A Novel Deep Learning Model by BiGRU with Attention Mechanism for Tropical Cyclone Track Prediction in the Northwest Pacific

TAO SONG,a,b YING LI,a FAN MENG,a PENGFEI XIE,a and DANYA XU,d

a College of Computer and Communication Engineering, China University of Petroleum, Qingdao, China
b Department of Artificial Intelligence, Faculty of Computer Science, Polytechnic University of Madrid, Madrid, Spain
c School of Geosciences, China University of Petroleum, Qingdao, China
d Guangdong Laboratory of Marine Science and Engineering, Zhuhai, China

(Manuscript received 31 December 2020, in final form 20 July 2021)

ABSTRACT: Tropical cyclones are among the most powerful and destructive meteorological systems on Earth. In this paper, we propose a novel deep learning model for tropical cyclone track prediction method. Specifically, the task is regarded as a time series predicting challenge, and then a deep learning framework by a bidirectional gate recurrent unit network (BiGRU) with attention mechanism is developed for track prediction. This proposed model can excavate the effective information of the historical track in a deeper and more accurate way. Data experiments are conducted on tropical cyclone best-track data provided by the Joint Typhoon Warning Center (JTWC) from 1988 to 2017 in the northwestern Pacific Ocean. Results show that our model performs well for tracks of 6, 12, 24, 48, and 72 h in the future. The prediction results show that our proposed combined model is superior to state-of-the-art deep learning models, including a recurrent neural network (RNN), long short-term memory neural network (LSTM), gate recurrent unit network (GRU), and BiGRU without the use of attention mechanism. In comparison with the methods used by the China Meteorological Administration, Japan Meteorological Agency, and the JTWC, our method has obvious advantages in the mid- to long-term track forecasting, especially in the next 72 h.

KEYWORDS: Tropical cyclones; Forecasting; Deep learning

1. Introduction

Tropical cyclones are warm cyclonic circulations that occur over the tropical and subtropical seas (Montgomery and Farrell 1993). The northwestern Pacific Ocean is the sea area with the most frequent, strongest, and most destructive tropical cyclones in the world. According to statistics, an average of about 35 tropical cyclones are generated in the northwestern Pacific every year, accounting for about 1/3 of the total number of tropical cyclones in the world (Guan et al. 2018). China is located on the west coast of the northwestern Pacific and is one of the countries where tropical cyclones frequently land (Xiao and Xiao 2010). When tropical cyclones approach and make landfall, they often bring violent winds, heavy rains, storm surges, and other meteorological disasters, which seriously restrict the economic development and human survival and development, social and economic construction, and ecological environmental protection.

Methods for forecasting tropical cyclone tracks can be mainly divided into two groups. One involves the statistical dynamic forecasting method, which is mainly based on regression technique (DeMaria and Kaplan 1999). The regression equation between tropical cyclone track and its correlation factors is used to solve the forecasting problem. The climatologically aware forecasting technique is one of the technologies that the statistical dynamic forecasting method relies on (Jeffries and Miller 1993). It is a subjective forecast method developed by the Joint Typhoon Warning Center (JTWC). Climatology and persistence statistical models (CLIPER) are the classic statistical prediction model (Neumann and Lawrence 1975). However, the tropical cyclone track data are nonlinear and random. The equations established by traditional statistical methods are simple, and the prediction results may not be ideal.

The other approach is a numerical weather prediction (NWP) model with physical dynamics equations, which simulates the change process of tropical cyclone track types through differential equations, which are calculated and solved with a supercomputer. For instance, Sanders proposed tropical cyclone forecasting models such as “SANBAR” and barotropic in the mid- to late twentieth century (Sanders and Burpee 1968; Sanders et al. 1975).

