

TWO-DIMENSIONAL SIGNIFICANT WAVE FORECAST BY USING CONVOLUTIONAL LSTM AND ERA5 DATA

Yuma Sato, Niigata University, f23e019f@mail.cc.niigata-u.ac.jp
 Ryota Nakamura, Niigata University, r-nakamura@eng.niigata-u.ac.jp

BACKGROUND

Wave forecasting has been used in various situations, such as feasibility decision of construction works. In addition, the demand for short-term wave forecasting for several hours to several days lead time is expected to increase under the situation where development of offshore wind farms in Japan.

On the other hand, in addition to the operational global ocean surface wave models, machine learning has been employed for wave forecasting as well. This is mainly because of lower computational cost compared with the numerical model. However, wave prediction using deep learning is mostly based on fixed-point prediction, and there are few researches regarding two dimensional field of surface wave forecasting. It should be noted that in the wind engineering field, Chen et al. (2022) has enabled the prediction of 2-D wind velocity fields. Therefore, it can be possible to conduct the two-dimensional fields of wave forecasting by using the machine learning.

In this study, we attempted to predict the 2D wave field using convolutional LSTM with training dataset based on the 5th generation ECMWF atmospheric reanalysis (ERA5). The evaluation of predicted significant wave height was carried out on semi-closed Japan sea area.

METHODS

ERA5 data was used for training data of the convolutional LSTM model. ERA5 is a fifth-generation atmospheric reanalysis by ECMWF. This study used the reanalysis data over a 20×20 mesh ranging from 131°E to 140°E and from 35°N to 44°N, covering the Sea of Japan (Fig. 1). The time resolution of the reanalysis data input to the model was 1 hour. Numerical results for 21 years from January 1, 2002, to December 31, 2022 were used in this study; data for the period from 2002 to 2021 were assigned to train the results, and data of 2022 were

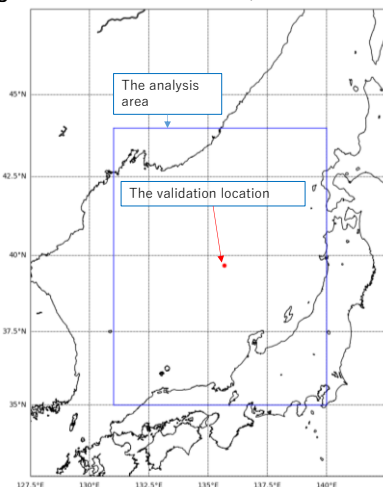


Figure 1 - The area of teacher data for the convolutional LSTM model.

used to test the model. In inputting the data into the model, the size was changed to 32 x 32 by bilinear interpolation.

Convolutional LSTM is a kind of LSTM which is capable of numerical prediction using time-series data with the added functionality of convolution, which enables feature extraction in two dimensions.

The number of layers in the convolutional LSTM model was set to 12 including input and output layers. The batch size was set to 80, and the number of epochs was set to 10. The data for training was converted into a sequence consisting of 6 time-series data and then input to the model.

In addition, to investigate the relationship between the number of training data and prediction accuracy, the model was trained with four different patterns of datasets: 2002-2021 (20 years), 2007-2021 (15 years), 2012-2021 (10 years) and 2017-2021 (5 years). The training conditions and conceptual figure are shown in Table 1 and Figure 2.

Table 1 - Setting of machine learning

Predict area	131° E - 140° E, 35° N - 44° N
Period of studying	January 1, 2002 – December 31, 2021
Period of prediction	January 1, 2022 – December 31, 2022
Data used for learning	Significant wave height (normalized from 0 to 1) (Sequence containing 6 time series data)
Predicted data	Significant wave height
Model	Convolutional LSTM
Loss functions and optimization functions	MSE, Adam
epochs / batch size	10 / 128

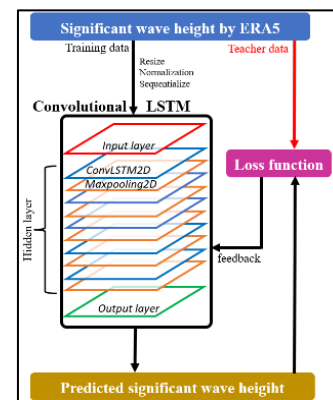


Figure 2 - Conceptual diagram of convolutional LSTM

After training the datasets, we predicted the significant wave heights using test data. We inputted the test data into LSTM and predicted the significant wave heights at 1, 3, 6, 12, and 24 hours from the last data in the sequence. We used the correlation coefficient and RMSE (root mean

square error) as a comparison and evaluation measure between the prediction results and the correct data.

