Article

A Hybrid Deep Learning Model for the Bias Correction of SST Numerical Forecast Products Using Satellite Data

Tonghan Fei [1], Binghu Huang [1], Xiang Wang [2,*], Junxing Zhu [2], Yan Chen [2], Huizan Wang [2] and Weimin Zhang [2]

1   College of Oceanography and Space Informatics, China University of Petroleum, Qingdao 266580, China; s19160016@s.upc.edu.cn (T.F.); huangbhu@upc.edu.cn (B.H.)
2   College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China; junxingzhu@nudt.edu.cn (J.Z.); chenyan16@nudt.edu.cn (Y.C.); wanghuizan17@nudt.edu.cn (H.W.); wmzhang@nudt.edu.cn (W.Z.)
*   Correspondence: xiangwangcn@nudt.edu.cn

Abstract: Sea surface temperature (SST) has important practical value in ocean related fields. Numerical prediction is a common method for forecasting SST at present. However, the forecast results produced by the numerical forecast models often deviate from the actual observation data, so it is necessary to correct the bias of the numerical forecast products. In this paper, an SST correction approach based on the Convolutional Long Short-Term Memory (ConvLSTM) network with multiple attention mechanisms is proposed, which considers the spatio-temporal relations in SST data. The proposed model is appropriate for correcting SST numerical forecast products by using satellite remote sensing data. The approach is tested in the region of the South China Sea and reduces the root mean squared error (RMSE) to 0.35 °C. Experimental results reveal that the proposed approach is significantly better than existing models, including traditional statistical methods, machine learning based methods, and deep learning methods.

Keywords: SST; bias correction; deep learning; ConvLSTM; 3D-C BAM

1. Introduction

Oceans take up almost 71% of the entire surface of the globe, and are closely related to human activities. Sea surface temperature (SST) is the water temperature near the surface of the ocean. SST is an important physical quantity for global climate studies [1], marine ecosystem studies, and related applications. The forecast accuracy of SST is essential for marine disaster prevention, navigation, ocean fishery [2], and other ocean-related cases. SST prediction methods can be classified into two major categories [3]. One category is the numerical model, which is based on physics [4]. The other category is data-driven models, based on data analysis. With the improvement and development of the numerical model, the accuracy of numerical prediction has been improved. However, the numerical model cannot completely describe various physical processes in the ocean [5,6]—an uncertainty of the initial field [7,8]—and calculation errors exist in the numerical solution process of the model. Therefore, the prediction results of numerical forecast products need to be further corrected.

Currently, there are mainly three kinds of methods for numerical forecast products correction: traditional statistical methods, machine learning based methods, and deep learning methods. Statistical post-processing [9] is the typical one, such as model output statistics (MOS) approach [10,11], Kalman filtering [12,13], and Bayesian probability decision [14], all of which have achieved some results. As the machine learning, deep learning development [15], and computing performance improved, data-driven approaches were introduced into numerical forecast product correction, such as SVM [16], BP neural network [17], and CNN [18]. However, current methods for numerical forecast products correction have

weaknesses, as they do not include the spatio-temporal relationships among the datasets. Meanwhile, observation data of buoys in the ocean have been lacking for a long time. Since the launch of satellites equipped with ocean observation sensors, ocean remote sensing data observed from satellites have been widely used in in coastal erosion calculation [19], offshore oil spill [20], disaster warning [21], and other related research. Therefore, we consider combining deep learning methods with numerical models [22] and applying satellite data into numerical prediction models for SST numerical forecast products correction.

In this paper, we propose a new hybrid SST correction model, which not only takes into account the influence of spatial distribution of the dataset, but also takes into account the importance of temporal information. This approach is inspired by the outstanding performance of the ConvLSTM in capturing the spatio-temporal relationships and the attention mechanism in improving feature utilization. Combining these novel methodologies together will create a more effective model to correct SST as it will create a greater synergy than the individual models on their own.

In recent years, with the rapid development of machine learning, deep learning methods have been widely used in many fields, such as natural language processing [23], audio classification [24], community detection [25], and image restoration [26]. Some researchers have already used these methods in areas related to our research. For example, Shi et al. [27] proposed the ConvLSTM method for precipitation prediction. D. Liu et al. [28] proposed a combination of empirical mode decomposition (EMD) algorithm and encoder decoder long short-term memory (EN-DE-LSTM) architecture for water flow prediction. Z.I. Petrou et al. [29] proposed an encoder–decoder network with a convolutional long short-term memory unit for sea ice prediction. Chen.R et al. [30] proposed the hybrid CNN-LSTM model for typhoon forecasting, which improved the accuracy of typhoon forecasting. A. Y. Winona et al. [31] use the so-called LSTM method to forecast the sea level and X. Kun et al. [32] proposed LSTM-Attention temperature prediction model I by combining LSTM with Attention mechanism in order to make full use of historical data and improve the accuracy of temperature prediction.

Our new hybrid SST correction model can be used to correct SST numerical forecast products more accurately. 3DCNN is used to determine the spatial relations of various marine variables. Simultaneously, 3D-CBAM model is used to improve the utilization of spatial features and marine environmental features. ConvLSTM is used to determine the spatio-temporal relationships of the data. The attention model is used to assign the weight of historical information. Our proposed model can effectively determine the spatio-temporal dependencies between SST field data, and at the same time introduce an attentional mechanism to correct the ConvLSTM output by learning the appropriate weights at each step, thus achieving high-precision SST correction. A series of experimental results show that the proposed method can achieve better accuracy in SST correction.

The contributions of this paper include:

1. We propose a new hybrid model for SST correction, which uses satellite remote sensing observation data and spatio-temporal data of sea surface variables. The performance of our model is then evaluated;

2. The attention mechanism is used to assign weights to the information in the dataset, which reflect the influence of spatio-temporal information on the SST correction, so that the key information is highlighted and thus we obtain better correction results;

3. Taking the South China Sea area (10°N–15°N, 125°E–130°E) as an example, the accuracy rate was improved by 41.9% after the correction. We analyze the influence of input sequence with different time steps, different model parameters and other variables on the correction effect through the experiments. Experiments on the dataset of the South China Sea show that our new hybrid model is more effective than existing methods, including some classical machine learning methods.

The paper is structured in the following manner: Section 2 takes a look at the current state of correction methods for numerical forecast products and deep learning in the discipline; Section 3 elucidates the central problems of this paper; Section 4 introduces

the new hybrid SST correction model; Section 5 introduces the evaluation scheme, the experimental set-up, and presents the experimental results; and, finally, Section 6 concludes this work and deals with recommendations for future work.

2. Related Work

Our research focuses on bias correction of SST numerical forecast products. In order to improve the accuracy of numerical forecast products, many scholars have proposed several methods to correct numerical prediction results. For example, Vannitsem et al. [10] and Tian et al. [11] used the mode output statistics (MOS) approach to establish a linear statistical relationship between model predictions and actual observations to improve SST forecasting accuracy, respectively. Krishnamurti et al. [33] used multiple regression to determine coefficients from multi-model forecasts and observations to improve weather and seasonal climate forecasts. Xu et al. [34] used the classical moving average method to analyze and correct the temperature forecast of the model, which improved the forecast accuracy to a certain extent. Libonati et al. [12] and Pelosi et al. [13] used Kalman filter to improve the quality of ensemble forecast in view of the existing deviation of ensemble forecast. X. Zhang et al. [35] proposed a method for correcting wave height prediction results of SWAN model based on Gaussian process regression (GPR).

With the continuous development of technology, the theory of machine learning has shown its extraordinary ability and great potential in the field of ocean and weather prediction and correction [36]. The correction model based on machine learning can capture the nonlinear variation [37] between the numerical model simulation results and the observation, so as to obtain more accurate model correction results. For example, J. Zeng et al. [16] used SVM to correct the weather forecast model, and the accuracy was effectively improved. Wang A et al. [38] designed a Random Forests-based adjusting method to correct the output of the WRF model and the RMSE of wind achieved an average decrease of 40% compared with the WRF model.

In addition, deep learning has injected fresh blood into artificial intelligence and machine learning. Deep learning is used to extract potential features and learn complex relationships in meteorological and oceanographic data, which provides a new idea for ocean and weather forecast and correction [39]. For example, Makarynskyy [40] improved wave parameter short-term forecasts based on artificial neural networks. Xu X. et al. [41] presented an ordinal Distribution Autoencoder (ODA) model, which can effectively correct numerical precipitation prediction based on ECMWF and SMS-WARMS model meteorological data. T. Wang et al. [42] proposed a residual single-hidden layer feedforward neural network, which is able to obtain effective corrections of numerical models. Rasp S et al. [43] proposed a flexible alternative based on neural networks to correct 2 m temperature. A. Sayeed et al. [18] used convolutional neural network (CNN) as post-processing technology to improve mesoscale weather research and prediction (WRF) daily simulated output. A. N. Deshmukh et al. [44] applied a wavelet neural network in improving numerical ocean wave predictions of significant wave height and peak wave period. It can be seen that deep learning has shown some potential in the temperature correction of model prediction, but it is still in the initial research and application stage. The above research also indicates the potential of machine learning in the correction of numerical model results.

However, there is little work to correct the forecast of SST. Zhang R used artificial neural network BP model to correct SST [17]. Yang X Q et al. [45] applied the prognostic trend (PT) correction method to reduce systematic errors in coupled GCM seasonal forecasts. Han Y.K. [46] proposed a new error-correction model based on the AR(p). Zhang P.J. [47] tried to correct numerical prediction SST product using GHRSST, and established a correction method for SST model prediction in the South China Sea—the effect of SST forecast correction was quite significant. The above methods for SST correction do not consider the temporal and spatial correlation between SST data. Therefore, we consider combining deep learning with the numerical model, using the deep learning method to

mine the temporal and spatial correlation of SST data and correcting forecast products to improve forecast accuracy.

