Analysis and Prediction of Significant Wave Height in the Beibu Gulf, South China Sea

Huan Wang¹, Dongyang Fu¹,², Dazhao Liu³,², Xiuchun Xiao¹,², Xianqiang He³, and Bei Liu¹

Key Points:
- Annual significant wave height had significantly increased over time. This trend began in 1984 and reached a significant level in 1995.
- The westward propagating swell generated by gales in the northeast of the South China Sea increased the significant wave height in the study area.
- Multiple sine function decomposition neural network was performed to forecast monthly significant wave height with high efficiency.

1. Introduction

The Beibu Gulf is located in the southwest of China, to the northwest of the South China Sea. This gulf is a typical semi-closed sea area and lies on the continental shelf. The mean water depth within the Beibu Gulf is about 46 m and the maximum value is no more than 100 m (Chen et al., 2011; Zhou et al., 2015). This gulf is also perennially exposed to the northeast and southwest monsoonal influences and typhoons pass by almost every year. The Beibu Gulf is managed jointly by China and Vietnam and is one of the best fishing grounds for the former and an important seaport in southwest China (Shao et al., 2006). Ocean waves exert significant impacts on fishing and oil platform safety within the Beibu Gulf (Yang et al., 2013, 2015). Waves are also essential factors for inducing disasters, based on wave data obtained from buoys within the Beibu Gulf between 2015 and 2018 and data show that ocean wave disasters refer to the wave height up to 6 m and mainly induced by typhoons and cold air took place 2.5 times per year on average and even four times in 2016 in the Beibu Gulf (Su et al., 2019). Wave height is an essential element that can be used to describe waves; its monitoring and acquisition are of great significance. Few studies have so far been performed on wave height prediction for the Beibu Gulf area. Studies aimed at forecasting significant wave height (SWH) within the Beibu Gulf are vital for fishery breeding and boat travel (Chen et al., 2009; Fu et al., 2019).

Ocean waves are common phenomena and vital components of physical oceanography. External factors, including kinetic and physical processes and boundary conditions, all affect the generation, development, and extinction of ocean waves. Several numerical models have, therefore, been established and used to calculate and predict waves state. The WAM model, a third-generation wave approach, was proposed in 1988 for ocean wave prediction (Group, 1988). Subsequent to the development of computer technology, WAVEWATCH III was then presented based on the WAM model (Tolman, 2009) and was applicable to open ocean field predictions. Another classical third-generation wave model, simulation waves near shore (SWAN), was also applicable to wave prediction in coastal regions (Booij et al., 1999). These three models all exhibit excellent performance in wave height forecasting (Chen et al., 2017; Jiang et al., 2010; Stanislav & Anna, 2016).
Results are significantly influenced by wind input and step set (Huang et al., 2013; Mao et al., 2016). Besides, using these models is highly computer-time demanding, which makes it difficult to utilize.

Subsequent to the development of neural networks and machine learning technologies, mathematical statistical methods have been utilized in wave height forecasting as well as other marine research fields (Li et al., 2020). In one example, Deo and Naidu (1998) employed a simple feed-forward network to forecast wave height in real-time. While Özger (2010) combined the wavelet method with fuzzy logic to predict wave height and period. Similarly, Altinkaynak (2013) employed a geno-multilayer perceptron to forecast wave height on Lake Okeechobee, Florida. In this case, wind values as supported data and the forecast results illustrated that height was consistent with observed data. Ali and Prasad (2019) then designed an elaborate and improved model that coupled an extreme learning machine model and improved complete ensemble empirical mode decomposition method with adaptive noise to forecast SWH across the eastern coastal zones of Australia. Some researches had also been carried out employing neural networks or machine learning methods to forecast wave height; those studies had yielded universally accurate results (Deka & Prahlada, 2012; Fernández et al., 2015; Gaur & Deo, 2008; Gopinath & Dwarakish, 2015; Kamranzad et al., 2011; Mandal & Prabaharan, 2006). This study utilizes a multiple sine function decomposition neural network (MSFDNN) to forecast SWH across the Beibu Gulf and its adjacent waters in the South China Sea. Especially, distinct from the numerical models that consider the structure of two-dimensional wave spectra, energy conservation equations and nonlinear transfer, the MSFDNN extracts the sine-characteristics of a time series and simply analyzes changing trends in data without considering mechanisms of formation. The MSFDNN does not use resources to solve complicated governing equations, but pays more attention to data change patterns. Additional researches also applied the MSFDNN to study time series as these have sinusoidal oscillation characteristics; again, outcomes of this method have been excellent (Lao et al., 2014; Wang et al., 2019; Zhang et al., 2016, 2015).

