



Article

Calibration Experiments of CFOSAT Wavelength in the Southern South China Sea by Artificial Neural Networks

Bo Li ^{1,2,3,4}, Junmin Li ^{1,2,3,4,*} , Junliang Liu ¹, Shilin Tang ^{1,2}, Wuyang Chen ¹, Ping Shi ^{1,2} and Yupeng Liu ¹

- ¹ State Key Laboratory of Tropical Oceanography, Key Laboratory of Science and Technology on Operational Oceanography, Guangdong Key Laboratory of Ocean Remote Sensing, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 511458, China; libo@scsio.ac.cn (B.L.); lerry710@scsio.ac.cn (J.L.); sltang@scsio.ac.cn (S.T.); chenwuyang@scsio.ac.cn (W.C.); pshi@scsio.ac.cn (P.S.); liuyupeng@scsio.ac.cn (Y.L.)
- ² Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China
- ³ Sanya Institute of Oceanology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Sanya 572025, China
- ⁴ Innovation Academy of South China Sea Ecology and Environmental Engineering, Chinese Academy of Sciences, Guangzhou 511458, China
- * Correspondence: jli@scsio.ac.cn

Abstract: The wave data measured by CFOSAT (China France Oceanography Satellite) have been validated mainly based on numerical model outputs and altimetry products on a global scale. It is still necessary to further calibrate the data for specific regions, e.g., the southern South China Sea. This study analyses the practicability of calibrating the dominant wavelength by using artificial neural networks and mean impact value analysis based on two sets of buoy data with a 2-year observation period and contemporaneous ERA5 reanalysis data. The artificial neural network modeling experiments are repeated 1000 times randomly by Monte Carlo methods to avoid sampling uncertainty. Both experimental results based on the random sampling method and chronological sampling method are performed. Independent buoy observations are used to validate the calibration model. The results show that although there are obvious differences between the CFOSAT wavelength data and the field observations, the parameters observed by the satellite itself can effectively calibrate the data. In addition to the wavelength, nadir significant wave height, nadir wind speed, and the distance between the calibration point and satellite observation point are the most important parameters for the calibration. Accurate data from other sources, such as ERA5, would be helpful to further improve the calibration results. The variable contributing the most to the calibration effect is the mean wave period, which virtually provides relatively accurate wavelength information for the calibration network. These results verify the possibility of synchronous self-calibration for the CFOSAT wavelength data and provide a reference for the further calibration of the satellite products in other regions.

Keywords: CFOSAT; wavelength; artificial neural network; calibration experiment; South China Sea



Citation: Li, B.; Li, J.; Liu, J.; Tang, S.; Chen, W.; Shi, P.; Liu, Y. Calibration Experiments of CFOSAT Wavelength in the Southern South China Sea by Artificial Neural Networks. *Remote Sens.* **2022**, *14*, 773. <https://doi.org/10.3390/rs14030773>

Academic Editor: Martin Gade

Received: 5 January 2022

Accepted: 4 February 2022

Published: 7 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The South China Sea (SCS) (Figure 1), the second-largest marginal sea in the Northwest Pacific, is one of the busiest shipping routes in the world and connects the Pacific Ocean and the Indian Ocean. Wave parameters are of great significance to marine engineering, marine transportation, fishing, renewable wave energy harvesting, typhoon disaster prevention, and military activities in the SCS. Many satellites have been widely applied in ocean wave observations [1]. Among them, wavelength (or wave period) parameters can be retrieved from altimetry wave height and wind speed parameters by empirical models with artificial neural network (NN), e.g., [2,3]. Since the launch of CFOSAT (China France Oceanography Satellite) on 29 October 2018, it provides a new source to obtain global wave

information. CFOSAT can not only sample wave height data such as standard altimeters but also measure the wave spectrum by off-nadir spectrometer SWIM (Surface Investigation and Monitoring) to obtain information on the wavelength and propagation direction [4–7].

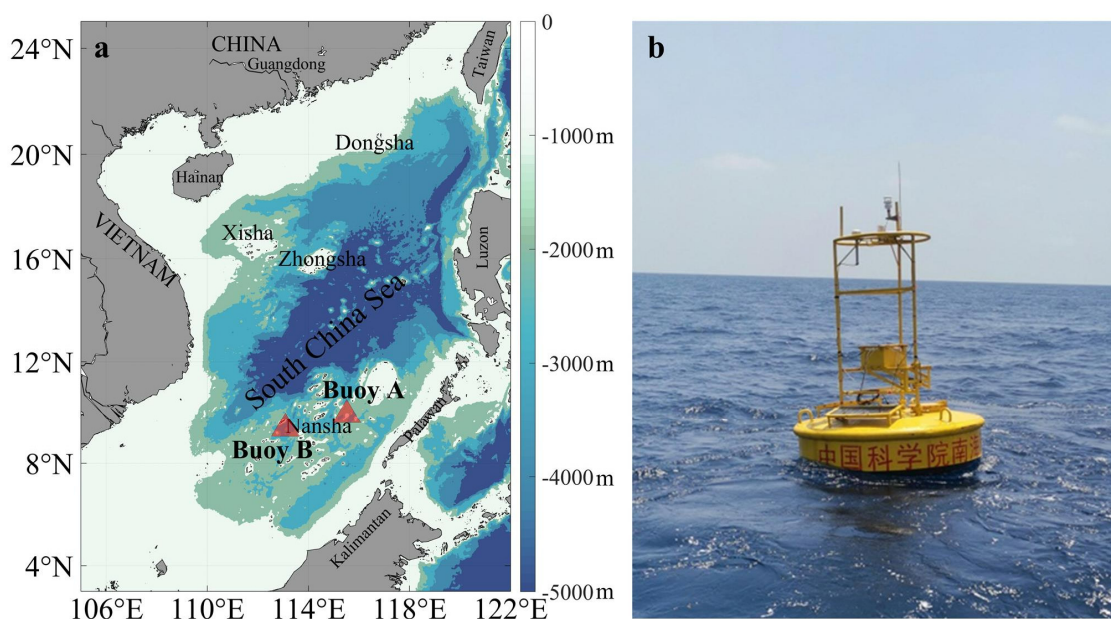


Figure 1. Geography and bathymetry of the South China Sea with the location of the buoys (red triangles) (a) and an in-situ photo of the buoy (b).

The performance of CFOSAT data has been examined mainly by comparing it with the ECMWF (the European Centre for Medium-Range Weather Forecasts) Ocean Wave Model (ECWAM) and altimetry products (Jason-3, AkliKa) on a global scale [4–6,8]. The results have shown that the wind and wave parameters retrieved from the nadir beam exhibit an accuracy similar to standard altimeter missions. However, SWIM spectra are perturbed by the speckle noise problem; thus, work is in progress to mitigate this effect [4,5]. It is also worth noting that few available in situ observations have been employed for data calibration, specifically for some local areas such as the SCS, especially for the wavelength parameters extracted directly from satellite observed spectra.

The wavelength information in the southern SCS (SSCS) is difficult to precisely capture by the satellite due to the complex topography of numerous reefs (Figure 1). Swell is the dominant component of waves in the SCS, although the propagation direction is affected by the monsoon [9]. The waves' components in the SSCS include not only swell propagated from open seas but also wave affected by local factors such as scattered islands and reefs [10–12]. However, the resolution of the CFOSAT wave spectra is $70 \text{ km} \times 90 \text{ km}$ [4,6–8], which is much larger than the size of most reefs. The wave parameters are retrieved from the box of such a big size, so that it may not be able to identify the local variations impacted by reefs. Moreover, only wavelengths in the range of 70 to 500 m are measurable by SWIM, among which the measured value of short wavelengths (e.g., 70 to 150 m) would be systematically larger than the true value, according to the product specifications [4]. Field observation is urgently needed to validate and calibrate the SWIM spectra here.

Much work has been done to calibrate satellite wave data against synchronous buoy observations, other satellite measurements, or reanalysis data combined with numerical models, e.g., [1,8,13,14]. The most common method is to establish the (multiple) linear regression relationship between the measured values and the corrected values, e.g., [15,16]. Recently, with the progress of artificial intelligence technology, the NN method has gradually become an important means to calibrate ocean observation and simulation results [17–20]. However, most of the existing studies focus on the calibration of wave

height, e.g., [15,16,19,20], and studies on using other synchronous parameters to calibrate wavelength information would be eagerly anticipated.

It is necessary to know whether self-calibration can be completed effectively by using the observation data of the satellite itself. If so, the observation information would be of great benefit to disaster early warning, maritime search and rescue, and other tasks with high timeliness. In contrast, if the calibration needs accurate data from other sources, the corrected information would be delayed. In addition, to calibrate the wavelength data of CFOSAT in the SSCS using the NN method, it is necessary to determine the critical parameters for the calibration. It is important to ascertain which parameters should be included in the calibration model. Usually, inputting more parameters can improve the effectiveness of the NN model [20–22]. However, if the input parameters are too many, it will inevitably increase the difficulty of data acquisition and data processing. It is necessary to extract the critical parameters used to calibrate to improve the computational efficiency of the model [20,23]. Therefore, preprocessing technologies, such as the clustering method and principal component analysis (PCA), were introduced to narrow down or reduce the sample space before applying an NN [21,24]. One can evaluate the impact degree of input neurons on the output in a network [25], and the critical input parameters can be extracted by mean impact value (MIV) analysis [26,27]. MIV analysis based on a well-trained NN model is theoretically capable of selecting the critical parameters [26]. The combination of dimension reduction and MIV analysis of the input variables would be beneficial to the acceleration and optimization of the NN calibrating the wavelength of CFOSAT.

