A_i(t_i^{'}) - A_i(t_i^{'})_j\right)^3 + \beta_{ij}(A_i(t_i^{'}) - A_i(t_i^{'})_j)^2 + \gamma_{ij}(A_i(t_i^{'}) - A_i(t_i^{'})_j)\Big) + \lambda_{ij} \quad (2)$$

where J refers to the number of the segments in total, j stands for the $j-th$ segment, and α , β , γ are respectively the coefficients of the polynomial curves, λ is a constant term, O represents the result of fitting, and the spacing of the amplitudes between the successive time points is:

$$h_{ij} = A_i(t_i^{'})_{j+1} - A_i(t_i^{'})_j \quad (3)$$

The interpolation of cubic spline constrains the function itself, as well as the first derivative and the second derivative, so the routine must ensure that $O_i\left(A_i(t_i^{'})\right)$, $O_i^{'}\left(A_i(t_i^{'})\right)$, and$O_i^{''}\left(A_i(t_i^{'})\right)$ are equal at the interior node points for adjacent segments. We substitute the coefficients by the new variable s

TABLE 2 | Statistics of 69 mesoscale eddies in study from 1993 to 2018.

Region	Character	Number	Average lifespan (week)	Average amplitude (cm)	Average radius (km)	Average speed (cm/s)
Izu-Ogasawara Ridge (130°E–170°E, 20°N–35°N)	Cyclonic	31	35 ± 7 (26, 51)	11 ± 6 (1, 39)	99 ± 28 (25,176)	25.8 ± 7.56 (7.47, 60.09)
	Anticyclonic	38	38 ± 8 (26, 52)	16 ± 11 (1, 23)	93 ± 30 (14,179)	32.54 ± 15.62 (5.83, 57.74)

TABLE 3 | Seasonal and polarized statistics of all mesoscale eddies in the ridge region (130°E–170°E, 20°N–35°N) from 1993 to 2018.

Region	Character	Attribute	Spring	Summer	Autumn	Winter
Izu-Ogasawara Ridge (139°E–143°E, 25°N–32°N)	Cyclonic	Amplitude (cm)	8.3	8.7	8.1	7.5
		Speed (cm/s)	23.17	23.98	23.23	22.11
		Radius (km)	84	83	84	84
	Anticyclonic	Amplitude (cm)	8.8	8.9	7.3	7.1
		Speed (cm/s)	25.25	24.92	21.08	20.54
		Radius (km)	82	86	85	81

for the polynomial's second derivative in each segment for simplification. For the $j-th$ segment, the governing equation is,

$$h_{i(j-1)}s_{i(j-1)} + \left(2h_{i(j-1)} + 2h_{ij}\right)s_{ij} + h_{ij}s_{i(j+1)} = 6\left(\frac{O_{i(j+1)} - O_{ij}}{h_{ij}} - \frac{O_{ij} - O_{i(j-1)}}{h_{i(j-1)}}\right) \quad (4)$$

In matrix form, the governing equation could be reduced to a tri-diagonal form.

$$\begin{bmatrix} 2(h_1+h_2) & h_2 & & \\ h_2 & 2(h_1+h_2) & \ddots & \\ & \ddots & \ddots & h_{J-2} \\ & & h_{J-2} & 2(h_{J-2}+h_{J-1}) \end{bmatrix}_i \begin{bmatrix} s_2 \\ s_j \\ \vdots \\ s_{J-1} \end{bmatrix}_i = 6\begin{bmatrix} \frac{O_3-O_2}{h_2} - \frac{O_2-O_1}{h_1} \\ \vdots \\ \frac{O_J-O_{J-1}}{h_{J-1}} - \frac{O_{J-1}-O_{J-2}}{h_{J-2}} \end{bmatrix}_i \quad (5)$$

where s_1 and s_J are zero for the natural spline boundary condition. In this way, the polynomial definition for each segment could be deduced as follows:

$$\begin{aligned} \alpha_{ij} &= \left(s_{i(j+1)} - s_{ij}\right)/6h_{ij} \\ \beta_{ij} &= s_{ij}/2 \\ \lambda_{ij} &= \frac{O_{i(j+1)}-O_{ij}}{h_{ij}} - \frac{2h_{ij}s_{ij}+h_{ij}s_{i(j+1)}}{6} \\ \lambda_{ij} &= O_{ij} \end{aligned} \quad (6)$$

Spatial Normalization

We further make a spatial normalization for those mesoscale eddies that partially travel the ridge, i.e., for any time $t_i^{'} \geq 0$, if all the geographical locations range within the ridge, $\left(x_i(t_i^{'}), y_i(t_i^{'})\right) \in r$, which means the mesoscale eddy does not completely go across the edge of the ridge region. Given the standard spatial distance $d_{\max}$ that the mesoscale eddy moves across the ridge with the maximum timespan $t_{*\max}^{'}$, we denote one spatial scaling index between the real distance d and the standard distance $d_{\max}$, and further modify their time intervals $t_{i\max}^{'}$ starting from the reference time 0 accordingly, to align the spatial distance for each given mesoscale eddy as follows,

$$t_{i\max}^{'} = \frac{d_i}{d_{\max}} \times \left(t_{*\max}^{'} + 1\right) \quad (7)$$

where $t_{i\max}^{'}$ could be regulated with the help of the above spatial scaling index, considering the maximum timespan duration over the ridge $t_{*\max}^{'}$, and the corresponding standard distance $d_{\max}$, i.e., the measurement from the arrival position to the departure position toward the entire region of the ridge. For the real distance d_i and the standard distance $d_{\max}$, the referred time domain are $t_i^{'} \in \left[0, t_{iend}^{'}\right]$ and $t_*^{'} \in \left(0, t_{*\max}^{'}\right)$. The cubic spline interpolation would hereby be applied again to normalize the initial amplitude time series $A_i(t_i^{'})$, $t_i^{'} \in \left[0, t_{iend}^{'}\right]$, into the newly updated one $A_i^{'}(t_i^{'})$, for each given mesoscale eddy that partially goes across the ridge.

Range Expansion

At last, we expand the domain of survival time in study to ranges out of the ridge region for all the mesoscale eddies, and follow the scaling principle derived from the ratio between the real and the corrected time duration in the ridge for all the mesoscale eddies completely or partially passing over it, by interpolation. For example, when we try to make the range

extension to the time domain [-30, 150] days after resetting the time series at the initialization stage, for the regulated amplitude $A_i^{'}(t_i^{'})$ of each given mesoscale eddy that completely goes across the ridge, we could similarly employ temporal regularity strategy by $t_i^{'} \times 30 \Big/ t_{*\max}^{'}$ and $t_i^{'} \times \left(150 - t_{*\max}^{'}\right) \Big/ t_{*\max}^{'}$ for the time domain before and after travailing into the ridge. For the mesoscale eddies that partially travel the ridge, we only need to extend the temporal regularity by $t_i^{'} \times 30 \Big/ t_{i\max}^{'}$.

Topographic Extraction

We collect the corresponding longitude and latitude along the way that mesoscale eddy trajectories pass, then apply the integrated bathymetric digital elevation of Izu-Ogasawara Ridge from ETOPO1, to discover the potential topographic effects on mesoscale eddies. Let the altitude at the geographical location $\left(x_i(t_i^{'}), y_i(t_i^{'})\right)$ along the trajectory of each given mesoscale eddy be $B_i(t_i)$. We follow the same principle mentioned above to extract the standardized time series of the amplitude, rotation speed, radius, altitude for the given mesoscale eddy, i.e., $A_i^{'}(t_i^{'})$, $S_i^{'}(t_i^{'})$, $R_i^{'}(t_i^{'})$, and $B_i^{'}(t_i^{'})$. We intuitively apply the mathematical statistics to all the mesoscale eddies in study by the aligned time averaging, for the above types of the potentially correlative time series, forming the mean curve of the altitude, the amplitude, rotation speed, radius of all the mesoscale eddy trajectories, $E_A(t^{'})$, $E_S(t^{'})$, $E_R(t^{'})$, and $E_B(t^{'})$, which sums up into the averaged variation curves to investigate how much of the possible impacts they might response to the topographic effects during the movement of the mesoscale eddies across the ridge.

Standardization

The zero-mean normalization will be further applied to all the above standardized mesoscale eddy trajectories in study with an extension over time, those completely spanning the ridge or those that demise within the ridge in comparison, in terms of the mesoscale eddy amplitude, rotation speed, radius and the altitude.

$$Z_A = \frac{E_A(t^{'}) - \mu_A}{\sigma_A}, Z_S = \frac{E_S(t^{'}) - \mu_S}{\sigma_S}, Z_R = \frac{E_R(t^{'}) - \mu_R}{\sigma_R}, \quad Z_B = \frac{E_B(t^{'}) - \mu_B}{\sigma_B} \tag{8}$$

where μ_A, μ_S, μ_R, and μ_B are respectively the mathematical expectation of $E_A(t^{'})$, $E_S(t^{'})$, $E_R(t^{'})$, $E_B(t^{'})$, σ_A, σ_S, σ_R, and σ_B are respectively the standard deviation of $E_A(t^{'})$, $E_S(t^{'})$, $E_R(t^{'})$, and $E_B(t^{'})$. Meanwhile, it is appropriate to identify those samples with the deviation that is highly probable to have an impact on assessment, and hereby to categorize as requiring quality control and eliminating singular values from time series of mesoscale eddy trajectories with and without passage.