The contents of each section are outlined there. Section 2 illustrates detailed information about data, study area, and research methodology. Section 3 initially discusses variation in annual and monthly SWH between 1979 and 2018 across the study area. Moreover, considering the influence of wind speed and sea surface temperature (SST) on SWH (Shimura et al., 2015; Wang et al., 2020), the relationships between SWH and wind speed as well as SST are discussed here. Besides, the performance of the MSFDNN at forecasting, monthly SWH is also presented in Section 3. Section 4 then summarizes the entire paper.

2. Materials and Methods

The detailed information about the study area, the Beibu Gulf and its adjacent waters in the South China Sea, and the data utilized in this present paper are depicted in this section. The method involved in monthly SWH forecasting for this area is presented in this section.

2.1. Study Area

The study area, the Beibu Gulf and its adjacent sea areas (17°N–22°N, 105°E–112°E) in the South China Sea, is plotted in the red dotted rectangle of Figure 1a. Additional details can be found in Figure 1b. Figure 1 is plotted using ETOPO1 data downloaded from the NOAA, which shows the South China Sea bathymetry conditions and the study area. Moreover, Figure 1b shows that the isobath across most of the study area is less than 1,000 m, except the southeast of the Hainan Island and the water depth becomes shallower as water gets closer to the land. The study area is mainly surrounded by Guangxi, Guangdong, Hainan provinces (China) and some cities in Vietnam, and an important harbor and a good fishing ground (Shao et al., 2018).

2.2. Data

A total of 40 years SWH reanalysis data from 1979 to 2018 were extracted and obtained from the European Center for Medium-Range Weather Forecasts (ECMWF) ERA-Interim data set. The temporal resolution is 6 h and the spatial resolution of the SWH data is 0.125° × 0.125°; this is a high resolution and meets the requirements of this forecasting experiment. As this paper also analyzed the relationships between SWH and wind speed as well as SST, we also downloaded the corresponding wind speed and SST from the same
2.3. Methods

2.3.1. Data Processing and Analysis Methods

Original 6-h SWH sampling data, wind speed, and SST across this study area from 1979 to 2018 were aggregated to annual average sequences across the entire region and monthly average grid point sequences for the whole area with spatial resolutions remaining unchanged first. In terms of analyzing trends and detecting change points in annual SWH, wind speed and SST, the Mann-Kendall (M-K) test, a widely used nonparametric statistical method in oceanography and meteorology (Patakamuri and Sridhar, 2020), was employed. This approach does not require data to follow a certain distribution and is unaffected by a few outliers and missing data.

For a time series $X$ with $n$ samples, the M-K test constructs a statistic $S_k$, the cumulative count of the number of values at the time $i$ greater than at time $j$, written as

$$s_k = \sum_{i=1}^{k} g_i \quad (k = 2, 3, \ldots, n),$$

where

$$g_i = \begin{cases} +1 & x_i > x_j \\ 0 & x_i \leq x_j \end{cases} \quad (j = 1, 2, \ldots, i).$$

The expectation $E(s_k)$ and variance $Var(s_k)$ of $s_k$ with the $x_1, x_2, \ldots, x_n$ is independent of each other and has the same continuous distribution can be calculated as

$$E(s_k) = \frac{n(n + 1)}{4},$$

$$Var(s_k) = \frac{n(n + 1)(2n + 5)}{72}.$$  

Then, the statistic $U_{ki}$ is defined under the assumption of random independence of time series:

Figure 1. Bathymetric conditions of (a) the South China Sea and its adjacent waters as well as (b) the Beibu Gulf study area and its adjacent waters.
where \( UF_k \) is normally distributed and which is calculated in the positive order of the time series \( X \). Additionally, \( UF_1 = 0 \). For a given significance level \( \alpha \), if \( UF_k > U_{\alpha} \) obtained from the standard normal cumulative distribution tables, indicating that there is an obvious trend change in the time series. Similarly, the statistic \( UB_k \) is obtained by repeating the above process in reverse order of time series \( X \), \( UB_k = -\overline{UF}_k \), \( k = n, n-1, \ldots, 1 \), \( UB_1 = 0 \). Moreover, statistic \( UF_k \) greater than 0 indicates an upward trend in the time series, while statistic \( UF_k \) less than 0 indicates a downward trend. When \( UF_k \) exceeds the given significance level, marking a significant upward or downward trend. If the \( UF_k \) and \( UB_k \) curves intersect at a significant level, the intersection point corresponds to the time of the onset of the change.

