Forecasting of Mesoscale Eddies in the Kuroshio Extension Based on Temporal Modes-Enhanced Neural Network

Haitong Wang, Yunxia Guo, Yuan Kong and Yong Fang *

Abstract: Mesoscale eddies are a common occurrence in the Kuroshio Extension (KE) that have a major impact on the levels of salinity and heat transport in the Northwest Pacific, the strength of the Kuroshio jet, and the fluctuations of the Kuroshio’s trajectory. In this study, a purely data-driven machine learning model, Temporal Modes-Enhanced Neural Network (TMENN), is proposed to forecast the spatiotemporal variation of mesoscale eddies based on daily sea surface height (SSH) data over a 20-year period (2000–2019) in the Kuroshio Extension. To reduce computational costs and facilitate faster forecasting, raw SSH data are decomposed into spatial modes and temporal modes (principal components, PCs) by empirical orthogonal functions (EOF) analysis, and the first 117 PCs (a total of 8384 PCs), wherein the cumulative variance contribution rate reaches 95%, are selected solely as the predictors of TMENN to train and forecast. Forecasting reconstruction results show that the model can reliably forecast the evolution of the eddy in the KE for about 30 days. Additionally, three classical mesoscale eddy processes are selected to verify the accuracy of the model, namely cold eddy attachment, warm eddy shedding, and attachment, and the results indicate that the model can well capture the evolution process of mesoscale eddies.

Keywords: forecasting; TMENN; mesoscale eddy; kuroshio extension

1. Introduction

The Kuroshio is a warm oceanic current that begins in the equatorial part of the Pacific Ocean. It divides into two branches near 35° N during the process of flowing northeastward along the Japanese Islands, and the main part, which turns eastward until 160°E, is called the Kuroshio Extension (KE). The KE region is a pivotal area for mid-latitude ocean–atmosphere interactions [1,2]. Mesoscale eddies are a common and highly energetic feature in this region [3], especially near the KE main axis. In recent years, there has been a great deal of research on the KE [4–7]. Lots of research has demonstrated the substantial impact of mesoscale eddies on multiple facets, encompassing the fluctuations of the Kuroshio trajectory in the KE region [8,9], the spatiotemporal alternative of the Kuroshio jet [10], the zonal transport of salinity and heat in the North Pacific Ocean [11], the storm axis and subtropical modal water changes in the North Pacific [12], and the local hydrological environment [13]. The KE region represents a complex and dynamic oceanic system, making it of significant scientific importance and practical relevance to develop accurate forecasts for the evolution of mesoscale eddies within this region. In addition, its study involves the intersection of several disciplinary fields. Therefore, it is essential to strengthen interdisciplinary cooperation and innovation, and develop more effective technical tools for the prediction research of mesoscale eddies.

An Artificial Neural Network (ANN) simulates information transmission and processing between neurons in the human brain, which enables it to perform intricate tasks and recognize patterns and shows clear advantages in many areas, such as identifying [14,15], predicting [16–21], and clustering [22]. In the field of physical oceanography, ANN has...
found widespread application [23–25], leading to numerous accomplishments. Pozzi [26] verified the effectiveness of ANN as an alternative to traditional methods in oceanography. Bhaskaran [27] applied ANN to derive temperature and salinity fields using monthly long-term means of temperature (T) and salinity (S) data for the Indian Ocean, with results indicating superior performance compared to traditional interpolation methods. Shao [28] used EOF, ANN, and other methods to forecast and analyze multivariate variables for instance sea surface height (SSH) and sea surface temperature (SST) in the South China Sea, and found that ANN was also effective in multivariate prediction. Overall, the ANN model is an effective tool for forecasting the physical elements of the ocean.

It is noteworthy that while ANN models have demonstrated high accuracy in predicting ocean physical elements, there still exist certain limitations, as highlighted by [29,30]. To address these limitations, particularly the challenges associated with hyperparameter selection, we introduce Empirical Orthogonal Function (EOF) analysis as an adjunct to the ANN framework for constructing a forecasting model for mesoscale eddies in the KE region. EOF analysis is adept at extracting the primary time-dependent principal components from raw data, representing the most significant features of the dataset. Incorporating these modes as inputs into the ANN facilitates quicker convergence and the acquisition of crucial data insights, potentially simplifying the task of identifying the optimal hyperparameter combinations and enhancing forecast accuracy.

Achieving long-term prediction skills in ocean models is a challenging goal. In this regard, researchers have employed various methods to enhance prediction accuracy. For example, Shao [31] utilized EOF and deep learning techniques to successfully predict sea surface height anomalies (SSHA) and SST in the South China Sea for a consecutive 10-day period. Additionally, they employed machine learning methods to extend the prediction horizon to 30 days [28]. Zeng [23] employed nonlinear autoregressive networks capable of forecasting backwards for up to 4 weeks, and even 5–6 weeks, for eddy dynamics changes. This demonstrates the feasibility and potential of long-term prediction. Furthermore, Wang [24] applied Long Short-Term Memory (LSTM) networks for prediction, extending the forecasting horizon to predict vortex dynamics (LCS) and SSH up to 9 weeks and even reaching 12 weeks. These research findings indicate that achieving long-term predictions in ocean models is feasible. Not only does it hold promise for improving prediction skills, but it also contributes to a deeper understanding of the evolution of ocean systems.