RESULTS

Predictions were made three times for each prediction time and number of input data. Table 2 shows the summary of 1, 2, 3, 6, 12 and 24 hours lead time significant wave forecast. The results of time series, comparison and spatial distribution of machine learning prediction compared with the ERA5 reanalysis are shown in Figs. 3, 4, and 5, respectively. The scatter plots and correlation coefficients are calculated in the center location of the simulations (Fig. 1)

Table 2 - The results of RMSE and correlation coefficient on each case

R : correlation coefficient

1 hour lead time forecast			6 hour lead time forecast		
Amount of data for training	RMSE[m]	R	Amount of data for training	RMSE[m]	R
For 5 years	0.1889	0.9937	For 5 years	0.5372	0.8808
10years	0.2010	0.9889	10years	0.4765	0.9616
15years	0.0575	0.9988	15years	0.7336	0.8309
20years	0.0970	0.9971	20years	0.5364	0.8828

2 hour lead time forecast			12 hour lead time forecast		
Amount of data for training	RMSE[m]	R	Amount of data for training	RMSE[m]	R
For 5 years	0.2454	0.9916	For 5 years	0.8554	0.6397
10years	0.1702	0.9888	10years	0.8971	0.5728
15years	0.1568	0.9905	15years	1.5391	0.5593
20years	0.1772	0.988	20years	0.9277	0.7121

3 hour lead time forecast			24 hour lead time forecast		
Amount of data for training	RMSE[m]	R	Amount of data for training	RMSE[m]	R
For 5 years	0.2195	0.9849	For 5 years	1.2387	0.4308
10years	0.3261	0.9750	10years	1.0429	0.3301
15years	0.2413	0.9752	15years	2.4910	0.2219
20years	0.2804	0.9749	20years	1.5168	0.2647

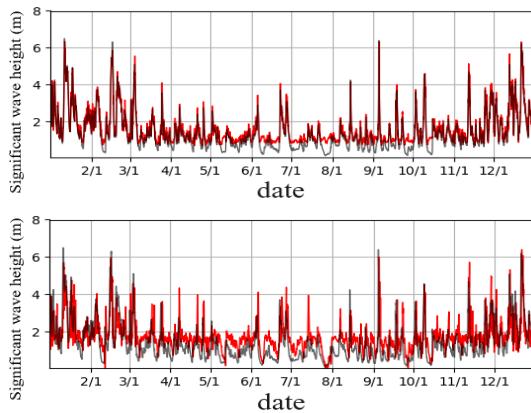


Figure 3 - The predicted significant wave height on time series diagrams (3h (upper), 12h (lower) lead time prediction). The results are shown at 39°41'15" N, 135°41'15" E (The center of analyzed domain).

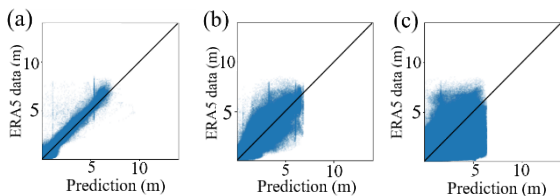


Figure 4 - Comparison of predicted and observed significant wave height ((a):1-h, (b):6-h, and (c):24-h lead time prediction).

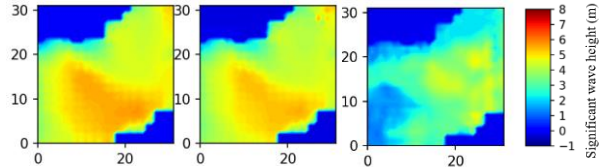


Figure 5 - Spatial distribution of significant wave height (ERA5 data, 3-h and 12-h lead time prediction)

Table 2 shows the short-term prediction is with high accuracy. For example, RMSE was 0.097 m and 0.280 m and R was 0.997 and 0.883 for 1- and 3-hour lead time prediction in the case using the 20 years training data sets. These results indicate that accurate prediction of significant wave heights is feasible. On the other hand, in the 6-hour forecast, the RMSE increased to 0.53 m, and the correlation coefficient decreased to 0.883 even in the case using the 20 years training data sets.

On the other hand, RMSE was 0.928 m and 1.517 m and R was 0.712 and 0.264 for 12- and 24-hour lead time prediction in the case using the 20 years training data sets (Table 2). RMSE and correlation coefficient indicate that the accuracy of predicted significant wave heights becomes lower beyond 12-h lead time even by using the 20 years training datasets and convolution LSTM model (Fig. 3, 4 and 5). The one of the possible reasons for low accuracy of the prediction was difficulty of capturing seasonality of wave prediction. The predicted significant wave height was underestimated in winter and overestimated in summer in the Japan Sea's side of Japan (Fig. 3). The Sea of Japan tends to experience high waves in winter due to strong monsoon winds from the Eurasian continent. On the other hand, in summer, the monsoon from the Pacific Ocean side blows. These characteristics cannot be well simulated in convolution LSTM because the predicted value was averaged over the annual year.

We also evaluated the relationship between the amount of data input to the model and forecast accuracy (Table 2). Generally, the amount of input data (training data) increases, the model would be able to read more patterns of wave field changes, which would improve its adaptability of the LSTM model. However, in the lead time becomes longer, the value of RMSE and R was not much improved in 12-h and 24-h lead time prediction. This result also show that the limitation of the two-dimensional significant wave forecast by using the Convolution-LSTM by using the large amount of data sets.

CONCLUSION

This study shows that short lead time (1-6 hours) significant wave height prediction on the Japan Sea can be feasible by using the convolution LSTM model. As the limitation, the wave height prediction over 12 hours produced lower accuracy. For the future task, climatic reanalysis data such as wind speed fields should be incorporated to increase the accuracy of wave prediction.

REFERENCES

Chen, Wang, Dong, Su, Han, Zhou, and Bao (2022): 2-D regional short-term wind speed forecast based on CNN-LSTM deep learning model, Energy Conversion and Management, Vol. 244, No.114451.