### 2.3.2. Forecast Method: MSFDNN

MSFDNN is a univariate prediction method, based on the historical data of the predicted variable itself. The key to the MSFDNN is to let number of different and appropriate sine functions to express how the time series changes. This is done initially by dividing the time series into \( M \) sine functions using the least squares estimation method (LSEM) in order. These are then added to obtain a mathematical expression \( F(t) \) Equation (6), which can give the value of the corresponding time series based on time \( t \) in the past, present, or future, as follows:

\[
F(t) = \sum_{i=1}^{M} \hat{L}_i(t) = \sum_{i=1}^{M} A_i \sin(B_i t + C_i).
\]

where \( t \) denotes time, while \( i = 1, 2, 3, \ldots, M \) is the ordinal number of sine function \( \hat{L}_i(t) \), and \( A_i, B_i \), and \( C_i \) are corresponding coefficients. These initial values of these are presented randomly.

A total of 40 years of monthly SWH time series were divided into two parts for learning and testing, respectively. One part was the learning series, which accounts for 75% of the time series and was denoted as \( L(t_i) \), where \( t_i \) represents the time between 1979 and 2008. We emphasize here that this is not the real number of years, but the time indexes of all months arranged by year. This means that the values of \( t_i \) are \([1, 2, 3, \ldots, 360]\) for monthly average SWH data. The remaining data were employed in verification data and comprised the series of the remaining 10 years from 2009 to 2018. This was denoted \( V(t_i) \) where \( t_i \) has the same meaning as \( t_i \) with its value listed as \([361, 362, 363, \ldots, 480]\) for monthly average data. The details of the MSFDNN method were described as follows:

The MSFDNN used in this analysis is shown in Figure 2. The original learning time series \( L(t_i) \) was assigned as \( L_1(t_i) \), the first decomposed object. Thus, following multiple iterations, the values of \( A_1, B_1 \), and \( C_1 \) were obtained using LSEM. This led to \( \hat{L}_1(t) \), the first sine function. In terms of second decomposition, the object \( L_2(t_i) \) became a new time series, the residual following the first decomposition. This was expressed as follows:

\[
L_2(t_j) = L_1(t_j) - \hat{L}_1(t_j).
\]

On the basis of the first decomposition process, the second sine function, \( \hat{L}_2(t) \), was obtained. Subsequent to the third decomposition object \( L_3(t_j) \) was obtained, and \( L_3(t) \) was assigned as follows:

\[
L_3(t_j) = L_2(t_j) - \hat{L}_2(t_j).
\]

Subsequent to this analogy, once the \( M \)th decomposition process was completed and the \( M \)th sine function was obtained, all \( M \) decomposed sine functions were added to obtain the mathematical expression \( F(t) \). The next step was then to verify the accuracy of the mathematical expression \( F(t) \). Calculating the mean square error (MSE) between the sequence \( F(t_k) \) was then calculated via the MSFDNN and verification sequence \( V(t_k) \). The value of \( \text{MSE}(F(t_k), V(t_k)) \) was greater than the desired value \( S \), and then so the mathematical
expression $F(t)$ should be recalculated by regulating $A_i$, $B_i$, and $C_i$. As the $\text{MSE}(F(t), V(t))$ was less than, or equal to, the desired value $S$, the mathematical expression $F(t)$ was finally output. The mathematical expression of SWH time series was obtained using the MSFDNN and predictions were then undertaken.

### 2.3.3. Statistic Methods

The mathematical evaluation terms involved in this study, including the MSE and correlation coefficient ($r$), were outlined here in order to judge the accuracy of predictions as well as to evaluate the performance of MSFDNN for forecasting SWH. The mathematical expressions used here were as follows:

$$\text{MSE}(X,Y) = \frac{\sum_{n=1}^{m} (X_n - Y_n)^2}{m}, \hspace{1cm} (9)$$

$$r(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}[X] \text{Var}[Y]}}, \hspace{1cm} (10)$$

where $X$ denotes the target sequence while $Y$ represents the predicted one, and $X_n$ and $Y_n$ are their $n$th value, respectively. In addition, $m$ denotes the length of both sequences, while $\text{Cov}(X,Y)$ indicates the covariance of $X$ and $Y$, $\text{Var}[X]$ indicates variance of $X$, and $\text{Var}[Y]$ indicates variance of $Y$.