In this study, we introduce a novel approach known as the Temporal Modes-Enhanced Neural Network (TMENN) model, which combines EOF analysis and an ANN model. Our aim is to develop a purely data-driven ocean forecasting model capable of forecasting mesoscale eddies in the KE region up to 30 days in advance. By first performing EOF analysis on the original data matrix to produce spatial modes (EOFs) and their related principal components (PCs), which are then utilized as input features for the ANN model, the technique reduces computational complexity.

The following parts of the essay are structured as follows. The methods used for this study and a detailed explanation of the satellite data are both included in Section 2. The TMENN model’s findings are shown in Section 3. Finally, Section 4 offers a conclusion summarizing the key insights from this research.

2. Data and Methods
2.1. Data Sources and Preprocessing

The data used in this paper are gridded, height-based SSH data, which are from Copernicus Marine Service Center (https://marine.copernicus.eu/, accessed on 17 March 2023) and are processed by the Data Unification and Altimeter Combination System (DUACS) multi-task altimeter data processing system from all altimeter missions. The SSH data adopted in this research cover the KE region (130–160° E, 25–45° N, Figure 1) from 1 January 2000 to 31 December 2019. The spatial precision is 1/4°, and the temporal precision is 1 day. Therefore, the dimensions of the raw SSH data matrix are 120 × 80 × 8384. We conduct dimensionality reduction on the raw SSH data. This process involves expanding
the original three-dimensional data matrix through the addition of columns (or rows), subsequently reconfiguring it into a two-dimensional matrix with dimensions of 9600 rows, representing spatial data points, and 8384 columns, symbolizing time series.

![Figure 1. Study area’s topographic and current field map (The base map’s color indicates the water depth; the monthly average flow field is represented by arrows).](image)

Simultaneously, we acquired topographic data from the National Oceanic and Atmospheric Administration (NOAA) National Geophysical Data Center (https://www.ngdc.noaa.gov/mgg/global/global.html, accessed on 17 March 2023). In the context of shelf seas, the collected data can still include unwanted interference from rapid phenomena like tides and internal waves, and they may also be affected by noise originating from land-based sources. Consequently, for this study, data obtained at depths shallower than 200 m were excluded or masked out. The terrain data of ETOPO1 developed in August 2008 with a resolution of 1' are adopted in this paper. In addition, the rows where the non-number (NAN) values in the data are removed to facilitate the establishment of the model later.

2.2. Temporal Modes-Enhanced Neural Network Model

The TMENN model integrates EOF analysis into ANN. Large spatial and temporal spanning datasets will bring too many input variables to the neural network, which makes direct prediction using the neural network unfeasible. Therefore, EOF analysis is used to reduce the dimensionality and capture the dominant temporal modes. ANN is trained to predict the PCs captured by the EOF analysis. The inclusion of EOF analysis can reduce computational costs and facilitate faster forecasting of the ANN model.

2.2.1. Empirical Orthogonal Function Analysis

EOF analysis is a powerful method used for dissecting structural characteristics within data and extracting key feature quantities from the dataset. It is widely used in research fields such as meteorology [32,33] and oceanography [34,35]. EOF analysis is mainly applied to data compression [36], dimensionality reduction, etc. The size of the artificial intelligence (AI) problem is greatly reduced by using EOF analysis to segment the spatiotemporal variability contained in remote sensing data [37].

The time and space spans of the SSH data used in this study are relatively large, and are difficult to forecast. Therefore, EOF analysis is adopted to decompose the SSH data, and then the extracted PCs are forecasted.
The calculation process for EOF can be divided into five steps. Firstly, give the matrix form (Equation (1)) of the observed SSH data.

\[
X = \begin{bmatrix}
  x_{11} & \cdots & x_{1n} \\
  \vdots & \ddots & \vdots \\
  x_{m1} & \cdots & x_{mn}
\end{bmatrix}, \quad (1)
\]

where \(x_{ij}\) represents the observation value at \(j\)-th moment on the \(i\)-th spatial point.

Secondly, create a new \(X_{m \times n}\) by normalizing the original data matrix \(X\). And then covariance matrix \(C_{m \times m} = \frac{1}{n}X \times X^T\) is obtained by calculation.

Thirdly, calculate the eigenvector \(V_{m \times m}\), that is spatial mode, and eigenroot \(\lambda_i (i = 1, \ldots, m)\) of \(C\), both of which satisfy Equation (2).

\[
C_{m \times m} \times V_{m \times m} = V_{m \times m} \times \Lambda_{m \times m}, \quad (2)
\]

where \(\Lambda\) is an \(m \times m\) dimensional diagonal matrix, that is:

\[
\Lambda = \begin{bmatrix}
  \lambda_1 & 0 & \cdots & 0 \\
  0 & \ddots & \vdots & \vdots \\
  \vdots & \ddots & \ddots & \vdots \\
  0 & \cdots & \cdots & \lambda_m
\end{bmatrix}. \quad (3)
\]

The eigenroot \(\lambda\) should be arranged in descending order, that is, \(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m\).