At present, the NWP models can be roughly divided into subjective and objective forecasting methods. Subjective forecasting methods used for tropical cyclone research are used by the China Meteorological Administration (CMA), the Japan Meteorological Agency (JMA), and the JTWC. Objective forecasting methods include global models such as NCEP GFS (NCEP 2012) and ECMWF IFS (ECMWF 2011) and regional models such as the Global Regional Assimilation and Prediction System–Tropical Cyclone Model (GRAPES-TCM) and Guangzhou numerical models. NWP models require a lot of time and resources to process complex dynamic equations. The solution process of NWP models is...
very complicated. Therefore, we want to try a new method to solve the problem of tropical cyclone prediction.

With the gradual establishment of meteorological satellites, ocean observatories, and ground stations, the amount of ocean and air data continues to accumulate. Combining deep learning models with big data environments provides us with a new opportunity to predict the track of tropical cyclones. In recent years, data-driven deep learning algorithms have been successfully applied in image processing, natural language processing, target detection, and other fields. Deep learning technology has strong nonlinear fitting capabilities, which have shown great advantages in processing large amounts of complex time series data. Song et al. (2005) used a support vector machine model to predict tropical cyclone tracks and improve prediction accuracy by reducing input dimensions and data. But it did not learn the inherent temporal correlations in the historical data of typical cyclones. Rüttgers et al. (2019) used a generative adversarial network to predict the track of typhoons with satellite images as input. Gao et al. (2018) developed a forecast model based on a long short-term memory (LSTM) neural network (Hochreiter and Schmidhuber 1997) to predict the track of tropical cyclone. Because of the limitations of LSTM network structure, the mid- to long-term prediction results are not so ideal.

In this paper, following this research line, we propose a novel tropical cyclone track prediction method by deep learning technology. Initially, we take the track task as a time series prediction challenge, and a novel deep learning model with a combination of a bidirectional gate recurrent unit network (BiGRU) and an attention mechanism is developed. Since BiGRU can find complete past and future track information from the current time sequence, our model can fully obtain and mine track information. With the attention mechanism, our model can capture the input time point that has the greatest impact on the prediction of the next step and enhance the importance of the input at that time point, thereby improving mining efficiency and accuracy.

The contributions of this work are mainly summarized as follows.

1) A novel tropical cyclone prediction method with an attention mechanism is proposed. More meaningful time-point information can be mined from massive track data, effectively improving the accuracy of the track prediction.

2) In comparison with other state-of-the-art deep learning methods, our model performs better. In medium- and long-term forecasting, it is better than the traditional forecasting mode. Especially, in 72-h prediction our model achieves a very prominent advantage.

2. Method

a. GRU

Recurrent neural networks (RNNs) take sequence data as input and perform recursion in sequence evolution direction. All of the recurrent units are connected by chain (Karpathy et al. 2015). A gated recurrent unit (GRU) is a kind of RNN, proposed in Cho et al. (2014), that has fewer parameters to convergence. The structure of GRU is shown in Fig. 1.

Each hidden layer in GRU can control the flow of information by a reset gate $r_t$ and update gate $z_t$. The reset gate $r_t$ is used to control how much information is retained from the previous hidden layer $h_{t-1}$ to the current candidate hidden layer $\tilde{h}_t$:

$$r_t = \sigma(\omega_0h_{t-1} + \omega_1x_t + b_r) \text{ and}$$

$$\tilde{h}_t = \tanh[\omega_0(r_t \cdot h_{t-1}) + \omega_1x_t + b_h],$$

where $x_t$ is the input value of the current moment, $b$ is the bias, $\omega$ is the coefficient, $\sigma$ is the sigmoid function, $\tanh$ is the hyperbolic tangent function, and $\cdot$ is the Hadamard product.