We attempt to use the deep learning method to mine the spatio-temporal relationship between SST forecast data and carry out the correction of daily mean SST in the study area. On the one hand, it is helpful to obtain more accurate prediction results, and on the other hand, it is also an exploratory application of the deep learning correction model in oceanography.

3. Problem Definitions

Our goal is to use historical ocean data and reanalysis data as truth values to modify model forecast data and establish an SST correction model. SST data is a time series data without considering spatial information. In order to analyze and obtain the time sequence relationship between the data, historical ocean data at multiple times should be used for correction. However, SST and other marine environmental variables are spatial fields at any time, so SST correction can be defined as a spatial-temporal series correction problem. Different from previous methods that take the SST of a single site as the model input data, this paper corrects the SST within the region as a whole, that is, a matrix, to facilitate the model to extract the temporal and spatial correlation of SST.

The input data with multiple elements can be represented as a matrix $W \times H \times C \times T$, where W and H represent longitude and latitude, C represents the number of elements, and T represents the length of time series. SST and sequence of marine environmental variables $T = T_1, T_2, \cdots, T$, where $|T|$ is the SST sequence length of time, $T_i(1 \leq i \leq |T|, i \in Z)$ is the marine environmental variables matrix of all the record points of day i in the region, which is a $W \times H \times C$ matrix. The sequence of these matrices is the input to the model. The SST correction problem can be defined as a series of historical marine environmental variables data of the previous N days $X_{t-n}(n = 1, 2, 3 \cdots)$, used to correct the SST at time t, where $X_{t-n}(n = 1, 2, 3 \cdots)$ is a sequential matrix, which is $W \times H \times C \times N$. Define the current moment to be corrected as t, and the SST and marine environment variable of the current moment to be corrected as X_t. Y_t is the corrected SST value and n is the previous days before the current time, each time step is 1 day. X_{t-n} is the grid data set of each variable at the predicted time and the previous n days.

The model can also be expressed as:

$$Y_t^{W \times H} = f\left(X_{t-n}^{W \times H \times C \times N}, X_t^{W \times H \times C}\right), n = 1, 2, 3 \cdots . \tag{1}$$

This is our target function, where f is the final model learnt by the historical data. On this basis, we design and train the deep learning model. During the training, the data is divided into two parts. First, we train our model with the training set, where the "truth value" is known, and use it to adjust the parameters in the model. Finally, we use the test set to evaluate the correction effect of the training model. Figure 1 shows the data structure of SST and the related variables.
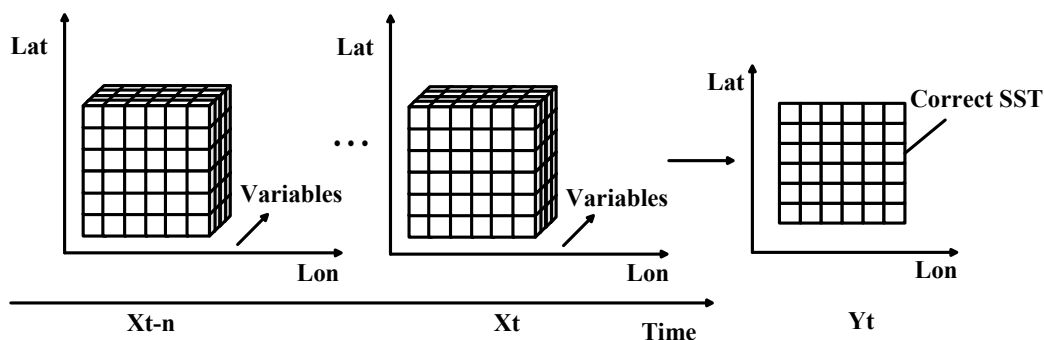

Figure 1. The structure of spatio-temporal variables sequence.

4. Method

In order to solve the problem of SST numerical prediction correction, we propose a new hybrid model for SST correction, which is based on ConvLSTM and the 3DCBAM model with attention mechanisms. It makes full use of the spatio-temporal information and marine environmental variables information.

4.1. The Framework of the New Hybrid SST Correction Model

The framework of the new hybrid SST correction model is shown in Figure 2, which is mainly divided into five stages: spatial feature extraction, spatial and channel attention mechanism, time-dependent learning, time attention mechanism, and output results. The main idea is to use convolution operation to extract and integrate spatial features of multiple variables, and use CBAM mechanism to improve the utilization rate of the spatial features of a 3D convolution network and show the importance of different environmental variables to the results. At the same time, ConvLSTM is used to learn the spatiotemporal relationship in the process of SST change, and attention mechanism is used to adjust the importance of information at different historical moments in variables. It not only considers the spatial correlation of SST field data, but also the time dependence between SST field data at different time and the interaction between marine environmental variables. Therefore, it can correct SST more accurately.
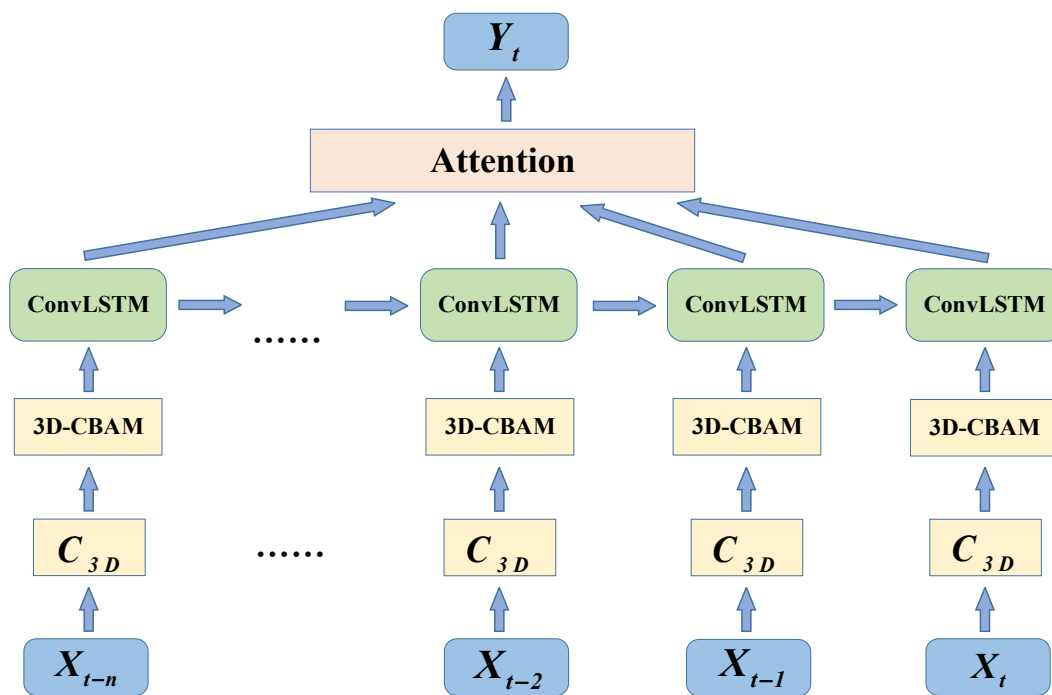

Figure 2. The framework of 3DCBAM-ConvLSTM method.

Therefore, the whole SST forecast revision model can be expressed as follows:

$$Y_t^{W \times H} = AT(\text{ConvLSTM}(\text{CBAM}(C_{3D}(X_t, X_{t-n})))), n = 1, 2, 3 \cdots. \tag{2}$$

For historical series data X, the X_t composed of SST data and other marine environmental variables at any time of t is grid data with $W \times H \times C$ specifications. Therefore, the input of the whole model is a five-dimensional tensor, which is expressed as $B \times T \times C \times W \times H$. Here B is the number of a batch of training samples, and T is the length of sequence data. W and H are the width and length of the SST field, and C is the number of marine environmental variables. In our experiments, length H and width W are longitude and latitude. The length of the time step can be obtained through the sliding window. For example, if the historical data of the past three days are used to correct the SST of the day,

then the length of the time step is 4, that is, the value of T. In the experiments, in addition to SST, salinity and water velocity u and water velocity v are added, so C is here 4. The five-dimensional tensor serves as input to the model.

As the correction of SST is a regression problem, this paper chooses MSE as the loss function. The calculation formula is shown in Equation (3), where n represents the number of points in grid data, $\hat{y}_i$ is the truth value of the point i, and y_i is the revised value of the point i. The training set is input into the model, and N iterations are carried out until the model converges.

$$\text{LOSS} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}. \tag{3}$$

4.2. Spatial Feature Extraction with 3D-CBAM

In the spatial feature extraction part, we use 3D convolution to extract spatial features from the input training data. 3D convolution is developed on the basis of 2D convolution [48]. 3D convolution is achieved by convolving a three-dimensional kernel with a cube formed by stacking multiple continuous matrices. Through this construction, the feature map of the convolution layer is connected with the previous layer to capture spatial information. The input of 3D convolution is sample X, $X \in R^{B \times T \times H \times W \times C}$. The 3D convolution operation C_{3D} mainly completes the spatial feature extraction and it can be computed as:

$$C_{3D}(X) = \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} \omega(p,q,r) * X, \tag{4}$$

where $*$ and $\omega(p,q,r)$ represent the convolution operation and kernel, and P, Q, R represent width, height, and temporal length of the data.