Fourthly, calculate the principal components \(PC_{m \times n}\) (Equation (4)). Temporal modes (PCs) are derived by projecting the spatial mode of EOF onto the original data matrix \(X\).

\[
PC_{m \times n} = V_{m \times m}^T \times X_{m \times n}, \quad (4)
\]

where each row of data in the \(PC_{m \times n}\) is the time coefficient corresponding to each eigenvector.

Ultimately, to determine the optimal number of PCs, it is essential to compute the variance contribution rate \(R_k\) and cumulative variance contribution rate \(G_k\), as illustrated in Equations (5) and (6).

\[
R_k = \frac{\lambda_k}{\sum_{i=1}^{m} \lambda_i} \quad [k = 1, 2, \ldots p (p < m)], \quad (5)
\]

\[
G_k = \frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{m} \lambda_i} \quad (p < m). \quad (6)
\]

The cumulative variance contribution rate of PCs serves as a valuable metric for ascertaining the appropriate number of PCs to employ in data reconstruction. A higher variance contribution rate indicates the greater significance of the corresponding PC, and utilizing the first PCs for data reconstruction effectively preserves the majority of information within the dataset. Investigating the temporal evolution of these PCs is equivalent to researching the temporal variations across the entire field.

2.2.2. Artificial Neural Network

The ANN, which has its roots in the fundamental neuron model initially presented by McCulloch and Pitts [38], is a powerful tool for machine learning that can mine complex rules hidden in multi-dimensional time series data by training on a large number of sample data [39]. If trained correctly using a neural network, its results can outperform traditional empirical and statistical linear methods, among others [40].

An ANN is made up of linked neurons or nodes that take input signals from other neurons as well as external sources. The incoming signals are given weights by these neurons, which send them through an activation function to create an output signal. Back Propagation (BP) Neural Network stands out as one of the most commonly utilized models.
within the field of ANNs. The input data are processed through a series of weighted connections and activation functions at the forward propagation stage before being sent to the output layer, where they are used to generate the model’s anticipated output. The network then computes the error between the anticipated output and the real goal during the backpropagation stage and employs a gradient descent technique to fine-tune each weight along the gradient’s direction to progressively minimize the error (Equation (7)). The iterative learning then continues until the weights and thresholds corresponding to the least error are found or a set number of iterations have been completed.

\[
E = \frac{1}{2} \sum_{k=1}^{z} (Y_k - O_k)^2 = \frac{1}{2} \sum_{k=1}^{z} e_k^2
\]

(7)

where \(Y_k\) indicates actual output, \(O_k\) intended target output, \(z\) indicates the number of output nodes, \(e_k\) is the error of the \(k\)th output.

An input layer, two hidden layers, and an output layer make up the four layers of the ANN model that we employed in our research. The input layer is composed of neurons that reflect previous PC values, whereas the output layer includes the values to be predicted. The two hidden layers’ neuron counts were experimentally set at 20 and 10, respectively. We use the momentum gradient descent optimization approach for training for building the ANN model. The initial learning rate was adjusted to 0.001 and the momentum factor was chosen as 0.9. The activation function used in the first hidden layer is specifically the tansig function (hyperbolic tangent sigmoid function), whereas the activation function used in the second hidden layer is linear.

For the purpose of our study, we employ PCs obtained through EOF analysis to train and test the ANN model. The BP neural network is trained using the PCs derived from the initial 85% (1 January 2000 to 31 December 2016) of the data, and the PCs from the remaining 15% (1 January 2017 to 31 December 2019) of the data are exclusively reserved for testing purposes.

If we want to forecast SSH data at time \(m + 1\), we only need to forecast PCs decomposed by EOF at time \(m + 1\). Since the PCs are independent of each other, the value of the \(j\)-th PC at time \(m + 1\) can be predicted by the values of previous \((l + 1)\) moments (Equation (8)).

\[
p_{m+1,j} = f(p_{m,j}, p_{m-1,j}, \ldots, p_{m-l,j})
\]

(8)

where \(p_{m,j}\) is the \(j\)-th PC of SSH at the \(m\)-th time, and \((l + 1)\) is the time delay. The training period will be greater the longer the time delay. We choose a 15-day time delay in our study. The \(j\)-th PC of the SSH at the \((m + 1)\)-th time obtained by ANN can be substituted into Equation (8) to forecast \(p_{m+2,j}\).

2.2.3. Temporal Modes-Enhanced Neural Network Model

The fundamental concept behind the TMENN model is to leverage EOF analysis for the extraction of primary temporal modes (PCs) from the original data. This approach aims to expedite the convergence rate of the BP neural network and reduce computational overhead.