Wavelet Coherence

Let $Y(t)$ be the mean curve of topography around the Izu-Ogasawara ridge. The cross-wavelet transform between the intrinsic attributes of the mesoscale eddies $X(t)$ and the topography can be represented as:

$$W_{XY}(a, \tau) = W_X(a, \tau)\, \bar{W}_Y(a, \tau) \tag{9}$$

where $W_X(a, \tau)$ and $W_Y(a, \tau)$ are respectively the wavelet transforms of $X(t)$ and $Y(t)$, and $\bar{W}_Y(a, \tau)$ indicates the complex conjugation of $W_Y(a, \tau)$, a is a scale factor, representing the period length of the wavelet, and τ is a shift factor, reflecting the shift in time. The covariance between the cross-wavelet power spectrum is then defined as follows:

$$|W_{XY}(a, \tau)| = |W_X(a, \tau)| \left|\bar{W}_Y(a, \tau)\right| \tag{10}$$

The correlation between the intrinsic attributes of the mesoscale eddies and the topography could be further measured via wavelet coherence (Rhif et al., 2019), as follows:

$$C_{XY}(a, \tau) = \frac{|s(W_{XY}(a, \tau))|}{\sqrt{s\left(|W_X(a, \tau)|^2\right) \cdot s\left(|W_Y(a, \tau)|^2\right)}} \tag{11}$$

where $s(W_{XY}(a, \tau))$ is the cross spectral density between $X(t)$ and $Y(t)$, with a smoothing operator s. The wavelet coherence coefficient reflects the synchronization similarity.

RESULTS

In our simulation experiment, mesoscale eddy trajectories of a life cycle more than 6 months, across the Izu-Ogasawara Ridge from east to west from 1993 to 2018, within the study area (130°E–170°E, 20°N–35°N), have been quantitatively investigated, to explore the potential spatio-temporal correlation with the topographic effects, with the average lifespan 37 weeks, the average amplitude 14 cm, the average radius 96 km, the average rotation speed 29.5 cm/s. **Figure 2** displays all the 69 mesoscale eddy trajectories, where 32 of them completely pass through the ridge from east to west, and 37 of them only partially leap over one side of the ridge and then demise inside it, with cyclonic eddies in black and anticyclonic eddies in white.

We screened out the solely mesoscale eddy which exhibited the maximum timespan duration within the ridge, 116 days, from all of the 32 mesoscale eddy trajectories that completely go through the ridge, and accordingly retrieved the distance 513.2 km that particular mesoscale eddy moves from the east to the west across the ridge. We standardized time series of the amplitude, rotation speed, radius for all the 32 mesoscale eddies in temporal regularity process.

We conducted the spatial normalization to all the 69 mesoscale eddy trajectories, to ensure all the trajectories of the same statistical significance across the aligned maximum timespan duration within the ridge, for not only those completely pass through the ridge but also those partially go across the ridge.

We further extended the time domain in study from [0, 116] to [-30, 150] days, i.e., to consider more the mesoscale eddy activities before and after reaching the edge of the ridge. We carried out the mathematical statistics to explore the potential spatio-temporal

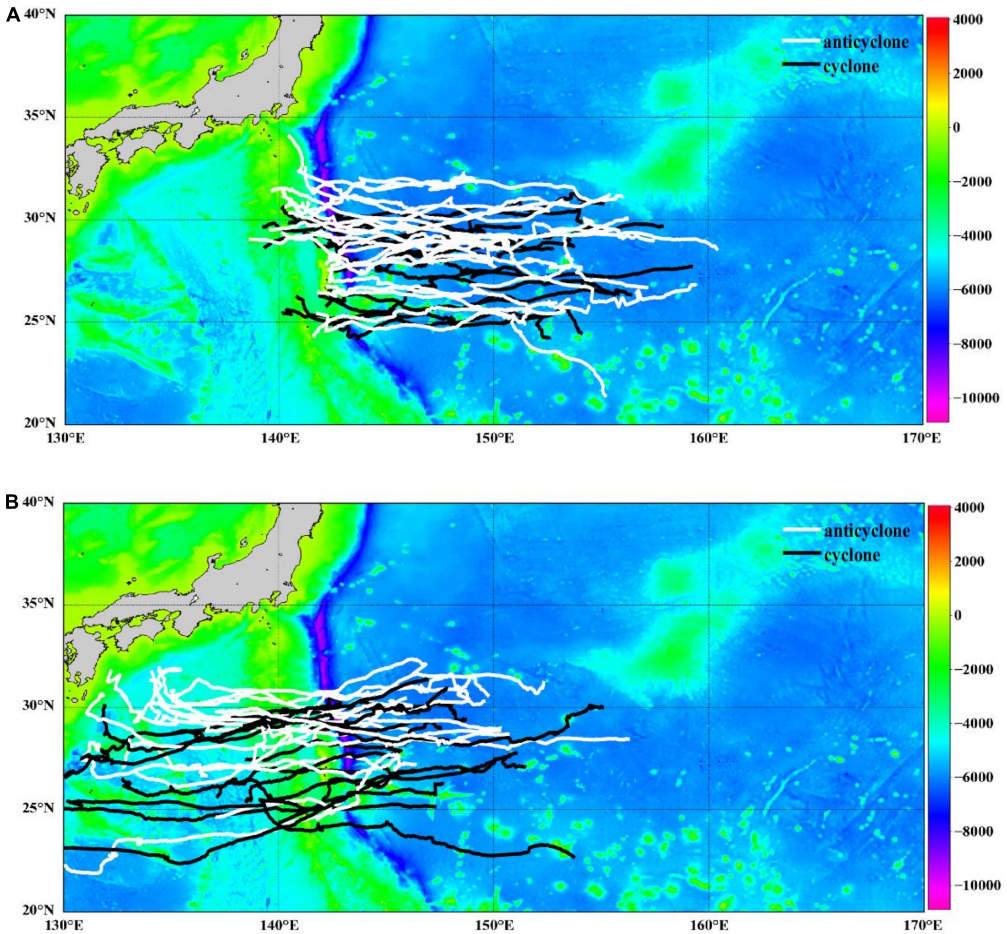


FIGURE 2 | Mesoscale eddy trajectories. **(A)** Mesoscale eddy trajectories that completely pass through the ridge, with cyclonic in black and anticyclonic in white. **(B)** Mesoscale eddy trajectories that partially travel across the ridge, with cyclonic in black and anticyclonic in white.

correspondence between the topography and the three attributes of mesoscale eddies, i.e., the amplitude, rotation speed, radius, by respectively synthesizing all the individual time series into a timely aligned averaged curve. The time series of the amplitude, rotation speed, radius for all the 69 mesoscale eddies after range extension have been individually illustrated in **Figure 3**, with the mean curve in red thick lines indicating the tendency as a whole.

The corresponding mean topography curve are averagely accumulated from the altitude at each geographical location over the ridge along all the mesoscale eddy trajectories referred. We collectively made a comparison between the mean curves of both the topography and three impact factors about the mesoscale eddies, i.e., the amplitude, rotation speed, radius, under the condition whether the mesoscale eddies go across the ridge, as is shown in **Figure 4**, with all the 69 mesoscale eddy trajectories in black, 32 completely passing in red, 37 partially passing in blue, in the top three rows, respectively, the gray shading indicating the transition before, during, and after the eddy-ridge interaction, and in the last row, the mean altitude has been listed, corresponding the eastward (143°E–146°E, 20°N–35°N), the interior (139°E–143°E, 20°N–35°N), the westward (136°E–139°E, 20°N–35°N) to the ridge. We also particularly plot the mean curves of the topography in black, mesoscale eddy amplitude in blue, radius in cyan, and rotation speed in red together after standardization with quality control, for all the 69 mesoscale eddy trajectories, as well as mesoscale eddy trajectories that completely pass through the ridge and partially travel across the ridge, regarding to the vortex polarity, with the vertical dashed lines indicating the beginning and end time points within the ridge region, as is shown in **Figure 5**. The zonal and meridional drifting speed have also been statistically observed along the extended range of the time domain daily on average in **Figure 6**, regarding to the vortex polarity and the passage of the ridge, with the statistics of all the 69 mesoscale eddy trajectories in black, the cyclonic eddies with passage in red, the anticyclonic eddies with passage in blue, the cyclonic eddies without passage in green, the anticyclonic eddies without passage in yellow.

We utilized the wavelet coherence to search for how frequently and how much extent the topography effects could exert possible impact to the variation in mesoscale eddy trajectories over time, in terms of the three relevant indices, i.e., the amplitude, rotation speed, radius, and described the lead-lag effects through the