### 3. Results and Discussion

#### 3.1. Data Analysis

##### 3.1.1. Annual Variation

The data series in Figure 3 shows annual SWH, wind speed, and SST over the past 40 years across the study area. To detect trends and change points in SWH, wind speed and SST time series, the M-K test was employed.
The data presented in Figure 3a show that SWH follows a significant upward trend with a growth rate of 0.12 cm/a based on a linear regression analysis from 1979 to 2018. This reached an annual average maximum value of 1.068 m in 2011, followed by 1.059 m in 1996 and 1.031 m in 1999. The minimum value of annual SWH was 0.863 m in 1982, followed by 0.882 m in 1979, 0.889 m in 1984, 0.899 m in 1997 and 0.909 m in 2014. The data presented in Figure 3c show a clear downward trend in the annual wind speed, with a decay rate of −0.44 cm/a. We also carefully observed variations in the broken lines plotted in Figures 3a and 3c, which show that variations in SWH and wind speed have remained almost synchronous over the past 40 years. These reached local maximum values in 1988, 1996, and 2011, as well as in other years, while the same thing happened with local minima in some cases. No marked upward or downward trends were seen in SST data over the past 40 years (Figure 3e). However, a negative correlation was present between SWH and SST; when the SWH reached a local maximum, SST got the local minimum, and there was another delay of a year or two in some cases. In other cases, the SWH went the local maximum in 2011, while SST reached the local minimum in 2011 and then again in 1996. In other years, SWH reached its local minimum in 1997, while SST continued to rise this year and did not reach its maximum until 1998, after a one-year delay.

The data presented in Figure 3b, 3d, and 3f describe SWH M-K test detection results as well as wind speed and SST, respectively. The red lines on these figures denote the sequential statistic $UF$, while the blue dotted lines indicate the reverse statistic $UB$ and the black dotted lines present the 0.05 significance level. Figure 3b therefore shows that SWH did not exhibit marked variation between 1979 and 1984 across the study area as there were multiple points of intersection between the statistic $UF$ and $UB$, and because their values remained between positive and negative significance levels throughout this period. Especially, from 1984 onward, SWH began to grow; this trend began to be significant from about 1995 and continued until 2018, exceeding the 0.05 significant level. There was also a sudden change in annual wind speed in 1999; from this point onwards, wind speed began to trend downwards, but this was not obvious because it did not exceed the significance level (Figure 3d). No obvious trend or abrupt change point was seen between 1979 and 2018 in annual SST data (Figure 3f).

Based on this analysis, we can conclude that SWH had exhibited significant growth; this began in 1984, reached a significant level in 1995, and reached a maximum in 2011. Variations in wind speed and SWH had been consistent, but the overall trend of SWH increased while wind speed decreased. This kind of phenomenon was very counterintuitive. Generally, high wind speed means correspondingly high wave height. Many research found that the study area, that is, the Beibu Gulf and its adjacent waters, was controlled by the East Asian monsoon, and the East Asian monsoon has been weakening since the 1970s and the trend will continue (Hori & Ueda, 2006; Wang, 2001). Also, Wang et al. (2020) systematically discussed the wave...
climate change in the South China Sea and its underlying mechanism. They found that the SWH and wave period in the South China Sea were increased significantly with wind speed had no significant change from 1979 to 2018 and they inferred the wave climate enhancement could be attributed to the growth of the swell associated with the increased frequency of gales whose speed is greater than class 6 (10.8 m/s) in the South China Sea. Moreover, the El Niño-Southern Oscillation (ENSO) activities influenced wave climate in the South China Sea by acting on local wind and frequency of gales and was of great significance to the wave climate change (Han et al., 2017). ENSO regulated the South China Sea wave’s inter-annual variation through the wind in the northeast of the South China Sea and the ocean propagation process (swell) pushed the associated signals to the southwest of the South China Sea (Wang et al., 2020). In other words, the increase in the SWH in the Beibu Gulf may be caused by the swell from the northern part of the South China Sea.

Causing the vital influence of swell induced by the northeast of the South China Sea gales on the variation of SWH in the study area, we calculated the frequency of gales and the variation of the South China Sea from 1979 to 2018 based on 6-h wind speed at 10 m above sea level. Additionally, the annual frequency of gales in this paper was calculated as

\[ F_a = \frac{N_g}{N_d}, \]

where, \( N_g \) indicates the number of gales whose speed is greater than class 6 (10.8 m/s) (Wang et al., 2020) in a year and \( N_d \) indicates the number of days in a year, 366 in a leap year and 365 in a normal year. Specifically, Figure 4a shows the 40 years average annual frequency of gales in the South China Sea, and Figure 4b indicates the change rate of \( N_g \) from 1979 to 2018. It can be seen from Figure 4a that gales occurred mainly in Luzon Strait and its western region and Taiwan Strait. Figure 4b shows that the west of the Luzon strait and the southwest of Taiwan Island were the regions with the fastest increase in gale frequency. Therefore, the west part of the Luzon Strait was selected as a representative gale-generating region of the South China Sea, as shown in the red dotted rectangle box (160°E—121°E, 18.5°N—21°N) in Figure 4. The wind speed data in red dotted rectangle box were averaged regionally and then integrated into monthly data for correlation analysis with the study area’s monthly SWH data in this research. The results are plotted in Figure 5.