The TMENN model typically encompasses the following key steps:

- The original SSH data are initially divided into PCs and spatial modes. Next, determine the main PCs and spatial modes that correspond to the cumulative variance contribution rate. The BP neural network receives the extracted primary temporal modes as inputs, which are subsequently trained to get final training results by further modifying the weights and biases.

- The specific computational process of the TMENN model is visually depicted in Figure 2.
Figure 2. The architecture of the TMENN model.

2.3. Evaluation Method

Correlation Coefficient (CC) and Root Mean Square Error (RMSE) are measurements we use in this study to evaluate the validity and precision of our model. A higher CC signifies increased forecast accuracy, while a smaller RMSE reflects higher forecasting accuracy.

The CC, calculated between the forecasted and actual values for each grid point at the \( n \)-th time step, is expressed by Equation (9):

\[
CC = \frac{\sum_{i=1}^{m} (SSH_{P,i,n} - \overline{SSH}_{P,n}) (SSH_{O,i,n} - \overline{SSH}_{O,n})}{\sqrt{\sum_{i=1}^{m} (SSH_{P,i,n} - \overline{SSH}_{P,n})^2} \sqrt{\sum_{i=1}^{m} (SSH_{O,i,n} - \overline{SSH}_{O,n})^2}},
\]  

(9)

where \( \overline{SSH}_{O,n} \) is the arithmetic mean of observed SSH, \( \overline{SSH}_{P,n} \) is the arithmetic mean of forecasted SSH. \( m \) is the total number of SSH data points in the study area. \( SSH_{P,i,n} \) is the
i-th SSH data point forecasted on day \( n \), and \( SSH_{O,i,n} \) is the i-th SSH data point observed on day \( n \).

And the RMSE between the forecasted and the true value of each grid point at the \( n \)-th time is defined as Equation (10):

\[
RMSE_{\text{pre}} = \sqrt{\frac{\sum_{i=1}^{m} (SSH_{P,i,n} - SSH_{O,i,n})^2}{m}}, \tag{10}
\]

Moreover, to comprehensively assess the performance of the forecasting model, we introduce a persistent approach for comparative analysis, following the methodology outlined by Oey in 2005 [41]. Calculating the RMSE between each grid point’s persistent value on day \( n \) and the observed SSH is a necessary step in the assessment process. Following is a definition of the RMSE calculation:

\[
RMSE_{\text{per}} = \sqrt{\frac{\sum_{i=1}^{m} (SSH_{O,i,0} - SSH_{O,i,n})^2}{m}}, \tag{11}
\]

among them, \( SSH_{O,i,0} \) represents the i-th SSH data point observed on day 0.

Finally, the skill score (SS) is used to validate the model [42], the higher the SS, the higher the accuracy of the model. So, the forecasted SS of SSH for each grid point on day \( n \) is defined as:

\[
SS_{\text{Pre}} = 1 - \frac{\sum_{i=1}^{m} (SSH_{P,i,n} - SSH_{O,i,n})^2}{\sum_{i=1}^{m} (SSH_{O,i,n} - SSH_{O,i,n})^2}, \tag{12}
\]

At the same time, the persistent SS of SSH at each grid point on the \( n \)-th day is defined as:

\[
SS_{\text{Per}} = 1 - \frac{\sum_{i=1}^{m} (SSH_{O,i,0} - SSH_{O,i,n})^2}{\sum_{i=1}^{m} (SSH_{O,i,n} - SSH_{O,i,n})^2}. \tag{13}
\]

2.4. Forecasting Process

In this study, the TMENN model is introduced for the forecasting of mesoscale eddies in the KE region. As a preliminary step aimed at reducing computational complexity and expediting the forecasting process, EOF analysis is employed to break down the SSH data into spatial modes and temporal modes. Secondly, ANN is adopted to train PCs with a sum variance equal to 95% on the training dataset [43]. Then, the PCs of the testing data are forecasted with the trained model. Finally, the forecasted PCs are combined with the spatial modes to reconstruct the SSH, which is convenient for subsequent analysis. The reconstructing formula is as follows:

\[
\begin{pmatrix}
\epsilon_{11} & \cdots & \epsilon_{1n} \\
\vdots & \ddots & \vdots \\
\epsilon_{kn} & \cdots & \epsilon_{nn}
\end{pmatrix}
\begin{pmatrix}
p_{m+1,1} & p_{m+1,2} & \cdots & p_{m+1,k}
\end{pmatrix}
= \begin{pmatrix}
x_{m+1,1} & x_{m+1,2} & \cdots & x_{m+1,n}
\end{pmatrix}, \tag{14}
\]

where \( p_{m+1,j} \) is the forecasted \( j \)-th PC value at the \((m + 1)\)-th time. \( x_{m+1,j} \) is the forecasted \( j \)-th SSH value at the \((m + 1)\)-th time, \( \epsilon_{ij} \) represents the corresponding eigenvectors of \( p_{m+1,j} \).