Then, we use 3D-CBAM attention mechanism to improve the utilization rate of the spatial features of the 3D convolution network and to show the importance of different environmental variables to the results. Convolutional Block Attention Module (CBAM) is a simple and effective attentional module that can be directly applied to a feedforward convolutional neural network, consisting of a channel attentional module and a spatial attentional module [49]. Figure 3 shows the structure of the 3D CBAM attention module.
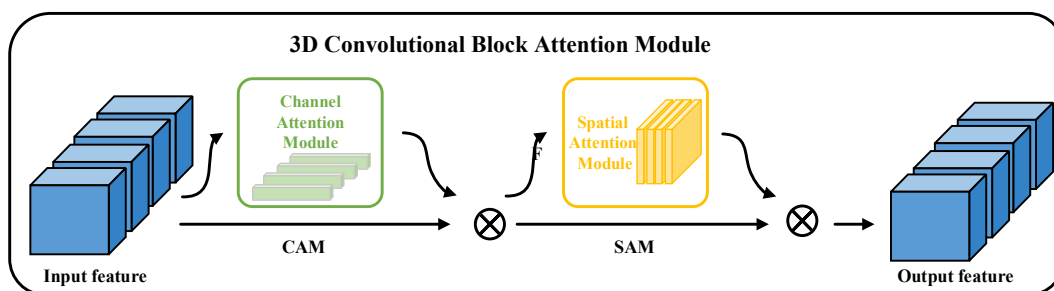

Figure 3. 3D convolutional block attention module.

The input of CBAM is $F = C_{3D}(X), X \in R^{B \times T \times H \times W \times C}$, the feature map from a previous 3D convolution layer. The 3D CBAM will apply channel attention module (CAM) and spatial attention module (SAM) in sequence to the input F. As shown in Figure 3, the CBAM can be designed as:

$$\text{CBAM}(F) = \text{SAM}(\text{CAM}(F)). \tag{5}$$

The channel attention module of 3D-CBAM pays attention to which feature plays a role in the final correction result. Firstly, we apply the global max pooling and global average pooling based on width, height, and time to the input feature matrix F, respectively, and we get F_{avg} and F_{max}. Both F_{avg} and F_{max} are one dimensional feature maps: $F_{avg} \in R^{B \times 1 \times 1 \times 1 \times C}$ and $F_{max} \in R^{B \times 1 \times 1 \times 1 \times C}$. Then, multilayer perceptron (MLP), a fully connected layer is used to efficiently combine the channel statistical information F_{avg} and F_{max}. To reduce

the parameter resources, the hidden size of MLP is set to $R^{C/r}$, where r is defined as the reduction rate, and the formula is shown below:

$$F_{\text{mlp\_avg}} = \text{MLP}(F_{avg}) = W_2(relu(W_1(F_{avg}))), \tag{6}$$

$$F_{\text{mlp\_max}} = \text{MLP}(F_{max}) = W_2(relu(W_1(F_{max}))), \tag{7}$$

where $W_1 \in R^{C/r \times C}{}_1$, $W_2 \in R^{C \times C/r}$ stands for the MLP weights and *relu* represents the active function ReLU, respectively. W_1 and W_2 are shared by both F_{avg} and F_{max}.

After obtaining the statistical information $F_{\text{mlp\_avg}}$ and $F_{\text{mlp\_max}}$ by MLP, the probability prediction matrix, which is the importance of each channel, can be obtained by element-wise summing and passing through the sigmoid function. Finally, the matrix generated by a sigmoid function is element-wise multiplied with the input matrix F to obtain the output, which is calculated by equation:

$$\text{CAM}(F) = F \times \sigma\left(F_{\text{mlp\_avg}} + F_{\text{mlp\_max}}\right), \tag{8}$$

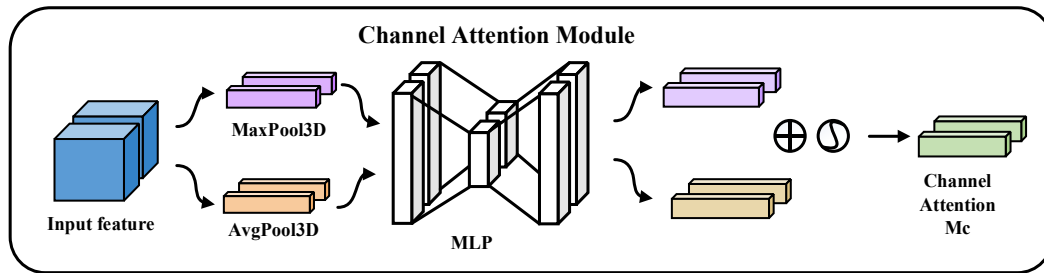where σ is the sigmoid function. Figure 4 shows the flowchart of CAM.

Figure 4. Channel attention module.

The feature matrix $F_c = \text{CAM}(F)$, which is output by the channel attention module, is taken as the input feature matrix of a spatial attention module. Firstly, we use global max pooling and global average pooling based on the channel to get two feature maps: $F_{c\_avg} \in R^{B \times T \times H \times W \times 1}$ and $F_{c\_max} \in R^{B \times T \times H \times W \times 1}$. Then, they are concatenated at the channel dimension and passed through a $3 \times 3 \times 3$ convolution to generate a feature descriptor. The spatial attention feature is generated through sigmoid activation function. Then, we multiply the spatial attention matrix with the input matrix F_c to obtain the output result, which is calculated by equation:

$$\text{SAM}(F_c) = F_c \times \sigma\left(f_{conv}^{3 \times 3 \times 3}\left([F_{c\_avg}; F_{c\_max}]\right)\right), \tag{9}$$

where σ is the sigmoid function. Figure 5 shows the flowchart of SAM.
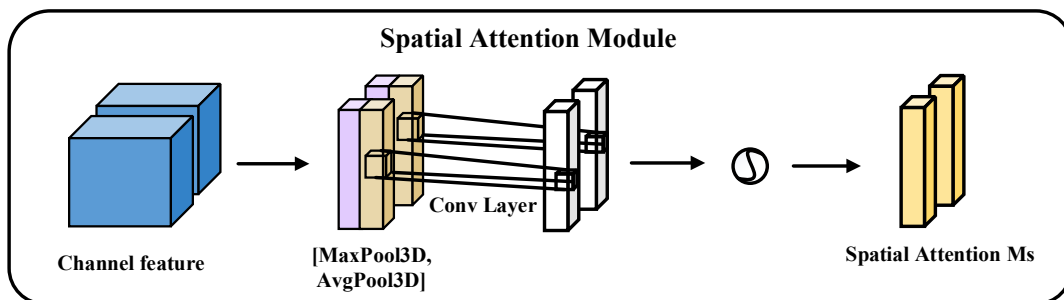

Figure 5. Spatial attention module.

4.3. Time Feature Extraction with Attention Mechanism

SST forecast correction is actually a spatio-temporal series problem with historical information as the input and revised SST as the output. LSTM has a strong ability to modeling time series data. ConvLSTM [27] inherits the merits of convolution operator and retains the advantages of LSTM to capture long-term memory, and can also reduce the redundancy of the fully connected structure. So that, ConvLSTM is used to model the temporal and spatial correlation of SST data. The input of the ConvLSTM in correction model is $X = \text{CBAM}(F)$, $X \in R^{B \times T \times H \times W \times C}$. The formula is shown in Equation (10):

$$H = \text{ConvLSTM}(X), \tag{10}$$

where H consists of the results h_t computed by ConvLSTM for each sample x of input data X, $x \in R^{T \times H \times W \times C}$. At each moment, since the interval time of data is one day, the ConvLSTM unit accepts the input x_t, $t = 1, 2, \cdots T$ at the moment of t, the state of the hidden layer at the last time h_{t-1}, and the state of the memory cell at the last time c_{t-1} as inputs, and outputs the hidden state h_t and the cell state c_t. The calculation process is as follows:

$$h_t = o_t \cdot tanh(c_t) , \tag{11}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot tanh(w_{xc} * x_t + w_{hc} * h_{t-1} + b_c). \tag{12}$$

As shown in Equations (10) and (11), $*$ and $\cdot$ denote the convolution operator and Hadamard product. w is the weight matrix, b_c is the offset, and *tanh* represents the activation function.

The ConvLSTM forgets and remembers the input information through four gates. The forgetting gate determines what information should be discarded from the c_{t-1} of the previous moment, the input gate determines what new information should be stored in the memory of the ConvLSTM, and the output gate determines what information should be selected from the c_t to be passed as output to the next ConvLSTM unit. The involved computation is given as follows in Equations (12)–(14):

$$i_t = \sigma(w_{xi} * x_t + w_{hi} * h_{t-1} + w_{ci} \cdot c_{t-1} + b_i) , \tag{13}$$

$$f_t = \sigma\left(w_{xf} * x_t + w_{hf} * h_{t-1} + w_{cf} \cdot c_{t-1} + b_f\right), \tag{14}$$

$$o_t = \sigma(w_{xo} * x_t + w_{ho} * c_{t-1} + w_{co} \cdot c_t + b_o) , \tag{15}$$

where i_t indicates the input gates, f_t indicates the forgotten gates, o_t indicates the output gates, and c_t indicates the cell state. In the above formulas, σ is the active function sigmoid, x_t represents the moment's input, ht−1 represents last time's hidden state, w is the weight matrix, and b is the offset from the input gate to the output gate, which are the characteristics that the ConvLSTM model must learn during training.