The update gate $z_t$ is used to control how much information needs to be forgotten from the previous hidden layer $h_{t-1}$:

$$z_t = \sigma(\omega_2h_{t-1} + \omega_3x_t + b_z).$$

The current hiding layer $h_t$ is determined by the update door $z_t$, the previous hidden layer $h_{t-1}$, and the current candidate hiding layer $\tilde{h}_t$:

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t,$$

b. BiGRU

In RNN and GRU, the transmission of state is one-way from forward to back, whereas in BiGRU it divides the neurons in regular GRU into the forward layer and the back layer (Bahdanau et al. 2015), which represent positive state and negative state, respectively. The topological structure of BiGRU is shown in Fig. 2. It is easy to capture the impact of input from the past and future on the current state (Goerse et al. 2004). BiGRU is usually used to learn and extract the feature expression of data more accurately, and to mine the deeper internal expression of historical data.

At time $t$, the hidden layer of BiGRU at the current moment is determined by the front hidden layer $\hat{h}_t^+$ and the back hidden layer $\hat{h}_t^-$ together. We denote by $G$ the activation function, and $x_t$ is the input at the current moment; $u_t$ and $v_t$ are the coefficients, and $b_t$ is the bias.
The hidden layer of two-way GRU at the current moment is determined by the front hidden layer and the back hidden layer together. The layers are calculated as follows:

$$
\overrightarrow{h_t} = G(x_t, \overrightarrow{h_{t-1}}), \\
\overleftarrow{h_t} = G(x_t, \overleftarrow{h_{t-1}}), \text{ and} \\
h_t = u_t \overrightarrow{h_t} + v_t \overleftarrow{h_t} + b_t.
$$ (5)

$$
\overrightarrow{h_t} = G(x_t, \overrightarrow{h_{t-1}}), \\
h_t = u_t \overrightarrow{h_t} + v_t \overleftarrow{h_t} + b_t.
$$ (6)

$$
y = \sum_{i=1}^{n} \omega_i h_i, \\
\omega_i = \exp(e_i), \text{ and} \\
e_i = \alpha_i \text{ATT}(\beta_i h_i + c_i).
$$ (7)

### c. Attention mechanism

The attention of the human brain focuses on one part at a time, reducing the attention on or even ignoring other parts. A human-like attention mechanism is involved in artificial intelligence, particularly in deep learning models, to imitate that attention behavior (Xu et al. 2015). Attention mechanisms can give different attention weights to characteristics at different time positions so as to capture the influence of different time characteristics. The structure of the attention mechanism is shown in Fig. 3.

After the attention module, the state of the hidden layer will be changed. We denote by $\omega_i$ the proportion of each input hidden layer according to the importance, and by $h_i$ the input of the initial hidden layer. The new hidden layer state $y$ is the sum of the $\omega_i$ and $h_i$ products:

$$
y = \sum_{i=1}^{n} \omega_i h_i, \\
\omega_i = \exp(e_i), \text{ and} \\
e_i = \alpha_i \text{ATT}(\beta_i h_i + c_i).
$$ (8)

$$
\omega_i = \exp(e_i), \text{ and} \\
e_i = \alpha_i \text{ATT}(\beta_i h_i + c_i).
$$ (9)

$$
e_i = \alpha_i \text{ATT}(\beta_i h_i + c_i).
$$ (10)

By $e_i$, we denote the state value of the hidden layer $h_i$ at time $t$, and symbols $\alpha_i$ and $\beta_i$ are the weight coefficient matrix, with $c_i$ being the coefficient. ATT represents the activation function.

### d. Our model

In forecasting tropical cyclone tracks, the latitude and longitude sequence of historical tracks reflects the changing laws and hides important information. The BiGRU model mines the time series information at a deeper level. It considers that the data at each moment in the input history track have the same impact on the next step prediction result, which is not the case in reality. In the long-term track sequence, not every moment of the information will play a role in the prediction results. There may be some interference with unimportant information that affects the learning and prediction of the neural network. This problem can be effectively solved by an attention mechanism. It calculates the influence degree of each time point of input on the predicted value, and assigns corresponding weights to each time point, making the model focus on the input of important time points. In addition, attention mechanisms can process longer time series, so that the characteristics of important information will not disappear with the increase of step size. Highlighting the influence of more important information makes it easier for the model to learn the long-term interdependent relationships in the sequence.