Furthermore, utilizing assessment methodologies including CC, RMSE, and SS, the model’s forecasting capacity for mesoscale eddies in the KE is evaluated. Figure 3 depicts the forecasting procedure used in this investigation.
3. Results

3.1. Sensitivity of PC Number Selection

Since EOF analysis has the advantage of easily concentrating the information of the variable field on several modes, it is possible to select PCs that cover most of the information for prediction, followed by data reconstruction. In this work, a sensitivity experiment is conducted to determine the optimal number of PCs (predictors), and the time range of the data collected in the experiment is from 1 January 2017 to 31 December 2019. According to the variance percentage occupied by different PCs in the EOF analysis, five groups of PCs are selected, corresponding to five variance percentage levels of 80%, 85%, 90%, 95%, and 96% for sensitivity testing. Simultaneously, to assess the forecast accuracy, we calculate the CC and RMSE as evaluation metrics for different forecast durations. Superior forecast performance is shown by a greater CC and a smaller RMSE. The findings are summarized in Table 1.

An analysis of Table 1 reveals several important observations. The cumulative variance contribution rate progressively climbs as the number of PCs does; however, the pace of increase slows down. For a fixed forecast duration, increasing the number of selected PCs leads to a gradual improvement in CC and a corresponding decrease in RMSE, signifying enhanced model forecast accuracy. However, there exists a threshold beyond which further PC inclusion does not significantly impact CC and RMSE. This is attributed to the fact that, after reaching a certain number of PCs, their variance contribution becomes marginal, and these additional PCs may introduce noise rather than meaningful information, rendering forecasting challenging. Additionally, when the number of selected PCs remains constant, CC tends to decrease gradually, while RMSE increases as the forecast time grows. This trend suggests that as the forecast duration increases, the model’s forecast accuracy diminishes. This phenomenon may arise from error propagation and accumulation within the model during the forecasting process.

According to Table 1, the experiment with 117 PCs has the largest CC and the smallest RMSE when the forecast days are thirty days (CC between the forecasted and raw PCs is up to 0.7035). In general, the higher the number of PCs, the better the performance, but when the number of PCs is 136, the performance of this experiment is slightly worse than...
when the number of PCs is 117, partly because the PCs with a small percentage of variance are noisier and more difficult to forecast. So, the number of PCs selected is 117 (a total of 8384 PCs), and the sliding forecast window is 30 days.

Table 1. Outcomes of PC number selection (The first column lists the number of PCs chosen, the second lists the associated variance contribution rate, and the other columns list the CC and RMSE, respectively, in blue and black (Bold)).

<table>
<thead>
<tr>
<th>PC No.</th>
<th>PC Percentage</th>
<th>Day 1</th>
<th>Day 5</th>
<th>Day 10</th>
<th>Day 15</th>
<th>Day 20</th>
<th>Day 25</th>
<th>Day 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>80%</td>
<td>0.8722</td>
<td>0.8709</td>
<td>0.8600</td>
<td>0.8303</td>
<td>0.7834</td>
<td>0.7357</td>
<td>0.6932</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0855</td>
<td>0.0860</td>
<td>0.0898</td>
<td>0.0988</td>
<td>0.1111</td>
<td>0.1216</td>
<td>0.1299</td>
</tr>
<tr>
<td>52</td>
<td>85%</td>
<td>0.9050</td>
<td>0.9033</td>
<td>0.8902</td>
<td>0.8560</td>
<td>0.8018</td>
<td>0.7443</td>
<td>0.6958</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0745</td>
<td>0.0752</td>
<td>0.0803</td>
<td>0.0918</td>
<td>0.1066</td>
<td>0.1198</td>
<td>0.1295</td>
</tr>
<tr>
<td>74</td>
<td>90%</td>
<td>0.9373</td>
<td>0.9352</td>
<td>0.9181</td>
<td>0.8761</td>
<td>0.8143</td>
<td>0.7504</td>
<td>0.6956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0610</td>
<td>0.0621</td>
<td>0.0695</td>
<td>0.0854</td>
<td>0.1035</td>
<td>0.1186</td>
<td>0.1296</td>
</tr>
<tr>
<td>117</td>
<td>95%</td>
<td>0.9701</td>
<td>0.9673</td>
<td>0.9472</td>
<td>0.8964</td>
<td>0.8264</td>
<td>0.7597</td>
<td>0.7035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0427</td>
<td>0.0447</td>
<td>0.0567</td>
<td>0.0787</td>
<td>0.1005</td>
<td>0.1167</td>
<td>0.1282</td>
</tr>
<tr>
<td>133</td>
<td>96%</td>
<td>0.9761</td>
<td>0.9729</td>
<td>0.9509</td>
<td>0.8989</td>
<td>0.8266</td>
<td>0.7524</td>
<td>0.6917</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0381</td>
<td>0.0407</td>
<td>0.0548</td>
<td>0.0781</td>
<td>0.1010</td>
<td>0.1190</td>
<td>0.1311</td>
</tr>
</tbody>
</table>