In order to improve the quality of the model by giving different weights to different parts of the model and make the model more focused on the parts that are more relevant to the task, a temporal attention layer is added after the ConvLSTM layer. To make full use of the hidden layer state of each step of the ConvLSTM model, we allocate the temporal attention weight to the hidden state of each time step, and adjust the final ConvLSTM output and thus obtain better correction results.

The attention [50] module assigns weight coefficients to the outputs of the ConvLSTM layer. It pays more attention to the features that contribute more to the important information and ignores useless information to reduce the calculations of the network and save storage space. The attention mechanism shown in Figure 6 provides an efficient way to aggregate the output sequence of ConvLSTM layer and it implements the following equation:

$$AT(h_t) = \frac{exp(W \cdot h_t)}{\sum_{t=1}^{T} exp(W \cdot h_t)}. \tag{16}$$
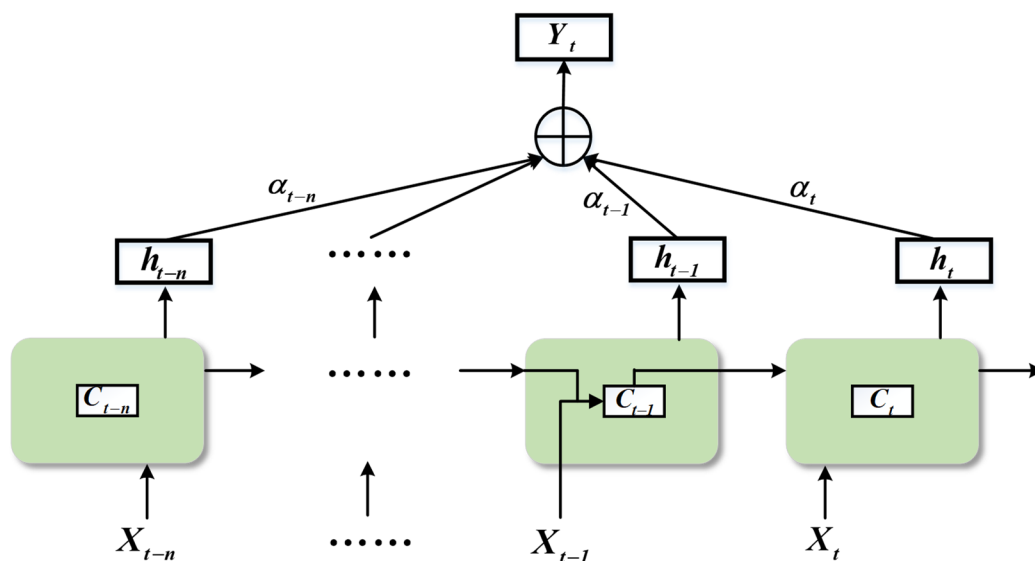
Figure 6. The ConvLSTM layer and attention layer.

The attention layer takes the output h_t of each iteration of ConvLSTM as input. At time t, normalized weights $AT(h_t)$ are computed by the softmax function through the weight W and the output h_t of the ConvLSTM, the calculation formula is shown in (16).

$$Y_t = \sum_{t=1}^{T} h_t AT(h_t). \tag{17}$$

Finally, the output Y_t can be obtained by multiplying the attention weight AT with the hidden layer state h_t. The calculation formula is shown in (17).

5. Experiments and Results

5.1. Data Preparation and Evaluation Metrics

The HYbrid Coordinate Ocean Model (HYCOM) [51] is a data-assimilative hybrid isopycnal-sigma-pressure (generalized) coordinate ocean model. The US Navy Operational Global Ocean Prediction System based on the HYCOM model is a relatively advanced and widely used ocean prediction system [52]. In the experiments, we use HYCOM model forecast product from National Oceanic and Atmospheric Administration (NOAA) as the prediction data to be corrected. The HYCOM model prediction product used in our experiments is a prediction product of 24 h in the future, which is reported every 3 h and includes ocean temperature, salinity, and current structure. Its horizontal resolution is $1/12°$ and the temporal resolution is 3 h.

There is a lack of ocean observation data to support our experiments, SST products data is a relatively good choice as the truth value. We use the NOAA OI SST [53] Analysis version 2(v2), which is acquired from the NOAA's National Climatic Data Center (NCDC) with a high spatial resolution of $0.25° \times 0.25°$ as truth values to evaluate correction accuracy. Liu et al. [54] showed that NOAA OI SST is the best one among the SST products when they were compared with in situ SST data This dataset was generated from several data sources including SST data from the Advanced Very High Resolution Radiometer (AVHRR), sea-ice data, and in situ data from ships and buoys. In order to unify the spatial and temporal resolution of the forecast data and remote sensing observation data, we average the HYCOM data daily and the daily average HYCOM model forecast data is interpolated to OI SST data grid points by using bilinear interpolation method. We select a dataset from January 2019 to December 2019, that covers the area from 8°N to 12°N in latitude and 110° E to 114° E in longitude. Figure 7 shows the location of test area.
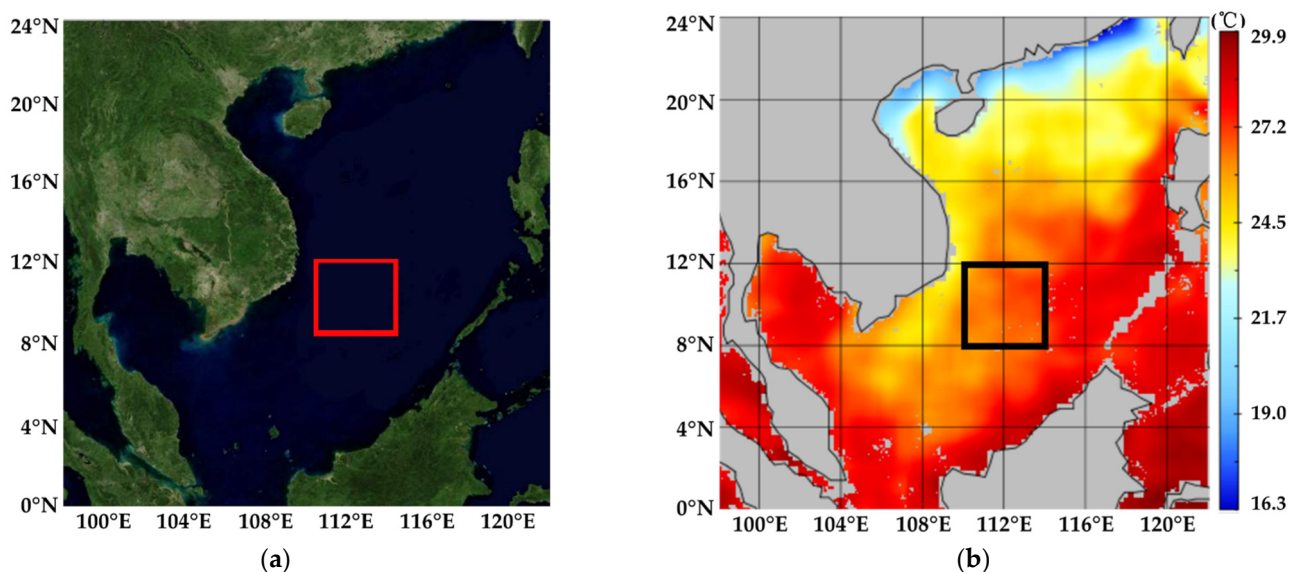
Figure 7. The location of test area: (**a**) Satellite image of the test area location, the box is the test area; (**b**) SST map of the test area location, the box is the test area.

Due to the obvious discrepancy of data value, we apply a normalization to each input sequence of data before inputting it into our model. The normalization operation can not only improve the convergence speed of the model, but also improve the accuracy of the model and prevent the gradient explosion of the model. The normalization function is shown in Equation (19):

$$X = \frac{X - X_{min}}{X_{max} - X_{min}}. \tag{18}$$

After correction, the output data and the truth would go through a de-normalization. The parameters of the de-normalization are based on the temperature span of the original input sequence of data.

In order to verify the validity of the new hybrid SST correction model, this study evaluated the model with four indexes, namely mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). MSE, RMSE, MAE, and MAPE can be defined as:

$$\text{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{19}$$

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{20}$$

$$\text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{21}$$

$$\text{MAPE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \tag{22}$$

where y_i represents the actual observed value, $\hat{y}$ represents the average of the actual observed values, and $\hat{y}_i$ represents the correction value.

The model is built by Pytorch. In order to prove the effectiveness of the model proposed in this paper, all SST data were divided into two parts. The dataset from January 2019 to September 2019 is used as the training set to train the parameters of the new hybrid SST correction model, and the remaining dataset from October 2019 to December 2019 is used as the verification set to verify the learning effect of the model. We adjusted the shape of the training data and the input data to the required Tensor format in the Pytorch framework. Then the parameters of the new hybrid SST correction model were defined,

including the input step length, the length of input sequence, the hidden layers, the length of output sequence, and the number of neurons in each layer. In our experiments, the convolution part of the model includes a Conv3D layer and a Batch Normalization layer. The main function of the Batch Normalization layer is to make the distribution of the input data of each layer in the network relatively stable, to accelerate the model learning speed, alleviate the problem of gradient disappearance, and have a certain regularization effect. When the network is set up, the size of the convolution kernel in Conv3D layer is $3 \times 3 \times 3$ and the size of the convolution kernel in the convolution attention in the 3D-CBAM part of the model is $3 \times 3 \times 3$. The activation function of all layers is 'relu', which can keep the convergence speed of the model in a steady state. In the ConvLSTM part of the model, due to the short sequence selected in the experiment, only a single layer ConvLSTM was selected, the number of neurons in the hidden layer was 32, and the number of neurons in the output layer was 1. After the model parameters are defined, we select MSE and Adam as the loss function and optimizer. Then, the appropriate training times are defined to start training the model. After the model training is completed, the test data is input into the model for testing, and the output results of the model are reversely normalized to obtain the deviation of SST. The effect of correction is tested by comparing the evaluation metrics before and after deviation correction.