In general, an NN is usually established and trained based on a large number of data samples [20,28]. However, CFOSAT has only been in operation for two years as of 2021, and the observation stations in the SSCS are still limited, so there are not enough satellite observation data that can match the site-specific field observations. Therefore, it is necessary to use appropriate methods, such as the Monte Carlo method [29], and a large number of repeated tests were carried out by random sampling to avoid occasional errors from individual sampling.

Therefore, this paper carries out systematic research on the practicability of calibrating CFOSAT wavelength data, through developing a relationship between satellite measurements and the accurate wavelength value by using artificial neural network methods. Two buoys were deployed and had run for two years in the central region of the SSCS (Figure 1), and the field-measured dominant wavelength was matched with synchronized CFOSAT data. Based on the Monte Carlo method, through a large number of random samplings, a set of NN model experiments for calibrating wavelength parameters were constructed. In the experiments, the dominant wavelength measured by the satellite was calibrated through three stages of NN models, and the key parameters needed for calibration were obtained through MIV analysis.

2. Data Sets

2.1. SWIM Products Measured by CFOSAT

The dominant wavelength data of CFOSAT used in this paper come from the SWIM product. SWIM is a Ku-band radar with a near-nadir scanning beam geometry. The SWIM instrument principle uses the fact that at a low incidence angle (approximately 8° – 10°), the normalized radar cross-section is sensitive to the local slopes related to the tilt of long waves but almost insensitive to small-scale roughness due to wind as well as to hydrodynamic modulations due to the interaction of short waves and long waves [7]. The tilt modulates the backscattering coefficient of the sea surface. The modulated Fourier spectrum is related to the directional spectrum of sea waves and contains information about the direction and wavelength of wave propagation [7]. Similar to standard satellite altimeter missions, the nadir beam also allows for estimating the wave parameters [5,7].

The SWIM data are inverted by the processing chain to retrieve three main kinds of Level 2 geophysical products, i.e., the backscattering coefficient profiles, the 2D directional wave spectra, and the nadir altimetry-like products [7]. The products are provided to

operational users (e.g., meteorology agencies) in nearly real-time and are available in less than four days to the scientific community through the AVISO+ website (<https://www.aviso.altimetry.fr/en/missions/current-missions/cfosat.html>, accessed on 31 December 2021) [5,7]. This paper uses the data from 29 July 2019, to 8 August 2021. To avoid the influence of rainfall on the data quality, the data during the rainfall period were removed.

To calibrate the dominant wavelength, nine parameters measured by CFOSAT were used (Table 1). There are physical reasons for choosing Table 1 parameters to be involved in the calibration process. Correct wave height values may provide important information relative to the wavelength, for example, significant wave height is often included to obtain or retrieve wave period data, e.g., [2,30], so swh_s and SWH_n should be considered in the calibration. Waves propagating to stations from different sources may experience different wave deformation processes, so dir_s is also considered an input parameter. The wind is the main factor in wave generation, and thus $Wind_n$ is also considered an input parameter. In addition, the distance and relative azimuth between satellite and field observation points may affect the difference between satellite and buoy measurements; therefore, the factors related to distance and azimuth, such as $dist_s$, $phase_s$, $Dist_n$, and $Phase_n$, should be fully considered. The calibration process using these parameters can be carried out synchronously with satellite observations, so it will have high timeliness.

Table 1. List of input parameters for the self-calibration of the dominant wavelength.

Symbol	Meaning	Unit
$wavelen_s$	Dominant wavelength estimated from the spectrum by the off-nadir spectrometer	m
swh_s	Significant wave height estimated from the spectrum by the off-nadir spectrometer	m
dir_s	Dominant wave direction estimated from the spectrum by the off-nadir spectrometer	°
$dist_s$	Distance between calibration point and the sampling point of the off-nadir spectrometer	m
$phase_s$	The azimuth of sampling point of off-nadir spectrometer relative to the calibration point	°
SWH_n	Significant wave height from nadir beam	m
$Wind_n$	Wind speed from nadir beam	m/s
$Dist_n$	Distance between the calibration point and the measured point of nadir beam	m
$Phase_n$	The azimuth of sampling point of nadir beam relative to the calibration point	°

2.2. Buoy Observations

There are few long-term in situ wave observation data in the SSCS. For example, National Data Buoy Center (NDBC) has no available long-term buoy data in the SCS. To obtain a long series of field observation data to verify the CFOSAT wave data in the SSCS, two wave buoys were deployed in the central region (buoy A at 10.0° N, 115.5° E and buoy B at 9.5° N, 113.0° E, Figure 1) of the SSCS. Buoys A and B were launched at a depth of 1200 m on 14 November 2018, and a depth of 1240 m on 17 October 2020, respectively. Wind data measured by buoy A have been used for the evaluation of CFOSAT scatterometer wind data [31]. A Triaxys wave sensor manufactured by AXYS Technologies Inc., Sidney, Canada, was equipped in each buoy to measure wave parameters every hour by accelerometers. Up to January 2021, more than 26 months of wave parameters were recorded by buoy A. Buoy B collected 10 consecutive months of wave parameters and wave spectrum data from October 2020 to August 2021. The dominant wavelength (L_p) is calculated by the peak wave period (T_p) through the dispersion equation:

$$L_p = \frac{gT_p^2}{2\pi} \tanh \frac{2\pi d}{L_p} \quad (1)$$

where d is the depth and g is the acceleration of gravity. The equation can be simplified as $L_p = gT_p^2/2\pi$ in deep water.

The spatial window with a radius $R = 150$ km and a temporal window of 1 h is used to match the CFOSAT parameters (Table 1) with the buoy data (see Figure 2a) using the inversed distance weighting interpolation method:

$$\overline{P_S} = \frac{\sum P_{S_i} w_i}{\sum w_i} = \frac{\sum P_{S_i} (1/d_i)}{\sum (1/d_i)} \quad (2)$$

where $\overline{P_S}$ is the mean satellite parameter value matched with the buoy, i is the number of satellite observations within the radius, P_{S_i} and w_i are the relative satellite parameter and its weight factor, respectively, and w decreases with increasing distance from the satellite sampling point to the buoy position:

$$d = \begin{cases} \text{dist}_s & (P_S = \text{wavelen}_s, \text{swh}_s, \text{dir}_s, \text{dist}_s, \text{phase}_s) \\ \text{Dist}_n & (P_S = \text{SWH}_n, \text{Wind}_n, \text{Dist}_n, \text{Phase}_n) \end{cases} \quad (3)$$

For comparisons between the buoy and satellite, a space domain of 0 to 150 km and a time domain of 0 to 1.5 h are typically acceptable under normal circumstances [1,3,32,33]. A relatively larger space window is chosen in the present study to avoid an insufficient number of data pairs. The comparison of the averaged SWIM data does not show a large gap under different space windows. For instance, the RMSE (root mean square error) between the averaged wavelength data within a space window of 150 km and 100 km is 24.5 m, which is significantly smaller than that compared to the in situ observation of buoys (i.e., 96.11 m in Figure 2b). Consequently, a total of 166 and 58 sets of matched data were obtained at the buoy A and buoy B stations, respectively.

As illustrated in Figure 2, the dominant wavelengths measured by CFOSAT are generally larger than those measured by buoys. The former is always no less than 100 m, while the latter is often less than 100 m. This difference may be due to the reasons described in the Introduction; that is, when waves pass through the SSCS, which is dotted with islands and reefs, they are prone to deformation due to refraction, diffraction, and shallowing [10–12]. It was shown that islands and reefs would influence the wave properties and dynamics by the shading effect [34–36]. In some harsh sea states, the islands could have a noticeable sheltering effect felt in various directional sectors and even at large distances from the islands [37]. Although the buoys are set in the relatively deep water, the waves here are still remarkably influenced by the complex topography of the nearby island groups.

However, because CFOSAT is better at observing fully developed swell whose wavelength range is relatively long [4], and the observation resolution of the CFOSAT has difficulty distinguishing islands and reefs [7], the measured dominant wavelengths in the SSCS would easily deviate from the true values. For instance, the wave spectrum data from buoy B shows a mixed wave pattern in the SSCS (Figure 3). Concerning the fact that the SSCS is dominated by the swell in general [9], this pattern may be influenced by local wind and also due to wave diffraction and refraction effects caused by island topography. The topography may affect the transformation of the waves, reducing the component of the swell and making disadvantage of the accuracy of SWIM measurement. The RMSE of SWIM wavelength is 84.79 m when the integrated wave spectrum density of the wind sea (0.15–0.385 Hz) is larger than that of the swell (<0.15 Hz). When conditions are opposite, the RMSE is reduced to 47.64 m. Thus, when the swell component is high, the waves are more fully developed, and the satellite observations are more in line with the in-situ observation values, and vice versa.