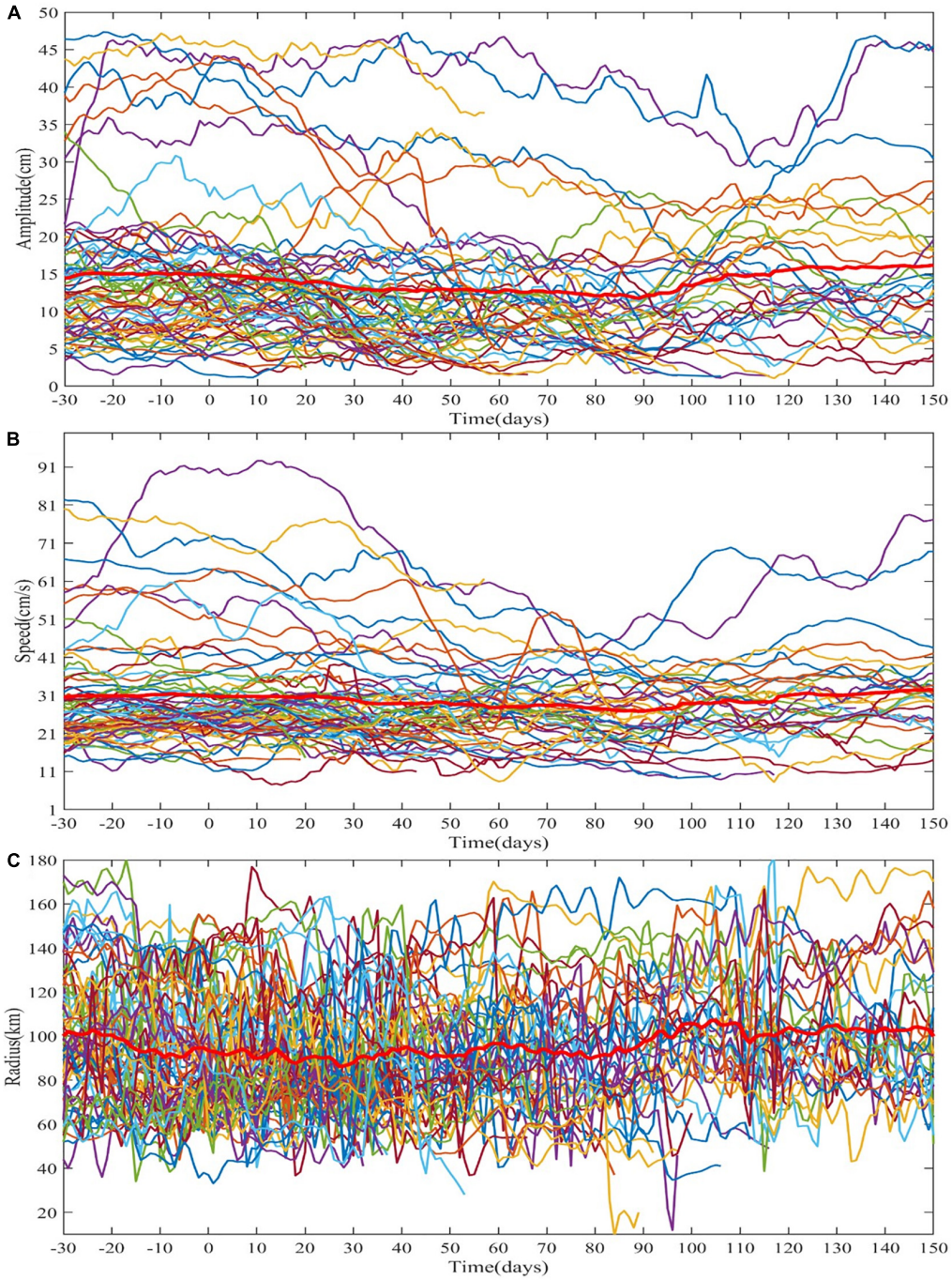


FIGURE 3 | The intrinsic attribute curves of 69 mesoscale eddies. **(A)** Amplitude curves of 69 mesoscale eddies after spatial normalization. **(B)** Rotation speed curves of 69 mesoscale eddies after spatial normalization. **(C)** Radius curves of 69 mesoscale eddies after spatial normalization.

wavelet phase difference. The wavelet coherence map between the mean altitude and the average amplitude, rotation speed, radius of mesoscale eddy trajectories over time are respectively shown in **Figure 6**, where wavelet correlation is affected by discontinuity, and the color bar corresponds to the relative intensity of each frequency. The stronger correlation tends to be in red, while weaker correlation in blue, from red to blue, the coherence spectrum values decrease in order. The cone of influence (COI), the region of the wavelet spectrum with edge effect, is denoted in thick black curve. Since the region in COI will be influenced by boundary distortion, reliable information could not be provided and should be removed. And 5 and 10% significance levels are respectively indicated in black thin line and black dashed line, with

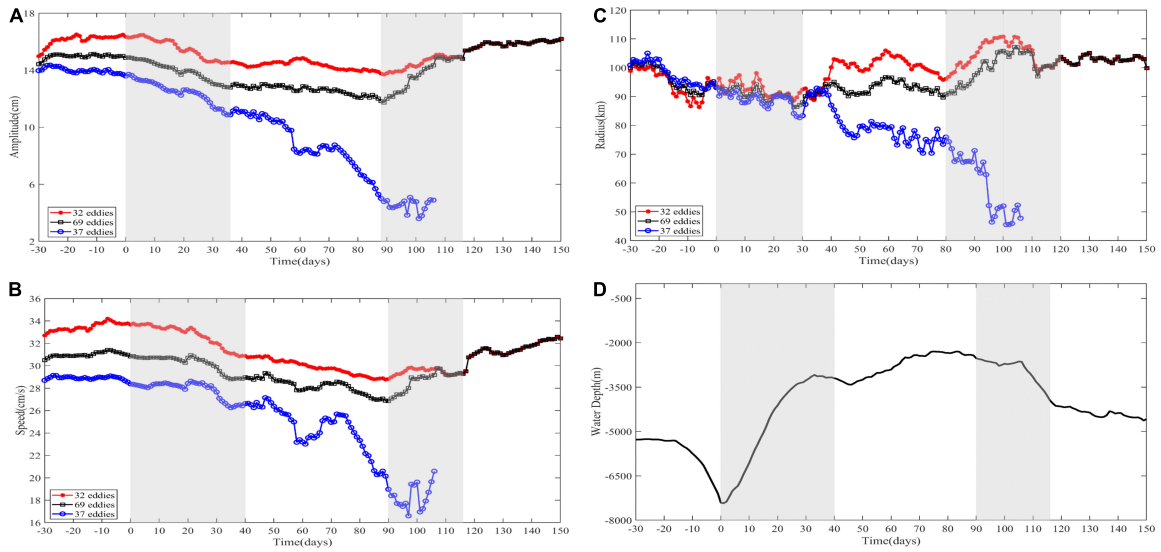


FIGURE 4 | Comparison of mesoscale eddy amplitude, rotation speed, radius with topography under conditions. **(A)** The mean curve of the amplitude for 69/32/37 mesoscale eddy trajectories. **(B)** The mean curve of the rotation speed for 69/32/37 mesoscale eddy trajectories. **(C)** The mean curve of the radius for 69/32/37 mesoscale eddy trajectories. **(D)** The mean curve of the topography for 69 mesoscale eddy trajectories (136°E–146°E, 20°N–35°N).

the significance value generated by Monte Carlo simulation (Aguiar-Conraria and Soares, 2011).

DISCUSSION

In this manuscript, we address to systematically and quantitatively investigate the topographic effects on evolution and propagation of mesoscale eddies across Izu-Ogasawara Ridge in Northwestern Pacific Ocean, in term of three relevant attributes, i.e., the amplitude, rotation speed, radius. We explore the potential topography effects through three phases of life cycle, i.e., before, during, and after the interaction with the ridge, by both the direct observation and the wavelet coherence map, from the perspective of the time-frequency domain. When the mesoscale eddies approach the ridge, and the steepness of the mean topography curve becomes higher, we observed from **Figures 5B,D** that most of the cyclonic eddies obviously tended to begin to decay or even demise, and the amplitude, rotation speed, radius of the mesoscale eddies decreased until some of the mesoscale eddies could move across the top of the ridge. The mean topography curve dropped, the three relevant attributes gradually increased, consistently coherent with the variation of the topography. On the contrary, it has been illustrated from **Figures 5C,E** that some of the anticyclonic eddies tended to intensify slightly and some didn't make a significant difference, regarding to the amplitude, rotation speed, radius, when encountering the upslope of the ridge. The general southwestward or northwestward mesoscale eddy trajectories tended to modify toward a more equatorward direction or oppositely by forcing the meridional components, as is shown in **Figure 6**, which is coherent to the previous studies (Beismann et al., 1999).

For topography vs. mesoscale eddy amplitude, it has been implied from the experimental results in **Figures 4A,D**, **5A–E** that there has been generally significant correlation between the two time series within the ridge region on the whole (139°E–143°E, 20°N–35°N). The mean amplitude value through life cycle of all 69 mesoscale eddies is 14.1 cm with the standard deviation of 9.5 cm, while the mean amplitude values of all the cyclonic eddies and all the anticyclonic eddies are respectively 11.6, 18.6 cm with passage and 10.4, 10.3 cm without passage over the ridge, implying that the mesoscale eddies that partially travel across the ridge often exhibit a tendency on the verge of demise, with relatively small amplitude values on a whole. The sudden changes with a range over 10% of the mean amplitude value in a 7-day period after standardization and quality control happen on the 92nd–97th days for 69 mesoscale eddies, the 13th–24th days and the 6th–13th, 28th–31st, 38th–46th, 59th–75th, 80th–89th days for cyclonic eddies with and without passage, the −30th to −22nd, 51st–55th, 104th–109th days and the 19th–26th, 52nd–58th, 75th–83rd, 88th–94th days for anticyclonic eddies with and without passage, along the time domain. In the 0–36th days, the mean topography curve gradually rises, with the depth rise from −7,409 to −3,134 m roughly, the mean amplitude curve gradually decreases from 14.3 to 9.5 cm for cyclonic eddies, with a relative decline of 33.6%, while the mean amplitude curve for anticyclonic eddies first rises from 12.7 to 13.6 cm, and then drops till 12.3 cm. To be exactly, the mean amplitude curve of cyclonic eddies falls from 15.6 ± 8.2 to 11.2 ± 5.4 cm with complete passage and from 13.0 ± 5.6 to 7.8 ± 4.9 cm with partial passage, on the contrary, the mean amplitude curve of anticyclonic eddies increases from 11.0 ± 2.5 to 13.4 ± 3.0 cm with complete passage, remains around 14.4 ± 9.8 cm for 21 days and then descends to 11.2 ± 11.1 cm with partial passage. When the mean topographic curve starts the downward trend, and