In Figure 4, we present a comparison of the first PC between the forecasted and observed SSH data from 1 January 2017 to 31 December 2019. As can be seen from Figure 4, it is clear that CC steadily drops from day one to day thirty, going from 1.00 to 0.9553. On the first day, the observed and forecasted SSH data align closely, yielding a CC of 1, highlighting the TMENN’s exceptional accuracy in predicting the first PC value on the initial day. As we progress to subsequent days, forecast errors begin to emerge and incrementally grow with the increasing forecast duration. By the time the forecast time reaches 30 days, these errors become more pronounced. This observation implies that while the model exhibits minimal discrepancies in its forecasts during the early stages of the forecast period, errors accumulate and magnify as time progresses. This phenomenon is often attributed to the propagation and accumulation of errors within the forecasting process, where minor inaccuracies at each forecasting step can accumulate, resulting in larger discrepancies as forecasting time increases.

3.2. SSH Forecasting Skill Assessment

We selected 117 PCs and a sliding forecast window of 30 days for the forecast. From January 2017 to December 2019, the CC and RMSE between forecasted and observed SSH (Figure 5) show obvious variations of forecasting performance. The CC is greater than 0.8 from the first day to the tenth day; from the 15th day, the CC gradually decreased and oscillated. The RMSE showed that the errors on the 1st, 5th, and 10th days are small (all less than 0.2), and the RMSE fluctuation began to increase from the 15th day. Although the forecasting performance gradually decreased from the first day to the 30th day, the CC was still mostly greater than 0.7 on the 30th day, which means that the model can be accurate at least 15 days up to 30 days ahead in most time periods to capture the evolution of the eddy.
Figure 4. Comparison between forecasted (blue line) and observation-derived (red line) first PC of SSH from 1 January 2017 to 31 December 2019. The correlation coefficient is in the upper right corner.

Figure 5. CCs and RMSEs of forecasted and observed SSH in the study area from 2017 to 2019 with 30 days ahead daily sliding forecast window using 117 PCs. Red lines represent RMSE points, while blue lines represent CC points. Three examples are shown in black lines (Case analysis in Section 3.3). a, b, c, d, e, f, and g, respectively, indicate that the forecast time is 1 day, 5 days, 10 days, 15 days, 20 days, 25 days, and 30 days.
As the forecast time lengthens, there are discernible changes in the model’s predicted performance, as seen in Figure 5. These oscillations can be attributed to three potential factors. Firstly, there are the propagation and accumulation of errors. Errors that arise at earlier forecasting stages can propagate through subsequent steps, accumulating along the way, and ultimately leading to fluctuations in forecast accuracy. Because the TMENN will generate errors when forecasting PCs and reconstructing SSH data, and the errors will propagate during the forecasting process, the errors will as the forecast time increases and accumulates gradually. As the prediction time lengthens, the final model’s oscillations will become more noticeable. Secondly, the nonlinear variation of mesoscale eddies can also affect the performance of the forecasting model. Finally, drastic weather changes can also reduce the accuracy of forecast, because drastic weather changes can cause transient changes in sea surface height.

A persistent approach is given to further measure the model’s predicting ability. Models are evaluated by comparing the RMSE of forecasted and persistent SSH between 1 January 2017 and 31 December 2019. The results are visualized in Figure 6. Figure 6 demonstrates that as the forecast time grows, the RMSE of the persistent approach and the forecasting model both gradually increase, but the growth rates of the two are different, and the RMSE of persistence increases faster. For short-term (up to five days) forecasts, the RMSE of the forecasting model is higher than the persistent RMSE; for medium and long-term (greater than five days) forecasts, the RMSE of forecasting is lower than the persistent RMSE. This shows that when making short-term forecasts, the persistent approach has higher forecast accuracy and better performance; the forecasting model has greater predictability and performs better for medium to long-term projections. This is because SSH changes slowly on the whole, so SSH does not change much in the short-term forecast period, and the persistent approach can obtain more accurate forecasting results.

The study also includes a comparison between the averaged RMSE of forecasted SSH and persistent SSH, as depicted in Figure 7. Notably, when the forecast time is less than five days, both the 30-day and 3-year averaged RMSE values of the forecasting model surpass that of the persistence model. This demonstrates that when it comes to short-term forecast accuracy, the persistent model outperforms the forecasting model. Conversely, as the forecast time beyond 5 days, the majority of the 30-day and 3-year averaged RMSE values fall below the persistent values. This signifies that the forecasting model achieves
higher precision than persistence in medium and long-term projections. It indicates that the results are consistent with the findings seen in Figure 6.

Figure 7. Averaged RMSE of SSH for forecasting (red) and persistence (gray). Monthly means are shown by thin lines, while forecasted 3-year means are represented by thick lines.

