

Research Article

Prediction of North Atlantic Oscillation Index Associated with the Sea Level Pressure Using DWT-LSTM and DWT-ConvLSTM Networks

Bin Mu , Jing Li, Shijin Yuan , and Xiaodan Luo

Department of Software Engineering, Tongji University, Shanghai, China

Correspondence should be addressed to Shijin Yuan; yuanshijin2003@163.com

Received 20 April 2020; Revised 26 July 2020; Accepted 24 August 2020; Published 5 October 2020

Guest Editor: Feng Zhang

Copyright © 2020 Bin Mu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The North Atlantic Oscillation (NAO), which manifests as an irregular atmospheric fluctuation, has a profound effect on the global climate change. The NAO index (NAOI) is the quantitative indicator that can reflect the intensity of the NAO events, and its traditional definition is the normalized sea level pressure (SLP) difference between Azores and Iceland. From the variation tendency of the NAOI, we found that it is difficult to predict the NAO with the characteristics of variability and complexity. As a data-driven approach, the deep neural network presents great potential in learning the mechanisms of climate forecasting. In this paper, we adopt long short-term memory (LSTM) and ConvLSTM to predict the NAO from two aspects, NAOI and SLP, respectively. In previous studies, LSTM has been regarded as a resultful method for time series prediction. ConvLSTM can capture both the temporal and spatial interdependencies of the SLP field; then, the NAOI can be calculated from the SLP output. In order to improve the prediction reliability, we utilize the discrete wavelet transform (DWT) as a preprocessing technique to decompose original data into different frequencies, considering the local time dependency. It can effectively preserve the features of high-frequency data and forecast extreme events more accurately. The proposed DWT-LSTM and DWT-ConvLSTM models are compared against multiple advanced models, such as LSTM, Holt-Winters, support vector regression (SVR), and gated recurrent unit (GRU). The results indicate that both DWT-LSTM and DWT-ConvLSTM perform better, particularly at peak values. As for the 31 NAO events from 2006 to 2015, our models achieve the lowest prediction error and the best stability. Compared with the forecast products of CPC named Global Forecast System (GFS) and the ensemble forecasts (ENSM), our models are much closer to observation in multistep forecasting.

1. Introduction

The North Atlantic Oscillation (NAO) is the most prominent mode that can reflect the climate variability of the North Atlantic region [1]. Quite a few extreme climate events are associated with the behaviour of the winter NAO [2] and would have a profound impact on several areas [3]. The NAO is a complex nonlinear process with unexplored influence factors, which are difficult to represent in kinetic equations. Besides the discovered factors, such as the geomagnetic activity intensity [4], stratospheric events [5], and sea surface temperature (SST) forcing [6], human influence is also detected as one of the causes of NAO events occurrence [7]. These uncertainties, along with the uncovered mechanism, make the reliable prediction hard to accomplish.

A number of studies have performed the NAO prediction based on numerical models. The winter-mean NAO index (NAOI) has been predicted using the multisystem ensemble combining the operational seasonal prediction systems (UKMO, CFSv2, and CMCC) [8]. The ASF-20C seasonal hindcast ensemble has been proposed to predict the multidecadal variability of the winter NAO [9]. The predicted winter NAOI is obtained via the dynamical seasonal prediction system forced by MPI-ESM-MR ensemble members [10]. However, the prediction accuracy of numerical models would be influenced by errors in initial conditions and model parameters. In addition to ocean-atmosphere models and ensemble products, statistical approaches have also been put into the prediction of climate events gradually. The multiple linear regression (MLR)

technique provided a robust prediction of the winter mean NAOI [11]. The k -nearest neighbor, which is based on the local linear regression, was proposed to estimate the low-frequency winter NAOI [12]. Hall et al. presented the interdecadal evaluation on the basis of the linear regression model [13]. Although the statistical outcome can roughly track the tendency of the NAO variation with averting the initial uncertainties, it is still hard to reflect the characteristics of the NAO via linear models.

Currently, machine learning approaches have become alternative techniques in researches of geoscience. In terms of identification in meteorology, a convolutional neural network called CloudNet was proposed for meteorological cloud classification with high precision [14]. The reliable classification of ice crystal habits has been performed using a deep CNN model named TL-ResNet152, and an ice crystal dataset has been built [15]. This kind of data-driven solution has also made significant progress in weather forecasting and climate prediction recently [16–19]. Thereinto, the long short-term memory (LSTM) and the convolutional LSTM (ConvLSTM) exhibit prominent performance on climate prediction since the climatic data is the typical time series. The structure of cell gates in LSTM helps to detect the inner connection between sequence points, and the convolutional layer is conducive to extract features from field data. The 6–24 h nowcasting of typhoon tracks with an improved precision was performed by the LSTM network, combining with data mining technologies [16]. The amount of rainfall was predicted by the deep learning model with ConvLSTM layers, and the two-stacked ConvLSTM network reduced RMSE by 23.0% compared to linear regression [18]. The Oceanic Nino Index (ONI) and sea surface temperature, which are closely relevant to ENSO forecasting, were trained by the ConvLSTM-RM model to provide a relatively reliable prediction [20]. As for the NAO, ConvLSTM network combined with ensemble empirical mode decomposition (EEMD) was adopted to forecast the NAOI sequence, and the result surpassed several state-of-the-art machine learning models [21]. Researchers have proved that the temporal or spatial dependencies among the climatic variable data can be identified by LSTM and ConvLSTM effectively.

In this paper, the NAO prediction is conducted with two schemes. One of them is to predict NAOI via the LSTM network, which is trained by the daily NAOI historical data derived from the Climate Prediction Center (CPC) website. The NAOI sequence is organized into training pairs with the rolling mechanism, and each piece of training input is handled with a smooth strategy of the grey model (GM). The best lead time of training pairs is determined by autocorrelation (ACF) and partial autocorrelation (PACF) within the e -folding time scale of the NAO. Since the typical NAOI is defined from sea level pressure (SLP) [22], the other scheme is to perform NAO prediction using SLP grid data, which is provided by National Oceanic and Atmospheric Administration (NOAA). The SLP frames are grouped in the same way as the NAOI series and fed into the ConvLSTM network. The spatial characteristics and temporal dependencies of the SLP series can be simultaneously captured by the ConvLSTM layers. Then the predicted NAOI can be calculated by the projection of normalized SLP between the

Azores and Iceland on the NAO anomaly pattern. However, the NAO prediction is more concerned with extreme events with extraordinarily high or low NAOI. Here, we adopt the discrete wavelet transform (DWT) method as the pre-processing operation of our models. The DWT is commonly used in singular signal analysis, noise elimination, and signal compression [23]. It can effectively improve the accuracy of the prediction model by decomposing the original sequence into several components, which are relatively stationary compared to the original time series [24]. The performances of our models are validated in the comparison of multiple statistics or machine learning models, including Holt-Winters, support vector regression (SVR), and gated recurrent unit (GRU). By reference to the observation, our models show less error and higher stability compared with the above methods for the NAO cases from 2006 to 2015. And more notably, our models behave especially well in peak values, and the multistep forecasting is more reliable than ensemble forecast products named Global Forecast System (GFS) and the ensemble forecasts (ENSM).

The rest of this paper is structured as follows. Section 2 introduces the dataset and describes the technical details of the proposed DWT-LSTM and DWT-ConvLSTM. The experiment procedures and results are presented in Section 3. In the last section, we conclude with a summary and discuss future work.

2. Materials and Methods

2.1. Dataset and Study Area. The daily NAOI data is provided by the Climate Prediction Center (CPC) of National Weather Service, with 25504 days from 1950-01-01 to 2019-10-31 (removing the missing data). The SLP daily grid data can be downloaded from the website of the Physical Sciences Division of National Oceanic and Atmospheric Administration (NOAA) and is from 1981-09-01 to 2019-10-31. The brief information about these two datasets is displayed in Table 1.

The region we mainly focus on is located in the North Atlantic region between $20^{\circ}N$ and $80^{\circ}N$ and between $90^{\circ}W$ and $40^{\circ}E$. The resolution of the grid is $2.5^{\circ} \times 2.5^{\circ}$ with 25×53 grid points in each pattern. Each frame is truncated into a matrix with a size of 24×52 , which has an even number of rows and columns, to facilitate decomposition.

2.2. DWT. Wavelet transform is an effective tool of spectrum analysis, and it can extract features in both frequency domain and time domain during a local time scale. The low-frequency part of climate data often carries the principal features of the signal, which would be helpful in researching for interdecadal variation. Capturing its trend is relatively easy. However, short-term prediction using low-frequency data always cannot suit the requirement of accuracy. The reason is that some climate data like NAOI fluctuates wildly within a short period, and the high-frequency component plays an important role on it. In fact, the NAO events with abnormal NAOI (< -1 or > 1) may not be detected because of the forecasting failure on high-frequency data. The significance of the DWT is to separate index sequence or field data with different scales, and different weights are trained and fixed for

TABLE 1: The basic information of the NAOI dataset (25504 days) and the SLP dataset (13939 days).