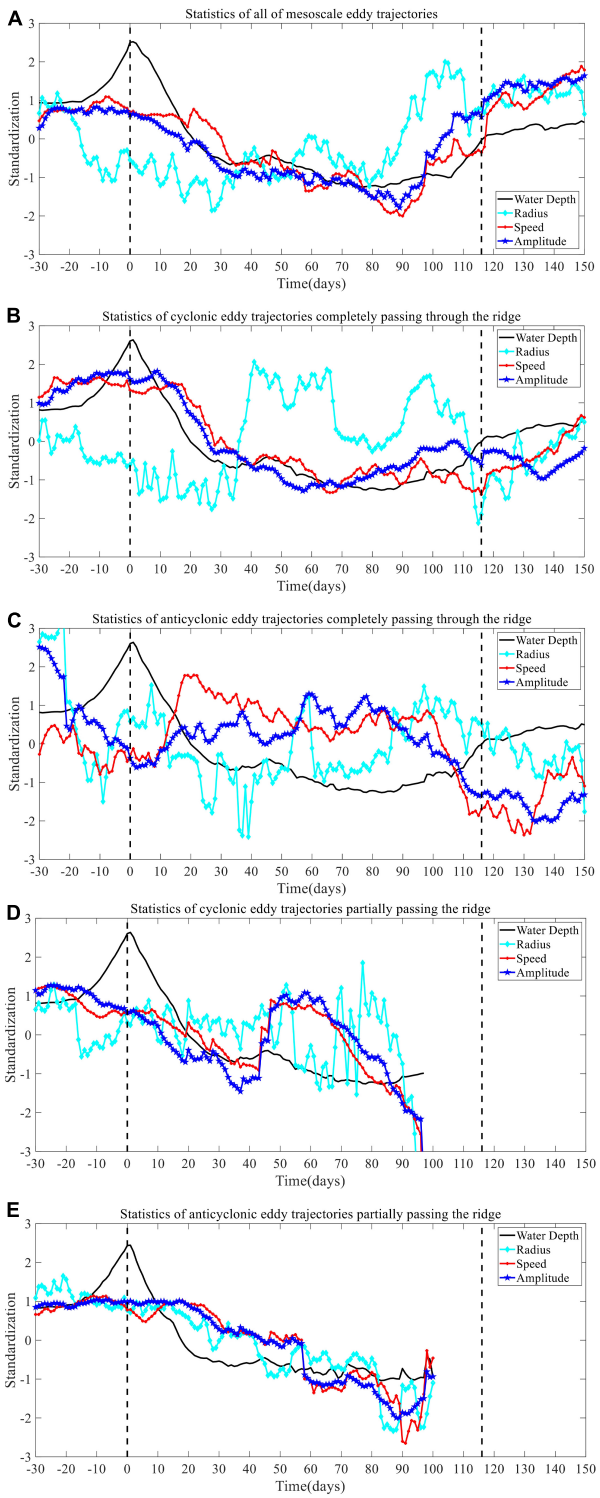


FIGURE 5 | Mean curve mapping of amplitude, rotation speed, radius, and topography after standardization. **(A)** The statistics in all of 69 mesoscale eddy trajectories. **(B)** The statistics in cyclonic eddy trajectories that completely pass through the ridge. **(C)** The statistics in anticyclonic eddy trajectories that completely pass through the ridge. **(D)** The statistics in cyclonic eddy trajectories that partially pass the ridge. **(E)** The statistics in anticyclonic eddy trajectories that partially pass the ridge.

accordingly the mean amplitude curve demonstrated the upward growth. Starting from the 88th day, the mean altitude value drops from roughly −2,410 to −3,792 m, and the corresponding mean amplitude has been recovered from 9.0 to 12.1 cm for cyclonic eddies, and from 7.9 to 9.0 cm for anticyclonic eddies. To be exactly, the mean amplitude curve of cyclonic eddies rises from 10.6 ± 4.9 to 12.1 ± 4.3 cm with complete passage over ridge, drops from 7.4 ± 3.7 to 1.4 ± 0 cm till the demise with no passage, in contrast, the mean amplitude curve of anticyclonic eddies decreases from 12.9 ± 3.0 to 9.0 ± 1.8 cm with complete passage, increases from 3.0 ± 0.9 to 7.0 ± 0 cm till dying with partial passage. After the range extension, it is found that around the region eastward to the ridge edge (143°E–146°E, 20°N–35°N), there tended to be a more stable variation of the mesoscale eddies, compared to that across the ridge region, with the average amplitude 14.8 cm for cyclonic eddies, and 14.0 cm for anticyclonic eddies. It seemed that no obvious evidence could be shown for the topography effects outside the east edge of the ridge. While the region westward to the ridge edge (136°E–139°E, 20°N–35°N) still involves the topography variation from roughly −3,954 to −4,583 m, the mean amplitude grew from 10.8 to 11.7 cm for cyclonic eddies, with an relative increase of 8.7%, and the mean amplitude decreased from 9.1 to 9.0 cm for anticyclonic eddies, with an relative drop of 1.1%. It has also been shown from the wavelet coherence map in **Figure 7A** that in the short-term frequency band of 2–5 days, the significantly strong correlation occurred on the 15–25th days and the 93–110th days, when the mesoscale eddies just reached the east edge of the ridge or approached to leave away from the ridge region. In the medium-term frequency band of 7–10 days, there has been significant correlation reflected on the 30–50th days. From the long-term frequency band of 16–25 days, the strong correlation lasted during the 80–110th days when the mesoscale eddies almost arrived at the west edge of the ridge.

For topography vs. rotation speed of mesoscale eddies, it has been demonstrated from the experimental results in **Figures 4B,D**, **5A–E** that around the region eastward to the ridge edge (143°E–146°E, 20°N–35°N), both the mean topography curve and the mean rotation speed behaved relatively stable for all of 69 mesoscale eddies, with the altitude value −5,279 and the rotation speed 31.0 cm/s. There have been no obvious evidences of the topography effects on the variation of the rotation speed before reaching the ridge. Within the region westward to the ridge edge (136°E–139°E, 20°N–35°N), the rotation speed for all of 69 mesoscale eddies increased from 29.3 to 32.4 cm/s on the whole, with a relative increase of 10.6%. However, it has been seen within the ridge region (139°E–143°E, 20°N–35°N) after standardization and quality control, in the 0–40th days, the mean rotation speed gradually decreases from 28.9 to 24.9 cm/s for cyclonic eddies with a relative reduction of 13.8%, and the mean rotation speed increases from 27.8 to 28.1 cm/s for anticyclonic eddies with a relative rise of 1.1%, when the mean topography curve gradually rises. To be exactly, the mean rotation speed curve of cyclonic eddies drops from 30.3 ± 10.7 to 26.3 ± 8.6 cm/s with complete passage and from 27.5 ± 6.9 to 23.5 ± 9.3 cm/s with partial passage, conversely, the mean rotation speed curve of anticyclonic eddies first grows from 25.9 ± 1.5 to 30.6 ± 0.7 cm/s

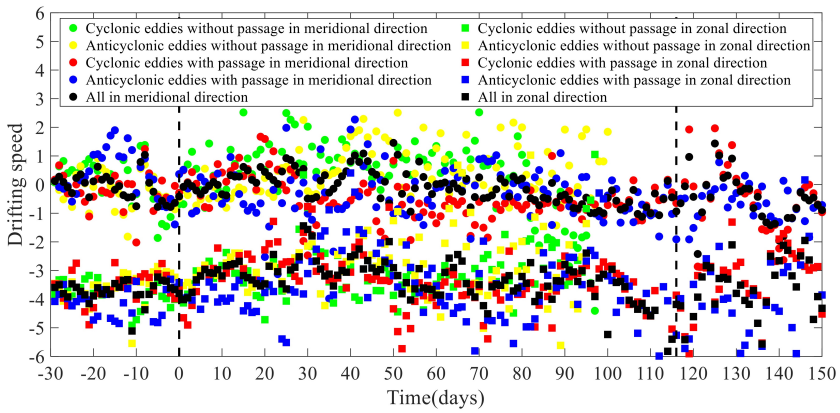


FIGURE 6 | Distribution of zonal and meridional drifting speed.

and then decreases to 28.9 ± 3.6 cm/s with complete passage, and first descends from 29.8 ± 14.2 to 28.5 ± 12.8 cm/s, rises to 30.7 ± 14.9 cm/s, and finally drops to 27.2 ± 12.3 cm/s with partial passage. Around the 90th day, when the altitude value is −2,427 m, the rotation speed decreases and reaches to the minimum 23.3 cm/s for cyclonic eddies, and gets to 22.0 cm/s for anticyclonic eddies, where the rotation speed values are respectively 25.4 ± 7.8, 28.1 ± 4.8 cm/s with passage and 21.2 ± 8.7, 16.0 ± 2.5 cm/s without passage over the ridge. In the 90–116th days, when the mean topographic curve drops from roughly −2,427 to −3,793 m, and the rotation speed gradually returned to 24.9 cm/s for cyclonic eddies, to 22.8 cm/s for anticyclonic eddies. To be exactly, the mean rotation speed curve of cyclonic eddies increases from 25.4 ± 7.8 to 26.7 ± 7.8 cm/s and then drops to 24.9 ± 7.9 cm/s with complete passage over ridge, falls from 21.2 ± 8.7 to 9.7 ± 0 cm/s till the demise with no passage, on the other hand, the mean rotation speed curve of anticyclonic eddies decreases from 28.1 ± 4.8 to 22.8 ± 0.5 cm/s with complete passage, increases from 16.0 ± 2.5 to 24.7 ± 0 cm/s till dying with partial passage. The mean rotation speed value through life cycle of all 69 mesoscale eddies is 29.8 cm/s with the standard deviation of 12.2 cm/s, while the mean rotation speed values of all the cyclonic eddies and all the anticyclonic eddies are respectively 27.1, 35.4 cm/s with passage and 25.6, 25.7 cm/s without passage over the ridge, implying that the mesoscale eddies that tend to approach dying over ridge mostly have smaller rotation speed values. Although we did not find out the dramatic conversions with a range over 10% of the mean rotation speed value in a 7-day period on the mean rotation speed curve for all 69 mesoscale eddies and cyclonic eddies with passage, they did appear on the 38th–43rd day for cyclonic eddies without passage, the 8th–13th day and the 52nd–57th, 84th–94th day for anticyclonic eddies with and without passage, along the time domain. The coherence between the topography and the rotation speed could be derived. It could also be seen from the wavelet coherence map in **Figure 7B** that in the short-term frequency band of 2–4 days, there have been significant correlation implied on the 0–10th days and the 95–105th days, when the mesoscale eddies first moved toward the east edge of the ridge and almost exceeded the west edge of the ridge. In the medium-term frequency band of 6–10 days, the strong correlation ranged between the 45–55th and 90–105th days.