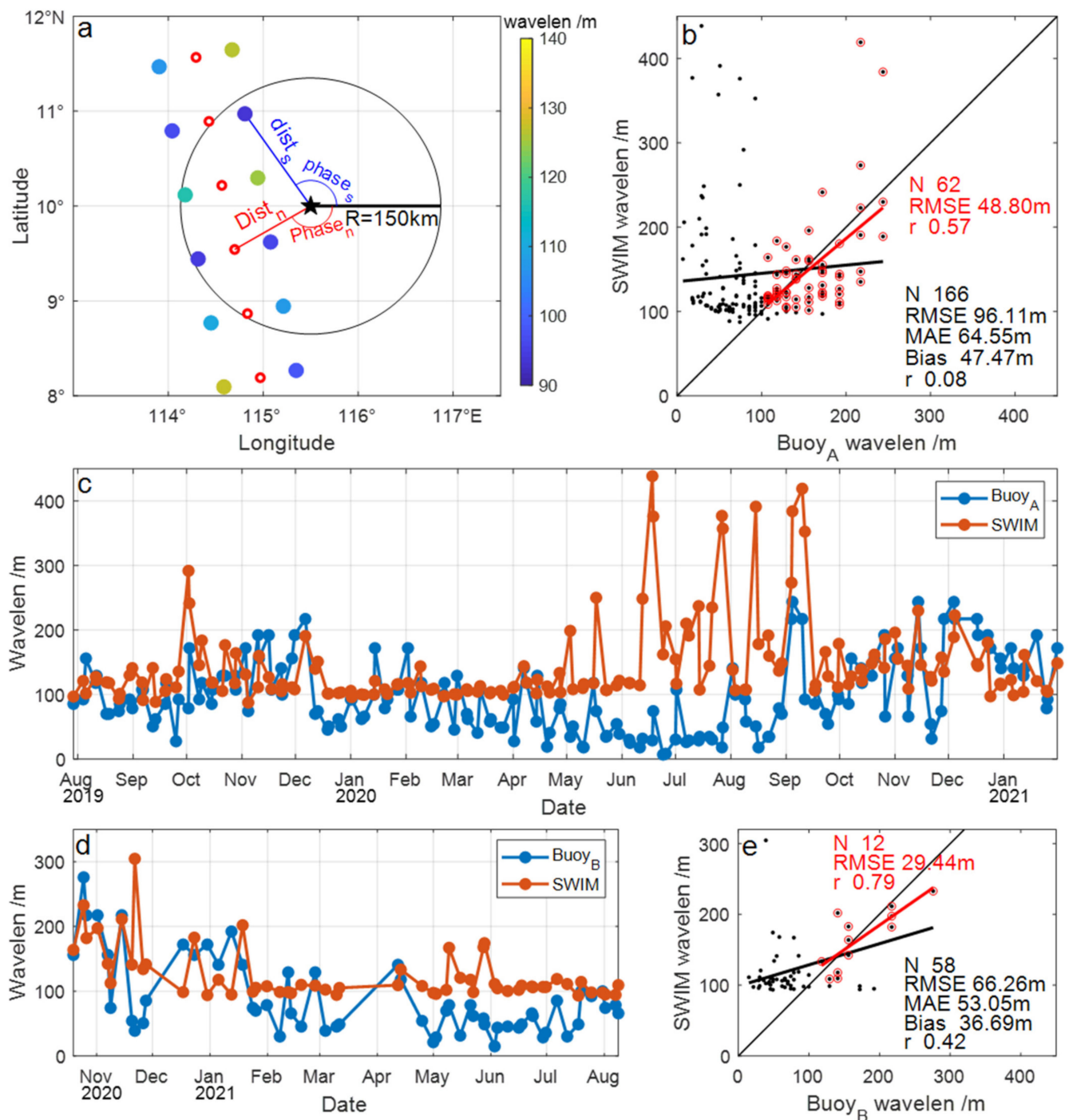


Figure 2. Space collocation and comparisons of the dominant wavelengths of CFOSAT and buoys; (a) shows the spatial matching results of position information recorded by buoy A and satellite data on 25 January 2021. The dark pentagram indicates the buoy position. The colored dots and the red circles represent the measuring points of the wave spectrum and the nadir beam, respectively; (b,c) show the comparison of dominant wavelengths from buoy A and CFOSAT in the forms of scatterplots and time series, respectively. (d,e) show the comparison of dominant wavelengths from buoy B and CFOSAT in the forms of time-series and scatterplots, respectively. The linear fitting curve, the sequence length (N), RMSE, mean absolute error (MAE), standard deviation (STD), and correlation coefficient (r) of the SWIM-observed dominant wavelengths are illustrated in black in (b,e). The scatters, curves, and legends in red in (b,e) are calculated from wavelengths exceeding 100 m.

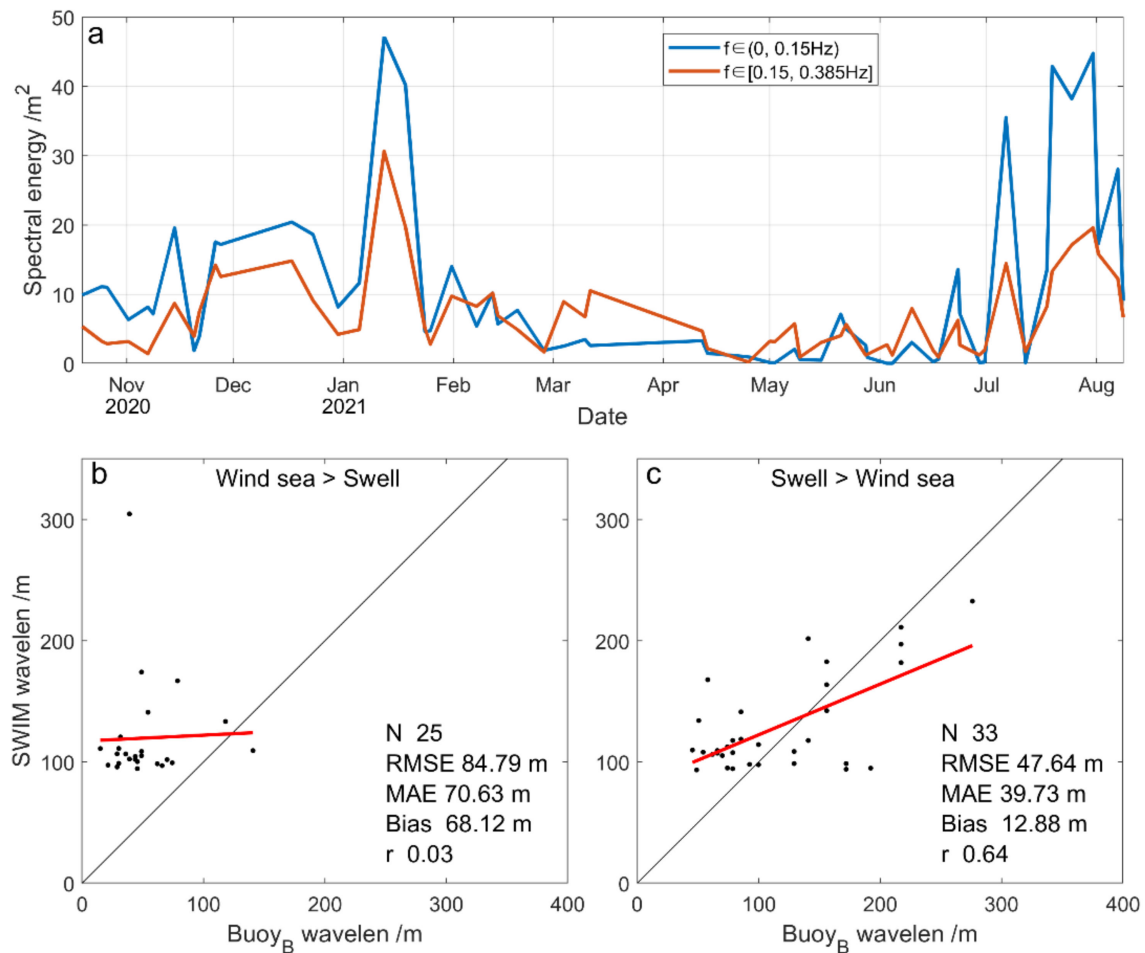


Figure 3. Comparisons of dominant wavelengths from buoy B and CFOSAT SWIM at different sea states; (a) shows the observed series of spectral energy of wind sea and swell, assuming 0.15 Hz is their frequency threshold; (b,c) are the scatterplots comparisons of dominant wavelengths when wind sea and swell dominates, respectively.

The largest differences between CFOSAT and buoy A appear in May–September 2020 (Figure 2). Such apparent errors can be further related to the sea state with low wave height (<1 m basically) and abnormally variable wave direction (Figure 4). In the figure, ERA5 data show that the significant wave height and wind speed in May–September 2020 are significantly lower and the dispersion of wave direction and wind direction is significantly higher compared with May–September 2019 or other months of 2020. The SWIM measures the wave spectrum by the modulated signals of the backscattering coefficient by sea surface tilt [6,7]. However, the backscattering coefficient itself is related to the sea surface roughness mainly linked to surface winds and waves [38], which is background noise for the modulated signals of sea surface tilt. If the wave height is too weak and the wave direction distribution is too scattered, the backscattering coefficient modulation signal of sea surface tilt would be fuzzy, which is not conducive to the wave spectrum retrieval. For instance, the satisfactory accuracy of the SWIM nadir significant wave height (SWH) referring to the altimeter observation can be achieved when the SWH is above 1 m [38]. A similar result is suggested by a comparison between the SWIM and NDBC buoy [8]. They stated the SWH of nadir from CFOSAT shows the best accuracy in the wave height within 2–3 m. Although the reason for the special sea state from May to September 2020 is still unclear, it indeed affected the accuracy of the retrieval accuracy of wave parameters from satellite measurements.