Name	Max value	Min value	Mean value	Standard deviation
NAOI	2.75	-3.25	0.02	0.81
SLP	1072.64	931.12	1013.85	10.15

these components, respectively. DWT would be conducive to preserve the feature of each component and avoid high-frequency signals vanish during the regression process.

The steps of the DWT can be described as follows. Firstly, take samples at the discrete points $(x = \dots, -(1/2^j), 0, (1/2^j), \dots)$, and the sampled signal can be regarded as the j -order scaling function:

$$f_j(x) = \sum_{k \in \mathbb{Z}} a_k^j \phi(2^j x - k), \quad (1)$$

where the discontinuity point is at $(k/2^j)$ and $a_k = f(k/2^j)$. j and k stand for resolution and translation, and $\phi(x)$ denotes the scaling function. In this work, we test the performance of multiple mother wavelets, including Haar, db, sym, and coif.

$$f_j(x) = \sum_{k \in \mathbb{Z}} a_{2k}^j \phi(2^j x - 2k) + a_{2k+1}^j \phi(2^j x - 2k - 1) = \sum_{k \in \mathbb{Z}} \left(\frac{a_{2k}^j - a_{2k+1}^j}{2} \right) \psi(2^{j-1} x - k) + \left(\frac{a_{2k}^j + a_{2k+1}^j}{2} \right) \phi(2^{j-1} x - k). \quad (5)$$

The approximate sequence and the detail sequence d are obtained via the above processes:

$$p_{j-1} = \sum_{k \in \mathbb{Z}} a_k^{j-1} \phi(2^{j-1} x - k), \quad \text{where } a_k^{j-1} = \frac{a_{2k}^j + a_{2k+1}^j}{2}, \quad d_{j-1} = \sum_{k \in \mathbb{Z}} b_k^{j-1} \psi(2^{j-1} x - k), \quad \text{where } b_k^{j-1} = \frac{a_{2k}^j - a_{2k+1}^j}{2}. \quad (6)$$

The procedure of the DWT on the index sequence is presented in Figure 1. Based on the result of preliminary experiments, the raw data is set to decompose with only one level. As for the two-dimensional DWT, the operations of high-pass filtering, low-pass filtering, and underclocking need to be conducted in two directions successively, as shown in Figure 2. The original grid data is broken up into approximation (LL), horizontal detail (LH), vertical detail (HL), and diagonal detail (HH). The approximation is the low-frequency component in

Finally, the Haar wavelet is selected to perform the DWT. $\phi(x)$ and wavelet function $\psi(x)$ in Haar wavelet are defined as follows:

$$\phi(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1, \\ 0, & \text{else,} \end{cases} \quad (2)$$

$$\psi(x) = \phi(2x) - \phi(2x - 1).$$

We can furthermore find that the relationship between scaling function and wavelet function can be written as follows:

$$\phi(x) + \psi(x) = 2\phi(2x). \quad (3)$$

$$\phi(x) - \psi(x) = 2\phi(2x - 1). \quad (4)$$

Then we decompose the scaling function in formula (1) into even and odd parts (see formula (3)). The even (odd) component is acquired by high (low) pass filter and downsampling filter:

both horizontal direction and vertical direction, containing the bulk of information in the raw patterns. Analogously, horizontal detail, vertical detail, and diagonal detail are the high-frequency data in the horizontal direction, vertical direction, and both directions, respectively.

After training and forecasting, these components need to be reconstructed to obtain the predicted sequence. The discrete function for decomposition can be described as the following sum form:

$$f_1(x) = p_0 + d_0 = \sum_{l \in \mathbb{Z}} a_l^1 \phi(2x - l), \quad \text{where } a_l^1 = \begin{cases} a_k^0 + b_k^0, & \text{if } l = 2k, \\ a_k^0 - b_k^0, & \text{if } l = 2k + 1. \end{cases} \quad (7)$$

The output is reestablished by fusing the parameters of wavelet (b^0) and scaling (a^0), upsampling operator U , and tower-type inverse composition (\tilde{L} and \tilde{H}):

$$a^1 = \tilde{L}Ua^0 + \tilde{H}Ub^0. \quad (8)$$

2.3. *LSTM and ConvLSTM.* The architectures of the NAO prediction using LSTM is shown in Figure 3. The subdivisions with different properties are handled separately with decomposition using DWT. The two sets of weights in neural networks are trained to fit the approximate components and

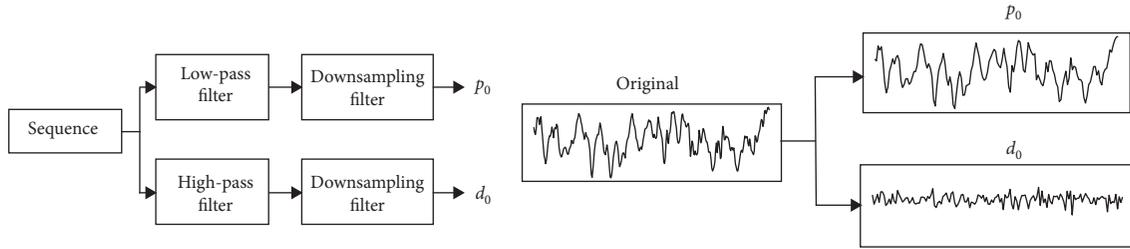


FIGURE 1: The flow and exemplification of decomposition for the NAO index sequence using DWT. The left subgraph shows the principle to obtain approximate component p_0 and the detail component d_0 in a one-level decomposition. The right subgraph represents the result of the DWT using a NAOI sequence as an example.

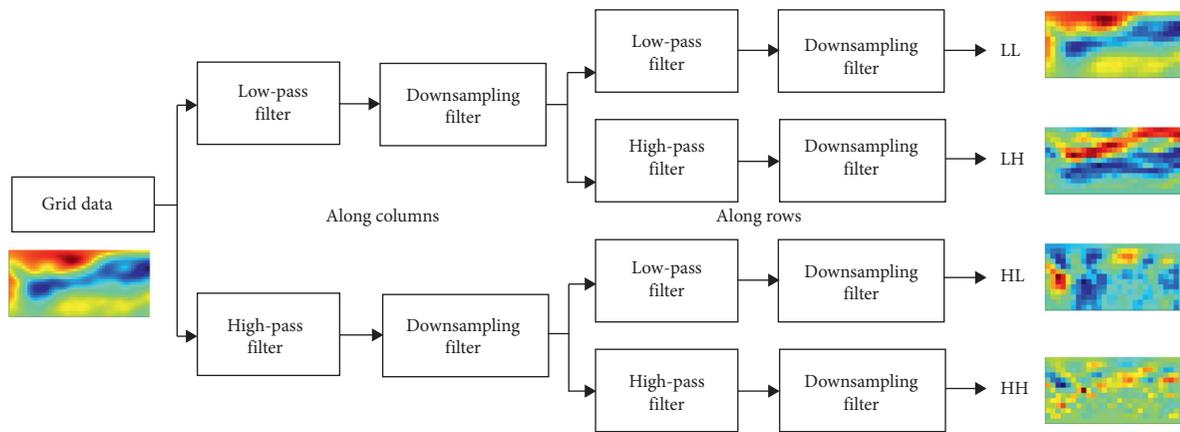


FIGURE 2: Similar to Figure 1, but for the SLP grid data using two-dimensional DWT. The sample grid data is filtered in the vertical direction at first, and then the acquired parts are decomposed in the horizontal direction, respectively. After these steps, we obtain four components: LL (approximation in two directions), LH (vertical approximation and horizontal detail), HL (vertical detail and horizontal approximation), and HH (detail in both directions).

