Thank you very much for your comments. We will correct the grammar issue in the next version of the manuscript. Next, we will answer a few questions you raised.

How many hidden layers are used in the configurations of the three single models?

Reply: For this question, we have shown in Table 3. Parameter of three single models. The number of hidden layers is all 10.

In terms of the methodology I am concerned about applying EMD to significant wave height time series, as it is not an oscillatory signal around zero like sea surface height would be. Significant wave height is a parameter obtained from a sea surface height time series by averaging the 1/3 largest waves or calculating the integral over the wave spectra. I wonder if it is appropriate to decompose significant wave height in IMF, which are oscillatory signals with zero-crossings, while significant wave height has no zero-crossing and it is, by definition, non-negative.

Reply: The purpose of EMD is to continuously extract the components of various scales that make up the original signal from high frequency to low frequency. Then the order of the characteristic mode functions obtained by decomposition is arranged in order of frequency from high to low, that is, the highest frequency component is obtained first. Then it is the sub-high frequency, and finally a residual component with a frequency close to 0 is obtained. In this article, we only use EMD to simplify the input, so as to better improve the prediction accuracy.

My understanding is that the model only considers 5 h of data to predict the following 1 to 3 h. physically speaking the evolution of waves very much depend on surface winds. Shouldn’t the model also consider information about surface winds as an input parameter?

Reply: Many researchers have proposed many methods for predicting the significant wave height, and the data input is also different, but most of them use the historical significant
wave height as the input. Of course, your suggestion is very helpful to us. The next thing we have to do is to consider more inputs to further improve the prediction accuracy.

Finally, thank you again for your suggestions.