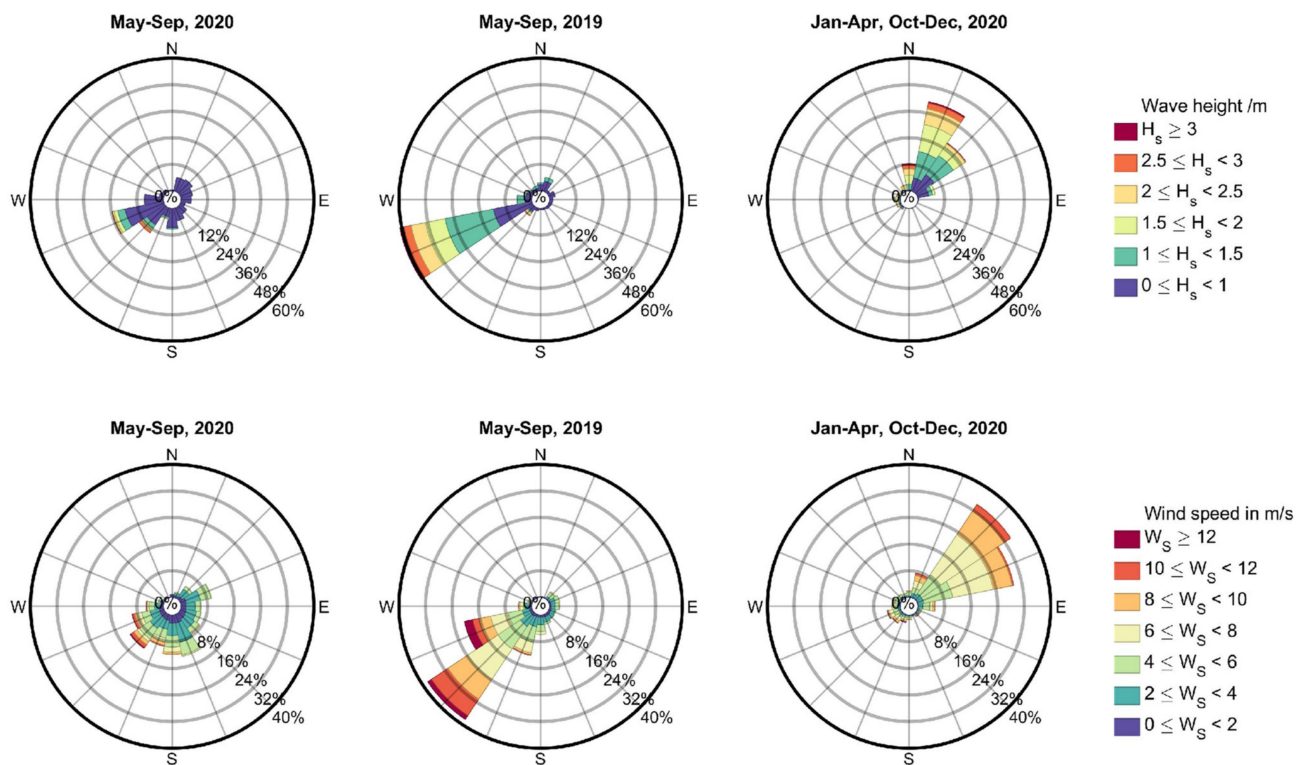


Figure 4. Comparison of the rose plot of SWH (**upper** panel) and wind speed (**lower** panel) between different periods at buoy A station based on ERA5 data.

Furtherly, the accuracy of the SWIM wavelength determined by SWH ranges is evaluated throughout the whole period of buoy employment. Both at the stations of buoys A and B, the SWIM wavelengths perform a significantly smaller RMSE when the wave height is within 2–3 m than when the wave height is below 2 m. That is, a more satisfactory correlation between SWIM and buoy wavelength is obtained after taking out all the data with low SWH. The accuracy of SWIM wavelength when the SWH exceeds 3 m is not evaluated because there are few samples in this range. Although the above differences in dominant wavelengths between the satellite observations and the buoy measured values are apparent, there is a clear correlation between them, especially when the in-situ observed wavelength is greater than 100 m (red circles in Figure 2b,e). Therefore, it is still possible to calibrate the wavelength data by using appropriate methods, such as NN models, to develop a relationship between SWIM measurement and the wavelength.

2.3. Reanalysis Data

To test the effect of additional input of other external data on the calibration of the dominant wavelength, this study further supplemented the European Center for Medium-Range Weather Forecasts (ECMWF) data. In the data correction and verification carried out by the CFOSAT research group, the wind and wave parameters outputted from the numerical model ECMWF are used to analyze the SWIM data [4,6]. ERA5 is the fifth generation ECMWF atmospheric reanalysis product of the global climate [39]. The data were obtained from the Climate Data Store of Copernicus Climate Change Service (<https://cds.climate.copernicus.eu>, 31 December 2021). Considering the close relationship between sea surface wind and waves, a total of six parameters (Table 2) of wave properties and wind fields were obtained. Similar to the buoy observations, the hourly ERA5 data at 10.0° N, 115.5° E and 9.5° N, 113.0° E were interpolated to match the SWIM measurement.

Table 2. List of external source parameters for the calibration of the dominant wavelength.

Symbol	Meaning	Unit
swh_e	Significant wave height from ERA5	m
dir_e	mean wave direction from ERA5	°
Tm_e	Mean wave period from ERA5	s
$u10_e$	10 m u-component of wind from ERA5	m/s
$v10_e$	10 m v-component of wind from ERA5	m/s

3. Methodology

3.1. Design of Calibration Experiments with Monte Carlo Sampling

Calibration experiments combining feed-forward backpropagation NN and Monte Carlo sampling methods are performed for the dominant wavelength. To select the critical input parameters for calibrating the dominant wavelength, NN models were constructed in three stages. In the first stage (denoted as NN1), all parameters were inputted to construct the NN model, and the critical parameters and the noncritical parameters were then selected by MIV analysis. In the second stage (denoted as NN2), the critical parameters were inputted to build the NN model. In the third stage (denoted as NN3), the principal components (PC) of noncritical parameters were calculated by kernel principal component analysis (KPCA), and both the critical parameters and the PC of noncritical parameters were inputted to construct the NN model. The overall process of the calibration experiment is shown in Figure 5. The establishment of NN models is shown in Section 3.2, and the MIV and KPCA methods are introduced in Section 3.3 below. To examine whether the data acquired by CFOSAT can complete the self-calibration of the dominant wavelength, two sets of data are used as input parameters in the experiment. The first set of input data includes nine parameters that can be acquired by the satellite itself (Table 1). Based on the former, the second set of input data includes external data (ERA5), with a total of fourteen parameters (i.e., Tables 1 and 2).

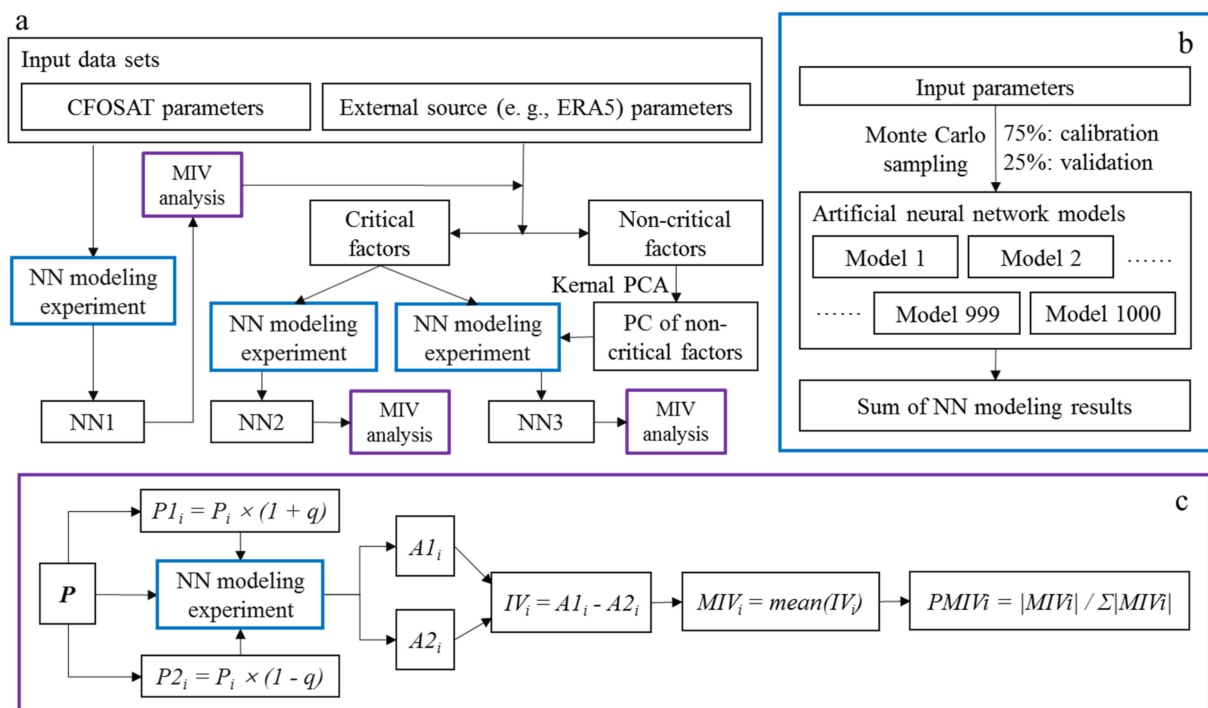


Figure 5. Flow charts of the whole calibration experiments (a), artificial neural network modeling with Monte Carlo sampling (b), and MIV analysis (c). The blue and purple borders indicate the nesting relationship among processes.

To avoid the uncertainty caused by insufficient data samples, the input parameters of each stage were randomly sampled 1000 times by the Monte Carlo method, and the calibration experiment was repeated. A total of 166 sets of data were obtained by matching buoy A with the satellite. As shown in Figure 5b, 75% of the data (125 groups) were randomly selected for the establishment of the wavelength calibration model, and 25% of the data (41 groups) were used for the verification of the model. Although the accuracy of SWIM wavelength differs according to the domination of wind sea or swell (Figure 3), and the ranges of wave height and wavelength itself (Figures 2b,e and 4), the calibration experiments do not separate these situations to establish calibration relationship in this study. One cannot judge the real sea states, or whether the SWIM wavelength data belong to a large value or a small value unless we obtain the real field observation value. So, matching samplings at each point were treated as an entirety to establish a calibration model suitable for a wide data range of waves. To further examine the validity and migration capability of the calibration method, the models established based on the data samples of buoy A were used to calibrate the wavelength data measured by the satellite at the buoy B location (58 groups in total). It is noted that the buoy B data are independent samples that are not involved in the model establishment process. The validation results will illustrate whether the model based on one station can be extended to a certain region, such as the SSCS.