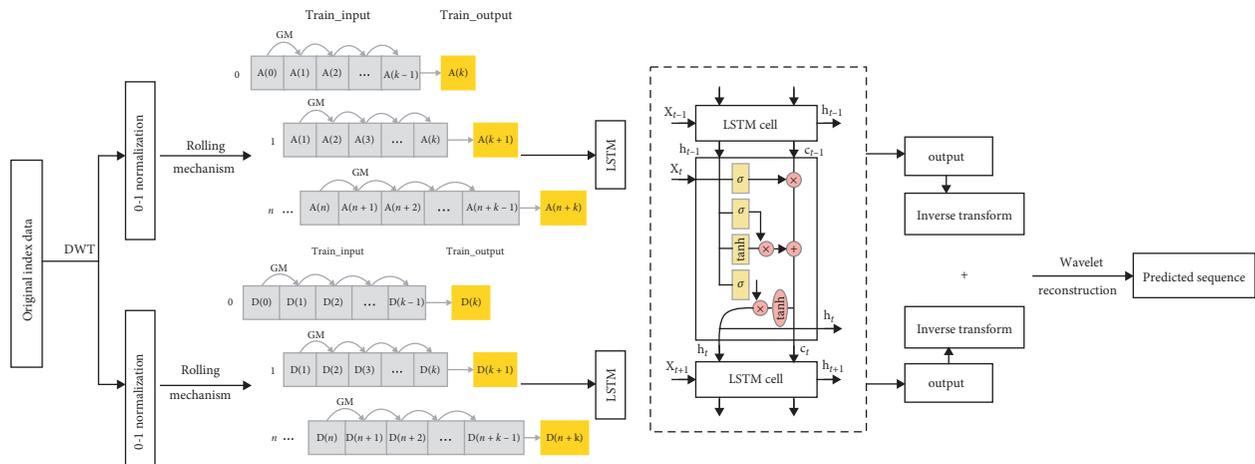


FIGURE 3: The architecture of the NAO prediction via the NAOI sequence using DWT-LSTM. The original NAO sequence is decomposed by DWT; then the two components are normalized into a range of $[0, 1]$. Their series are grouped with rolling mechanism and the smooth scheme of the GM and are fed into LSTM. After the inverse transformation and wavelet reconstruction, the predicted sequence is obtained:

detail components. It can avoid the adverse effect of conflicting information against the prediction result. These two components are restricted in the range $[0, 1]$ by Min-Max normalization:

$$v' = \frac{v - \min(x)}{\max(x) - \min(x)}. \quad (9)$$

Then, the normalized data is organized with a rolling form, which incorporates data with the fixed time horizon into the network, with the following expression:

$$\{X_{n-k}, \dots, X_{n-2}, X_{n-1}\} \longrightarrow X_n, \quad (10)$$

where k is the lagged coefficient, which denotes the scale of the time window. Since the NAO can be regarded as the nonlinear process with an e-folding time scale in about two weeks [25], we adopt ACF and PACF to determine the optimal lagged coefficient within 14 days. ACF describes the correlation of a given time series and a lag version within a continuous time interval, while PACF reports the correlation of two independent points, like X_{n-k} and X_n , ruling out the effects caused by $k-1$ points between them:

$$\text{ACF}_k = \frac{\sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad (11)$$

$$\text{PACF}_k = \frac{E[(X_n - \hat{E}X_n)(X_{n-k} - \hat{E}X_{n-k})]}{E[(X_{n-k} - \hat{E}X_{n-k})^2]},$$

where $\hat{E}X_n = E[X_n | X_{n-1}, \dots, X_{n-k+1}]$ and

$$\hat{E}X_{n-k} = E[X_{n-k} | X_{n-1}, \dots, X_{n-k+1}]. \quad (12)$$

Figure 4 illustrates the ACF and PACF of the NAOI sequence in 14 days, and the confidence interval is set to 95%. We can see that the dependency reflected by the ACF becomes weaker as the lagged coefficient k increases, containing the synthesis relation of the previous sequence. The PACF reveals the direct association between observation and the lagged version. When k is larger than 5, the PACF is trending to zero. Thus, k is set to 5 as the optimal lagged coefficient.

Then, the sequence is grouped into training pairs with highlighting the features by generation method in the grey model. Following this, these components are fed into our model, which has the bidirectional LSTM layer, dropout layers, and fully connected layers. The LSTM layer adopted in this paper is based on the structure with forget gates:

$$\begin{aligned} i_t &= \sigma(W_{xi}X_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \\ f_t &= \sigma(W_{xf}X_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \\ c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}X_t + W_{hc}h_{t-1} + b_c), \\ o_t &= \sigma(W_{xo}X_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \\ h_t &= o_t \tanh(c_t), \end{aligned} \quad (13)$$

where i_t , f_t , c_t , o_t , and h_t are the vectors of input gates, forget gates, cell activation, output gates, and hidden layer at time t , and W_x , W_h , and W_c stand for the weights from cells, hidden vectors, and cell activation to another component. σ and \tanh denote the logistic sigmoid and hyperbolic tangent, respectively. The rectified linear unit (ReLU) is adopted as the activation function in the LSTM layer.

The function of the dropout layer is to avoid overfitting, and the drop rate is set to 0.2. The linear transfer operation is performed by the fully connected layer, and its activation function is sigmoid. Mean square error is selected as the loss function of this model, and a nadam optimizer with a learning rate of 0.002 is used to minimize the cross-entropy loss.

As for the SLP prediction in Figure 5, the preprocessing operation is similar to the NAOI sequence. Since the spatiotemporal signals are mutually dependent, we adopt the convolution and recurrent network to extract the inter-coupling features. The model for SLP data consists of ConvLSTM layers, BatchNormalization layers, MaxPooling layers, UpSampling layers, and the Conv3D layer.

The ConvLSTM layer is able to detect both the temporal and spatial characteristics of the seasonal pressure field. It adopts convolution operation as the feedforward method instead of the matrix multiplication in LSTM. The updates of internal gates and outputs in ConvLSTM are shown as follows:

$$\begin{aligned} i_t &= \sigma(W_{xi} * X_t + W_{hi} * h_{t-1} + W_{ci} \circ c_{t-1} + b_i), \\ f_t &= \sigma(W_{xf} * X_t + W_{hf} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f), \\ c_t &= f_t \circ c_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * h_{t-1} + b_c), \\ o_t &= \sigma(W_{xo} * X_t + W_{ho} * h_{t-1} + W_{co} \circ c_t + b_o), \\ h_t &= o_t \circ \tanh(c_t), \end{aligned} \quad (14)$$

where i_t , f_t , and o_t are the gates of ConvLSTM represented by 3D tensors. $*$ denotes the convolution operator and \circ denotes the Hadamard product. The kernel size is set to 3×3 , which is the appropriate size. The role of the BatchNormalization layer is to accelerate the training process in deep neural networks. It can also alleviate the problems of gradient vanishing and gradient exploding. The MaxPooling layer is applied to simplify the calculation, while the UpSampling layer is responsible for the retrieval of dimensions. The nadam optimizer is also selected in this model.

The prediction SLP, which is reconstructed using the inverse transform of the DWT2D, can be utilized to calculate the NAOI. NAOI has multiple definitions related to SLP [26–28], and we select the block index as the NAOI in this work [29]:

$$\text{NAOI} = \frac{\text{SLP}_{\text{NAO}}, \text{SLP}_d}{\text{SLP}_{\text{NAO}}, \text{SLP}_{\text{NAO}}}, \quad (15)$$

where SLP_{NAO} is the anomaly pattern, which is the first mode decomposed by EOF analysis on 30-year historical SLP data. As shown in Figure 6, the NAO anomaly mode

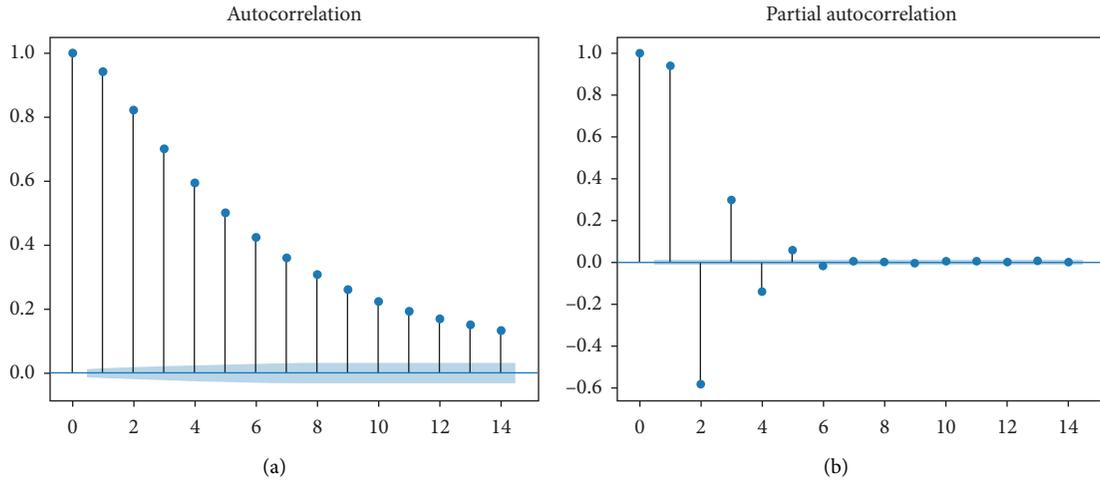


FIGURE 4: The ACF and PACF for daily NAOI sequence within 14 days with confidence bounds of 95%. In the right subgraph, the PACF shows k -order censored, and the lagged coefficient k can be used to determine the window size of the training data. (a) Autocorrelation. (b) Partial autocorrelation.