5.2. Comparison of Correction Methods

In order to prove the validity of the proposed new hybrid SST correction model, the experimental results will be compared with two traditional machine learning methods for SST correction. They are Linear Regression (LR) and Support Vector Regression (SVR), respectively. The LR model has the advantages of strong anti-interference ability and fast training speed, but it cannot simulate nonlinear relations, as the accuracy is not very high, and it is easy to lack fitting. The SVR model has good generalization performance, it is not easy to overfit, and it can achieve good performance with less data. However, SVR is sensitive to missing data, parameters, and kernel function.

The process of realizing the two methods in this paper is to expand all samples into the form that the algorithm can handle, and use the machine learning package, sklearn, for correction analysis. In order to match the input form of these methods, SST and ocean variables were generally regarded as independent features, thus the spatial-temporal relationship between variables could not be considered in these methods. For both LR and SVR, we combined vector of SST, SSS, and water current u and v of HYCOM forecasts for days *n*, *n*-1, *n*-2, *n*-3 as 4×4 features for one-day correction. The performances of these two methods are shown in Table 1. For SVR, we use the Radial Basis Function (RBF) kernel for correction, which can realize nonlinear mapping with few parameters.

In addition, we set up the comparison experiment, which only considers the temporal relationship without considering the spatial relationship, namely, compared with the traditional sequence model LSTM to enhance the contrast. For the LSTM network, we set the learning rate = 0.01, epochs = 300, timestep = 3, and use SST, SSS, and water current u and v for one-day correction. Our new hybrid correction model also compares the traditional sequence model LSTM and its improved model ConvLSTM, which considers temporal relations and spatial relations.

Furthermore, we develop and compare a series of models with the new hybrid SST correction model (3DCNN-CBAM-CONVLSTM-AT), because there were fewer methods previously used to correct SST. These include an improved ConvLSTM model that combines 3D-convolution, a ConvLSTM model that only adds temporal attention mechanism (AT), and a ConvLSTM hybrid model where both 3D-convolution and AT are added. Here, we set the experimental parameters learning rate = 0.01 and epochs = 300, the size of convolution kernel in Conv3D layer is $3 \times 3 \times 3$, and used three days of historical data for SST correction. The input data form of these models is consistent with our new model input data.

Table 1. The experimental results of SST correction. Bold entries show the best results.

	MAPE	MAE	MSE	RMSE	Improve
Forecast	1.6118	0.4587	0.3600	0.6000	
Linear Regression (LR)	1.4592	0.4075	0.3005	0.5482	8.67%
Support Vector Regression (SVR)	1.3767	0.3832	0.2536	0.5036	16.17%
LSTM	1.2781	0.3553	0.2115	0.4599	23.35%
CONVLSTM	1.1679	0.3312	0.1842	0.4292	28.47%
CONVLSTM-AT	1.1071	0.3139	0.1623	0.4028	32.92%
3DCNN-CONVLSTM-AT	1.0033	0.2839	0.3600	0.3690	38.5%
3DCNN-CBAM-CONVLSTM-AT	**0.9546**	**0.2641**	**0.1239**	**0.3520**	**41.33%**

Table 1 shows the experimental results of different correction methods for SST correction. When we use the traditional machine learning method to correct SST, the accuracy of SVR is higher than Linear Regression; the RMSE value is 0.5036 and the MAE value is 0.3832. The accuracy is not greatly improved after the correction.

Among other deep learning models, 3DCNN-ConvLSTM-AT has the best results; the RMSE value is 0.3690 and the MAE value is 0.2839. However, our new hybrid correction model can achieve a level where the MSE value is 0.3520 and the MAE value is 0.2641 in the correction experiment, which is better than the other models.

It can be seen from Table 1 that the effects of LSTM are better than traditional machine learning methods, which illustrates the importance of time correlation in SST data. However, the original LSTM does not consider the spatial relations in data. The result of ConvLSTM, an improved method, is better than the result of LSTM, which verifies the importance of spatial correlation to SST correction. Experimental results show that ConvLSTM-AT model, which adds attention mechanism, has better performance than ConvLSTM. Attention mechanism can assign different weights to historical data, allowing the model to focus more on the parts that are more important, thus improving the quality of the model. We compared the results of 3DCNN-CONVLSTM-AT with CONVLSTM-AT, in which 3DCNN-CONVLSTM-AT added a convolution layer. With the same parameters and the same input, the RMSE of ConvLSTM-AT and 3DCNN-CONVLSTM-AT were 0.4028 and 0.3690, respectively. 3DCNN-ConvLSTM-AT has a higher correction accuracy than ConvLSTM-AT. The experimental result shows that the addition of the convolution layer can improve the accuracy of SST correction to a certain extent. The main reason for this is that the local features extracted from input data through ConvLSTM's own convolution operation is not obvious enough. A convolution layer is added into the model, which improves the feature extraction ability of the model and makes the spatial features of the data more obvious in the ConvLSTM model, which is beneficial to improve the accuracy of SST correction.

After adding 3D-CBAM attention mechanism on the basis of the 3DCNN-CONVLSTM-AT model, the RMSE index is 0.3520, and the correction effect is the best in our experiments. 3DCBAM mechanism and AT mechanism were used based on ConvLSTM in our new hybrid correction model to improve the utilization rate of spatial features, environmental variables, and historical time series information.

To further prove the effectiveness of our new hybrid SST correction model, we visualize the correction results, forecast results, and the truth in Figure 8, which shows the comparison of the revised SST of several models. To put things into places in the overall view, there is high similarity between the correction that is shown in Figure 8i and the truth that is shown in Figure 8a. Combined with Figures 8 and 9 and Table 1, it can be seen intuitively from the figure that the result of the new hybrid SST correction model is closest to the truth value. The new hybrid correction model further extracts spatial features and adds weights to environmental information and spatial features to improve information utilization, making the model closer to reality and containing more comprehensive information, and finally improving the accuracy of SST prediction. In conclusion, compared with LR, SVR, and other traditional machine learning correction methods, as well as deep

learning methods LSTM, ConvLSTM, and ConvLSTM-AT, the new hybrid correction model has the best performance in SST correction, which verifies the effectiveness of this method.