As an additional means of evaluating the forecasting model’s performance, a comparison is made between the SS values of the forecasting model and those of the persistent method. In this assessment, greater SS values are indicative of enhanced forecast accuracy and superior modeling performance. The outcomes of this comparison are presented in Figure 8. The SS of the forecasting model is higher than that of the persistent approach in most time periods (medium and long-term), indicating that the accuracy of the forecasting model is better than that of the persistent approach. The SS of both the forecasting model and the persistent approach gradually declines as forecast time increases, but their rates of decline differ. The persistence SS declines more quickly, proving that the forecasting model is superior to persistence and produces more accurate results.

Figure 8. SS comparison of SSH between forecasting (red) and persistence (black).
A comparative analysis of the averaged SS between the forecasting model and the persistent approach is conducted, with the results being presented in Figure 9. The following summarization is derived from our findings: both the forecasting model and the persistent approach exhibit a gradual decrease in averaged SS as the forecast time increases. When the forecast time is less than 5 days, whether considering the 30-day or 3-year averaged SS, the forecasting model yields lower averaged SS values compared to the persistent approach. This shows that when making short-term forecasts, the persistent approach has higher forecast accuracy and better performance. Conversely, the averaged SS of the forecasting model surpasses that of the persistent method when the prediction time is greater than 5 days. This shows that the forecasting model’s exceptional performance in achieving long and medium-term forecast accuracy. It is clear that the forecasting model outperforms the persistent technique in terms of accuracy when it comes to medium- and long-term projections. Consistent with the previous conclusions, the averaged SS also indicates that the forecasting model is more accurate and has better forecasting performance than the persistent approach in medium and long-term forecasts.

![Figure 9](image_url)

**Figure 9.** Averaged SS of SSH for forecasting (red) and persistence (gray).

### 3.3. Case Analysis

To further assess the predicting capabilities for mesoscale eddies and establish the correctness of the forecasting model, a comparison of the observed SSH and the projected SSH is made over the period of 1 January 2017 to 31 December 2019. Therefore, three classical mesoscale eddy processes are selected as case studies in this study, and we compare the forecasted and observed values of the three eddy events. Since eddy shedding and reabsorption often occur in the KE flow axis, cold eddy shedding, warm eddy shedding, and attachment are selected as case studies, respectively. Due to only the first 117 PCs (a total of 8384 PCs) being selected to forecast and reconstruct the SSH, some raw information is lost. Since the forecasted values are typically weaker than the observed values, it is challenging to identify the appropriate outlines that indicate the forecasted eddy’s limits. After many experiments, the −0.25 m, 0.22 m, and 0.28 m contours are selected to represent the observed and forecasted boundaries of the three eddy events, respectively. To demonstrate that the forecasting model can follow the development of the eddies, the evolution of three mesoscale eddies will be examined in detail below.

#### 3.3.1. Cold Eddy Attachment

Figure 10 shows a comparison of SSH forecasted and observed values from 19 December 2017 to 17 January 2018 during which there is a cold eddy attachment. According to the observation results, it can be seen that there were two relatively strong cold eddies in the...
upstream area of the KE (152–157° E, 30–35° N) on 19 December 2017. With the increase in
time, the two cold eddies gradually merged, and finally, a strong cold eddy was formed
on 2 January 2018. At the same time, the intensity of the cold eddy gradually decreases
with time, and the scope of influence is also reduced, but the shrinking rate is very slow.
According to the forecast results, two cold eddies also existed in the upstream area of
the KE from 9 December 2017 to 17 January 2018, and the intensity and size are weaker
than the observed results. The predicted cold eddy’s position matches the position of
the one observed prior. Two cold eddies merge, and a stronger cold eddy is finally formed
in the forecast results. It is shown that the forecasting model can accurately capture the
location, size change and evolution of the cold eddy. The discrepancy between the two in
Figure 10, the disparity between forecasted and observed findings, steadily increases over
time and concentrates on the location of the eddy. The forecasting model can still properly
depict the development and position of the cold eddy in this region, even if the observed
and anticipated outcomes differ. This shows that the forecasting model provides accurate
prediction results.

Figure 10. SSH forecasted and observed during a cold eddy attachment are compared (The observed
values are shown in the first row, the forecasted values are shown in the second row, and the difference
between the two is shown in the third row).

3.3.2. Warm Eddy Shedding

Figure 11 shows the warm eddy shedding that occurred in the upstream region of the
KE (152–157° E, 34–39° N) from 27 May 2018 to 25 June 2018 and the SSH comparison of
forecasted and observation values. It can be seen from the observation results that there
was a strong warm eddy in this area on 27 May 2018, and the eddy shedding phenomenon
occurred in this warm eddy within a week to form two relatively strong warm eddies on
5 June 2018. According to the forecast results, the warm eddy shedding phenomenon also
occurred from 27 May 2018 to 25 June 2018, and the evolution of the forecasted eddy is
consistent with the observed result. The eddy shedding phenomenon is demonstrated to
be precisely captured by the forecasting model.