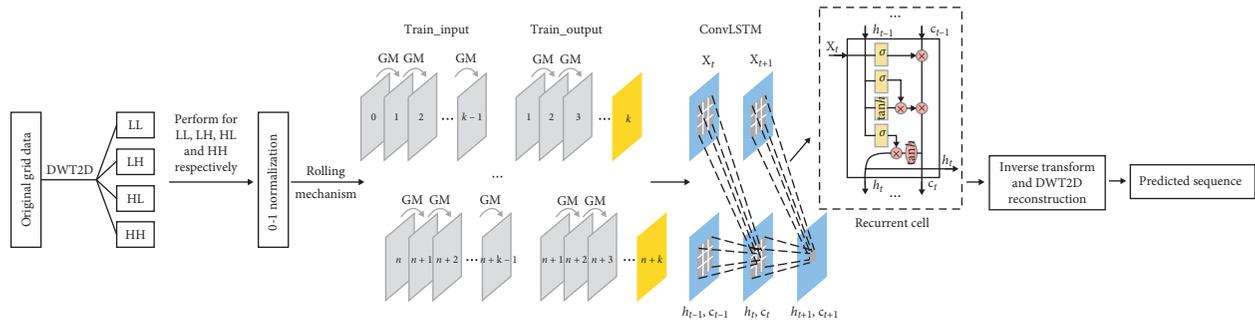


FIGURE 5: Similar to Figure 3, the difference is that the raw data is SLP grid data located in the North Atlantic sector, and the prediction model is DWT-ConvLSTM.

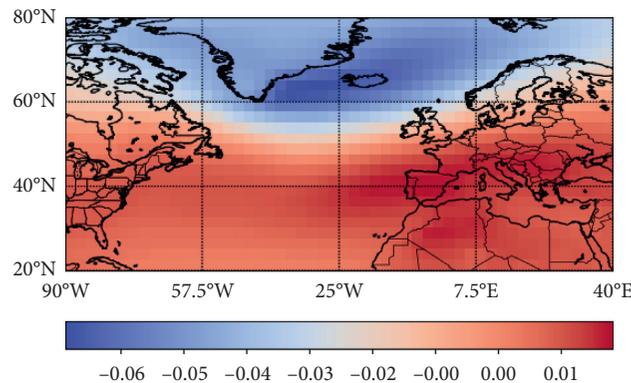


FIGURE 6: The NAO anomaly pattern decomposed by EOF analysis of SLP historical data located in the North Atlantic region between $90^{\circ}W-40^{\circ}E$ and $20^{\circ}N-80^{\circ}N$. The anomaly pattern manifests the typical structure of the NAO^+ event.

manifests the typical dipole structure, with low pressure in Iceland and the high pressure located in the North Atlantic subtropical. SLP_d is acquired by subtracting the climatological mean from the SLP data.

2.4. Model Settings. Our models are developed based on a high-level neural network API called Keras with a Python version of 3.6. The DWT-LSTM is constructed with a bidirectional LSTM layer, a flatten layer, two dropout

layers, and two fully connected layers, totaling 69121 trainable parameters. The DWT-ConvLSTM is composed of 14 layers with 406681 tunable parameters, as reported in Table 2. It contains 4 ConvLSTM layers and adopts the Maxpooling layer to enhance the stability of feature recognition.

We conduct experiments on the Tianhe-2 supercomputer, which is located in the National Supercomputer Center in Guangzhou, China. NVIDIA Tesla K80 GPUs are applied for training acceleration in this work.

3. Experiments and Results

3.1. NAO Prediction Using NAOI Sequence. To visualize the performance of DWT-LSTM, we compare its prediction result with multiple models, which contain GRU, near neighbour regression (NNR), LSTM, SVR, and Holt-Winters. As with DWT-LSTM and LSTM, GRU is a kind of recurrent neural network for solving long-term memory and backpropagation. NNR and Holt-Winters are statistical models; thereinto, NNR is a nonparametric method based on the feature space, while Holt-Winters takes the concept of exponential smoothing estimation. SVR transforms machine learning to solve a convex quadratic programming problem by the optimization theory. Using these approaches, a 100-day prediction is performed during the wintertime from Oct 2018 to Feb 2019, which is illustrated in Figure 7. To simplify the problem, assume that an NAO event occurs when $NAOI > 1.0$ or $NAOI < -1.0$. It is defined as an NAO^+ (NAO^-) event when the NAOI is positive (negative) during the persistent episode. By the above definition, five NAO events are recognized at the date of 2018-10-26 ($NAO^- - 1.412$), 2018-11-26 ($NAO^- - 1.404$), 2018-12-07 ($NAO^+ 1.539$), 2019-01-01 ($NAO^- - 1.061$), and 2019-01-10 ($NAO^- - 1.18$). Although the tendency of LSTM, GRU, and SVR fits the observation data closely, they overestimate (underestimate) the peak value distinctly. For instance, the NAOI provided by the above-mentioned models surrounding Jan 11, 2019, is heavily skewed, and the error is even greater than 1.5. However, the DWT-LSTM behaves better, especially at the peak points enlarged in round boxes.

We select the root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) as the evaluation indicators. The formulas are given as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |X_o - X_p|^2},$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_o - X_p|, \quad (16)$$

$$R^2 = \frac{\sum_{i=1}^n (X_p - \bar{X}_o)^2}{\sum_{i=1}^n (X_o - \bar{X}_o)^2}$$

TABLE 2: The structure (layer, shape, and parameter numbers) of the DWT-ConvLSTM.

Layer	Shape	Number of parameters
ConvLSTM	(None, none, 12, 26, 40)	59200
BatchNormalization	(None, none, 12, 26, 40)	160
MaxPooling	(None, none, 6, 13, 40)	0
ConvLSTM	(None, none, 6, 13, 40)	115360
BatchNormalization	(None, none, 6, 13, 40)	160
Dropout	(None, none, 6, 13, 40)	0
ConvLSTM	(None, none, 6, 13, 40)	115360
BatchNormalization	(None, none, 6, 13, 40)	160
Dropout	(None, none, 6, 13, 40)	0
ConvLSTM	(None, none, 6, 13, 40)	115360
BatchNormalization	(None, none, 6, 13, 40)	160
Dropout	(None, none, 6, 13, 40)	0
UnSampling	(None, none, 12, 26, 40)	0
Conv3D	(None, none, 12, 26, 1)	1081

where n denotes the number of samples, and the X_p stands for the predicted values. X_o is the observation, and \bar{X}_o is the mean value of the observation data. To give a visual representation, we plot the indicator values corresponding to the predicted flows in Figure 7, as shown in Figure 8. The RMSE and MAE of the DWT-LSTM are slightly lower than other models, and the R^2 is higher. Since the closer the R^2 is to 1, the more reliable the model is, the result of DWT-LSTM is more similar to the ground truth as a whole. Notably, the DWT-LSTM's RMSE (0.2669), MAE (0.2016), and R^2 (0.8051) improve by around 21.3%, 5.0%, and 17.5%, respectively, in comparison to the pure LSTM. Combining the trend of flows in Figure 7, it shows that errors of other models are mainly concentrated at peaks, and the DWT-LSTM can alleviate the problem to some extent.

The previous research counted the NAO events of two phases (NAO^+ and NAO^-) from 2006 to 2015, totaling 31 cases, 22 for NAO^+ (red), and 9 for NAO^- (blue) [30]. The distribution of these cases is shown in Figure 9, with aggregate persistent periods. We can see that the cases are mainly concentrated in wintertime (DJF), and a case may span two months. The durations of these cases vary from 3 to 33 days, and some of them even last for the best part of a month. For instance, the NAO^- event that started in Feb 2010 carried over into the next month, and it would deeply affect adjacent weather.

We also select some metrics to evaluate these cases. In addition to RMSE and MAE, other error evaluation indices are adopted to verify the performance of our model, such as mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE), variance of absolute percentage error (VAPE), and root mean square percentage error (RMSPE), which are defined as

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{X_p - X_o}{X_o} \right| * 100\%. \quad (17)$$

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|X_p - X_o|}{(|X_p| + |X_o|)/2} * 100\%. \quad (18)$$

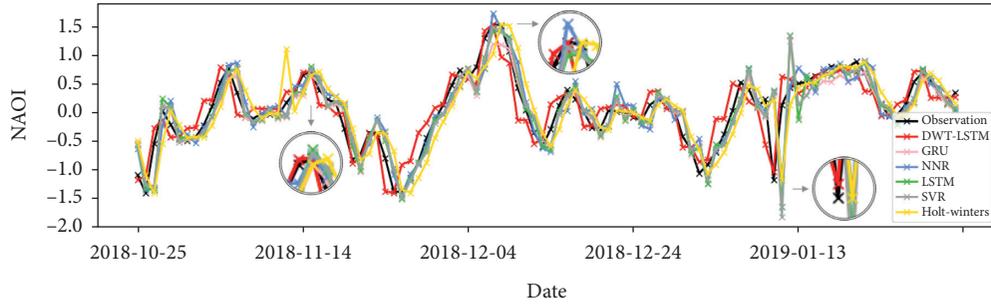


FIGURE 7: The one-step NAOI prediction for 100 days from 2018-10-25 to 2019-02-02 with multiple models. The red line denotes DWT-LSTM, and the black line denotes the observation data. The peak values (at 2018-11-15, 2018-12-07, and 2019-01-10) in this period are zoomed in to show the result more plainly.