For SST correction, Zhang et al. [47] proposed a new bias correction model for sea surface temperature in 2020, which used satellite remote sensing data for correction of the numerical forecast model on SST in the South China Sea as well. After being corrected, the RMSE of the SST forecast results was dropped from 0.8 °C to 0.5 °C, reducing by 37.5%, whereas the RMSE of our model is approximately 0.35 °C after being corrected, reducing by 41.33%. The SST correction by our new hybrid SST correction offers higher accuracy.
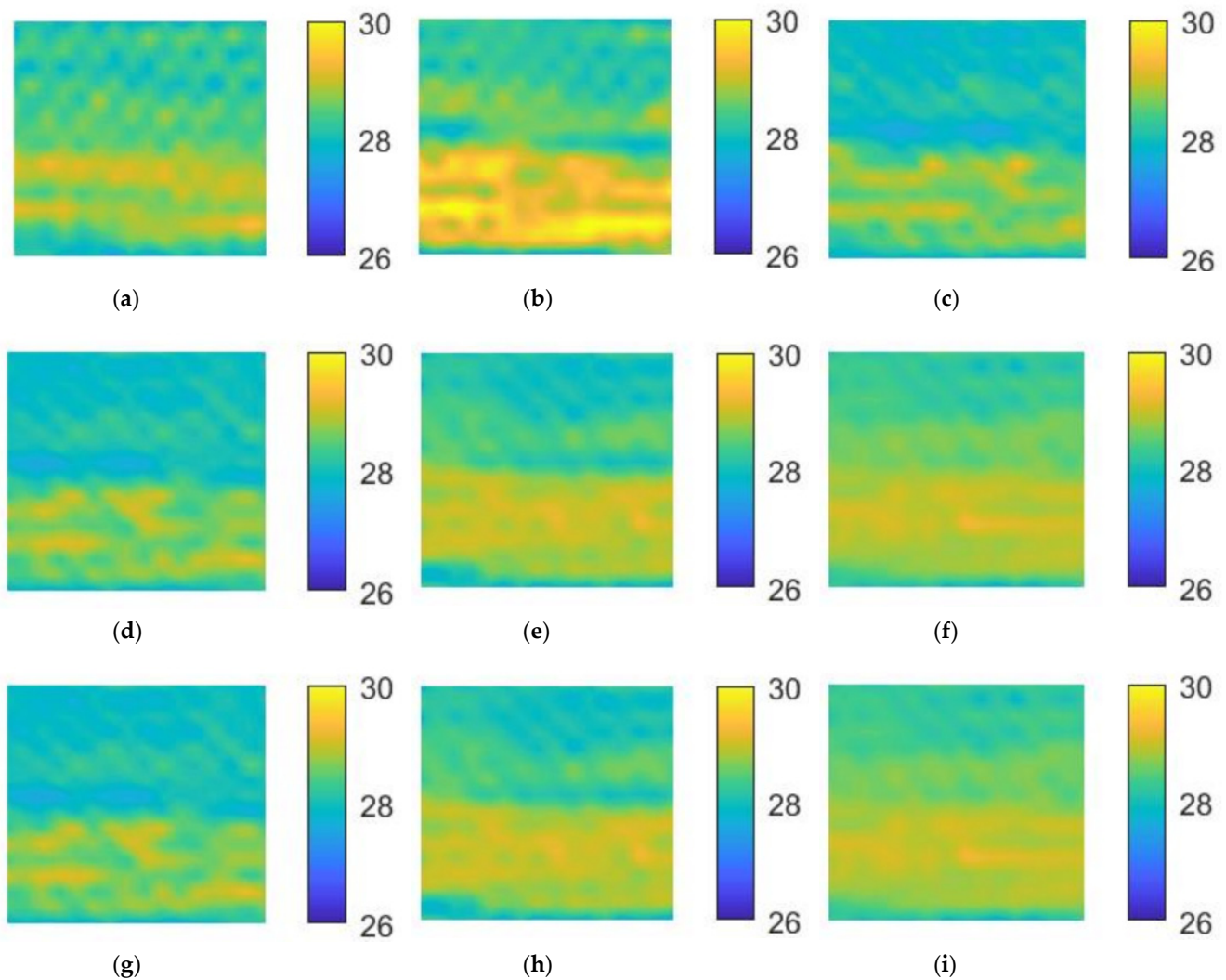
Figure 8. The experimental results of different methods for SST correction. (**a**) Truth; (**b**) forecast; (**c**) linear regression; (**d**) SVR; (**e**) LSTM; (**f**) CONVLSTM; (**g**) CONVLSTM-AT; (**h**) 3DCNN-CONVLSTM-AT; (**i**) 3DCNN-CBAM-CONVLSTM-AT.
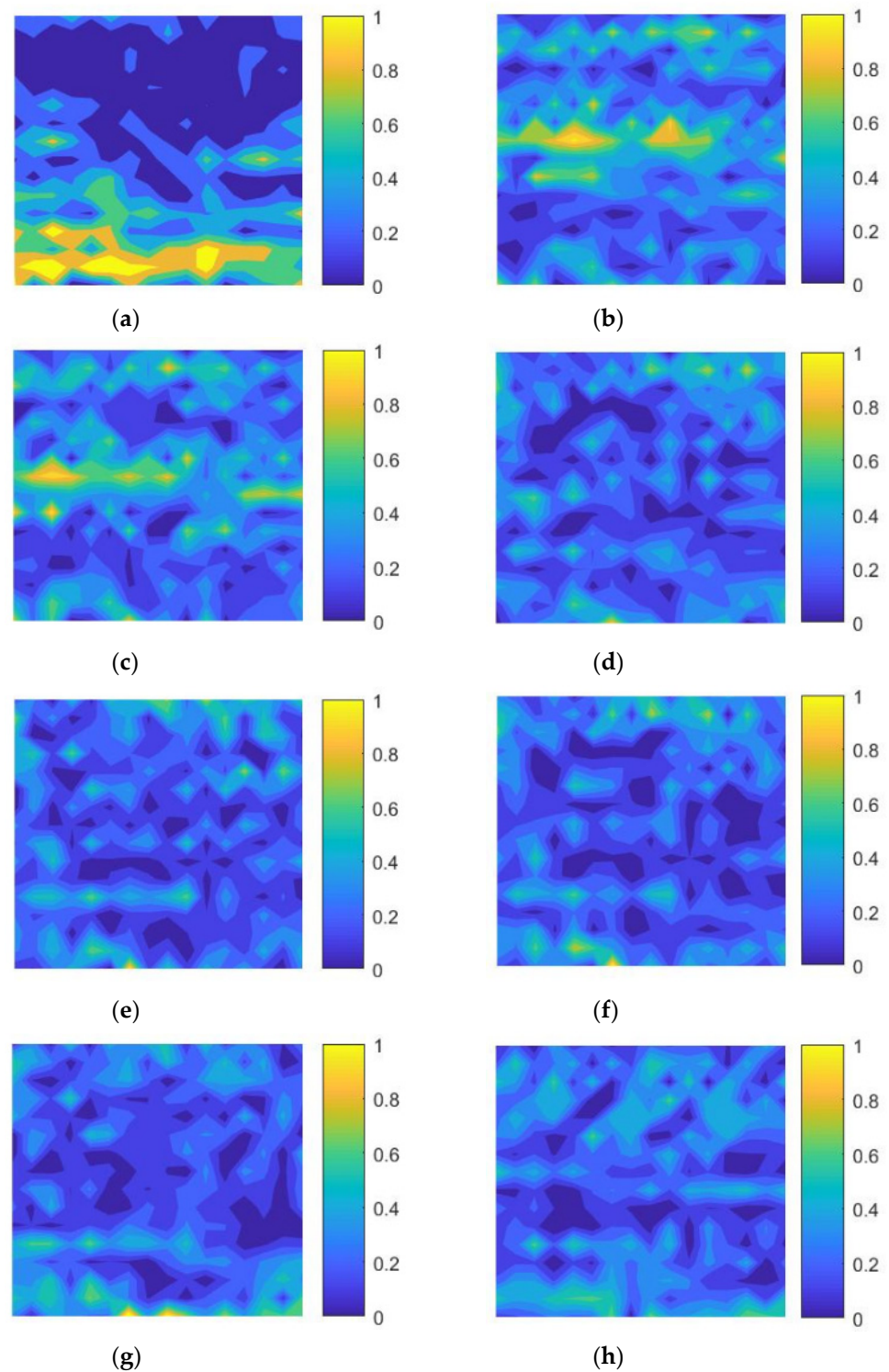
Figure 9. The comparisons of difference between the truth and the correction output. (**a**) Difference between the truth and the forecast; (**b**) difference between the truth and the linear regression result; (**c**) difference between the truth and the SVR result; (**d**) difference between the truth and the LSTM result; (**e**) difference between the truth and the CONVLSTM result; (**f**) difference between the truth and the CONVLSTM-AT result; (**g**) difference between the truth and the 3DCNN-CONVLSTM-AT result; (**h**) difference between the truth and the 3DCNN-CBAM-CONVLSTM-AT result.

5.3. Complexity and Training Time Analysis

The experimental environment is Windows10, Intel Core i5 11, 2.4 GHz, 16G RAM, with algorithm implementation using python3.

Table 2 lists the training time and the parameters of models used in the experiment. It can be found that the training parameters of the new hybrid SST correction model are about three times less than those of ConvLSTM, which makes the training much faster and more suitable for practical application. Our proposed new hybrid SST correction model consumes the least time and has fewer parameters. The parameter of 3DCNN-ConvLSTM-AT model is close to that of the new hybrid SST correction model, indicating that the 3D-CBAM module is very small and the training time of the model is reduced. Our proposed new hybrid SST correction model consumes the least time and has fewer parameters, and it has good performance.

Table 2. The number of network parameters and training time for each model.

	Parameters	Train(s)	Test(s)
LSTM	13,601	271.52	0.55
CONVLSTM	44,993	236.98	0.46
CONVLSTM-AT	46,079	437.67	0.94
3DCNN-CONVLSTM-AT	13,197	272.57	0.55
3DCNN-CBAM-CONVLSTM-AT	13,560	223.15	0.43

5.4. Parameters Analysis

5.4.1. Time Step Analysis

In the previous experiments to determine the model structure, the previous three days of data is used to correct the SST according to expert empirical knowledge. Time step is an important parameter for the model to learn time series character. Considering that the size of timestep has an impact on the accuracy of SST correction, timestep = 1, 3, 5, 7, 10, 15 is used to correct SST in our experiments to determine the appropriate timestep for SST correction.

Timestep represents the information of the time dimension, which has an impact on the performance of the model. Figure 10 shows the variation of the model of several evaluation indicators with the timestep size. When timestep = 3, RMSE is 0.35, which is better than others when timestep = 1, 5, 7, 10, and 15. It is obviously seen from the figure that timestep = 3 works best to revise SST. When the timestep is greater than 10, the results of correction tend to be stable, and the time information has less influence on the revised results. When correcting SST, the information of temporal dimensions should be moderate, as too much or too little will affect the performance of the model. To sum up, timestep = 3 is used in this paper to correct SST.