3.2. Artificial Neural Network for Wavelength Calibration

A multilayer feed-forward backpropagation NN is established to calibrate the dominant wavelength. Although the optimum number of layers is often the subject of research, a three-layer structure has been proven to be sufficient to approximate any practical function, given enough neurons in the hidden layer [40]. Therefore, a network composed of an input layer, a hidden layer, and an output layer is used in this research to form a neural model. Such a type of NN is an efficient and mature algorithm that is widely used in the calibration, modeling, and prediction of ocean hydrological parameters [21,23,41–43].

Bayesian regularization is adopted as a back-propagation algorithm for sample training by using the MATLAB neural network toolbox. This algorithm updates the weight and bias values according to Levenberg–Marquardt optimization. It minimizes a combination of squared errors and weights and then determines the correct combination to produce a network that generalizes well [44,45].

The number of hidden layer neurons s for the neural models is determined by $s = 2n + 1$, where n represents the number of nodes in the input layer, which varies with the experimental settings. The logsig function is selected as the transfer function from the input layer to the hidden layer of the model, and the Purelin function is selected as the transfer function from the hidden layer to the output layer of the model. The initial learning rate is set as 0.005, and the iterative training times are set as 5000. The output neurons depend on the goal of the NN model, i.e., the dominant wavelength in this study.

3.3. Mean Impact Value Analysis and Kernel Principal Component Analysis

In the structural design of the NN network, taking all parameters as the input layer will consume many computer resources, and the introduction of nonimportant parameters will lead to overfitting or non-convergence of the model. Therefore, MIV analysis and KPCA are used to screen the key parameters and reduce the dimension of inputs in this study.

The impact of input neurons on the output neurons can be obtained by examining the internal weight matrix value [25]. Following this concept, MIV is introduced to evaluate the importance of input parameters to output in a neural network [26,27]. As illustrated in Figure 5c, the procedures of MIV analysis are as follows: First, a well-trained NN model is established using training dataset P . Subsequently, each input parameter i in P increases and decreases by a small percentage of q (equal to 10% in the present study) to create two new datasets $P1_i$ and $P2_i$, which are later imported into the trained NN model to produce the simulated outputs $A1_i$ and $A2_i$ (as a sequence). The difference between $A1_i$ and $A2_i$ is the impact value (IV_i). The average value of IV_i is denoted by MIV_i . A

positive (negative) value of MIV indicates a direct (inverse) relationship between the results (dominant wavelength) and the input parameters. The contribution proportion is calculated according to $PMIV_i = |MIV_i| / \sum |MIV_i|$, and the input parameters are ranked according to $PMIV_i$. The parameters ahead that contribute more than 90% cumulatively are regarded as critical parameters, and the rest are noncritical parameters.

For the noncritical parameters, one important multivariate analysis to reduce the data dimension is PCA, which is a kind of “linear” method based on the linear or approximate linear structure in high-dimensional space. However, wave variation and its interactions with different hydrological and meteorological parameters are typically nonlinear processes. Therefore, normal PCA may not be able to capture nonlinear features from various responses in wave dynamics. As a result, the present study adopted KPCA to realize nonlinear dimension reduction of noncritical parameters. KPCA is an enhanced PCA that incorporates a kernel function, thereby facilitating the solution of nonlinear problems [24]. By the use of integral operator kernel functions, one can efficiently extract PC in high-dimensional feature spaces related to input space by a nonlinear map [46]. In this research, the Gauss function is selected as the kernel function.

4. Results

4.1. Overall Results of the Calibration Experiments with Monte Carlo Sampling

The calibration experiment shown in Figure 5 was carried out. To reduce the consumption of computer resources and avoid the negative impact of unimportant parameters on the performance of NN models, the models were established in three stages (NN1, NN2, and NN3), and the critical parameters influenced the calibration results were obtained. In the experiments, two sets of input parameters (i.e., Tables 1 and 2) were used to test their effects on the calibration of the dominant wavelength.

The MIV analysis and the statistical indexes of the 1000 NN models with only input CFOSAT parameters (Table 1) are shown in Figures 6 and 7, respectively. By calculating the MIV and PMIV of the input parameters, the critical parameters that affect the performance of the calibration model can be extracted. As shown in Figure 6, MIV analysis of NN1 results shows that PMIV of five parameters, dominant wavelength ($wavelen_s$), the distance of the off-nadir sampling point from the calibration point ($dist_s$), the distance of the nadir sampling point from the calibration point ($Dist_n$), significant wave height from nadir beam (SWH_n), and wind speed from nadir beam ($Wind_n$) accounts for 93.12% contribution in total absolute MIV. Both distance-related parameters, $dist_s$, and $Dist_n$, contribute significantly to MIV. However, considering that the two parameters have similar meanings (see Figure 2a), if they are selected as critical parameters at the same time, the weight of the distance factor in the calibration process may be too high. Therefore, only the first parameter ($dist_s$) is selected as a critical parameter. Then, the critical parameters were used as input parameters to establish model NN2, and the relative MIV and PMIV were calculated again. The order of the four critical parameters can be further determined as follows: SWH_n , $Wind_n$, $dist_s$, and $wavelen_s$. Finally, these critical parameters and the first two PCs (accounting for more than 66.01%) of the noncritical parameters were used as inputs to construct the NN3 model. After calculating MIV and PMIV again. The results of NN3 confirm that the noncritical parameters can be ignored, considering their low PMIV values.

For the verification results of each sampling, RMSE, MAE, STD, and r were calculated. Therefore, a total of 1000 groups of RMSE, MAE, STD, and r can be obtained. Figure 7 can be summarized by further counting these statistics. As shown in the figure, when only the parameters obtained from CFOSAT are used for calibration, the RMSE decreases from 80–110 m to 40–50 m, the MAE decreases from 60–70 m to 30–40 m, the r increases significantly from 0–0.2 to 0.5–0.7, and the STD is closer to the buoy-observed STD. The calibration performance of the model with critical parameters (NN2) is better than that of the model with all parameters (NN1) and the model with critical parameters and the PCs of noncritical parameters (NN3). The RMSE and MAE of NN2 are approximately 4 m lower than those of NN1, and r is approximately 0.05 higher than those of NN1 (Table 3). By

extracting the critical parameters and using the critical parameters to build the calibration model, both the calculation efficiency and calibration performance can be improved.

Table 3. Median RMSE, MAE, r, and STD of the dominant wavelength from the SWIM and NN models based on different input parameters.

	Buoy	SWIM	CFOSAT Parameters Only			CFOSAT + ERA5 Parameters		
			NN1	NN2	NN3	NN1	NN2	NN3
RMSE (m)	0	94.54	48.23	44.35	46.88	36.00	32.24	33.33
MAE (m)	0	63.58	36.55	33.78	35.49	26.63	23.47	24.35
r	1	0.08	0.59	0.64	0.61	0.78	0.83	0.82
STD (m)	56.02	66.41	47.94	44.20	47.84	50.01	50.14	49.55

For the NN models that both CFOSAT and ERA5 parameters are input (Tables 1 and 2), five key parameters affecting the calibration performance can be obtained (see Figure 8). In addition to SWH_n , $Wind_n$, $dist_s$, and $wavelen_s$, based on MIV analysis of NN1, the mean wave period from ERA5 (Tm_e) was also included as a critical parameter. It is worth noting that in the MIV analysis of NN2, Tm_e has a great influence on the calibration performance, and its PMIV contribution is 76%, while the total contribution of the other four parameters is only 24%. The critical parameters and the first two PCs of noncritical parameters were input to construct the NN3 model. The calculated MIV and PMIV also indicate that the noncritical parameters contribute little to the model output, which is similar to the results in which only CFOSAT parameters are included (Figure 6).