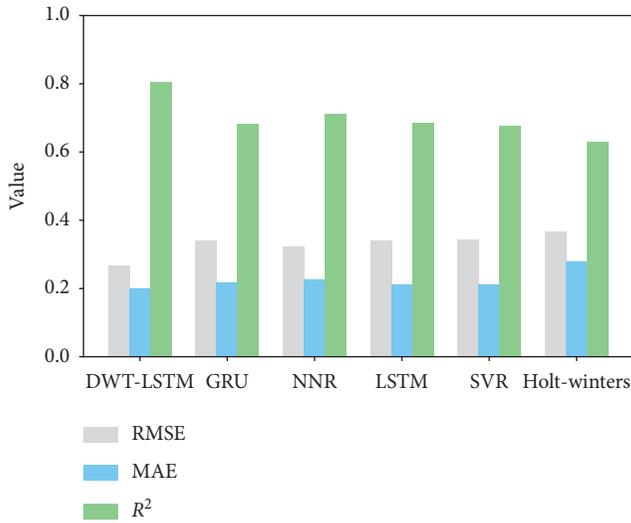


FIGURE 8: The RMSE (grey), MAE (blue), and R^2 (green) of the 100-day prediction with multiple models from 2018-10-25 to 2019-02-02 in Figure 7.

$$\text{VAPE} = \text{variance} \left(\frac{X_p - X_o}{X_o} \right) * 100\%. \quad (19)$$

$$\text{RMSPE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{X_p - X_o}{X_o} \right)^2}. \quad (20)$$

Figure 10 presents the distribution of indices for these cases predicted by multiple models. From formulas (17) and (18), the MAPE and SMAPE would increase sharply when the model produces false negatives or false positives. The statistical methods have comparatively larger deviation at extreme points and result in larger MAPE and SMAPE. Among them, the DWT-LSTM has relatively smaller steady-state errors and better stability with a smaller range of metric values. Besides, lower outliers also display occurrences of a higher positive correlation between the DWT-LSTM's result and the observation.

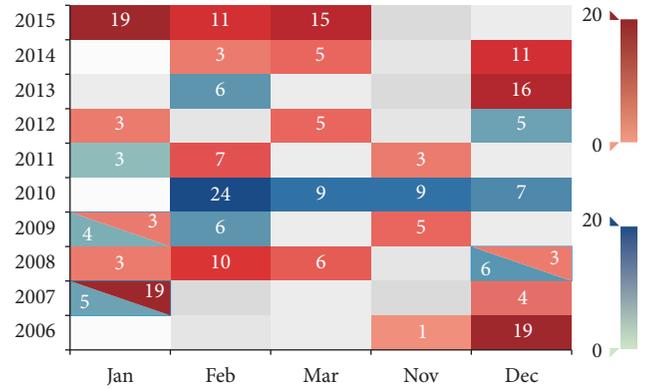


FIGURE 9: The selected 31 NAO events from 2006 to 2015 with an aggregate persistent episode in each month. The red blocks denote the NAO^+ events, and the blue blocks denote the NAO^- . The color depth of blocks indicates the duration also the intensity of events, and the blocks having two colors means occurring both NAO^+ and NAO^- in this month (not at the same time). The text in block refers to the duration of the NAO event.

3.2. NAO Prediction Using SLP Grid Data. As for SLP prediction, we perform a one-step prediction from 2019 to 01-10 to 2019-01-16, as illustrated in Figure 11. From the previous subsection, we know that these frames are in a state of the event with a negative phase. Due to dramatic changes in the NAOI, it is inherently difficult to predict. However, these problems are focal points of the NAO prediction and have been paid much attention. In Figure 11, the first row displays observed patterns of daily SLP, which are rather changeable and have the property of stochasticity. Therefore, it is relatively hard to fit the sequential variation. The rows below illustrate the results of ConvLSTM and DWT-ConvLSTM, and the difference sequences between each of them and the ground truth are also displayed below. We can find that the pressure structures predicted by both ConvLSTM and DWT-ConvLSTM are very close to the observation data. Especially for the DWT-ConvLSTM, differences of grids are mainly distributed around 0 in a light green color. As for the ConvLSTM, the errors focus on the pressure centers, indicating the underestimation of extreme values.

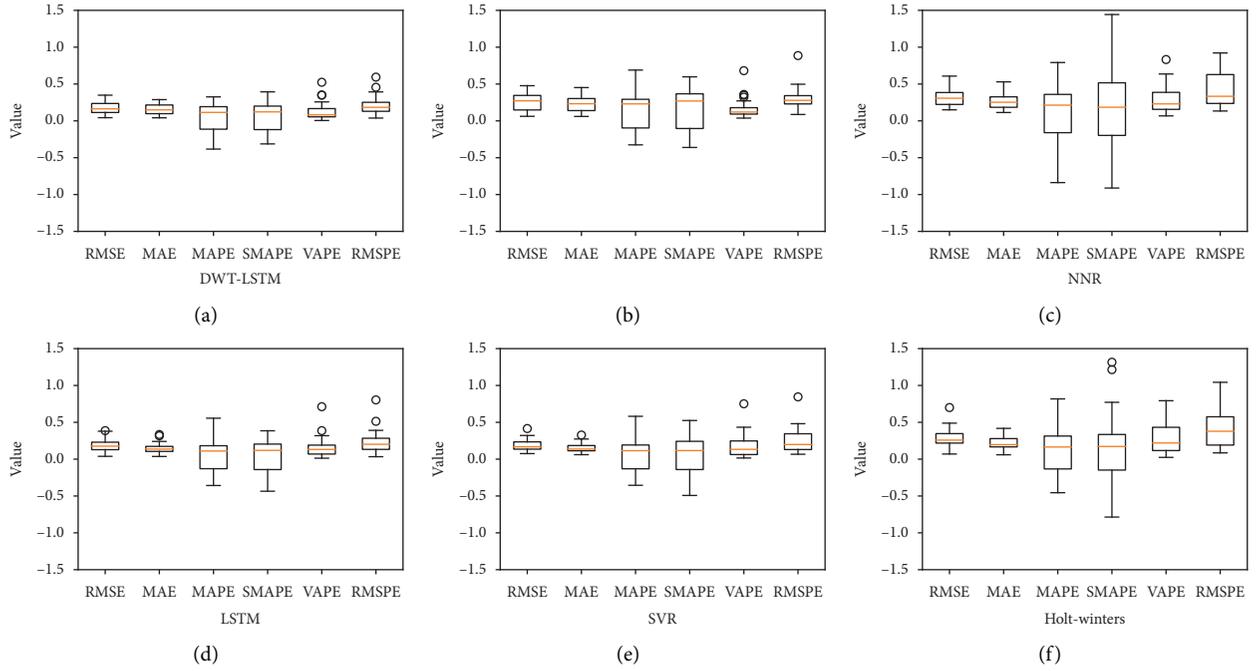


FIGURE 10: The error distribution for 31 NAO events predicted by multiple models. Each box extends from the lower (Q_1) to the upper quartile (Q_3), and the orange line denotes the median. The upper whiskers extend from Q_3 to $Q_3 + 1.5IQR$, where $IQR = Q_3 - Q_1$, and the lower whiskers extend from Q_1 to $Q_1 - 1.5IQR$. The hollow circles in this figure are all greater than the upper limit and refer to the discrete values larger than $Q_3 + 1.5IQR$. (a) DWT-LSTM. (b) GRU. (c) NNR. (d) LSTM. (e) SVR. (f) Holt-Winters.