Figure 5 shows the relationships between the monthly SWH in the study area and the monthly wind speed in the western Luzon Strait (Figure 5a) as well as the study area (Figure 5b). The results are shown in Figure 5a were to calculate the correlation coefficients between the regional average monthly wind speed and
the monthly SWH of each grid point from 1979 to 2018. The results are shown in Figure 5b to calculate the correlation coefficients between the monthly wind speed and the monthly SWH at the corresponding grid point. In the lower right areas of the Hainan Island as shown in Figure 5a, the correlation is 0.90 or above, gradually decreasing as it moves into the bay and near the shore. This suggests that the westerly propagation of swell induced by the gales in the west of Luzon Strait had a significant influence on the SWH in the Beibu Gulf region. Due to the energy attenuation, the frequency dispersion and angular spreading, the swell gradually weaken (Jiang et al., 2017). Figure 5b shows that the correlation between the upper left and lower right areas of Hainan Island reached more than 0.95. Except for the lower-left area of Hainan Island, the correlation between the SWH and the local wind speed in the whole study area reached 0.90 or more. This suggests that the local wind had a significant influence on the SWH in the Beibu Gulf. To further explore the variation mechanism, the line between points 107°E, 17°N and 112°E, 22°N was selected as the dividing line, as shown in the blue dotted line in Figure 5. The study area was divided into two sub-regions, left and right, and then the M-K test of the regional annual SWH of the two sub-regions was conducted. The results were shown in Figure 6.

Figure 6 plots the left sub-region's M-K test results, which were mainly affected by local wind, as mentioned above. It can be concluded from Figure 6a that there is no significant variation of SWH in the left sub-region over the past 40 years from 1979 to 2018 due to the statistic UF and statistic UB do not change significantly. On the contrary, statistic UF plotted in Figure 6b, the M-K test results of right sub-region, which was subject to the double action of local wind and swell from the Luzon Strait, shown a significant increase and exceeded the significance level in 1999. In conclusion, the increase of SWH in the study area was mainly caused by the invasion of swell spreading westward caused by gales that occurred in the northeast of the South China Sea. Additionally, the values of SWH in the right sub-region were much larger than the values on the left, as shown in Figure 9; therefore, the SWH of the whole study area presented a significantly increasing trend.

### 3.1.2. Monthly Variation

Figure 7 provides an overview of variations of monthly average SWH, wind speed and SST across the whole study area from 1979 to 2018. The X-axis of Figure 7 indicates the year while the Y-axis of which indicates the month. It is clear that SWH values were lower between April and September, generally below 1 m, and higher between January and March, and between October and December, usually above 1 m. Combined with Table 1, it is clear that over the last 40 years, SWH had increased at a rate of 0.35 cm/a and 0.54 cm/a in January and December, respectively. The contour lines at 1 and 1.2 m in Figure 7a show an upward trend for October and November, moving the contour lines at 1.4 m upward and indicating a downward trend for October and November. Regression analysis showed that rates of decline were −0.20 cm/a and −0.31 cm/a, respectively. Comparisons between Figures 7a, 7b, and 7c show the same change between monthly SWH and wind speed, while an opposite change pattern was present between monthly SWH and SST annually. The contour line at 4.5 m/s tended to be extended downwards in the lower half of Figure 7b as time increased and upwards in the upper half over the last 40 years. This illustrated that overall wind speed had followed a downward trend, the same as the conclusion reported in Section 3.1.1.
Figure 8 is presented as a series of box-plots to display the distribution of monthly SWH, wind speed, and SST across the whole study area for the last 40 years from 1979 to 2018. As shown in Figure 8, the SWH and wind speed variations with the month were like a parabola opening upward, and the variation of SST was like a parabola opening downward. In a year, SWH and wind speed were lower from April to September and higher in the remaining months, while the SST was higher from May to November and lowered in the remaining months. Over the past 4 decades, it can also be concluded that SWH fluctuated across the widest range in December, between 1.097 m and 1.932 m. SWH was most variable in October; four outliers were presented during this month (Figure 8a). Figure 9 summarizes the spatial distribution of monthly SWH across the study area over the past 40 years, from 1979 to 2018. This figure shows that, on the whole, SWH in the bay was smaller than to the north of Hainan Island. The SWH values between April and September were generally smaller and more uniformly distributed, exhibiting relatively gentle changes. SWH values over the remaining months were quite different in and out of the bay. Change in SWH in December largest, followed by November.