We propose here a novel prediction method by combination of BiGRU with attention mechanisms to improve the accuracy of tropical cyclone track prediction. First of all, the reset gate in GRU intends to capture the short-term dependencies of the tropical cyclone track sequence, and the update gate helps to capture the long-term dependencies. In our model, BiGRU is used to dynamically model the input time series data. An attention mechanism is involved to give different probability weights to the BiGRU hidden layer so as to strengthen the influence of important time point information on the tropical cyclone track prediction. Then, the dimension is adjusted through the dense layer to output the predicted value.

The structure is shown in Fig. 4 with the flowchart of our model. Tropical cyclone track prediction is regarded as a time series prediction problem, which can be predicted by its historical change principle. Mathematically, the problem is described by inputting the historical track value from time $T_i$...
to time $T_r$, and then calculating the track value at the time $T_{t+n}$. In the model, $T$ is the historical tropical cyclone track on which the prediction depends, and $n$ indicates the predicted time point in the future.

3. Data experiments

a. Environment experiment

Data experiments are run under the environment of Intel Core i5-10210U CPU at 1.60 GHz, with a 16-Byte RAM operating system. All deep learning methods were implemented using Python 3.6.5, Tensorflow 2.3.1, and Keras 2.4.3.

b. Data preprocessing

The northwestern Pacific Ocean is an area that is prone to tropical cyclones and therefore has sufficient data. The track data of tropical cyclones in the northwestern Pacific provided by JTWC are used in the experiment (JTWC 2018). The data observed during the 30 years from 1988 to 2017 were selected, and a total of 25124 records covering 907 tropical cyclones were collected to form the dataset of this experiment. All of

![Flowchart of the proposed method.](image)

![All tropical cyclone data used in the experiment.](image)
the tropical cyclone data used in the experiment are shown in Fig. 5. Each record has an interval of 6 h, with the latitude and longitude coordinates of the tropical cyclone position.

We found that many of the same tropical cyclone track data appeared multiple times in the downloaded data, and we deleted duplicate values. The time interval of each record in our dataset is 6 h, and the time record points are 0000, 0600, 1200, and 1800 UTC. Some records at 0300, 0900, and 1500 UTC were found in the data. We cleaned up these redundant values so as not to affect the experimental results. Because of the large difference between latitude and longitude data, the data were normalized. A normalization function was applied to map their values between 0 and 1. Normalization of the input value can effectively help the neural network to train better.

The whole dataset was divided into two parts: a training set and a test set. An initialized model was trained by the training set. During model training, the data of the training set go through all the steps in Fig. 4, from the input, to the BiGRU, attention mechanism, and dense layer, and finally to the output layer. The model parameters in the initial model were set randomly. The parameters were continuously optimized through the back propagation of errors until the optimal state of the model was reached, which represented the end of the model training.

The test set was used for model evaluation. We made predictions on the model in the test set. The data in the test set were input into the trained model, and after the BiGRU, attention mechanism, and dense layers were applied, the output result was the predicted value. The test set was not used for training. It comprised data the model had never seen in training before. The content of our result analysis comes from the results of the model on the test set. The training set is data from 1988 to 2008, and the test set is data from 2009 to 2017.

c. Training processing

In this paper, the optimization algorithm Adam (adaptive moment estimation) is used to conduct training optimization for the model parameters. Adam is a first-order optimization algorithm that can replace traditional stochastic gradient descent (Kingma and Ba 2015). The mean square error (MSE) function is used as the loss function:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (P_i - R_i)^2. \quad (11)$$

4. Results and analysis

a. Evaluation indices and parameters

For evaluation, mean absolute error (MAE), root-mean-square error (RMSE), mean absolute percentage error (MAPE), and mean absolute position error (APE) are selected as evaluation indices to measure the difference between the predicted track value $P_i$ and the real track value of model $R_i$.