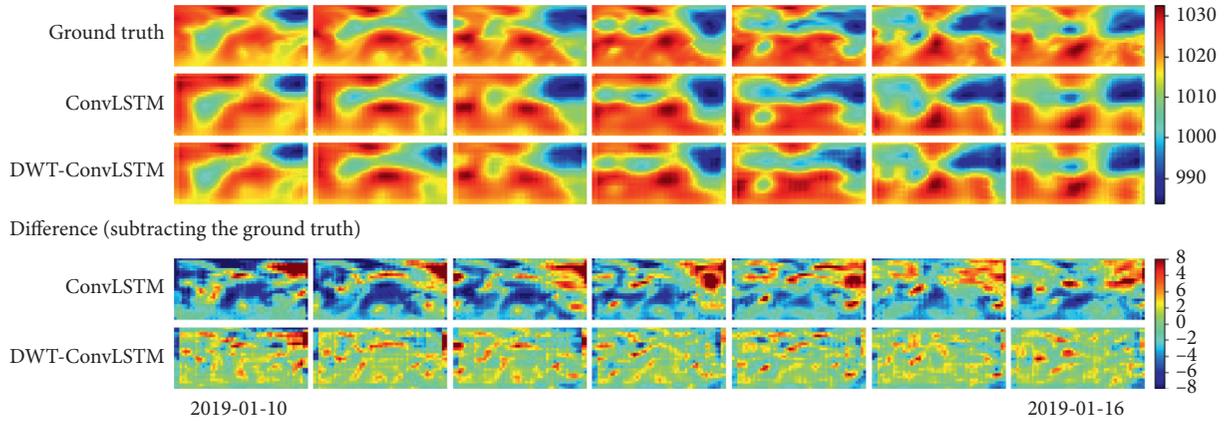


FIGURE 11: The comparison of ground truth along with predicted SLP pattern using ConvLSTM and DWT-ConvLSTM from 2019 to 01-10 to 2019-01-16. The first row is the observation SLP data, and the second and the third row are prediction results of ConvLSTM and DWT-ConvLSTM. The differences between the predicted sequence and the ground truth are displayed in the fourth and the fifth row.

To observe the accuracy directly, we plot some quantitative metrics for prediction results in Figure 11. Despite RMSE, MAE, and R^2 , we also select several indicators related to pattern similarity, such as the universal quality index (UQI) and anomaly correlation coefficient (ACC). UQI has been widely used to measure the image quality in the domain of computer vision, and ACC refers to the spatial correlation between the verifying analysis anomaly and forecast anomaly. UQI and ACC can be formulated as

$$UQI = \frac{2\overline{X_o} * \overline{X_p} * \sigma_{op}}{(\overline{X_o^2} + \overline{X_p^2})(\sigma_o^2 + \sigma_p^2)},$$

$$ACC = \frac{\sum_{i=1}^n (X_p - X_m - M_{jc})(X_o - X_m - M_{ac})}{\sqrt{\sum_{i=1}^n (X_o - X_m - M_{jc})^2 \sum_{i=1}^n (X_o - X_m - M_{ac})^2}} \quad (21)$$

where $M_{jc} = (1/n) \sum_{i=1}^n (X_p - X_m)$, $M_{ac} = (1/n) \sum_{i=1}^n (X_o - X_m)$, and σ_{op} denotes the covariance of the predicted data and the observation, σ_o (σ_p) stands for the standard deviation of the observation (prediction), and X_m is the climatological mean. The closer UQI and ACC are to 1, the higher reliability the model has. In Figure 12, bars report the mean values of the corresponding indicators in Figure 11, and error bars (standard deviation) describe the stability of these evaluation values. The DWT-ConvLSTM's errors are significantly lower than ConvLSTM, and its RMSE and MAE reduce by 43.7% and 48.1%. The R^2 , UQI, and ACC of the DWT-ConvLSTM are all greater than 0.9 and are improved by 24.4%, 4.6%, and 40.7% due to the application of the DWT. The result of our model has stronger relevance with actual values. High robustness is also reflected from the shorter error bars of the DWT-ConvLSTM. Taking errors in Figure 12 as a reference, other extreme events achieve similar outcomes to the NAO⁻ event in Figure 11.

The DWT-ConvLSTM is also tested with NAO events shown in Figure 9, with 31 cases and 255 frames. Figure 13 displays the distribution of RMSE using DWT-ConvLSTM and ConvLSTM, and RMSE in most of the cases are within 10. The RMSE of events predicted by DWT-ConvLSTM chiefly centers on 1 to 4, while the distribution is more dispersed in the RMSE of ConvLSTM, which is generally larger than 4. The better result is obtained by the DWT-ConvLSTM, suggesting an essential effect of decomposition and respective training on the forecasting performance. Each training frame is divided into four parts with a size of 12×26 , and these parts can be fitted accurately with tunable parameters. This ensures that the characteristics of the abnormal signal we care about can be well preserved. Despite the use of the dropout layer, overfitting is inevitable in some cases whose intensity is exorbitant.

3.3. Multistep Prediction for NAOI Comparison against Ensemble Forecast. In general, ensemble forecast products always provide multistep NAOI predictions to meet the practical requirement. GFS and ENSM, which are authoritative products of NAOI prediction, present the 7-day, 10-day, and 14-day forecasts on the CPC website (<http://www.cpc.ncep.noaa.gov>), as shown in Figure 14. The first row is the ground truth, and the following lines show the predicted flows. Overall, the predicted NAOI sequences are basically consistent with the ground truth. However, we can find that the accuracy decreases markedly when the lead time increases, and the 14-day forecasting cannot even identify the probable NAO events in the next two weeks.

To make our prediction have more practical implications, we perform the multistep prediction with a rolling mechanism. The principle of rolling forecasting is presented in Figure 15. The predicted value X_0 is obtained by feeding the input sequence with lagged length, which is set to 5 in this work. Then, we stitch it behind the previous input sequence and remove the first element of the sequence, resulting in generating a new input. After that, the prediction

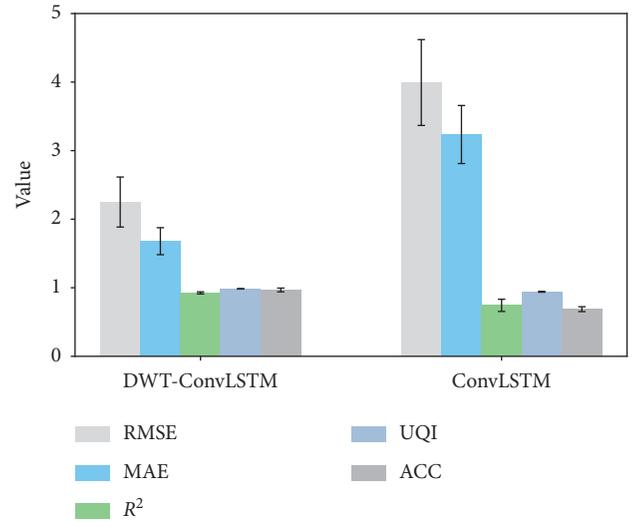


FIGURE 12: The comparison of multiple evaluation indicators (RMSE, MAE, R^2 , UQI, and ACC) between DWT-ConvLSTM and ConvLSTM, whose prediction period is the same as Figure 11.

is conducted again. After the loop is repeated n times, a prediction sequence with a length of n can be acquired. However, the error would be accumulated if the rolling forecasting has been performed too many times. Therefore, we apply this method to implement short-term forecasts within two weeks.

Figure 16 plots the multistep prediction result along with the observation data. According to the situation described in Figure 14, the 7-day predictions of both GFS and ENSM achieve better compared with the longer lead time. In the early part, almost all methods can attain relatively high accuracy, while the 14-day predictions of GFS and ENSM always maintain a state of underestimation. In contrast with ENSM, the prediction flow provided by GFS has a similar trend with the actual values. The proposed model, DWT-LSTM and DWT-ConvLSTM, also obtain a conviction result. Notably, our models are more accurate at some points with high NAOI, like 2020-01-05 and 2020-01-08. When the prediction time becomes longer, it has a significant deviation between our predictions and observation, and even a false positive is caused at 2020-01-12. Furthermore, a NAO⁺ event appears during the presented period, and GFS_14, ENSM_10, and ENSM_14 do not discern the occurrence of the event. The results indicate that proposed models can simulate the main trends of the NAOI variation within 14 days.

Table 3 reports the error corresponding to the prediction in Figure 16. The longer the lead time that the ensemble forecast has, the worse the forecasting performance, except the GFS_10. The models of GFS are generally reliable than that of the ENSM, consistent with the description for Figure 16. As for neural networks, DWT-LSTM and DWT-ConvLSTM have slightly smaller errors than other models in this case. It is proved that our models have the potential capacity for the NAOI short-term prediction.