For topography vs. mesoscale eddy radius, it has been shown from the experimental results in **Figure 4C,D**, **5A–E** that to the eastern side of the ridge (143°E–146°E, 20°N–35°N), the mean radius curve exhibited a downward tendency for all of 69 mesoscale eddies on the whole, from 100.0 to 93.4 km, with a relative reduction of 6.6%, when the mean topography remains stable to about −5,279 m. Out of the western side of the ridge (136°E–139°E, 20°N–35°N), the mean radius curve was relatively stable on the whole for all of 69 mesoscale eddies, between 100.0 and 105.1 km, when the mean topography curve actually drops from −3,646 to −4,582 m. Therefore, there was no direct evidence of the mutual coherence between the topography and the radius of mesoscale eddies outside the ridge. It has been seen within the ridge region (139°E–143°E, 20°N–35°N) after standardization and quality control, in the 0–30th days, the mean radius for cyclonic eddies first rises from 93.7 to 96.5 km, and then drops till 93.1 km, and the mean radius for anticyclonic eddies first decreases from 107.1 to 81.4 km, and then increases to 93.9 km, when the mean topography curve rises from −7,109 to −3,140 m. To be exactly, the mean radius curve of cyclonic eddies declines from 97.2 ± 17.6 to 88.1 ± 27.1 km and goes up to 95.3 ± 25.9 km with complete passage, and from 90.1 ± 19.9 km down to 82.6 ± 18.6 km and up to 91.0 ± 14.0 km with partial passage, whereas the mean radius curve of anticyclonic eddies grows from 120.0 ± 39.0 to 135.4 ± 19.4 km, drops to 80.9 ± 10.4 km, and recovers to 105.8 ± 35.0 km with complete passage, and descends from 94.2 ± 27.8 to 81.9 ± 20.8 km during a vibration process with partial passage over the ridge. In the 80–100th days, when the mean topography curve is about to decline, the mean radius curve begins to demonstrate an upward tendency for cyclonic eddies, from 98.4 km to nearly 113.5 km, with a relative increase of 15.3%, and from 88.6 km to nearly 94.7 km for anticyclonic eddies. To be exactly, the mean radius curve of cyclonic eddies climbs from 99.9 ± 23.0 to 113.5 ± 14.1 km with complete passage over ridge, dives from 96.8 ± 8.9 to 46 ± 0 km till the demise with no passage, meantime, the mean

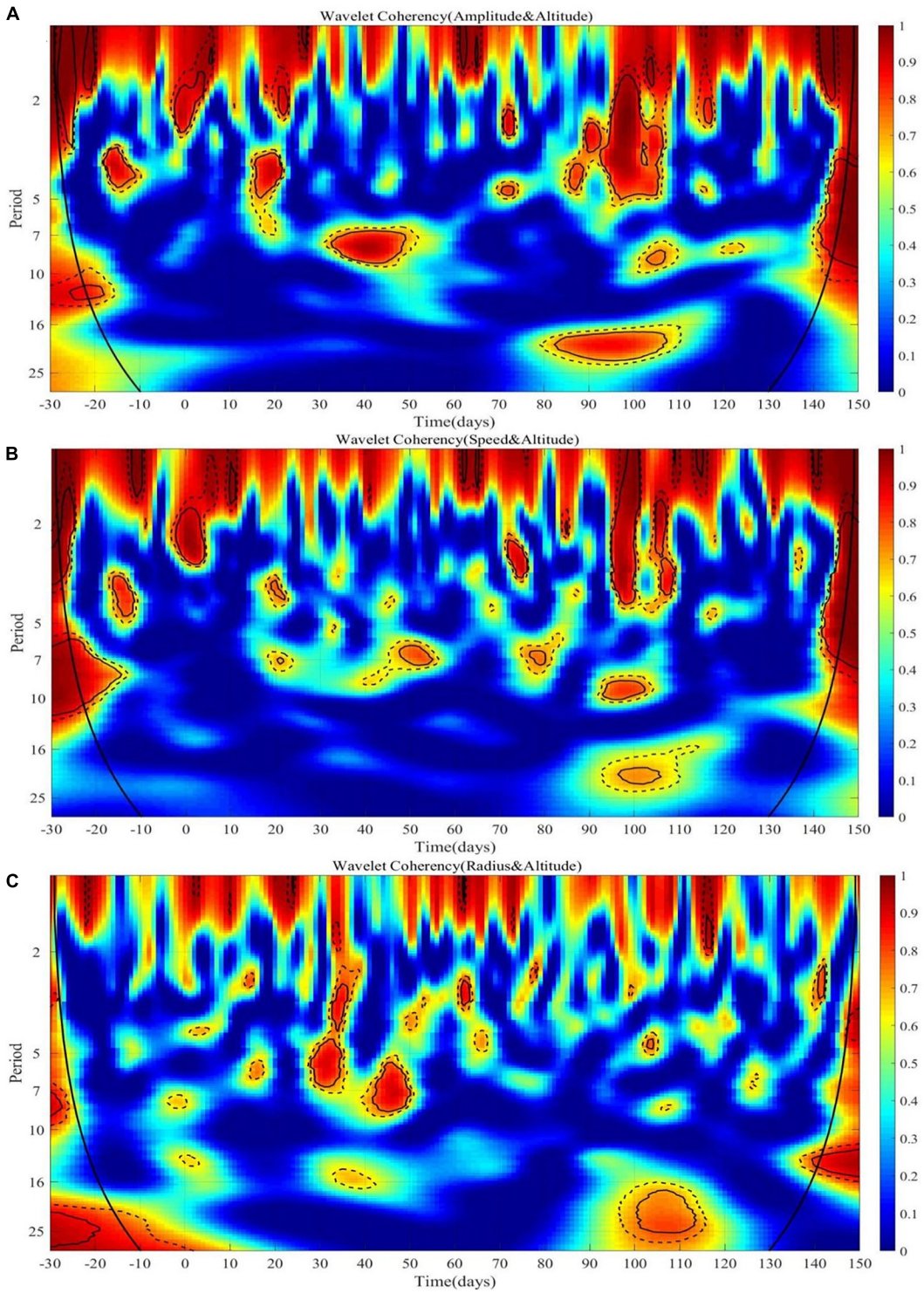


FIGURE 7 | Wavelet coherence mapping. **(A)** Wavelet coherence map between mesoscale eddy amplitude and topography. **(B)** Wavelet coherence map between rotation speed of mesoscale eddy and topography. **(C)** Wavelet coherence map between mesoscale eddy radius and topography.

radius curve of anticyclonic eddies also surges from 102.4 ± 27.0 to 134.9 ± 16.1 km and then decays to 124.4 ± 18.4 km with complete passage, sinks from 74.8 ± 16.0 to 45.7 ± 23.5 km till dying at 65.0 ± 0 km with partial passage. In the 100–120th days, when the topography curve drops from about −2,751 m to about −4,147 m, the mean radius curve of cyclonic eddies

declines from 113.5 ± 14.1 to 85.2 ± 25.5 km and then goes up to 95.3 ± 18.2 km with complete passage, while the mean radius curve of anticyclonic eddies descends from 124.4 ± 18.4 to 122.1 ± 17.6 km, rises to 128.4 ± 23.8 km, and drops to 106.3 ± 33.2 km in a repeatedly vibration process with complete passage over the ridge. Unlike the other two relevant attributes, there has been no obviously significant correlation directly from the eddy-ridge interaction on mesoscale eddy radius, but more implications on whether the mesoscale eddies could long survive tend to be revealed from its variation. The mean radius value through life cycle of all 69 mesoscale eddies is 96.4 km with the standard deviation of 27.4 km, while the mean radius values of all the cyclonic eddies and all the anticyclonic eddies are respectively 100.3, 98.2 km with passage and 84.7, 78.7 km without passage over the ridge, demonstrating that the mesoscale eddies that are only able to partially go over the ridge behave a blink of decay with the relatively small radius values. Although we did not find out the sharp alteration with a range over 10% of the mean radius value in a 7-day period on the mean radius curve for 69 mesoscale eddies, they did emerge on the 32nd–38th, 107th–111th days and the −21st to −17th, 84th–91st days for cyclonic eddies with and without passage, the −27th to −18th, −10th to −6th, 6th–11th, 18th–27th, 47th–59th days, and the 38th–43rd, 48th–52nd, and 77th–92nd days for anticyclonic eddies with and without passage, along the time domain. It has also been illustrated from the wavelet coherence map in **Figure 7C** that in the short-term frequency band of 4–7 days, there existed a correlation on the 28–35th days. In the medium-term frequency band of 7–10 days, the correlation was reflected on the 42–50th days. From the long-term frequency band of 16–25 days, the correlation lasted during the 100–115th days when the mesoscale eddies almost departed the west edge of the ridge.

For topography vs. daily drifting speed of mesoscale eddies, it has been indicated from the experimental results in **Figure 6** that there exists a generally significant correlation within the ridge region on the whole (139°E–143°E, 20°N–35°N) as well. The mean daily drifting speeds in both the meridional and zonal direction through life cycle of all 69 mesoscale eddies across the ridge are 0.2 km southward and 3.5 km westward, with the meridional drifting speed and the zonal drifting speed, respectively, 0.1, 0.1, 0.5 and 3.7, 3.4, 3.6 km before, during, and after the interaction with the ridge in the three phases of life cycle. To be exactly, the daily meridional drifting speed of the cyclonic eddies averagely rises from 0.2 km southward to 0.5 km southward, then drops to 0.4 km southward with passage, and increases from 0.1 km northward to 0.4 km northward till demise without passage, meanwhile, the daily meridional drifting speed of the anticyclonic eddies averagely switches from 0.4 km northward to 0.4 km southward, then surges to 0.6 km southward with passage, and shifts from 0.2 km southward to 0.5 km northward till demise without passage. Most of the mesoscale eddies proceeding toward northward disappear on the top of the ridge while those marching to southward mostly tend to survive over the ridge. Almost all of the mesoscale eddy trajectories has a tendency with the increased meridional drifting speeds when reaching the upslope of the ridge. The daily zonal drifting speed of the cyclonic eddies oriented to westward averagely declines from 3.6 to 3.4 km, then down to 3.3 km with passage, and decreases from 3.8 to 3.1 km till decease without passage, on the other hand, the daily zonal drifting speed of the anticyclonic eddies toward westward averagely descends from 4.4 to 3.7 km, then comes up to 4.3 km with passage, and reduces from 3.6 to 3.0 km till the end without passage. Almost all of the mesoscale eddy trajectories exhibits the decayed zonal drifting speeds when emerging the east edge of the ridge. The westward daily zonal drifting speed of anticyclonic eddies looks slightly higher than that of cyclonic eddies on average, owing to the higher drifting performances in anticyclonic eddies that completely pass across the ridge.