3.2. Forecast Results
3.2.1. Monthly SWH Forecast Results

Monthly forecast results for SWH using the MSFDNN were reported based on the ERA-Interim data set from 1979 to 2018 across the study area. Thus, out of 75% of 40 years monthly SWH data, those form the first 30 years from 1979 to 2008 were used to learn and obtained the mathematical expression \( F(t) \) to describe these changes. The remaining 25% over the last 10 years from 2009 to 2018 were employed to verify the accuracy of the expression \( F(t) \). The number of sine functions, \( M \), was utilized to decompose SWH time series \( L(t) \) drawn from every grid point across the study area. It was set to 140 as MSE between \( V(t) \) and \( F(t) \) was limited to 0.12 to prevent forecasting results from being too divergent after several iterations. MSE values between test data \( V(t) \) and forecast data \( F(t) \) over the period between 2009 and 2018 with \( M \) equal to 140 across the study area were shown in Figure 10. These data indicate that larger \( MSE(F(t), V(t)) \) values appeared in the southeastern part of the Hainan Island while the inside Beibu Gulf values were small. A similar distribution was seen between \( MSE(F(t), V(t)) \) and topography across
the study area. The deeper the water depth, the greater the error while the shallower the water depth, the smaller the error. One of the reasons for this phenomenon can be attributed to the fact that SWH in the bay was relatively small and changed smoothly compared with that in the northeastern region of Hainan Island. Overall variation in SWH across the bay was not significant, while learning and forecast remain relatively accurate and straightforward. In addition, we know that the SWH in the southeastern waters of Hainan Island was affected not only by local wind, but also by the swell from the northeast part of the South China Sea based on the previous analysis in Section 3.1.1. These swells were induced by gales and

Figure 7. Monthly values of (a) SWH, (c) wind speed, and (e) SST from 1979 to 2018 across the study area. SST, sea surface temperature; SWH, significant wave height.

| Table 1 | Growth Rates in Monthly SWH Over 40 years in the Study Area |
|---------|--------------------------------------------------------|
|         | January | February | March | April | May    | June   |
| Linear regression coefficient (cm/a) | 0.35    | 0.17     | 0.08  | 0.11  | −0.004 | 0.11   |
|         | July    | August   | September | October | November | December |
| Linear regression coefficient (cm/a) | 0.12    | 0.11     | 0.21  | −0.20 | −0.31  | 0.54   |
were primarily influenced by ENSO events, which were hard to predict (Wang et al., 2020). The MSFDNN utilized sine functions to decompose the historical data of SWH and learn the data change periodic law; it might not be sensitive to the swell caused by ENSO events. This might explain why the prediction error of SWH in the southeastern sea area of Hainan Island was high. The SWH in this area was controlled by the local wind and greatly affected by the swell from outside, and MSFDNN was not sensitive to this impact factor caused by ENSO events.

To further illustrate the effectiveness of the MSFDNN in forecasting SWH across the study area, we extracted predicted results at points with the minimum and maximum MSE. Minimum MSE was 0.0062 at 21°N, 107.125°E (Site 1), while maximum MSE was located at 18°N, 111.75°E with a value of 0.1120 (Site 2) as shown in Figure 10. The predicted results for these two points were plotted in Figure 11 and Figure 12, respectively. Additionally, the deep blue lines indicate original data $L(t_j)$ from the ERA-Interim data set for learning, while the red dotted lines indicate learning results $F(t_j)$ based on the MSFDNN. The sky blue lines denote test data $V(t_k)$, while brick-red dotted lines are corresponding predicted SWH $F(t_k)$ from the MSFDNN. Obviously, in the learning process, $F(t_j)$ followed $L(t_j)$ well and these two had a good consistency. Throughout the validation process of the MSFDNN, $F(t_k)$ was a little bit gentler than $V(t_k)$, and there was a good degree of consistency in both Sites 1 and 2. The forecasting performance of MSFDNN was outstanding at Sites 1 and 2; moreover, the correlation coefficient between the test data and forecast data was 0.7086 across the whole study area. In a word, utilizing MSFDNN to forecast SWH across the study area had been both feasible and successful.