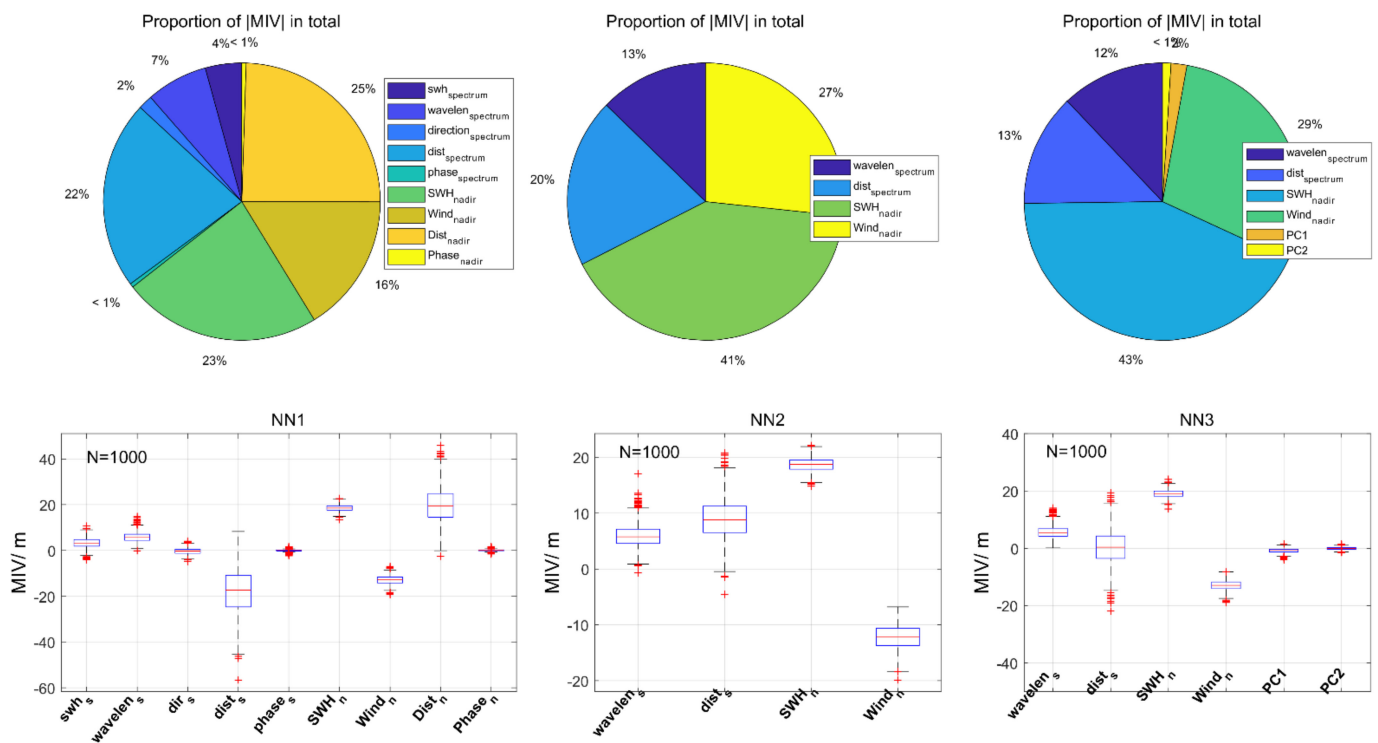


Figure 6. Proportions of PMIV with the median RMSE (upper panels) and boxplots of the MIV were calculated from the three stages NN modeling results (NN1, NN2, and NN3) with 1000 samples from the CFOSAT parameters (Table 1) (lower panels). PC1 and PC2 are the first two PC of the non-critical parameters. Points in red “+” are drawn as outliers if they are larger than the maximum ($Q3 + 1.5 \times (Q3 - Q1)$) or smaller than the minimum ($Q1 - 1.5 \times (Q3 - Q1)$), where Q1 and Q3 are the 25th and 75th percentiles, respectively.

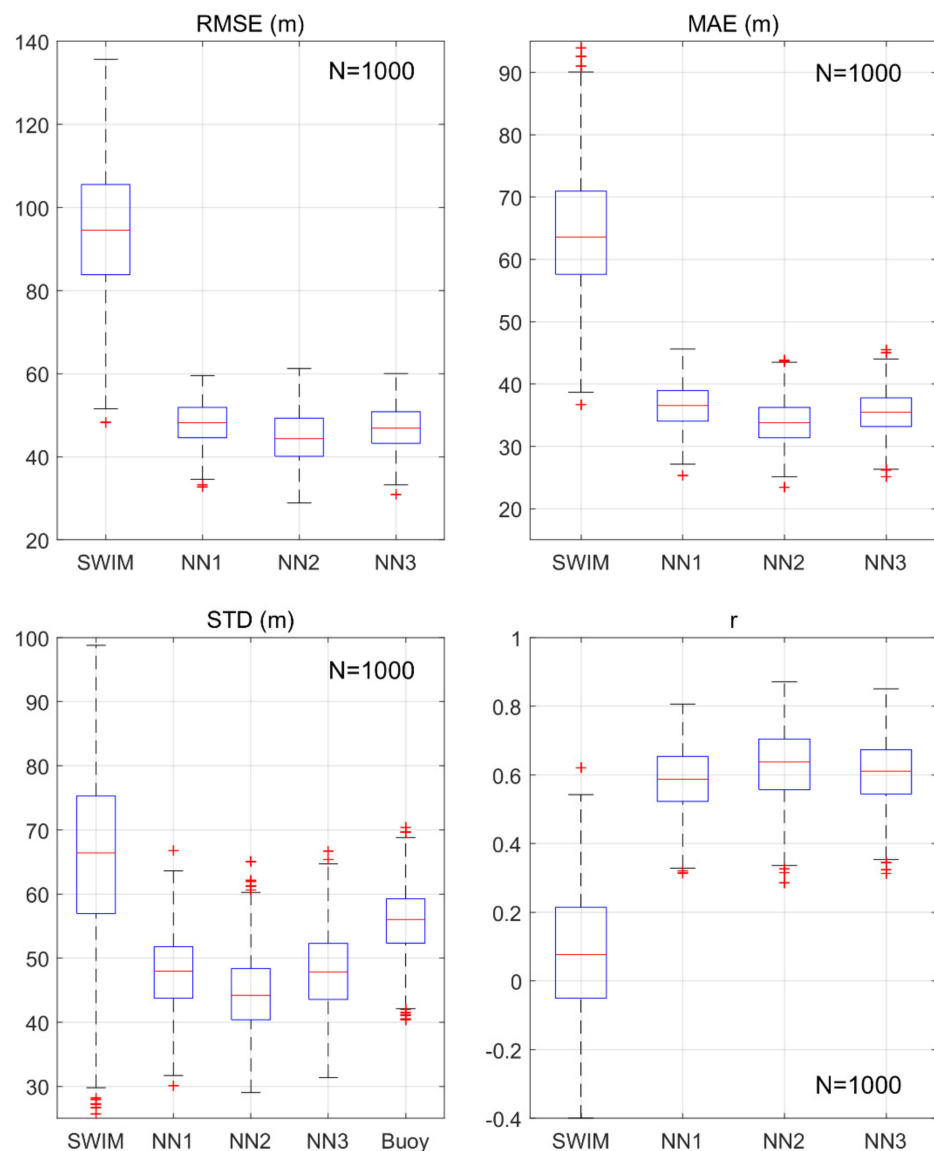


Figure 7. Boxplots of the RMSE, MAE, STD, and r of the SWIM observed dominant wavelengths and the validation results in the three-stage NN models (NN1, NN2, and NN3) based on self-calibration in which only CFOSAT parameters (Table 1) were used. Points in red “+” are drawn as outliers if they are larger than the maximum ($Q3 + 1.5 \times (Q3 - Q1)$) or smaller than the minimum ($Q1 - 1.5 \times (Q3 - Q1)$), where $Q1$ and $Q3$ are the 25th and 75th percentiles, respectively.

Compared to the calibration results based only on the parameters obtained from CFOSAT (Figure 7), if the ERA5 parameter is supplemented, the performance will be further improved (Figure 9). The RMSE will be reduced to 30–40 m, the MAE will be reduced to 20–30 m, r will be increased to 0.7–0.9, and the STD will be closer to the buoy-observed STD. By comparing the results of the three modeling stages, the models with critical parameters (NN2) are better than those with all parameters (NN1) or with critical parameters and the PC of noncritical parameters (NN3). As shown in Figure 9 and Table 3, the RMSE and MAE of NN2 are approximately 4 m lower than those of NN1, and r is approximately 0.05 higher than those of NN1. It can also be seen that selecting critical parameters to actuate an NN model can greatly accelerate the calculation efficiency with a more optimized result.

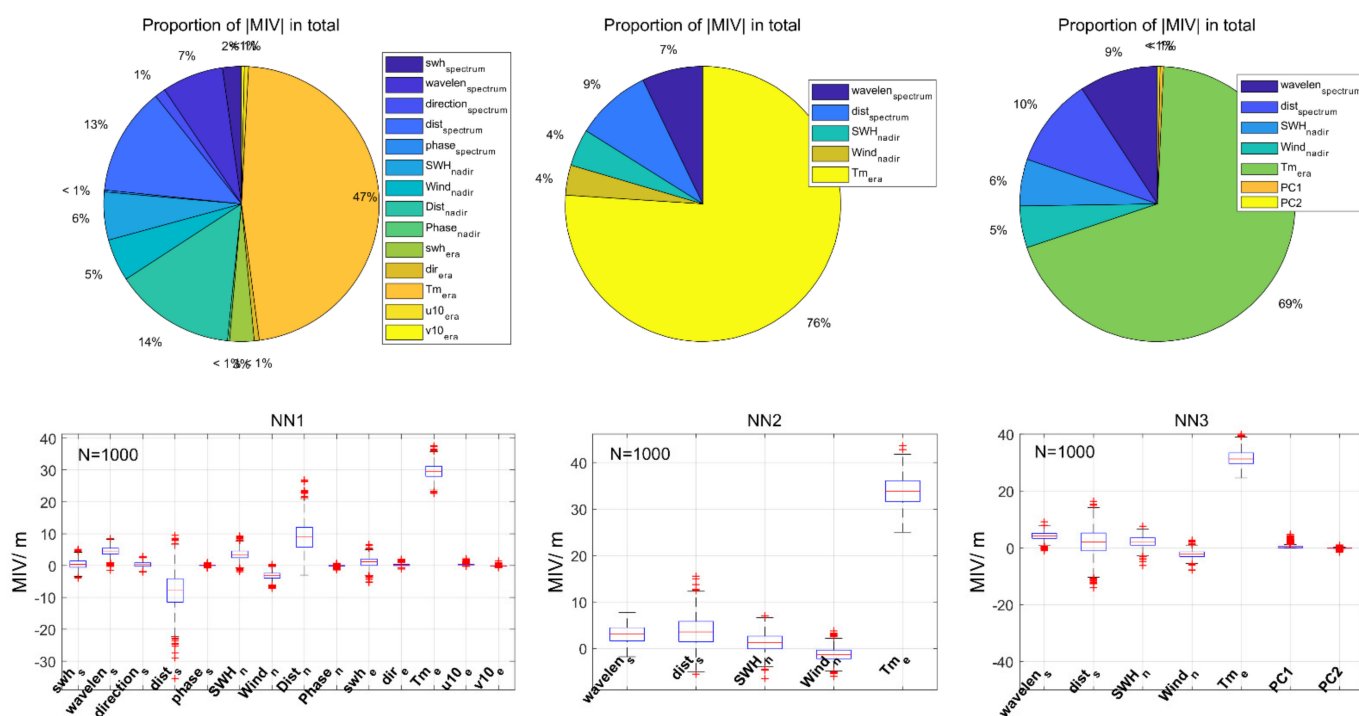


Figure 8. Similar to Figure 6 but for CFOSAT + ERA5 parameters (Tables 1 and 2) are included in the NN modeling. Points in red “+” are drawn as outliers if they are larger than the maximum ($Q3 + 1.5 \times (Q3 - Q1)$) or smaller than the minimum ($Q1 - 1.5 \times (Q3 - Q1)$), where $Q1$ and $Q3$ are the 25th and 75th percentiles, respectively.