5.4.2. Learning Rate Analysis

Learning rate is an important hyperparameter, which determines whether and when the objective function converges to the local minimum. The proper learning rate can make the objective function converge to the local minimum in the proper time. Then we adjust the learning rate and other hyperparameters within the fixed model frame. The first step is to drop from 0.1 to 0.001, at a speed of 10. Then, when the learning rate is at the level between 0.01 and 0.001, the training and validation loss of the model will be in a steady state. The experiment is conducted by adjusting the learning rate, and the experimental results are shown in Figure 11. The figure shows that RMSE, MAPE, and other indicators change with the learning rate. According to RMSE, the optimal learning rate is 0.01. According to MAPE, the optimal LR is 0.004. Thus, the best learning rate in our data set is at the level between 10^{-2} and 10^{-3}.
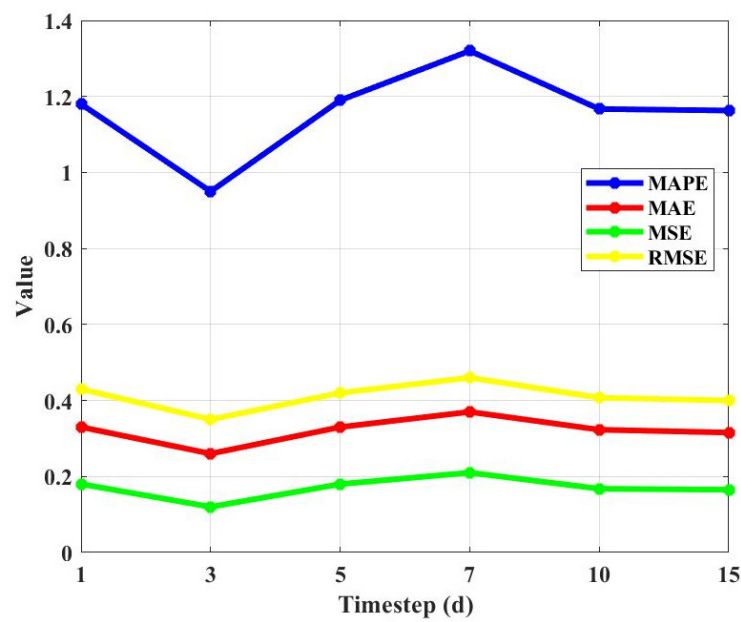
Figure 10. The experimental results of the new hybrid SST correction model in different timesteps. The units of RMSE, MAE, and MSE are °C, the unit of MAPE is %, and the unit of timestep is day.
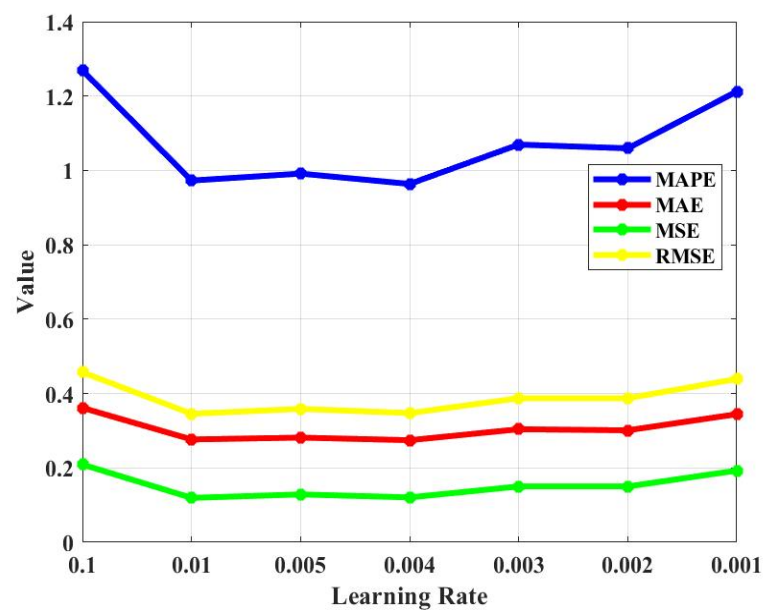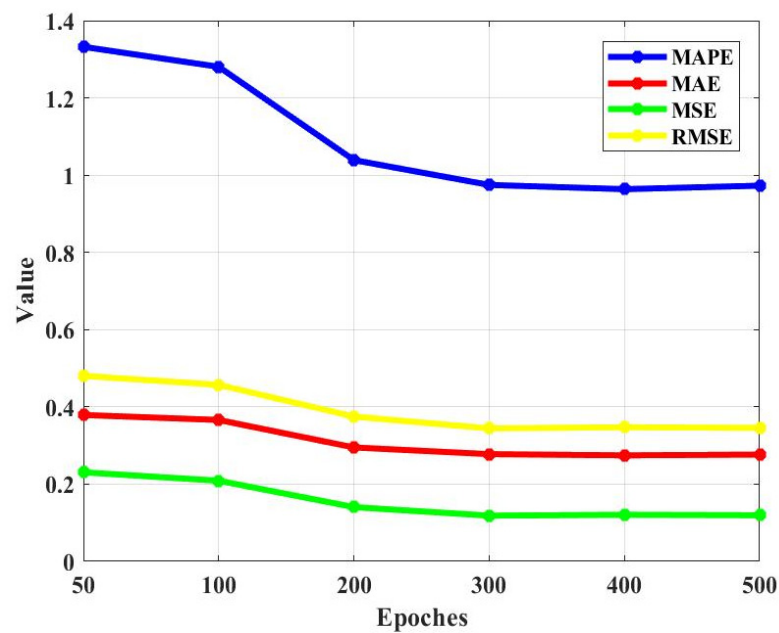

Figure 11. The experimental results of the new hybrid SST correction model in different learning rate. The units of RMSE, MAE, and MSE are °C and the unit of MAPE is %.

5.4.3. Epochs Analysis

In order to determine the best epochs for the dataset, different epochs were set for the experiments. The experimental results are shown in Figure 12. The figure shows that RMSE reaches a stable state at 300 epochs. Therefore, 300 epochs are suitable for our experiments with consideration of model accuracy and performance.

Figure 12. The experimental results of the new hybrid SST correction model in different epochs. The units of RMSE, MAE, and MSE are °C and the unit of MAPE is %.

6. Conclusions

In this paper, the new hybrid SST correction model is applied to correct the HYCOM forecasts and it is evaluated for its performance. Our proposed model combines spatio-temporal information and marine environmental variables information to correct the SST forecast and improve the accuracy of the SST forecast. The model defines the SST correction problem as the spatio-temporal series regression problem, which mainly consists of three parts: first, 3D convolution and 3D-CBAM are used to improve the utilization rate of spatial features and marine environmental variables. Secondly, time and space characteristics of SST were extracted by ConvLSTM. Thirdly, the attention mechanism is used to enhance the historical temporal information. What is more, the new hybrid SST correction has a better correction effect than the other models we compared in this paper, and it can reduce the RMSE of the HYCOM forecast results by 41.33%.

As for future development, further refinements to the new hybrid SST correction model will be undertaken. Our study only corrects the temperature of the sea surface, but the subsurface temperature in the inner ocean is much more important. Therefore, in the next step, we consider extending the model to three-dimensional space to realize the forecast correction of ocean internal temperature. Meanwhile, in this paper, we only revised the forecast data for the next day due to the limitation of forecast data. For future development, our correction model can be considered to improve and apply to correct the forecast of three days, five days, or one month into the future.

Author Contributions: Conceptualization, X.W. and B.H.; methodology, T.F.; validation, J.Z. and T.F.; formal analysis, T.F., X.W., B.H., J.Z., H.W., Y.C. and W.Z.; supervision, X.W.; data curation, H.W. and T.F.; writing—original draft preparation, T.F.; writing—review and editing, X.W., B.H. and Y.C.; visualization, T.F.; project administration, W.Z.; funding acquisition, W.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: SST OI data sets were obtained from https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.highres.html (accessed on 10 July 2021); HYCOM forecast data sets were obtained from https://www.ncei.noaa.gov/thredds-coastal/catalog/hycom_sfc/catalog.html (accessed on 3 July 2021).