Figure 8. Box plots of monthly (a) SWH, (b) wind speed, and (c) SST data across the study area over the last 40 years. SST, sea surface temperature; SWH, significant wave height.
4. Conclusion

SWH data from the ERA-Interim data set of ECMWF was analyzed at a temporal resolution with 6-h and a spatial resolution with 0.125° × 0.125° for the Beibu Gulf and its adjacent waters in the South China Sea from 1979 to 2018. Then, the variations of the annual and monthly SWH were analyzed, as well as the relationship between SWH and wind speed, SST across the study area were discussed. Moreover, the mechanism of SWH variation and the possible causes of prediction errors were also discussed. After that, the MSFDNN was employed to forecast monthly SWH in the study area from 2019 to 2028 based on historical monthly SWH data and the forecast results performance was discussed.

Figure 9. The spatial distribution of monthly SWH in the study area over the past 40 years from 1979 to 2018. SWH, significant wave height.
The analyzed results of annual SWH across the study area illustrated that the annual SWH increased significantly from 1979 to 2018. Specifically, the annual SWH had grown since 1984, reached a significant level in 1995 and reached a maximum 1.068 m in 2011 based on the M-K test method. Besides, the monthly SWH has shown that the low SWH, generally less than 1 m, was appeared from April to September, and the high SWH had occurred in the remaining months with a maximum value in December over the past 40 years. Especially, SWH in December also exhibited the widest variation range from 1.097 to 1.932 m. Additionally, when the SWH reached a local maximum, SST got the local minimum, and there was another delay of a year or two in some cases. The predicted results of monthly SWH from 2019 to 2028 show that the MSFDNN possessed a satisfying predictive performance across the Beibu Gulf and its adjacent waters in the South China Sea. The MSE between the predicted and test data was less than 0.12 and the correlation coefficient between the test data and forecast data was 0.7086 across the whole study area. The prediction of SWH is of great significance for wave energy assessment, coastal engineering protection and offshore oil platform maintenance (Ali & Prasad, 2019).

The variation of SWH in the study area, that is, the Beibu Gulf and its adjacent waters, was influenced by local wind and invasion swell. This swell was mainly from the northeast part of the South China Sea and induced by gales associated with the ENSO activities. Bounded by the Hainan Island, the SWH in the left sea area was mainly affected by local wind, due to land blocking with the swell energy attenuation was obvious and the influence of swell on the change of SWH was not obvious. On the other hand, because the East Asian monsoon controlled the whole study area, with the global warming, the East Asian monsoon weakens and then the variation of SWH in this area was not obvious or even weakened. The sea area on the right side of Hainan Island was affected by local wind and influenced by the swell. Additionally, the increasing frequency of gales in the South China Sea would make the SWH increased. Generally, since the SWH of the left of Hainan Island was small and the changing trend was not obvious, while the SWH of the right has an obvious growth trend, the SWH of the whole study area was increasing over the past 40 years from 1979 to 2018. In terms of the MSFDNN, it is clear that the predicted results were relatively smooth and stable compared with verification data. While some outliers cannot be predicted. As a result, this method exhibited better prediction for time series with obviously similar sinusoidal oscillation, but was unable to capture the longer-term variability generated by the ENSO and the other interannual variability.

![Figure 10. MSE between test \( V(t_k) \) and forecast data \( F(t_k) \) based on the MSFDNN in the study area from 2009 to 2018. Minimum MSE was 0.0062 at 21°N, 107.125°E (Site 1), maximum MSE was 0.1120 at 18°N, 111.75°E (Site 2).](image)

![Figure 11. Forecast SWH results of Site 1 at 21°N and 107.125°E with minimum MSE 0.0062 between \( V(t_k) \) and \( F(t_k) \) based on the MSFDNN method.](image)
Figure 12. Forecast SWH results of Site 2 at 18°N and 111.75°E with maximum MSE 0.1120 between $V(t)$ and $F(t)$ using the MSFDNN method.

Data Availability Statement

The SWH data and wind speed data utilized in this analysis can be downloaded from the European Centre for Medium-Range Weather Forecasts (https://www.ecmwf.int/). Terrain data can be downloaded from https://www.ngdc.noaa.gov/mgg/global/global.html. All data processing and prediction processes were carried out using MATLAB.