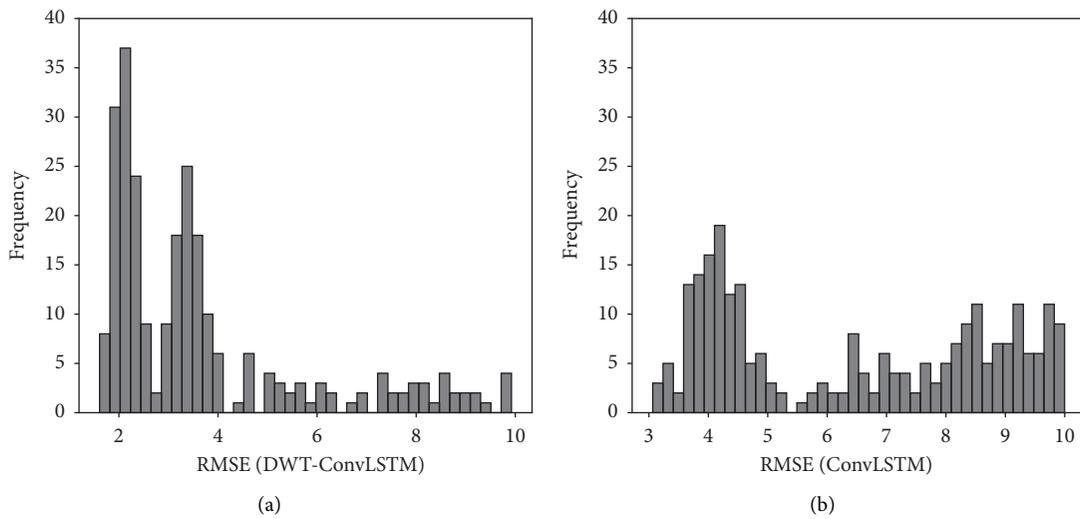


FIGURE 13: The frequency of RMSE for predicting NAO cases in Figure 9 with DWT-ConvLSTM and ConvLSTM. 31 NAO events in Figure 9 consist of 255 days in total, and the RMSE is calculated for each day.

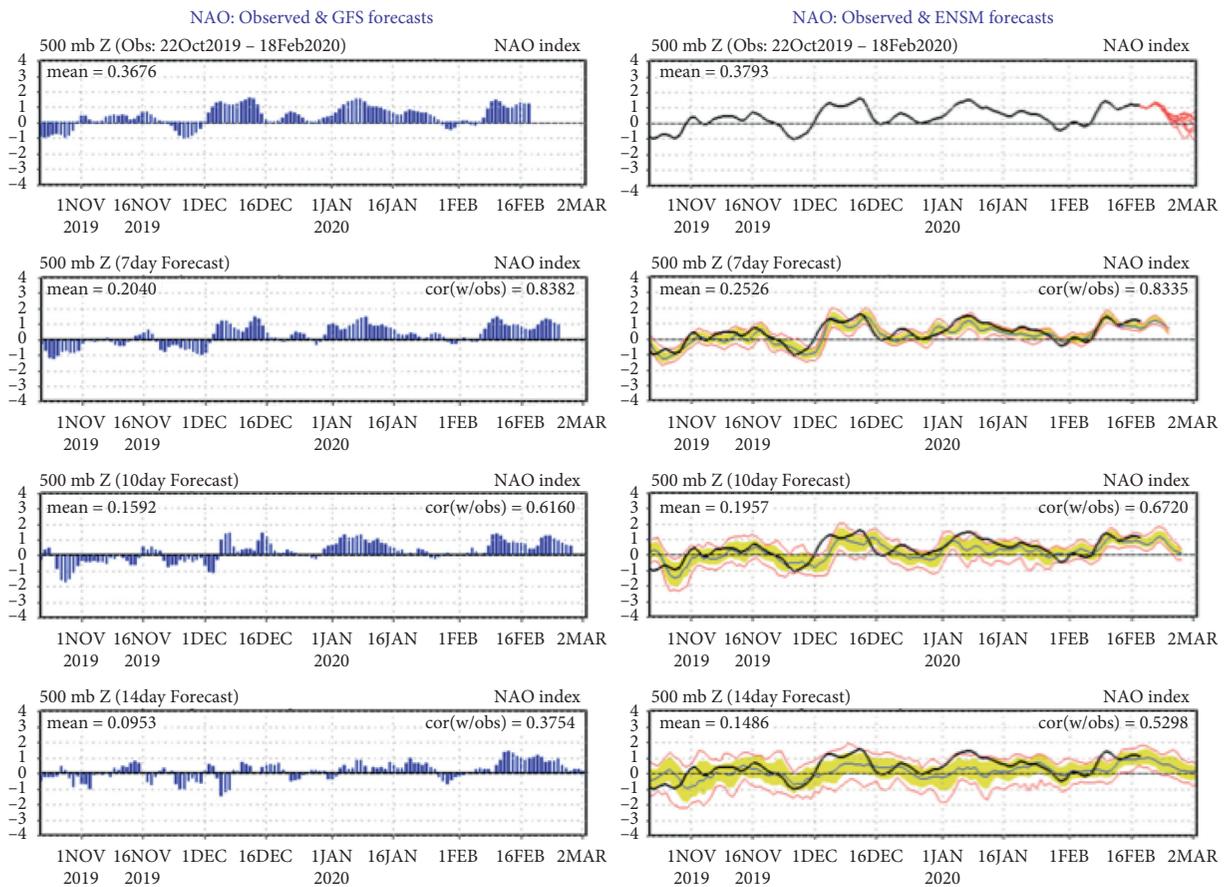


FIGURE 14: Recent NAOI forecast result of GFS and ENSM. This figure is derived from the CPC website and includes forecasts of 7 days, 10 days, and 14 days. The NAOI is calculated from the SLP on 500 mb.

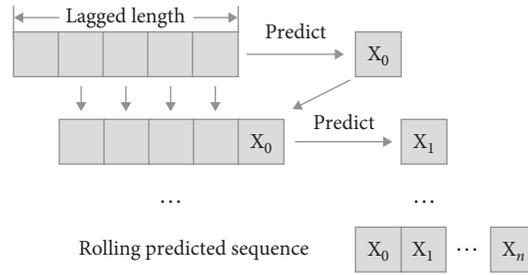


FIGURE 15: The sketch map of the rolling forecasting, which is applied in multistep prediction. Each piece of training data's length is set to the lagged coefficient determining in Section 2.3 ($k = 5$). Suppose that we put the first piece of training data into the model and obtain the predicted value X_0 ; then the next input is generated by removing the first frame and appending X_0 to the end of input. Putting the next input into the model, we obtain a new predicted value X_1 . Repeating the above steps, the predicted sequence is $\{X_0, X_1, \dots, X_n\}$.

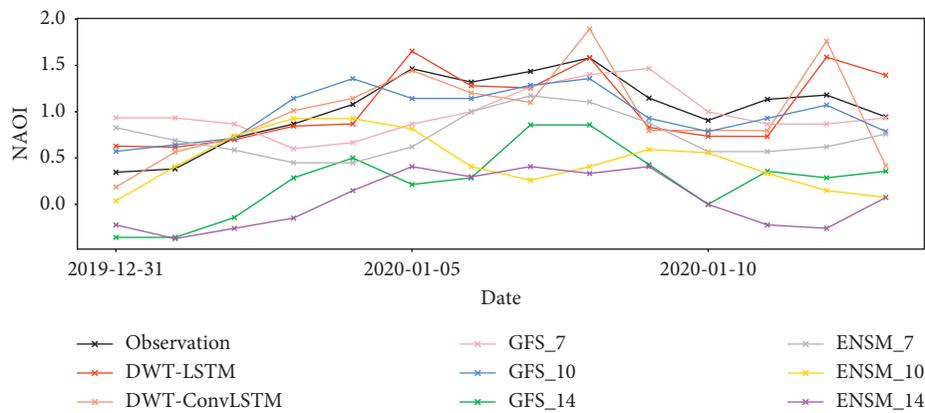


FIGURE 16: A comparison of multistep prediction between proposed models (DWT-LSTM and DWT-ConvLSTM) and ensemble forecast products (GFS and ENSM) from 2019-12-31 to 2020-01-13. The black line stands for the observation data, the red and orange line refer to the predicted flow of DWT-LSTM and DWT-ConvLSTM, respectively. Both GFS and ENSM include 7-day, 10-day, and 14-day products.

TABLE 3: The errors of the proposed models and ensemble forecast products.

Model	RMSE	MAE	MAPE (%)	SMAPE (%)	VAPE
DWT-LSTM	0.2547	0.2089	25.7524	23.1516	0.2433
DWT-ConvLSTM	0.2103	0.1945	23.3784	21.3440	0.1975
GFS_7	0.3496	0.3025	41.4826	33.8523	0.5028
GFS_10	0.2893	0.2317	24.8254	26.8585	0.1875
GFS_14	0.8030	0.7808	89.7068	87.6454	0.5045
ENSM_7	0.4533	0.4132	46.5721	49.4401	0.3246
ENSM_10	0.7085	0.5775	51.9023	82.7799	0.3337
ENSM_14	1.0195	0.9935	107.0605	1090.4483	0.3851

4. Conclusions

The NAO is the most dominant atmospheric circulation in the North Hemisphere during the wintertime, and it has a profound effect on the weather of Eurasia, even the global climate. Owing to the drastic change of climate and human activities in the recent two decades, the cause of the climate phenomenon becomes unstable and complicated. In

addition, the undetected physical mechanism also makes it more difficult to conduct a reliable prediction. As the alternative solution for time series prediction, the neural network displays its tremendous potential on climate forecasting.