It has been concluded that, as is shown in **Figures 4–6**, regardless of the amplitude and rotation speed, 37 mesoscale eddies were more prone to be significantly influenced by the topography effects, most of which followed the decline tendency, starting from the arrival location until demise over the ridge. 32 mesoscale eddies that completely pass through the ridge exhibited significant negative correlation with the mean topography curve for cyclonic eddies during the evolution process that began at the east boundary of the ridge and lasted until passing over the ridge, while there occurred a little intensified for some of anticyclonic eddies. In total, it could be demonstrated in **Figure 5** that for all the 69 mesoscale eddies, except the radius, there exist a highly degree of agreement among the other three attributes after standardization, which further proves the existence of the topographical effects along the life cycle. The highly significant correlation between the topography and the amplitude of the mesoscale eddies could be particularly identified here. When the mesoscale eddies approach the ridge and the water depth gets shallower, the amplitude of the mesoscale eddies decrease for cyclonic eddies and increase a bit for some of anticyclonic eddies accordingly. Then, when the mesoscale eddies leave the ridge and the water depth increase, the amplitude of the mesoscale eddies recover almost to the level before the encounter. This is consistent with the expectation by PV conservation: The vorticity of the eddy and the area of the water column evolve in a compensational way. The vortex column squeezing will lead to the decreasing of the eddy vorticity for cyclonic eddies, and the stretching leads to intensification, conversely, for anticyclonic eddies. The coherence between the mesoscale eddies and the topography impairs the initial stability of the mesoscale eddy structure. This is due to the westward shift of the eddy by the planetary β effect which causes rising up to the shallower region over the eastern slope of the ridge. Namely, as the mesoscale eddies shift to a shallower region, the large frictional spin-down, which is inversely proportional to the lower layer thickness, and the generation of negative vorticity by the conservation of the potential vorticity are carried out. These two phenomena would weaken the eddy structure. Conversely, the westward shift of the cyclonic eddies has a small influence from the topographic effect of the bottom slope. Almost all of the mesoscale eddy trajectories manifests the rise in the daily meridional drifting speed and the decline in the daily zonal drifting speed, at the moment originally outstretching the ridge, no matter with the passage or not. In addition, out of 69 mesoscale eddies, only 32 of them could manage to go across the ridge. It has been revealed from our experimental results that the average horizontal scale of 32 mesoscale eddies

that completely pass over the ridge tended to exhibit a relatively large radius on the whole, compared to 37 mesoscale eddies that only partially pass the ridge, whereas 32 mesoscale eddies that completely move across the ridge provided visually dominant clues on both the amplitude and the rotation speed as well, showing the attribute divergence between the mesoscale eddies with passage and without passage over the ridge. We did see from the evaluation of the daily meridional drifting speed that the mesoscale eddies oriented to southward were more likely to survive over the ridge when proceeding toward the upslope of the ridge. It has been reported that the baroclinic or barotropic structure of mesoscale eddies could provide some explanation in the passage or not over the meridian ridge (Kamenkovich et al., 1996). We expect to extend our current work to involve more evidences on the vertical structure of mesoscale eddies over the meridional ridge in the future.

The above potential spatio-temporal dependency could quantitatively reveal how the relevant attributes of mesoscale eddies response to the topography variability over the Izu-Ogasawara Ridge, and inspire to identify the fundamental roles that possible dominantly correlative with the time-varying evolution process. Concerning the self-correlated properties inherently underlined in the propagation of mesoscale eddies, together with the time-frequency interaction imposed from the topography variability, we wish to make an initial attempt to capture the potentially patterns that mesoscale eddy trajectories might dynamically vary during the evolution process over the meridional ridge in the future, via deep learning, such as Long Short Term Memory (LSTM) architecture. Inspired by the real observation statistics in our research, the intrinsic and extrinsic attributes, such as the amplitude, rotation speed, radius, as well as the latitude and longitude at the geographical location, the zonal displacement and meridional displacement, the drifting velocity, and the variation of the bathymetric topography, could engage in deep learning selectively as the multivariate input for the predictive model, which allows to capture and retain long-term dependencies for the variation of mesoscale eddy trajectories, and meanwhile concludes the dominantly mutual linkage between the historic and current mesoscale eddy trajectories. The evaluation on whether and when the topographic attribute involves as input could substantially benefit for the prediction accuracy, will be further comprehensively estimated, when we try to intentionally feed with the topography variation as the additional inputs, rather than only the relevant indices of the amplitude, rotation speed, radius, drifting velocity, supposing the topography effects play relatively leading roles in determining the evolution and propagation of mesoscale eddy trajectories to be predicted.

The life cycle of ocean mesoscale eddies, including generation, decay, dissipation processes, is of great scientific value in understanding the propagation and evolution characteristics. Previous numerical experiments, carried out on the f-plane (Herbette et al., 2003), and on the β-plane (Herbette et al., 2005), with a seamount located in the bottom layers, below the isopycnes in which the surface eddy core resides, showed that surface eddies could get strongly eroded by deformation effects induced by bottom eddies, generated near the topography. Large intense eddies, known to have a long life time on the f-plane, no longer stayed coherent. At the same time, the presence of other poles made the exact path and structure of the eddy, as it impinged on the seamount, hypersensitive to details of the configuration, such as initial eddy position and numerical viscosity. Splitting, and subsequent erosion, of surface eddies became extremely sensitive to the initial conditions.

In contrast to previous studies, the present scheme provides more experimental evidences and hopeful solution with the promise of the end-to-end data-driven modeling of the topographic effects from the real observation data. The survival lifespan, generation date, dissipation date, geographical location, of the mesoscale eddy trajectories are all diverse. The differences of individual mesoscale eddies could easily affect the eddy-ridge interaction evaluation as a whole. So we propose to systematically examine the spatio-temporal dependency from mathematical statistics and wavelet coherence analysis, with the help of temporal regularity, spatial normalization and range expansion, to ensure the understanding of mesoscale eddy trajectories could be identically aligned to the geographical meaning from the standardization perspectives, and to synthesize all the time series statistically, regarding to the polarity. On the premise that all the time series are projected into the same domain, it will be more convenient to carry out the subsequent prediction with sufficient mesoscale eddy samples to be considered for training in deep learning, which could further prove the underlined correlation that the relevant attributes of mesoscale eddies might hold through their life cycle, including the amplitude, rotation speed, radius, and drifting velocity, in combination with the topography effects.

SUMMARY AND CONCLUSION

Ocean mesoscale eddies, involve significant amounts of water movements with irregular, discrete occurrences. Evidences have shown that the variation of mesoscale eddy trajectories would be to some extent attributed to the topographic effects when encountering with mid-ocean ridges, seamounts, bottom slopes, or over a variety of topography, while both the theoretical understanding and the survey from available observational data still remains incomplete. In this manuscript, we propose to make an attempt to explore whether and how the topographic effects of one meridional ridge, could exert considerable influences on the evolution and propagation process of mesoscale eddies through life cycle, directly from the perspectives of real observation statistics. We quantitatively investigate the known variability of ocean mesoscale eddies in the study area (130°E–170°E, 20°N–35°N) from 1993 to 2018, of a life cycle more than 6 months, on the basis of the trajectory atlas retrieved from AVISO satellite altimeter and the integrated bathymetric digital elevation sourced from ETOPO1, including 69 mesoscale eddy trajectories in total, with 32 completely traveling across the ridge, and 37 partially leap over the ridge, to discover the possible eddy-ridge interaction, by both observation statistics and wavelet coherence, with respect to the intrinsic attributes, namely, the amplitude, the rotation speed, the radius. A series of correlative steps, including the initialization, temporal regularity, spatial normalization, range

expansion, topographic extraction, wavelet coherence map, have been particularly designed along time-frequency domain to trace back mesoscale eddy trajectories in a variety of origins, location, lifespan, polarity, either completely or partially passing over the ridge, and to facilitate the standardization in statistics across three phases of their life cycle, i.e., before, during and after the interaction with the ridge. It has been revealed in our experiment that three intrinsic attributes of mesoscale eddies within 25 years, all demonstrated significant correlation with the variation of topographic relief over the ridge. The bathymetric topography reflects its relationships with mesoscale eddies through an integrated mode of a short-term, a mid-term, a long-term, respectively in the frequency band with a period of 2–7, 6–10, 16–25 days. We observed that most of the cyclonic eddies obviously tended to begin to decay or even demise, while on the contrary, some of the anticyclonic eddies preferred to intensify slightly, or making no significant difference when encountering the upslope of the ridge until climbing across the top, basically consistent with the expectation of potential vorticity (PV) conservation. The drifting velocity agreed with the tendency that the direction would be more probably modified toward equatorward or poloidally by forcing to the meridional component, with the zonal component reduced at the beginning. The mesoscale eddies with the passage over the ridge exhibited the relatively high average horizontal scales, as well as the relatively high average amplitude and rotation speed on the whole, compared to those with only partially passage. We expect to identify the relevant attributes dominantly correlative with the time-varying evolution process of mesoscale eddies during the eddy-ridge interaction, employ a deep learning architecture, to capture the predictability of mesoscale eddy trajectories on a global scale, when selectively feed the amplitude, rotation speed, radius, the latitude and longitude at the geographical location, the zonal displacement and meridional displacement, the drifting velocity, the variation of the bathymetric topography as the input variables for training. The developed scheme could integrate more evidences on how mesoscale eddies response to the topographic effects during their evolution and propagation process, and help provide opportunities to potentially identify the underlying dynamic patterns and mechanism that mesoscale eddies engage in ocean dynamics with the promise of the end-to-end data-driven solution.