References

Ali, M., & Prasad, R. (2019). Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renewable and Sustainable Energy Reviews*, 104, 281–295. https://doi.org/10.1016/j.rser.2019.01.014

Altunkaynak, A. (2013). Prediction of significant wave height using geno-multilayer perceptron. *Ocean Engineering*, 58, 144–153. https://doi.org/10.1016/j.oceeng.2012.08.005

Booij, N., Ris, R. C., & Holthuijsen, L. H. (1999). A third-generation wave model for coastal regions: 1. Model description and validation. *Journal of Geophysical Research*, 104(C4), 7649–7666. https://doi.org/10.1029/98JC02622

Chen, C., Zhao, Y., Xing, J., Zhu, J., Liu, J., Huang, X., et al. (2017). Validation of the wave model SWAN against altimeter data from Jason-2 satellite. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 3385–3388. https://doi.org/10.1109/IGARSS.2017.8127724

Chen, S., Li, Y., Hu, J., Zheng, A., Huang, L., & Lin, Y. (2011). Multiparameter cluster analysis of seasonal variation of water masses in the southern China Seas. *Acta Oceanologica Sinica*, 36(5), 44–50. https://doi.org/10.1007/s13131-017-1059-2

Chen, Z., Xu, S., Qiu, Y., Lin, Z., & Jia, X. (2009). Modeling the effects of fishery management and marine protected areas on the Beibu Gulf using spatial ecosystem simulation. *Fisheries Research*, 100(3), 222–229. https://doi.org/10.1016/j.fishres.2009.08.001

Deka, P. C., & Prahlada, R. (2012). Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. *Ocean Engineering*, 43, 32–42. https://doi.org/10.1016/j.oceaneng.2012.01.017

Deo, M., & Naidu, C. S. (1998). Real time wave forecasting using neural networks. *Ocean Engineering*, 26(3), 191–203. https://doi.org/10.1016/S0029-8018(97)00225-7

Fernández, J. C., Salcedo-Sanz, S., Gutiérrez, P. A., Alexandre, E., & Hervás-Martínez, C. (2015). Significant wave height and energy flux range forecast with machine learning classifiers. *Engineering Applications of Artificial Intelligence*, 43, 44–53. https://doi.org/10.1016/j.engappai.2015.03.012

Fu, D., Wang, H., Xiao, X., Liao, S., & Jin, L. (2019). Adaptive zeroing-gradient controller for ship course tracking with near singularity considered and zero theoretical tracking error. *IEEE Access*, 7, 38205–38212. https://doi.org/10.1109/ACCESS.2019.2904993

Gaut, S., & Deo, M. (2008). Real-time wave forecasting using genetic programming. *Ocean Engineering*, 35(11–12), 1166–1172. https://doi.org/10.1016/j.oceeng.2008.04.007

Gopinath, D. I., & Dwarkish, G. (2015). Wave prediction using neural networks at New Mangalore Port along west coast of India. *Aquatic Procedia*, 4, 143–150. https://doi.org/10.1016/j.aqpro.2015.02.020

Group, T. W. (1988). The WAM model third generation ocean wave prediction model. *Journal of Physical Oceanography*, 18(12), 1775–1810. https://doi.org/10.1175/1520-0485(1988)018<1775:TWMTGO>2.0.CO;2

Han, S., Fan, Y., Dong, Y., & Wu, S. (2017). A study on the relationships between the wave height and the El Niño in the north area of the south china sea. *Acta Oceanologica Sinica*, 39(5), 44–50. https://doi.org/10.1007/s13131-017-1059-2

Hori, M. E., & Ueda, H. (2006). Impact of global warming on the east Asian winter monsoon as revealed by nine coupled atmosphere-ocean GCMs. *Geophysical Research Letters*, 33(3). https://doi.org/10.1029/2005GL024961

Huang, Y., Weisberg, R. H., Zheng, L., & Zijlema, M. (2013). Gulf of Mexico hurricane wave simulations using SWAN: Bulk formula-based drag coefficient sensitivity for Hurricane Ike. *Journal of Geophysical Research: Oceans*, 118(8), 3916–3938. https://doi.org/10.1002/jgo.20283

Jiang, H., Rahmani, A. V., Liu, Q., Stropa, J. E., Chapron, B., & Chen, G. (2017). Can contemporary satellites estimate swell dissipation rate? *Remote Sensing of Environment*, 201, 24–33. https://doi.org/10.1016/j.rse.2017.08.037

Jiang, L. F., Zhang, Z. X., & Qi, Y. Q. (2010). Simulations of SWAN and WAVESWITCH III in northern South China Sea. *The Twentieth International Offshore and Polar Engineering Conference*. Beijing, China. https://onepetro/ISOPEIOPEC/proceedings-abstract/ISOPE100-All-ISOPE10/ISOFE-1-10-637-11740