**Table 1. Performances of the proposed model with different batch sizes.** Here and in subsequent tables, the boldface font indicates the best performance.

<table>
<thead>
<tr>
<th>Batch size</th>
<th>Lat MAE</th>
<th>Lat RMSE</th>
<th>Lat MAPE</th>
<th>Lon MAE</th>
<th>Lon RMSE</th>
<th>Lon MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>0.0274</td>
<td>0.0611</td>
<td>0.1224</td>
<td>0.0143</td>
<td>0.0423</td>
<td>0.0201</td>
</tr>
<tr>
<td>128</td>
<td>0.0219</td>
<td>0.0543</td>
<td>0.1002</td>
<td>0.0161</td>
<td>0.0427</td>
<td>0.0189</td>
</tr>
<tr>
<td>256</td>
<td>0.0249</td>
<td>0.0578</td>
<td>0.1132</td>
<td>0.0178</td>
<td>0.0431</td>
<td>0.0190</td>
</tr>
</tbody>
</table>

The RMSE is the square root of the ratio of the error to the number of observations between the predicted value and the true value. The smaller the RMSE value is, the better is the model performance. Its expression is shown by

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (P_i - R_i)^2}. \quad (12)$$

MAE represents the average of the absolute value of the error between the predicted value and the true value. The greater the error is, the worse is the model performance. It can be expressed as

$$\text{MAE} = \frac{1}{n} \sum_{i=0}^{n-1} |P_i - R_i|. \quad (13)$$

MAPE represents the average error rate of the actual value. It is the ratio of the error between the predicted value and the actual value. The value of MAPE is between 0 and 1. The closer the value is to 0, the better is the model. The calculation is

$$\text{MAPE} = \frac{1}{n} \sum_{i=0}^{n-1} \left| \frac{P_i - R_i}{R_i} \right|. \quad (14)$$

The mean absolute position error (APE) measures the distance between the real position and the predicted position in kilometers. The smaller the APE value is, the better is the model. The calculation is

$$\text{APE} = R \arccos(\sin(Lat_R) \sin(Lat_P) \cos(Lon_R - Lon_P)) + \cos(Lat_R) \cos(Lat_P) \frac{\pi}{180}. \quad (15)$$

**Table 2. Performances of the proposed model with different numbers of neurons.**

<table>
<thead>
<tr>
<th>No. of neurons</th>
<th>Lat MAE</th>
<th>Lat RMSE</th>
<th>Lat MAPE</th>
<th>Lon MAE</th>
<th>Lon RMSE</th>
<th>Lon MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.0199</td>
<td>0.0502</td>
<td>0.0992</td>
<td>0.0121</td>
<td>0.0412</td>
<td>0.0190</td>
</tr>
<tr>
<td>30</td>
<td>0.0221</td>
<td>0.0504</td>
<td>0.1165</td>
<td>0.0155</td>
<td>0.0437</td>
<td>0.0206</td>
</tr>
<tr>
<td>40</td>
<td>0.0298</td>
<td>0.0547</td>
<td>0.1198</td>
<td>0.0172</td>
<td>0.0498</td>
<td>0.0203</td>
</tr>
</tbody>
</table>
There is no clear prior experience with either the values of parameter batch size or the number of neurons in the hidden layer. We predict the tropical cyclone track in the next 6 h, and we selected the optimal value of batch size and number of neurons according to the evaluation indices.

For the determination of the batch size, we chose several values that people usually choose when training the model, which are 64, 128, and 256. Experimental results are shown in Table 1. If the batch size is 128, only one index is not optimal. We set the value of batch size of 128 for our follow-up experiments. We run experiments on different number of neurons in hidden layers. As shown in Table 2, we found that when the number of neurons in hidden layers was set to 20, the model worked best.