Among the 1000 models based on random samples, the model in which the RMSE of validation sample of the SWIM wavelength is closest to median was selected as a representative for analysis. As shown in Figure 10, even if samples are randomly divided for calibration and verification, different NN models are capable of calibrating the SWIM wavelengths. In this sampling scheme, the RMSE for the validation samples is 94.6 m (corresponding to the red line of SWIM boxplot in the upper-left panels of Figures 7 and 9), and the RMSE for the calibration samples is 100.0 m. In the calibration phase, the NN2 models based on CFOSAT self-calibration and adding ERA5 parameters reduced RMSE to 34.0 m and 21.0 m, respectively, with other indexes also improved. In the validation phase, the NN2 models based on CFOSAT self-calibration and adding ERA5 parameters reduced RMSE to 33.5 m and 31.0 m, MAE to 27.5 m and 20.2 m, with r rising to 0.77 and 0.82, respectively.

4.2. Typical Calibration Results in the Chronological Order

In the above section, model experiments were randomly sampled by the Monte Carlo method, and the critical parameters of the model were selected from the statistical point of view, which verified the feasibility of using the NN model to calibrate the CFOSAT dominant wavelength. However, in the practical application of model calibration parameters, the periods of modeling and application are not randomly distributed but usually use available data to establish a model and then apply the model to an unknown period. To consider this situation, the present section adopts the corresponding sampling method; that is, the former 75% of the data sequence (late July 2019 to early September 2020) is used for model training (calibration), and then the remaining 25% (mid-September 2020 to January 2021) is used for model validation. We repeat the above three stages of NN model experiments. Considering that the effect of NN modeling also has randomness, we use the same sample and parameter settings to model 20 times and take the group of results with RMSE as the median for analysis.

Taylor diagrams of models are shown in Figure 11. It can be seen from the figure that the accuracy of the satellite observed wavelength can be greatly improved by using NN models with different input parameters in each stage. In the calibration period, the centered pattern root-mean-square difference (RMSD) is reduced from more than 80 m to 20–40 m, and r is increased from 0 to 0.6–0.9. In the verification period, the RMSD decreases from 70 m to 30–50 m, r increases from 0.4 to 0.5–0.8, and the STD difference also decreases. Furthermore, it is found that the performance of the model based on the combination of CFOSAT and ERA5 parameters is better than that based only on the CFOSAT parameters. Moreover, comparing the NN models in the three stages, it is found that the performance of the NN2 model is relatively better. This shows that the effect of the calibration model with critical parameters is better than that with all parameters.

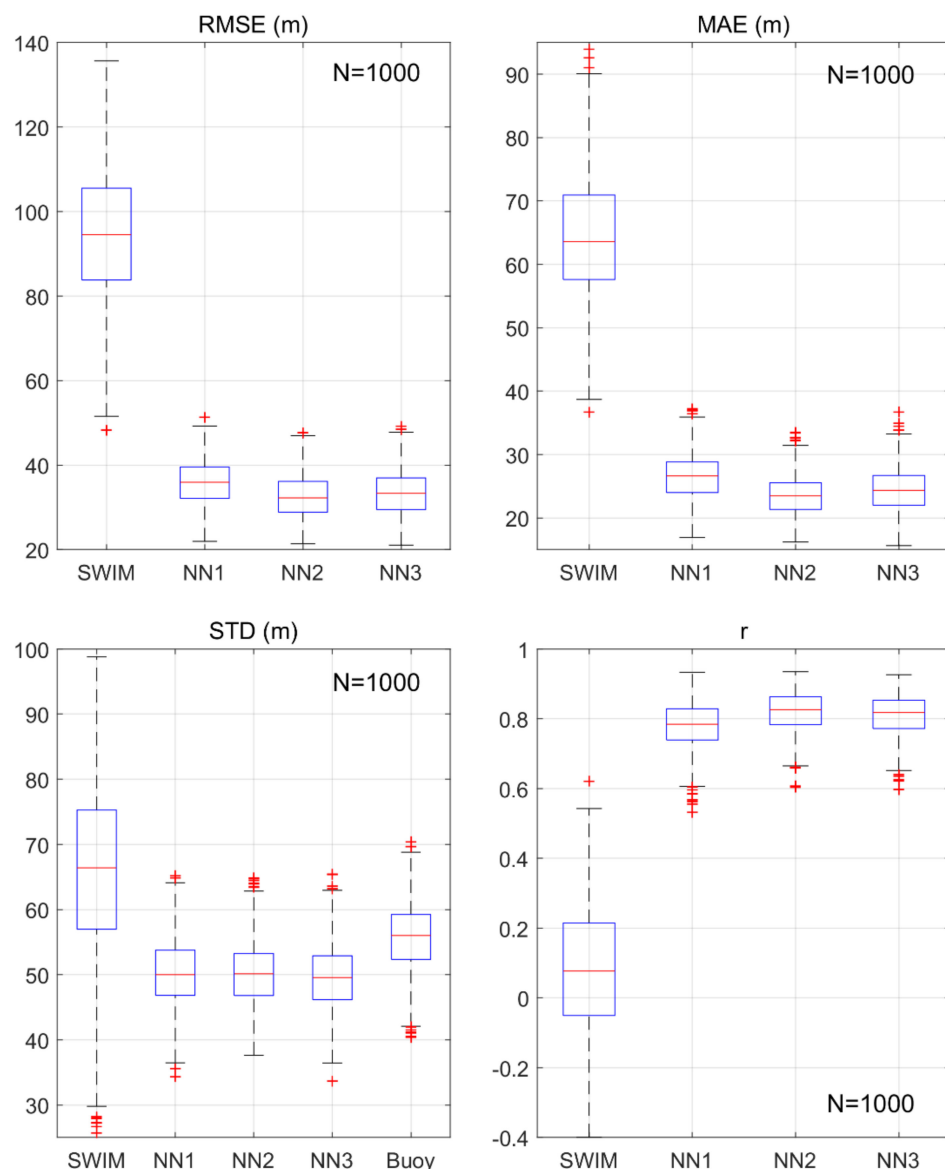


Figure 9. Same as Figure 7 but for external-source calibration, ERA5 parameters are also included (i.e., Tables 1 and 2) in the NN models. Points in red “+” are drawn as outliers if they are larger than the maximum ($Q3 + 1.5 \times (Q3 - Q1)$) or smaller than the minimum ($Q1 - 1.5 \times (Q3 - Q1)$), where $Q1$ and $Q3$ are the 25th and 75th percentiles, respectively.

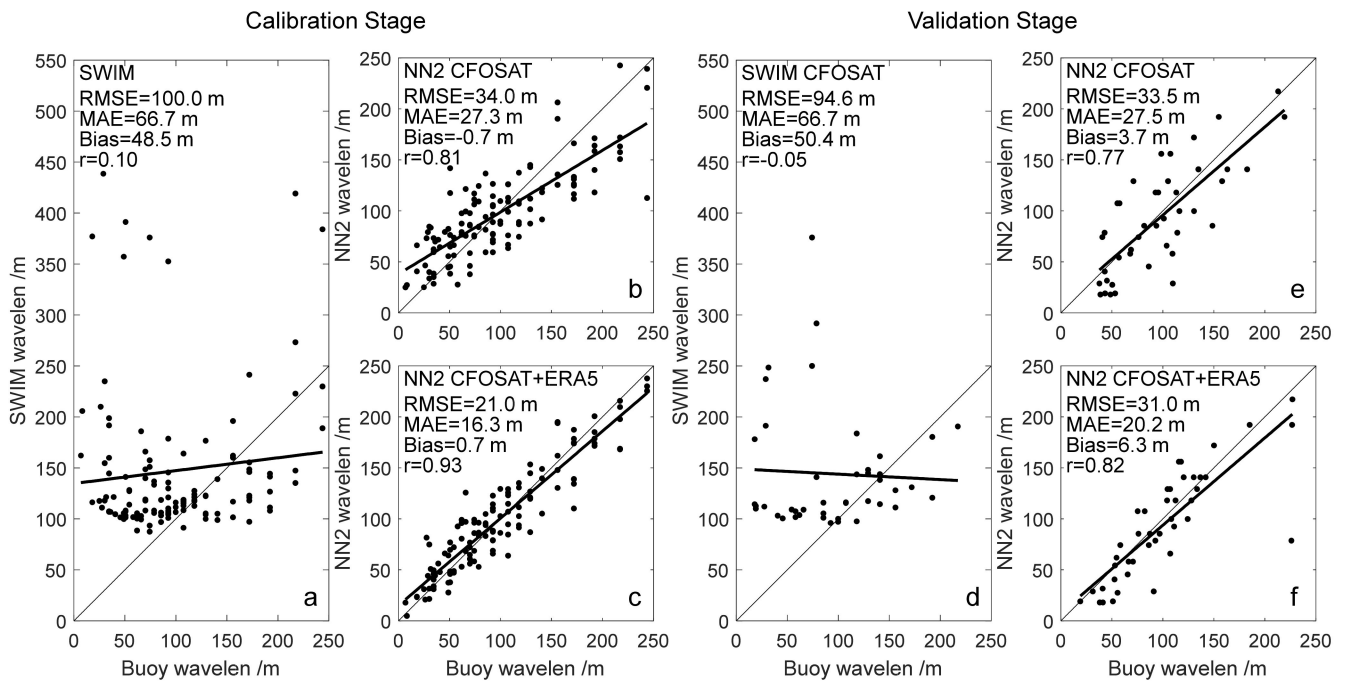


Figure 10. Typical calibration (left part, (a–c)) and validation (right part, (d–f)) example of the NN2 model in which the RMSE of validation sample of the SWIM wavelength is closest to median out of 1000 samplings (i.e., the red lines of the SWIM boxplots in the upper-left panels of Figures 7 and 9); (a,d) are the comparisons of wavelengths by buoy A and SWIM in scatters; (b,e) and (c,f) are the scattering comparisons of wavelengths from buoy A and NN2 model with CFOSAT parameters and CFOSAT + ERA5 parameters, respectively.