Figure 11. SSH forecasted and observed during a warm eddy shedding are compared.
3.3.3. Warm Eddy Attachment

Figure 12 shows a comparison of SSH forecasted values between and observed from 16 July 2018 to 14 August 2018. According to the observation results, it can be seen that there were two strong warm eddies in the upstream area of the KE region (150–155° E, 34–39° N) on 16 July 2018, and the strength of the two warm eddies increased with time, and the high latitude warm eddy gradually moved to the southwest, approaching the low latitude warm eddy, and finally formed a strong warm eddy on 30 July 2018. At the same time, the intensity of the warm eddy gradually increased with time, and the scope of influence changed from time to time. A strong warm eddy center formed on 9 August 2018. According to the forecast results, there was also a warm eddy in the upstream area of the KE from 16 July 2018 to 14 August 2018, but the intensity and size were weaker than the observed results. The forecast results also show a process of a warm eddy approaching, merging and forming a stronger warm eddy. It is also shown that the forecasting model has reliable forecast results.

3.3.2. Warm Eddy Shedding

Figure 11 shows the warm eddy shedding that occurred in the upstream region of the KE from 27 May 2018 to 25 June 2018 and the SSH comparison of forecasted and observed values. It can be seen from the observation results that there was a strong warm eddy in this area on 27 May 2018, and the eddy shedding phenomenon was precisely captured by the forecasting model. The disparities between observed and forecasted SSH values become increasingly pronounced as the forecast time extends. At the same time, the intensity of the eddy occurs. While dissimilarity between forecasted and actual values exists, the forecasting model demonstrates a commendable capacity in accurately discerning the spatial positioning and temporal evolution of mesoscale eddies. This signifies the forecasting model’s robust predictive performance. The disparities between observed and forecasted SSH values predominantly concentrate within the region of eddy occurrence, which is probably caused by the mesoscale eddy’s nonlinear development.

Figure 11. SSH forecasted and observed during a warm eddy shedding are compared.

Figure 12. SSH forecasted and observed during a warm eddy attachment are compared.

From Figures 10–12, it is evident that the forecasted values of SSH data consistently exhibit lower magnitudes compared to their observed counterparts for the three mesoscale eddies phenomena under investigation. Moreover, this disparity between forecasted and observed SSH values becomes increasingly pronounced as the forecast time extends. At
the same time, it is found that the difference mainly appears in the area where the eddy occurs. While dissimilarity between forecasted and actual values exists, the forecasting model demonstrates a commendable capacity in accurately discerning the spatial positioning and temporal evolution of mesoscale eddies. This signifies the forecasting model’s robust predictive performance. The disparities between observed and forecasted values predominantly concentrate within the region of eddy occurrence, which is probably caused by the mesoscale eddy’s nonlinear development.

4. Conclusions

TMENN is used to forecast mesoscale eddies in the KE in this study. SSH data from the KE region from 1 January 2000 to 31 December 2019 are used for the study. Given the substantial data volume, employing an ANN directly for forecasting becomes challenging and cost-intensive. Therefore, EOF analysis is adopted to break down the original data, which lowers the computing burden and complexity of prediction.

The current SSH is first divided into spatial and temporal modes using EOF analysis. Then, ANN is deployed to forecast the decomposed PCs. Finally, future SSH is constructed by multiplying the spatial modes of the SSH with the predicted PC.

Through sensitivity testing, the top 117 PCs, which together add up to 95% of the variance, are selected as predictors for the model. This selection effectively retains the majority of the information within the dataset. As a result, the data from 1 January 2000 to 31 December 2016 were identified as the training set in order to promote learning and model training. This training dataset serves as the foundation for developing a forecasting model and training a neural network. The remaining data are used as a test set for the forecasting model.

A 30-day sliding forecast window is used to forecast PCs from 1 January 2017 to 31 December 2019. The model is evaluated by calculating the CC and RMSE between the forecasted and observed SSH data, and the outcomes demonstrate the forecasting model’s excellent accuracy. The forecasting model outperforms the persistent approach for producing medium to long-term projections, as shown by comparisons of the RMSE and SS between the two methods. In addition, the model is evaluated by studying three classical mesoscale eddy processes in the study area. The findings illustrate the forecasting model’s capability to effectively discern the location and evolution of mesoscale eddies in this area, underscoring the model’s reliability towards accurate forecasts.

Furthermore, the forecasting model exhibits potential utility in predicting various other marine physical parameters, including but not limited to sea surface temperature and sea surface height anomaly. It can also be applied to predict other oceanic phenomena and mesoscale eddies in other sea areas.

Author Contributions: Conceptualization, Y.F. methodology, H.W.; software, H.W.; validation, Y.G. and Y.K.; data curation, Y.K.; writing—original draft preparation, H.W.; writing—review and editing, Y.G.; supervision, Y.G.; project administration, Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shandong Provincial Natural Science Foundation Youth Project (ZR2021QD108), and National Natural Sciences Foundation of China (No. 42306233).

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.


Conflicts of Interest: The authors declare no conflict of interest.
References


42. Murphy, A.H. Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. *Mon. Weather Rev.* 1988, 116, 2417–2424. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.