The division ratio of the training set and the test set also has a certain impact on the prediction effect of the experiment. The dataset is arranged according to the time when the tropical cyclone was generated, and one year's data compose a set. This is also the arrangement when downloading data from JTWC. The whole dataset was divided into ratios of 6:4, 7:3, 8:2, and 9:1 by the year. We tried training with different division ratios for the tropical cyclone path predictions for the next 12, 24, 48, and 72 h. As shown in Fig. 6, the mean APE error of the division ratio of 7:3 is the least among the four time steps. In other words, the model works best when the training set is data from 1988 to 2008 and the test set is data from 2009 to 2017.

b. Comparison with other methods of deep learning

In this section, we compare our method with other deep learning methods. To ensure the fairness and credibility of the comparative experiment, all the models use historical data of the previous 72 h to predict the track information of the next 6 h; that is, they use the previous 12 time steps to predict the information for the next time step. In this article, other deep models used for comparison are the same as our proposed model training method, training set, and test set. The parameters have reached the best state during training. The results of MAE, RMSE, and MAPE are used for comprehensive analysis. The comparison results are shown in Table 3, with the results with the best performance shown in boldface type.

It is found that our model performs the best in predicting tracks among all the metrics, and the RNN has the worst performance. BiGRU is indeed superior to regular GRU by comparing the results. According to the result of BiGRU and BiGRU with attention mechanism, we found that the performance of BiGRU has been enhanced. It shows that adding an attention mechanism plays a role in improving the prediction accuracy.

c. Comparison with traditional methods

With regard to the assessment of tropical cyclone forecast errors, people tend to pay more attention to the forecast results for the next 24 h and longer. In this section, we analyze and compare the proposed model with the results of traditional tropical cyclone forecasting models. The prediction time steps are 24, 48, and 72 h.

We verify the 2009–17 data in the test set, based on the best-track dataset compiled based on the Shanghai Typhoon Institute of the CMA (Lu et al. 2021; Ying et al. 2014). The error of the test set is calculated, which is the average error over the nine years from 2009 to 2017. The track data released by the three official agencies of CMA, JMA, and JTWC are selected for comparison. The CMA, JMA, and JTWC results come from the statistical analysis of tropical cyclones in the northwestern Pacific by Zhan et al. (2010), Tang et al. (2011), Chen and Cao (2014), and Chen et al. (2012, 2013, 2015, 2017, 2018, 2019).

In this study, the historical track information of the previous 72 h is used to predict the track in the future of 24, 48, and 72 h. As shown in Table 4, JTWC performs the best among all the methods in predicting tropical cyclones of the coming 24 h. The proposed BiGRU prediction model with an attention mechanism performs better than the above-mentioned methods in forecasting the tracks for the next 48 and 72 h. This indicates that our model is more suitable for mid- to long-term tropical cyclone time scale
forecasting. The model with the attention mechanism can fully mine and extract 72 h of historical data. Moreover, with the extension of the prediction time, the errors of other models will accelerate and accumulate over time. The method we proposed is relatively stable, as the model performance declines only slowly. The attention mechanism is used to better extract key features with greater impact from long-term sequences, effectively alleviating the problem of rapid increase in error with the prediction time step.

Furthermore, we analyze and compare the errors of each year. We calculated the error of each year in the test set and compared it with the corresponding error in CMA, JMA, and JTWC during the same period. It can be seen from Fig. 7 that our method does not perform very well in the next 24-h prediction. But in the next 72-h forecast, we have an absolute advantage. The error is about 170 km. The result is much lower than the prediction error of the CMA, JMA, and JTWC.

**d. Visualization of prediction results**

In practice, we need to pay more attention to the prediction effect of the prediction model on the tropical cyclone track of the mid- to long term in the future. The forecast results in this time-step range are more in line with actual needs. To further verify the true validity of the model proposed in this paper, we did a visual analysis of the prediction results of the four strong tropical cyclones in the next 72 h. They were tropical cyclone Malakas in 2016, Matmo in 2014, Koppu in 2015, and Khanun in 2017. These strong tropical cyclones all caused considerable damage to the area at that time. We also visualized the results predicted by other deep learning methods for comparison. The results are shown in Fig. 8. The blue line in the Fig. 8 represents the actual tropical cyclone track value, and the other lines represent the prediction result of deep learning methods. It can be seen from Fig. 8 that our model can better predict the track. The predicted track was relatively close to the real. Although the position coordinates had some errors, our forecast result was the best among all deep learning methods. The tropical cyclone track is affected by many factors, including landform, geostrophic force, cold air, and so on (Roy and Kovordányi 2012). Therefore, the errors exist for certain understandable reasons.