DATA AVAILABILITY STATEMENT

AVISO satellite altimeter database can be available in the AVISO website (https://www.aviso.altimetry.fr/en/data.html). ETOPO1 database is available from NOAA website (http://www.ngdc.noaa.gov/mgg/global/).

AUTHOR CONTRIBUTIONS

RN has brought up the design of the study, together with analysis and interpretation of data, drafted the manuscript, and made final consent about the manuscript to be published. XG has performed the main experimental work, the analysis and interpretation of data, and drafted the manuscript. ZZ has brought up the conception of the study, provided guidance on data analysis and interpretation. MY, ZF, HX, and HY helped to draft the manuscript. QL helped the analysis and drafted the manuscript. HH and QY helped the statistics. CS has helped with the design and the experiments. LZ has assisted in the experimental work. BH has discussed and helped the experiments. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key R&D Program (2019YFC1408304), the Natural Science Foundation of People's Republic of China (42022041), the Natural Science Foundation of People's Republic of China (41876001), the National Key R&D Program (2016YFC0301400), the National High-Tech R&D 863 Program (2014AA093410), the Natural Science Foundation of People's Republic of China (31202036), the National Program of International S&T Cooperation (2015DFG32180), the National Science and Technology Pillar Program (2012BAD28B05), and the Natural Science Foundation of People's Republic of China (41376140).

ACKNOWLEDGMENTS

We would like to acknowledge the team members: Yu Cai for his guidance in data analysis.

REFERENCES

Adduce, C., and Cenedese, C. (2004). An experimental study of a mesoscale vortex colliding with topography of geometry in a rotating fluid. *J. Mar. Res.* 62, 611–638. doi: 10.1357/0022240042387583

Aguiar-Conraria, L., and Soares, M. J. (2011). Oil and the macroeconomy: using wavelets to analyze old issues. *Empir. Econ.* 40, 645–655. doi: 10.1007/s00181-010-0371-x

Beismann, J.-O., Käse, R. H., and Lutjeharms, J. R. E. (1999). On the influence of submarine ridges on translation and stability of *Agulhas rings*. *J. Geophys. Res.* 104, 7897–7906. doi: 10.1029/1998JC900127

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. Hoboken, NJ: John Wiley and Sons. doi: 10.2307/3008255

Braakmann-Folgmann, A., Roscher, R., Wenzel, S., Uebbing, B., and Kusche, J. (2017). Sea level anomaly prediction using recurrent neural networks. *arXiv* [Preprint]. arXiv:1710.07099

Chelton, D. B., Gaube, P., Schlax, M. G., Early, J. J., and Samelson, R. M. (2011). The influence of nonlinear mesoscale eddies on near-surface oceanic chlorophyll. *Science* 334, 328–332. doi: 10.1126/science.1208897

Daubechies, I. (1991). *Ten Lectures on Wavelets, CBMS-NSF Series in Applied Mathematics*. Philadelphia, PA: SIAM. doi: 10.1137/1.9781611970104

Ebuchi, N., and Hanawa, K. (2001). Trajectory of Mesoscale eddies in the Kuroshio recirculation region. *J. Oceanogr.* 57, 471–480. doi: 10.1023/A:1021293822277

Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1016/0364-0213(90)90002-E

Falcini, F., and Salusti, E. (2015). Friction and mixing effects on potential vorticity for bottom current crossing a marine strait: an application to the Sicily Channel

(central Mediterranean Sea). *Ocean Sci.* 11, 391–403. doi: 10.5194/os-11-391-2015

Freedman, D. A. (1981). Bootstrapping regression models. *Ann. Stat.* 9, 1218–1228. doi: 10.1214/aos/1176345638

Frenger, I., Gruber, N., Knutti, R., and Münnich, M. (2013). Imprint of Southern Ocean eddies on winds, clouds and rainfall. *Nat. Geosci.* 6, 608–612. doi: 10.1038/ngeo1863

Gangopadhyay, T., Tan, S. Y., Jiang, Z., Meng, R., and Sarkar, S. (2021). "Spatiotemporal attention for multivariate time series prediction and interpretation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Piscataway, NJ: IEEE), 3560–3564. doi: 10.1109/ICASSP39728.2021.9413914

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi: 10.1007/s10994-006-6226-1

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, (Red Hook, NY: Curran Associates, Inc), 2672–2680. doi: 10.1145/3422622

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2016). LSTM: a search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* 28, 2222–2232. doi: 10.1109/TNNLS.2016.2582924

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Piscataway, NJ: IEEE), 770–778. doi: 10.1109/CVPR.2016.90

Herbette, S., Morel, Y., and Arhan, M. (2003). Erosion of a surface vortex by a seamount. *J. Phys. Oceanogr.* 33, 1664–1679. doi: 10.1175/2382.1

Herbette, S., Morel, Y., and Arhan, M. (2005). Erosion of a surface vortex by a seamount on the β plane. *J. Phys. Oceanogr.* 35, 2012–2030. doi: 10.1175/JPO2809.1

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi: 10.1126/science.1127647

Hochreiter, S., and Schmidhuber, J. (1997). "LSTM can solve hard long time lag problems," in *Advances in Neural Information Processing Systems*, eds M. C. Mozer, M. I. Jordan, and T. Petsche (Cambridge, MA: MIT Press), 473–479. doi: 10.5555/2998981.2999048

Hu, J., and Zheng, W. (2020). Multistage attention network for multivariate time series prediction. *Neurocomputing* 383, 122–137. doi: 10.1016/j.neucom.2019.11.060

Hu, J., Zheng, Q., Sun, Z., and Tai, C. K. (2012). Penetration of nonlinear Rossby eddies into South China Sea evidenced by cruise data. *J. Geophys. Res. Ocean* 117:C03010. doi: 10.1029/2011JC007525

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Piscataway, NJ: IEEE), 4700–4708. doi: 10.1109/CVPR.2017.243

Ihara, C., Kagimoto, T., Masumoto, Y., and Yamagata, T. (2002). Eddy formation near the Izu-Ogasawara Ridge and its link with seasonal adjustment of the subtropical gyre in the Pacific. *J. Korean Soc. Oceanogr.* 37, 134–143.

Jacob, J. P., Chassignet, E. P., and Dewar, W. K. (2002). Influence of topography on the propagation of isolated eddies. *J. Phys. Oceanogr.* 32, 2848–2869.

Jing, Z., Wu, L., Li, L., Liu, C., Liang, X., Chen, Z., et al. (2011). Turbulent diapycnal mixing in the subtropical northwestern Pacific: spatial-seasonal variations and role of eddies. *J. Geophys. Res. Oceans* 116:C10028. doi: 10.1029/2011JC007142

Kalchbrenner, N., Danihelka, I., and Graves, A. (2015). Grid long short-term memory. *arXiv* [preprint]. arXiv:1507.01526 doi: 10.3390/biology9120441

Kamenkovich, V. M., Leonov, Y. P., Nechaev, D. A., Byrne, D. A., and Gordon, A. L. (1996). On the influence of bottom topography on the Agulhas eddy. *J. Phys. Oceanogr.* 26, 892–912.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105. doi: 10.1145/3065386

Labat, D. (2005). Recent advances in wavelet analyses: Part 1. A review of concepts. *J. Hydrol.* 314, 275–288. doi: 10.1016/j.jhydrol.2005.04.003

Li, W., Qin, B., and Zhang, Y. (2015). Multi-temporal scale characteristics of algae biomass and selected environmental parameters based on wavelet analysis in Lake Taihu, China. *Hydrobiologia* 747, 189–199. doi: 10.1007/s10750-014-2135-7

Li, Y., and Wang, F. (2012). Spreading and salinity change of North Pacific Tropical Water in the Philippine Sea. *J. Oceanogr.* 68, 439–452. doi: 10.1007/s10872-012-0110-3

Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *arXiv* [preprint]. arXiv:1506.00019

Maraun, D., and Kurths, J. (2004). Cross wavelet analysis: significance testing and pitfalls. *Nonlinear Process. Geophys.* 11, 505–514. doi: 10.5194/npg-11-505-2004

Meyers, S. D., Kelly, B. G., and O'Brien, J. J. (1993). An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of Yanai waves. *Monthly Weather Rev.* 121, 2858–2866.