In this paper, we adopt the DWT-LSTM and the DWT-ConvLSTM to perform the short-term NAO prediction. The DWT-LSTM is proposed to predict the daily NAOI sequence directly, while the predicted NAOI can also be obtained by the DWT-ConvLSTM via SLP prediction. The DWT is used as the preprocessing method to enhance the prediction skill of peak values. By decomposing the signal into components with different frequencies, the features of the anomaly signal that we care about can be retained successfully. Compared with other effective models and ensemble forecast products, both of the DWT-LSTM and the DWT-ConvLSTM achieve higher reliability. Besides, our models have smaller errors and yield better results in the prediction of NAO events from 2006 to 2015.

In the future, we tend to predict the NAO with other related meteorological variables, such as zonal wind and sea surface temperature (SST). Moreover, overfitting and error accumulation need to be improved in the following work.

Data Availability

The original daily NAO index data can be accessed on the CPC website (<https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml>), and the daily SLP grid data can be downloaded from NOAA's service (https://www.esrl.noaa.gov/psd/cgi-bin/db_search/DBSearch.pl?Variable=Sea+Level+Pressure&group=0&submit=Search).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The training process of this work was performed on the Tianhe-2 supercomputer. Thanks are due to the support of the National Supercomputer Center in Guangzhou (NSCC-GZ). This study was supported by the Fundamental Research Funds for the Central Universities of China (Grant 22120190207) and the National Key Research and Development Program of China (Grant 2018YFC1506402).

References

- [1] J. W. Hurrell, "Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation," *Science*, vol. 269, no. 5224, pp. 676–679, 1995.
- [2] J. W. Hurrell and H. Van Loon, "Decadal variations in climate associated with the North Atlantic oscillation," in *Climatic Change at High Elevation Sites*, pp. 69–94, Springer, Dordrecht, Netherlands, 1997.
- [3] N. Dunstone, D. Smith, A. Scaife et al., "Skilful predictions of the winter North Atlantic Oscillation one year ahead," *Nature Geoscience*, vol. 9, no. 11, pp. 809–814, 2016.
- [4] V. Bucha, "Geomagnetic activity and the north Atlantic oscillation," *Studia Geophysica et Geodaetica*, vol. 58, no. 3, pp. 461–472, 2014.
- [5] F. Hansen, R. J. Greatbatch, G. Gollan, T. Jung, and A. Weisheimer, "Remote control of North Atlantic oscillation predictability via the stratosphere," *Quarterly Journal of the Royal Meteorological Society*, vol. 143, no. 703, pp. 706–719, 2017.
- [6] M. Hoerling, K. Wolter, J. Perlwitz et al., "Northeast Colorado extreme rains interpreted in a climate change context," *Bulletin of the American Meteorological Society*, vol. 95, no. 9, pp. S15–S18, 2015.
- [7] F. Di Rita, F. Lirer, S. Bonomo et al., "Late Holocene forest dynamics in the Gulf of Gaeta (central Mediterranean) in relation to NAO variability and human impact," *Quaternary Science Reviews*, vol. 179, pp. 137–152, 2018.
- [8] P. J. Athanasiadis, A. Bellucci, A. A. Scaife et al., "A multi-system view of wintertime NAO seasonal predictions," *Journal of Climate*, vol. 30, no. 4, pp. 1461–1475, 2017.
- [9] A. Weisheimer, N. Schaller, C. O'Reilly, D. A. MacLeod, and T. Palmer, "Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution," *Quarterly Journal of the Royal Meteorological Society*, vol. 143, no. 703, pp. 917–926, 2017.
- [10] M. Dobrynin, T. Kleine, A. Düsterhus, and J. Baehr, "Skilful seasonal prediction of ocean surface waves in the Atlantic ocean," *Geophysical Research Letters*, vol. 46, no. 3, pp. 1731–1739, 2019.
- [11] L. Wang, M. Ting, and P. J. Kushner, "A robust empirical seasonal prediction of winter NAO and surface climate," *Scientific Reports*, vol. 7, no. 1, pp. 1–9, 2017.
- [12] T. Lee, T. B. M. J. Ouarda, and S. Yoon, "KNN-based local linear regression for the analysis and simulation of low flow extremes under climatic influence," *Climate Dynamics*, vol. 49, no. 9–10, pp. 3493–3511, 2017.
- [13] R. J. Hall, A. A. Scaife, E. Hanna, J. M. Jones, and R. Erdélyi, "Simple statistical probabilistic forecasts of the winter NAO," *Weather and Forecasting*, vol. 32, no. 4, pp. 1585–1601, 2017.
- [14] Z. Jing, L. Pu, Z. Feng et al., "CloudNet: ground-based cloud classification with deep convolutional neural network," *Geophysical Research Letters*, vol. 45, no. 16, pp. 8665–8672, 2018.
- [15] H. Xiao, F. Zhang, Q. He et al., "Classification of ice crystal habits observed from airborne cloud particle imager by deep transfer learning," *Earth and Space Science*, vol. 6, no. 10, pp. 1877–1886, 2019.
- [16] S. Gao, P. Zhao, B. Pan et al., "A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network," *Acta Oceanologica Sinica*, vol. 37, no. 5, pp. 8–12, 2018.
- [17] Y.-L. Hu and L. Chen, "A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and Differential Evolution algorithm," *Energy Conversion and Management*, vol. 173, pp. 123–142, 2018.
- [18] S. Kim, S. Hong, M. Joh et al., "DeepRain: ConvLSTM network for precipitation prediction using multichannel radar data," 2017, <https://arxiv.org/abs/1711.02316>.
- [19] F. Wang, Y. Yu, Z. Zhang, J. Li, Z. Zhen, and K. Li, "Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting," *Applied Sciences*, vol. 8, no. 8, p. 1286, 2018.
- [20] B. Mu, C. Peng, S. Yuan et al., "ENSO forecasting over multiple time horizons using ConvLSTM network and rolling mechanism," in *Proceedings of the 2019 international joint conference on neural networks (IJCNN)*, pp. 1–8, Budapest, Hungary, July 2019.
- [21] S. Yuan, X. Luo, B. Mu, J. Li, and G. Dai, "Prediction of North Atlantic Oscillation index with convolutional LSTM based on ensemble empirical mode decomposition," *Atmosphere*, vol. 10, no. 5, p. 252, 2019.
- [22] H. C. Andersson, "Influence of long-term regional and large-scale atmospheric circulation on the Baltic sea level," *Tellus A*, vol. 54, no. 1, pp. 76–88, 2002.
- [23] F. Ali, M. Kabir, M. Arif, Z. N. Khan Swati, Z. U. Khan, M. Ullah et al., "DBPPred-PDSD: machine learning approach for prediction of DNA-binding proteins using Discrete Wavelet Transform and optimized integrated features space," *Chemometrics and Intelligent Laboratory Systems*, vol. 182, pp. 21–30, 2018.
- [24] Y. Liao, X. Zeng, and W. Li, "Wavelet transform based convolutional neural network for gearbox fault classification," in *Proceedings of the 2017 Prognostics and System Health Management Conference (PHM-Harbin)*, pp. 1–6, IEEE, Harbin, China, July 2017.
- [25] S. B. Feldstein, "The timescale, power spectra, and climate noise properties of teleconnection patterns," *Journal of Climate*, vol. 13, no. 24, pp. 4430–4440, 2000.

- [26] R. L. Wilby, G. O'hare, and N. Barnsley, "the North atlantic oscillation and British isles climate variability, 1865-1996," *Weather*, vol. 52, no. 9, pp. 266–276, 1997.
- [27] U. Schneider and C. D. SCHONWIESE, "Some statistical characteristics of El Niño/southern oscillation and North Atlantic oscillation indices," *Atmosfera*, vol. 2, no. 3, 2009.
- [28] J. C. Rogers, D. J. Bathke, E. Mosley-Thompson et al., "Atmospheric circulation and cyclone frequency variations linked to the primary modes of Greenland snow accumulation," *Geophysical Research Letters*, vol. 31, no. 23, 2004.
- [29] Q. Liu, "On the definition and persistence of blocking," *Tellus A*, vol. 46, no. 3, pp. 286–298, 1994.
- [30] G. Dai, M. Mu, and Z. Jiang, "Evaluation of the forecast performance for North Atlantic oscillation onset," *Advances in Atmospheric Sciences*, vol. 36, no. 7, 2019.