**5. Conclusions**

In this paper, we proposed a novel neural network model based on the combination of BiGRU and an attention mechanism for tropical cyclone track prediction. The BiGRU layer was used to extract the deeper nonlinear and complex time characteristics of tropical history cyclone data from the time series dataset. The attention mechanism was adopted to help the model capture the characteristic influence degree of different input time. The model focused on the input of important time points so as to improve the forecast accuracy of tropical cyclone tracks. Historical data on tropical cyclones in the northwestern Pacific provided by JWTC during 1988–2017 were used for model training and testing. The final results verify that the proposed BiGRU prediction model with attention mechanism has better

<table>
<thead>
<tr>
<th>Methods</th>
<th>24 h</th>
<th>48 h</th>
<th>72 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>China Meteorological Admin.</td>
<td>93.88</td>
<td>166.51</td>
<td>255.91</td>
</tr>
<tr>
<td>Japan Meteorological Agency</td>
<td>99.11</td>
<td>167.23</td>
<td>273.86</td>
</tr>
<tr>
<td>Joint Typhoon Warning Center</td>
<td>93.39</td>
<td>165.31</td>
<td>257.17</td>
</tr>
<tr>
<td>BiGRU + attention</td>
<td>147.379</td>
<td>164.8218</td>
<td>171.1925</td>
</tr>
</tbody>
</table>

**FIG. 7.** The error results of different years and different prediction time steps with NWP and our model: (a) forecasting the error of the next 24 h, (b) forecasting the error of the next 48 h, and (c) forecasting the error of the next 72 h.
FIG. 8. The visualization of prediction results of the four strong tropical cyclones in the future 72 h, showing the track predictions of tropical cyclone (a) Koppu (2015), (b) Matmo (2014), (c) Malakas (2016), and (d) Khanun (2017).
performance than deep learning models such as RNN, LSTM, GRU, and BiGRU.

In addition, when compared with the CMA, JMA, and JTWC, our model achieves better performance in the track prediction for 48 and 72 h. Some specific tropical cyclone track prediction results are visualized to show the prediction effect of our proposed model more intuitively. Although there are still some errors, the prediction of the trend of the tropical cyclone track is accurate.

The main reason why we choose previous 72-h track data to predict the future 72-h track is that after several data experiments we found that this scheme is associated with the best performance. The physical reason is hard to explain at this moment, but it is a very good research direction in both deep learning and physical oceanic models.

In general, the method proposed in this paper can learn the effective information of tropical cyclone track in historical data and can accurately predict the future tropical cyclone track in the mid- to long term. As the results show, we are not very effective in short-term forecasting. We suspect that atmospheric environmental factors have a greater impact on short-term forecasts. Therefore, for future work, we intend to proceed from the following aspects. First, some environmental variables, such as terrain and geostrophic deflection force, can be added into the model so as to make a more comprehensive track prediction. The second is to increase the prediction variables of the model, which can predict the tropical cyclone effective information such as tropical cyclone intensity, rainfall, and wind speed while predicting the tropical cyclone track. Third, only tropical cyclone data in the northwestern Pacific are discussed in this paper. The next step can be to combine global tropical cyclone track data to design and build a global multifluence information fusion neural network for prediction.

Acknowledgments. This work was supported by the National Key Research and Development Program (2018YFC1406201) and the Natural Science Foundation of China (Grant U1814164). The project was supported by the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311020008).

REFERENCES