Morrow, R., Birol, F., Griffin, D., and Sudre, J. (2004). Divergent pathways of cyclonic and anti-cyclonic ocean eddies. *Geophys. Res. Lett.* 31:L24311. doi: 10.1029/2004GL020974

Nian, R., Xu, Y., Yuan, Q., Feng, C., and Lendasse, A. (2021a). Quantifying time-frequency co-movement impact of COVID-19 on US and China stock market toward investor sentiment index. *Front. Public Health* 9:727047. doi: 10.3389/fpubh.2021.727047

Nian, R., Yuan, Q., He, H., Geng, X., Su, C., He, B., et al. (2021b). The identification and prediction in abundance variation of Atlantic cod via long short-term memory with periodicity, time–frequency co-movement, and lead-lag effect across sea surface temperature, sea surface salinity, catches, and prey biomass from 1919 to 2016. *Front. Mar. Sci.* 8:629. doi: 10.3389/fmars.2021.665716

Nian, R., Cai, Y., Zhang, Z., He, H., Wu, J., Yuan, Q., et al. (2021c). The identification and prediction of Mesoscale eddy variation via memory in memory with scheduled sampling for sea level anomaly. *Front. Mar. Sci.* 8:753942. doi: 10.3389/fmars.2021.753942

Nycander, J., and Lacasce, J. H. (2004). Stable and unstable vortices attached to seamounts. *J. Fluid Mech.* 507, 71–94. doi: 10.1017/S0022112004008730

Ohara, Y., Tokuyama, H., and Stern, R. J. (2007). Thematic section: geology and geophysics of the Philippine Sea and adjacent areas in the Pacific Ocean. *Island Arc* 16, 319–321. doi: 10.1111/j.1440-1738.2007.00596.x

Pineda-Sanchez, M., Riera-Guasp, M., Perez-Cruz, J., and Puche-Panadero, R. (2013). Transient motor current signature analysis via modulus of the continuous complex wavelet: a pattern approach. *Energy Conv. Manag.* 73, 26–36. doi: 10.1016/j.enconman.2013.04.002

Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., and Cottrell, G. (2017). "A dual-stage attention-based recurrent neural network for time series prediction," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17)*, (Palo Alto, CA: AAAI Press), 2627–2633. doi: 10.24963/ijcai.2017/366

Qiu, B., and Lukas, R. (1996). Seasonal and interannual variability of the North Equatorial Current, the Mindanao Current, and the Kuroshio along the Pacific western boundary. *J. Geophys. Res. Ocean* 101, 12315–12330. doi: 10.1029/95JC03204

Qiu, B., and Miao, W. (2000). Kuroshio path variations South of Japan: bimodality as a self-sustained internal oscillation. *J. Phys. Oceanogr.* 30, 2124–2137.

Reboredo, J. C., Rivera-Castro, M. A., and Ugolini, A. (2017). Wavelet-based test of co-movement and causality between oil and renewable energy stock prices. *Energy Econ.* 61, 241–252. doi: 10.1016/j.eneco.2016.10.015

Rhif, M., Ben Abbes, A., Farah, I. R., Martínez, B., and Sang, Y. (2019). Wavelet transform application for/in non-stationary time-series analysis: a review. *Appl. Sci.* 9:1345. doi: 10.3390/app9071345

Rudnick, D., Jan, S., Centurioni, L., Lee, C. M., Lien, R. C., Wang, J., et al. (2011). Seasonal and mesoscale variability of the Kuroshio near its origin. *Oceanography* 24, 52–63. doi: 10.5670/oceanog.2011.94

Sapankevych, N. I., and Sankar, R. (2009). Time series prediction using support vector machines: a survey. *IEEE Comput. Intell. Mag.* 4, 24–38. doi: 10.1109/MCI.2009.932254

Sarangi, R. K. (2012). Observation of oceanic eddy in the northeastern Arabian Sea using Multisensor remote sensing data. *Int. J. Oceanogr.* 2012:531982. doi: 10.1080/01490419.2011.637848

Sekine, Y. (1989). Topographic effect of a marine ridge on the spin-down of a cyclonic eddy. *J. Oceanogr. Soc. Japan* 45, 190–203. doi: 10.1007/BF02123463

Shahbaz, M., Loganathan, N., Zeshan, M., and Zaman, K. (2015). Does renewable energy consumption add in economic growth? An application of auto-regressive distributed lag model in Pakistan. *Renew. Sustain. Energy Rev.* 44, 576–585. doi: 10.1016/j.rser.2015.01.017

Shi, X., and Yeung, D. (2018). Machine learning for spatiotemporal sequence forecasting: a survey. *arXiv* [Preprint]. arXiv:1808.06865

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., and Woo, W. C. (2015). Convolutional LSTM network: a machine learning approach for precipitation nowcasting. *arXiv* [preprint]. arXiv:1506.04214 doi: 10.5555/2969239.2969329

Shu, Y., Chen, J., Li, S., Wang, Q., Yu, J., and Wang, D. (2018). Field-observation for an anticyclonic mesoscale eddy consisted of twelve gliders and sixty-two expendable probes in the northern South China Sea during summer 2017. *Sci. China Earth Sci.* 61, 451–458. doi: 10.1007/s11430-018-9239-0

Steel, E. A., and Lange, I. A. (2007). Using wavelet analysis to detect changes in water temperature regimes at multiple scales: effects of multi-purpose dams in the Willamette River basin. *River Res. Appl.* 23, 351–359. doi: 10.1002/rra.985

Su, C. W., Cai, X. Y., Qin, M., Tao, R., and Umar, M. (2021a). Can bank credit withstand falling house price in China? *Int. Rev. Econ. Finance* 71, 257–267. doi: 10.1016/j.iref.2020.09.013

Su, C. W., Huang, S. W., Qin, M., and Umar, M. (2021b). Does crude oil pricestimulate economic policy uncertainty in BRICS? *PacificBasin Finance J.* 66:101519. doi: 10.1016/j.pacfin.2021.101519

Su, C. W., Qin, M., Rizvi, S. K. A., and Umar, M. (2021c). Bank competition in China:a blessing or a curse for financial system? *Econ. Res. Ekon. IstraŽivanja* 34, 1244–1264. doi: 10.1080/1331677X.2020.1820361

Su, C. W., Qin, M., Tao, R., and Umar, M. (2020). Does oil price really matter for the wage arrears in Russia? *Energy* 208:118350. doi: 10.1016/j.energy.2020.118350

Su, C. W., Qin, M., Zhang, X. L., Tao, R., and Umar, M. (2021d). Should Bitcoin be held under the US partisan conflict? *Technol. Econ. Dev. Econ.* 27, 511–529. doi: 10.3846/tede.2021.14058

Su, C. W., Wang, X. Q., Tao, R., and Oana-Ramona, L. (2019b). Do oil prices drive agricultural commodity prices? Further evidence in a global bio-energy context. *Energy* 172, 691–701. doi: 10.1016/j.energy.2019.02.028

Su, C. W., Khan, K., Tao, R., and Nicoleta-Claudia, M. (2019a). Does geopolitical risk strengthen or depress oil prices and financial liquidity? Evidence from Saudi Arabia. *Energy* 187:116003. doi: 10.1016/j.energy.2019.116003

Sutyrin, G., Carton, X., and Perrot, X. (2008). Integrable motion of a vortex dipole in an axisymmetric flow. *Phys. Lett. A* 372, 5452–5457. doi: 10.1016/j.physleta.2008.06.038

Sutyrin, G., Herbette, S., and Carton, X. (2011). Deformation and splitting of baroclinic eddies encountering a tall seamount. *Geophys. Astrophys. Fluid Dyn.* 105, 478–505. doi: 10.1080/03091929.2011.566566

Tao, R., Su, C. W., Xiao, Y., Dai, K., and Khalid, F. (2021). Robo advisors,algorithmic trading and investment management: wonders of fourth industrial revolution in financial markets. *Technol. Forecast. Soc. Change* 163:120421. doi: 10.1016/j.techfore.2020.120421

Thierry, V., and Morel, Y. (1999). Influence of a strong bottom slope on the evolution of a surface-intensified vortex. *J. Phys. Oceanogr.* 29, 911–924.

Torrence, C., and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 79, 61–78.

Torres, H. S., and Gomez-Valdes, J. (2017). Erosion of a California Undercurrent eddy by bottom topography. *J. Geophys. Res. Oceans* 122, 3715–3735. doi: 10.1002/2016JC011870

Trodahl, M., and Isachsen, P. E. (2018). Topographic influence on baroclinic instability and the Mesoscale eddy field in the Northern North Atlantic Ocean and the Nordic Seas. *J. Phys. Oceanogr.* 48, 2593–2607. doi: 10.1175/JPO-D-17-0220.1

Wan, R., Mei, S., Wang, J., Liu, M., and Yang, F. (2019). Multivariate temporal convolutional network: a deep neural networks approach for multivariate time series forecasting. *Electronics* 8:876. doi: 10.3390/electronics8080876

Wang, Y., Wu, H., Zhang, J., Gao, Z., Wang, J., Yu, P. S., et al. (2021). PredRNN: a recurrent neural network for spatiotemporal predictive learning. *arXiv* [preprint]. arXiv:2103.09504.

Yang, G., Wang, F., Li, Y., and Lin, P. (2013). Mesoscale eddies in the northwestern subtropical Pacific Ocean: statistical characteristics and three-dimensional structures. *J. Geophys. Res. Oceans* 118, 1906–1925. doi: 10.1002/jgrc.20164

Yuan, D., and Wang, Z. (2011). Hysteresis and dynamics of a western boundary current flowing by a gap forced by impingement of mesoscale eddies. *J. Phys. Oceanogr.* 41, 878–888. doi: 10.1175/2010JPO4489.1

Zhang, Q., Wang, H., Dong, J., Zhong, G., and Sun, X. (2017). Prediction of sea surface temperature using long short-term memory. *IEEE Geosci. Remote Sens. Lett.* 14, 1745–1749. doi: 10.1109/LGRS.2017.2733548

Zhang, W. Z., Ni, Q., and Xue, H. (2018). Composite eddy structures on both sides of the Luzon Strait and influence factors. *Ocean Dyn.* 68, 1527–1541. doi: 10.1007/s10236-018-1207-z

Zhang, Y., Zhang, Z., Chen, D., Qiu, B., and Wang, W. (2020). Strengthening of the Kuroshio current by intensifying tropical cyclones. *Science* 368, 988–993. doi: 10.1126/science.aax5758

Zhang, Z., Qiu, B., Klein, P., and Travis, S. (2019). The influence of geostrophic strain on oceanic ageostrophic motion and surface chlorophyll. *Nat. Commun.* 10:2838. doi: 10.1038/s41467-019-10883-w

Zhang, Z., Wang, W., and Qiu, B. (2014). Oceanic mass transport by mesoscale eddies. *Science* 345, 322–324. doi: 10.1126/science.1252418

Zhang, Z., Zhang, Y., Wang, W., and Huang, R. X. (2013). Universal structure of mesoscale eddies in the ocean. *Geophys. Res. Lett.* 40, 3677–3681. doi: 10.1002/grl.50736

Conflict of Interest: XG was employed by the company JD.com, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership