PREDICTION OF OCEAN WAVE HEIGHTS BASED ON A CONVOLUTIONAL-LSTM

Donghwi Son, Korea Institute of Ocean Science and Technology, dhson@kiost.ac.kr
Jeseon Yoo, Korea Institute of Ocean Science and Technology, jyoo@kiost.ac.kr
Sung-Hwan Park, Korea Institute of Ocean Science and Technology, spark@kiost.ac.kr

ABSTRACT
Recent climate change has led to rising sea levels, increasing the importance of disaster prevention and prediction, particularly in coastal areas, for events such as wave overtopping, storm surges, and beach erosion. In the field of time series data prediction, recent advancements in artificial intelligence technologies have primarily focused on models like Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM). However, these models have limitations in capturing spatial characteristics. In light of this, our study aims to establish a grid-based wave prediction model using a combination of Convolutional Neural Network (CNN) and LSTM, known as Convolutional LSTM (ConvLSTM), to account for spatial features in time series data prediction. Furthermore, our focus is on enhancing wave prediction performance with a simpler algorithm compared to previous research efforts. Additionally, we explored the potential for improving time series data prediction through the decomposition of temporal components within the input data. Overall, the application of time series decomposition elements in model training led to an improvement in accuracy of approximately 23% across the entire study area.

METHOD
As for the study area, we examined the East Asian region with a focus on the Korean Peninsula, and assessed the model's predictive accuracy by categorizing it into the Yellow Sea (YS), East China Sea (ECS), East Sea (ES), and the entire area, adjacent sea of Korea (ASK), based on the criteria defined in the Large Marine Ecosystems framework.

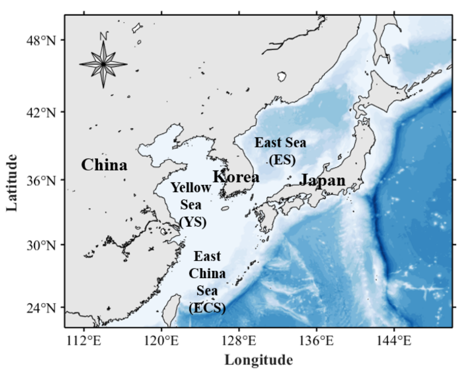

Figure 1 - Study Area (Adjacent Sea of Korea)

To train and construct the model, ERA5 reanalysis data with a spatial resolution of 0.5° and a temporal resolution of 1-hour intervals was utilized (Hersbach et al., 2020). Data for training and validation were collected from 2016 to 2020 (5 years), and the performance evaluation of the generated model was conducted using data from 2021 (1 year).

Figure 2 depicts the network configuration, which is composed of only three consecutive ConvLSTM layers. This design is intended to be simpler and more straightforward compared to the network structures in prior studies (Xiao et al., 2019; Zhou et al., 2021; Kim et al., 2022).

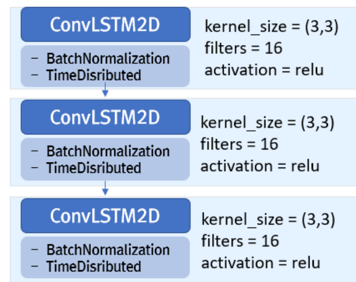

Figure 2 - Summary of the ConvLSTM Network Structure

To verify the effectiveness of component decomposition in time series data prediction, we categorized the experimental cases as shown in Figure 3 below. Case 1 represents predictions of the next 1 hour using 12-hour input data with only raw data, without utilizing any time series component decomposition factors. Case 2 corresponds to predictions of the next 12 hours using a 12-hour input with only raw data. Case 3 represents predictions of the next 12 hours using a 12-hour input with the incorporation of both raw data and three different time series decomposition datasets as part of the input data: raw data, Long-term trend (LT), Seasonal trend (ST), and Residual components (RE).

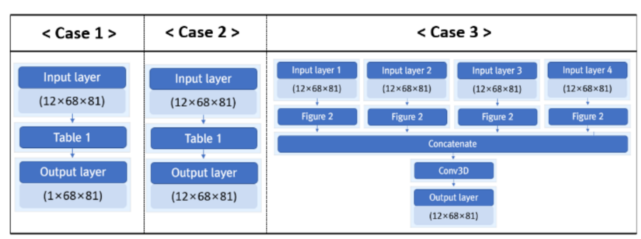

Figure 3 - Network Structures of the Experimental Cases

RESULTS
Figure 4 below presents the results of time series component decomposition for the entire study area (ASK). While the actual decomposition was performed individually at each grid point, the figure displays the average results for all grid points. The component decomposition involved separating the data into three constituent elements: Long-term trend, Seasonal trend, and Residual components.

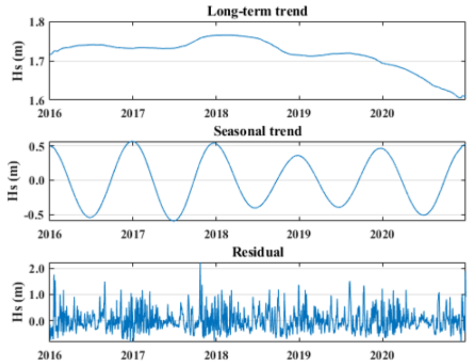

Figure 4 - Results of the time series decomposition in the entire study area (ASK)

Figure 5 displays the experimental results for different maritime regions conducted using the 2021 (1 year) data. As the prediction time step increased from 1 hour to 12 hours based on the past 12-hour input data, the relative RMSE also increased. It also shows that the model predictions with the application of time series component decomposition (Case 3) exhibited higher accuracy compared to those without it (Case 1, Case 2). This signifies that the decomposition of time series data components in the input data can contribute to accuracy improvement in predicting time series data.

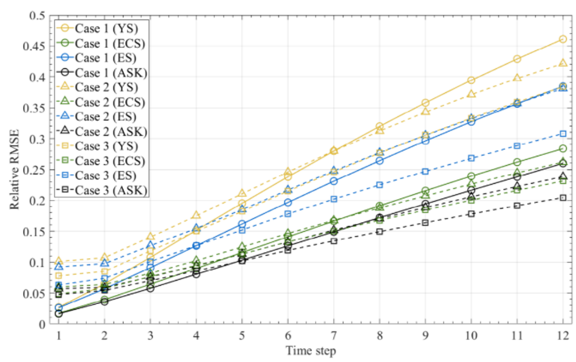

Figure 5 - Results of Accuracy Improvement using Time Series Component Decomposition Data

When comparing the extent of accuracy improvement by region, the entire area (ASK) exhibited an approximate 23% increase in accuracy, while the Yellow Sea (YS), which had relatively higher errors compared to other regions, experienced an approximate 15% improvement. This can be attributed to the model's lower learning effectiveness in the Yellow Sea region, possibly due to its location between the Korean Peninsula and the Chinese mainland, taking into account the surrounding topography.

REFERENCES

Hersbach et al., (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049.
Kim et al., (2022). STG-OceanWaveNet: Spatio-temporal geographic information guided ocean wave prediction network. Ocean Engineering, 257, 111576.
Xiao et al., (2019). A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. Environmental Modelling & Software, 120, 104502.
Zhou et al., (2021). ConvLSTM-based wave forecasts in the South and East China seas. Frontiers in Marine Science, 8, 680079.