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

References

1. Funk, C.C.; Hoell, A. The leading mode of observed and cmip5 enso-residual sea surface temperatures and associated changes in indo-pacific climate. *J. Clim.* **2015**, *28*, 150202132719008. [CrossRef]
2. Solanki, H.U.; Bhatpuria, D.; Chauhan, P. Integrative analysis of altika-ssha, modis-sst, and ocm-chlorophyll signatures for fisheries applications. *Mar. Geod.* **2015**, *38* (Suppl. 1), 672–683. [CrossRef]
3. Yang, Y.; Dong, J.; Sun, X.; Lima, E.; Mu, Q.; Wang, X. A cfcc-lstm model for sea surface temperature prediction. *IEEE Geosci. Remote Sens. Lett.* **2017**, *15*, 207–211. [CrossRef]
4. Stockdale, T.N.; Balmaseda, M.A.; Vidard, A. Tropical atlantic sst prediction with coupled ocean-atmosphere gcms. *J. Clim.* **2006**, *19*, 6047. [CrossRef]
5. Song, Z.; Qiao, F.; Yang, Y.; Yuan, Y. An improvement of the too cold tongue in the tropical pacific with the development of an ocean-wave-atmosphere coupled numerical model. *Prog. Nat. Sci.* **2007**, *17*, 576–583.
6. Xu, Z.; Li, M.; Patricola, C.M.; Ping, C. Oceanic origin of southeast tropical atlantic biases. *Clim. Dyn.* **2014**, *43*, 2915–2930. [CrossRef]
7. Peng, S.Q.; Xie, L. Effect of determining initial conditions by four-dimensional variational data assimilation on storm surge forecasting. *Ocean Model.* **2006**, *14*, 1–18. [CrossRef]
8. Li, X.; Wang, Q.; Mu, M. Optimal initial error growth in the prediction of the kuroshio large meander based on a high-resolution regional ocean model. *Adv. Atmos. Sci.* **2018**, *35*, 1362–1371. [CrossRef]
9. Hemri, S.; Scheuerer, M.; Pappenberger, F.; Bogner, K.; Haiden, T. Trends in the predictive performance of raw ensemble weather forecasts. *Geophys. Res. Lett.* **2014**, *41*, 9197–9205. [CrossRef]
10. Vannitsem, S. Dynamical properties of mos forecasts: Analysis of the ecmwf operational forecasting system. *Weather Forecast.* **2010**, *23*, 1032–1043. [CrossRef]
11. Tian, D.; Martinez, C.J.; Graham, W.D.; Hwang, S. Statistical downscaling multimodel forecasts for seasonal precipitation and surface temperature over the southeastern united states. *J. Clim.* **2014**, *27*, 8384–8411. [CrossRef]
12. Libonati, R.; Trigo, I.; Dacamara, C.C. Correction of 2m-temperature forecasts using kalman filtering technique. *Atmos. Res.* **2008**, *87*, 183–197. [CrossRef]
13. Pelosi, A.; Medina, H.; Bergh, J.V.D.; Vannitsem, S.; Chirico, G.B. Adaptive kalman filtering for postprocessing ensemble numerical weather predictions. *Mon. Weather Rev.* **2017**, *145*, 4837–4854. [CrossRef]
14. Wang, J.; Chen, C.; Long, K.; Feng, L. Temporal and spatial distribution of short-time heavy rain of Sichuan Basin in summer. *Plateau Mt. Meteorol. Res.* **2015**, *35*, 16–20.
15. Zhang, Q.; Yu, Y.; Zhang, W.; Luo, T.; Wang, X. Cloud detection from fy-4a's geostationary interferometric infrared sounder using machine learning approaches. *Remote Sens.* **2019**, *11*, 3035. [CrossRef]
16. Zeng, J.; Zhang, C.; Wang, H.; Chu, H. Correction model for the temperature of numerical weather prediction by SVM. *Second Target Recognit. Artif. Intell. Summit Forum* **2020**, *11427*, 114270Z.
17. Zhang, R.; Yu, Z.H.; Jiang, Q.R. Neural network bp model approximation and prediction of complicated weather systems. *Acta Meteorol. Sin.* **2001**, *15*, 105–115.
18. Sayeed, A.; Choi, Y.; Jung, J.; Lops, Y.; Eslami, E.; Salman, A.K. A deep convolutional neural network model for improving WRF forecasts. *Atmos. Environ.* **2020**, *253*, 118376. [CrossRef]
19. Kupilik, M.; Witmer, F.D.W.; MacLeod, E.-A.; Wang, C.; Ravens, T. Gaussian Process Regression for Arctic Coastal Erosion Forecasting. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1256–1264. [CrossRef]
20. Brekke, C.; Solberg, A. Oil spill detection by satellite remote sensing. *Remote Sens. Environ.* **2005**, *95*, 1–13. [CrossRef]
21. Yu, Y.; Yang, X.; Zhang, W.; Duan, B.; Cao, X.; Leng, H. Assimilation of sentinel-1 derived sea surface winds for typhoon forecasting. *Remote Sens.* **2017**, *9*, 845. [CrossRef]
22. Chen, R.; Zhang, W.; Wang, X. Machine learning in tropical cyclone forecast modeling: A review. *Atmosphere* **2020**, *11*, 676. [CrossRef]
23. Xi, X.F.; Zhou, G.D. A survey on deep learning for natural language processing. *Acta Autom. Sin.* **2016**, *42*, 1445–1465.
24. Lee, H.; Pham, P.T.; Largman, Y.; Ng, A.Y. Unsupervised feature learning for audio classification using convolutional deep belief networks. *Adv. Neural Inf. Process. Syst.* **2009**, *22*, 1096–1104.
25. Sattar, N.S.; Arifuzzaman, S. Community Detection using Semi-supervised Learning with Graph Convolutional Network on GPUs. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 5237–5246.

26. Jain, V.; Murray, J.F.; Roth, F.; Turaga, S.; Zhigulin, V.; Briggman, K.L.; Helmstaedter, M.N.; Denk, W.; Seung, H.S. Supervised Learning of Image Restoration with Convolutional Networks. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007; pp. 1–8.

27. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. *Convolutional Lstm Network: A Machine Learning Approach for Precipitation Nowcasting*; MIT Press: Cambridge, MA, USA, 2015.

28. Liu, D.; Jiang, W.; Mu, L.; Wang, S. Streamflow Prediction Using Deep Learning Neural Network: Case Study of Yangtze River. *IEEE Access* **2020**, *8*, 90069–90086. [CrossRef]

29. Petrou, Z.I.; Tian, Y. Prediction of sea ice motion with convolutional long short-term memory networks. *IEEE Trans. Geosci. Remote Sens.* **2019**, *99*, 1–12. [CrossRef]

30. Chen, R.; Wang, X.; Zhang, W.; Zhu, X.; Li, A.; Yang, C. A hybrid cnn-lstm model for typhoon formation forecasting. *GeoInformatica* **2019**, *23*, 375–396. [CrossRef]

31. Winona, A.Y.; Adytia, D. Short Term Forecasting of Sea Level by Using LSTM with Limited Historical Data. In Proceedings of the 2020 International Conference on Data Science and Its Applications (ICoDSA), Bandung, Indonesia, 5–6 August 2020; pp. 1–5.

32. Kun, X.; Shan, T.; Yi, T.; Chao, C. Attention-based long short-term memory network temperature prediction model. In Proceedings of the 2021 7th International Conference on Condition Monitoring of Machinery in Non-Stationary Operations (CMMNO), Guangzhou, China, 11–13 June 2021.

33. Krishnamurti, T.N.; Kishtawal, C.M.; LaRow, T. Improved weather and seasonal climate forecasts from multimodel superensemble. *Science* **1999**, *285*, 1548–1550. [CrossRef]

34. Xu, Z.; Wang, Y.; Fan, G. A two-stage quality control method for 2-m temperature observations using biweight means and a progressive eof analysis. *Mon. Weather Rev.* **2013**, *141*, 798–808. [CrossRef]

35. Zhang, X.; Gao, S.; Wang, T.; Li, Y.; Ren, P. Correcting Predictions from Simulating Wave Nearshore Model via Gaussian Process Regression. In Proceedings of the Global Oceans 2020: Singapore—U.S. Gulf Coast, Biloxi, MS, USA, 5–30 October 2020; pp. 1–4.

36. Doroshenko, A.; Shpyg, V.; Kushnirenko, R. Machine Learning to Improve Numerical Weather Forecasting. In Proceedings of the 2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 25–27 November 2020.

37. Wang, X.; Li, X.; Zhu, J.; Xu, Z.; Yu, K. A local similarity-preserving framework for nonlinear dimensionality reduction with neural networks. In Proceedings of the The 26th International Conference on Database Systems for Advanced Applications (Dasfaa 2021), Tai Pei, China, 11–14 April 2021.

38. Wang, A.; Xu, L.; Li, Y.; Xing, J.; Zhou, Z. Random-forest based adjusting method for wind forecast of WRF model. *Comput. Geosci.* **2021**, *55*, 104842. [CrossRef]

39. Zheng, G.; Li, X.; Zhang, R.H.; Liu, B. Purely satellite data–driven deep learning forecast of complicated tropical instability waves. *Sci. Adv.* **2020**, *6*, eaba1482. [CrossRef] [PubMed]

40. Makarynskyy, O. Improving wave predictions with artificial neural networks. *Ocean Eng.* **2004**, *31*, 709–724. [CrossRef]

41. Xu, X.; Liu, Y.; Chao, H.; Luo, Y.; Chu, H.; Chen, L. Towards a precipitation bias corrector against noise and maldistribution. *arXiv* **2019**, arXiv:1910.07633.

42. Wang, T.; Gao, S.; Xu, J.; Li, Y.; Li, P.; Ren, P. Correcting Predictions from Oceanic Maritime Numerical Models via Residual Learning. In Proceedings of the 2018 OCEANS—MTS/IEEE Kobe Techno-Ocean. (OTO), Kobe, Japan, 28–31 May 2018; pp. 1–4.

43. Rasp, S.; Lerch, S. Neural networks for post-processing ensemble weather forecasts. *Mon. Weather Rev.* **2018**, *146*, 3885–3900. [CrossRef]

44. Deshmukh, A.N.; Deo, M.C.; Bhaskaran, P.K.; Nair, T.; Sandhya, K.G. Neural-network-based data assimilation to improve numerical ocean wave forecast. *IEEE J. Ocean. Eng.* **2016**, *4*, 944–953. [CrossRef]

45. Yang, X.Q.; Anderson, J.L. Correction of systematic errors in coupled gcm forecasts. *J. Clim.* **2000**, *13*, 2072–2085. [CrossRef]

46. Han, Y.K.; Dan, Y.U.; Shen, X.Y.; Zhou, Y.Y. Study on the correction of SST prediction of HYCOM. *Mar. Forecast.* **2018**, *35*, 5. (In Chinese)

47. Zhang, P.J.; Zhou, S.H.; Liang, C.X. Study on the correction of SST prediction in South China Sea using remotely sensed SST. *J. Trop. Oceanogr.* **2020**, *39*, 59–67. (In Chinese)

48. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3d convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 221–231. [CrossRef]

49. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.

50. Mnih, V.; Heess, N.; Graves, A.; Kavukcuoglu, K. Recurrent models of visual attention. *Adv. Neural Inf. Processing Syst.* **2014**, *2*, 2204–2212.

51. Bleck, R. An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. *Ocean Modeling* **2002**, *4*, 88. [CrossRef]

52. Metzger, E.J.; Smedstad, O.M.; Thoppil, P.G.; Hurlburt, H.E.; Cummings, J.A. US Navy Operational Global Ocean and Arctic Ice Prediction Systems. *Oceanography* **2014**, *27*, 32–43. [CrossRef]

53. Reynolds, R.W.; Smith, T.M.; Liu, C.; Chelton, D.B.; Casey, K.S.; Schlax, M.G. Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *J. Clim.* **2007**, *20*, 5473–5496. [CrossRef]

54. Liu, Y.; Weisberg, R.H.; Law, J.; Huang, B. Evaluation of Satellite-Derived SST Products in Identifying the Rapid Temperature Drop on the West Florida Shelf Associated With Hurricane Irma. *Mar. Technol. Soc. J.* **2018**, *52*, 43. [CrossRef]