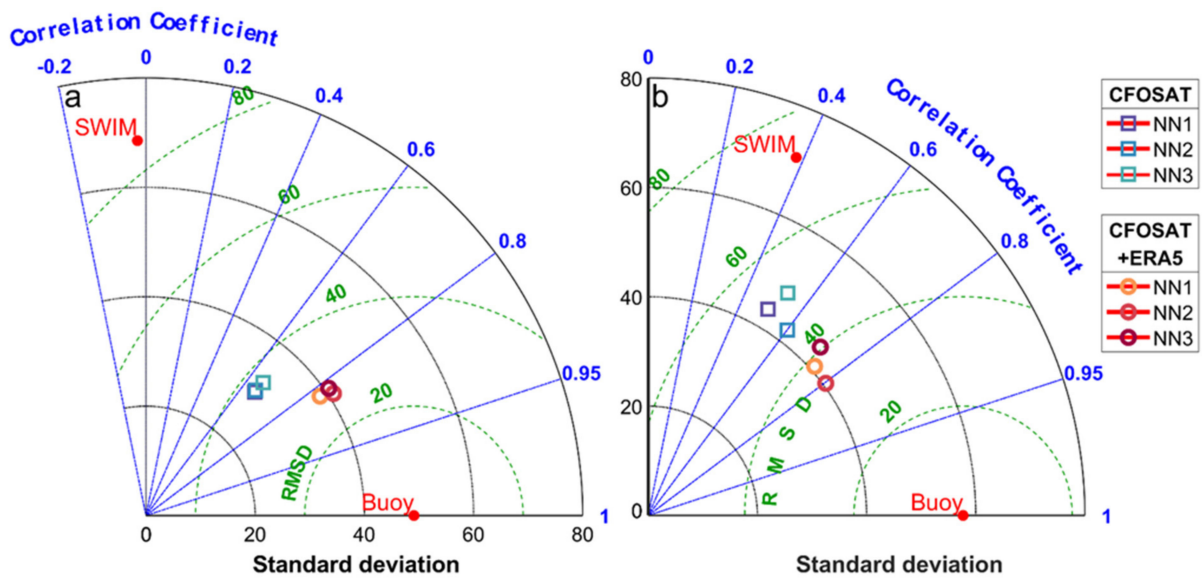


Figure 11. Taylor diagrams of the calibration (a) and validation (b) results for the three-stage NN models (NN1, NN2, and NN3) using the typical sampling method (former 75% for calibration and latter 25% for validation); squares represent the results in which only CFOSAT parameters (Table 1) were used, while circles represent the results in which ERA5 parameters (Tables 1 and 2) were additionally included. RMSD in the Taylor diagram indicates the centered pattern root-mean-square difference.

Specifically, the NN2 model is further analyzed, and the results of the model built with the CFOSAT parameters and the CFOSAT + ERA5 parameters are shown in Figure 12. From the comparison of time series and scatter points, it can be found that the NN2 model can effectively calibrate the satellite-observed results of the dominant wavelength. In the calibration period, more than one year of data is used, among which the wavelength observation errors in summer and autumn are large. The calibration model can be effectively corrected to improve many model indexes. Although the verification period is in winter and spring when the satellite observation errors are relatively small, the model results can still bring the dominant wavelength closer to the observation values. In the scatterplots where calibrated wavelengths are compared with observations, most of the matching points are not far from either side of the 1:1 line, and the fitting lines almost coincide with the 1:1 line. Under this sample division scheme, the performance of using the CFOSAT + ERA5 parameters to build the model is also better than using the CFOSAT parameters only, and the matching point is closer to the 1:1 line.

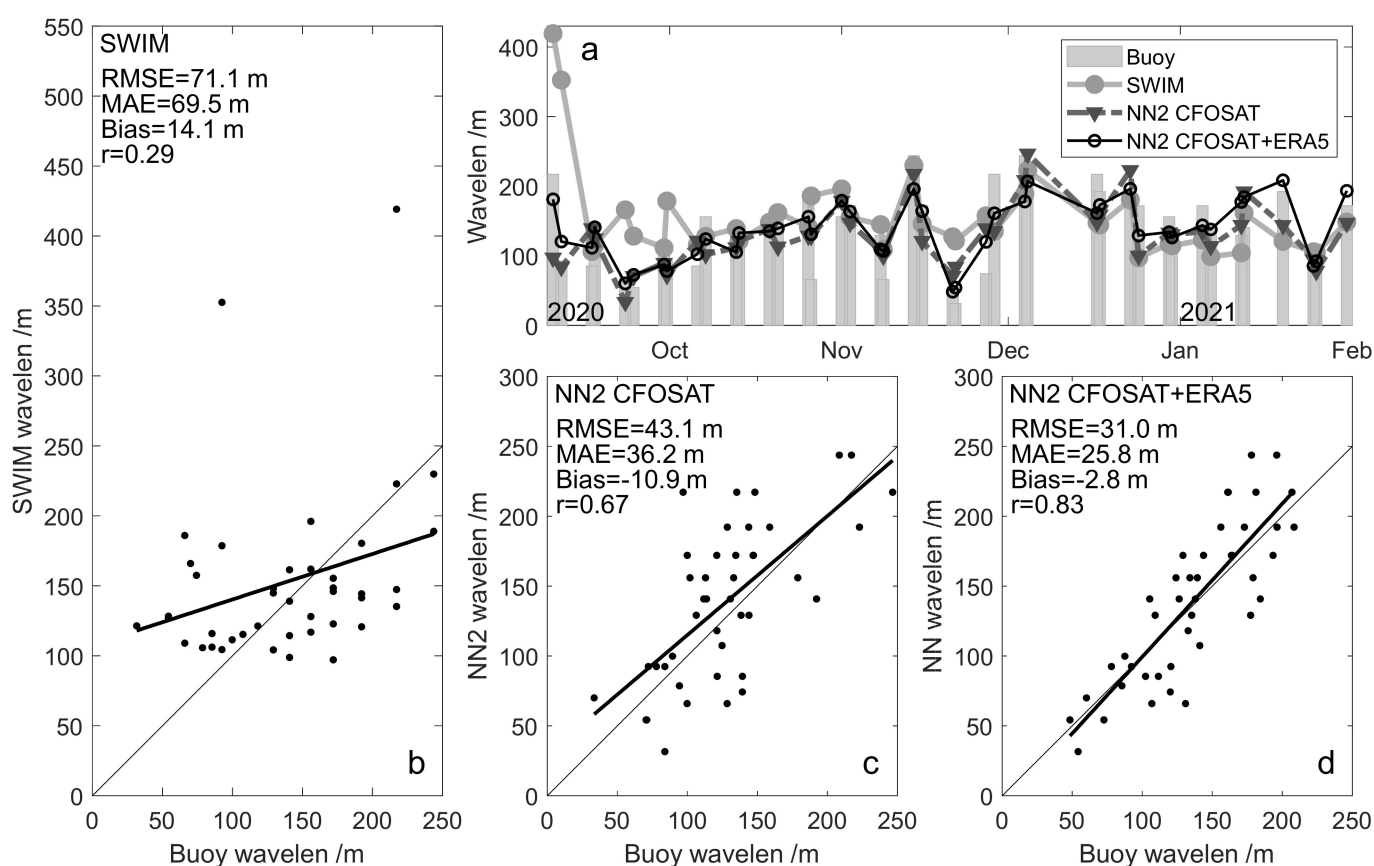


Figure 12. Validation results of the NN2 model with a typical sampling method (former 75% for calibration and latter 25% for validation); (a) comparisons of wavelengths by SWIM, NN2 model, and buoy A in time series; (b) comparisons of wavelengths by SWIM and buoy A in scatters; (c,d) scattering comparisons of wavelengths from buoy A and NN2 model with CFOSAT parameters and CFOSAT + ERA5 parameters, respectively.

4.3. Validation Results against an Independent Buoy Dataset

The above section shows the calibration and verification results of SWIM data at buoy A station. As mentioned, this study collected buoy observations at buoy B for 10 months and compared satellite results. Due to a small number of matching data (58 groups), independent modeling was not possible. Using buoy B data, the model established based on the samples of buoy A was validated at a different site. The results showed that all the statistics improved significantly based on different modeling schemes (Figure 13).

The results not only further prove the validity of the proposed calibration model for the wavelength data of satellite measurements at specific stations but also prove that the calibration method has good migration capability.

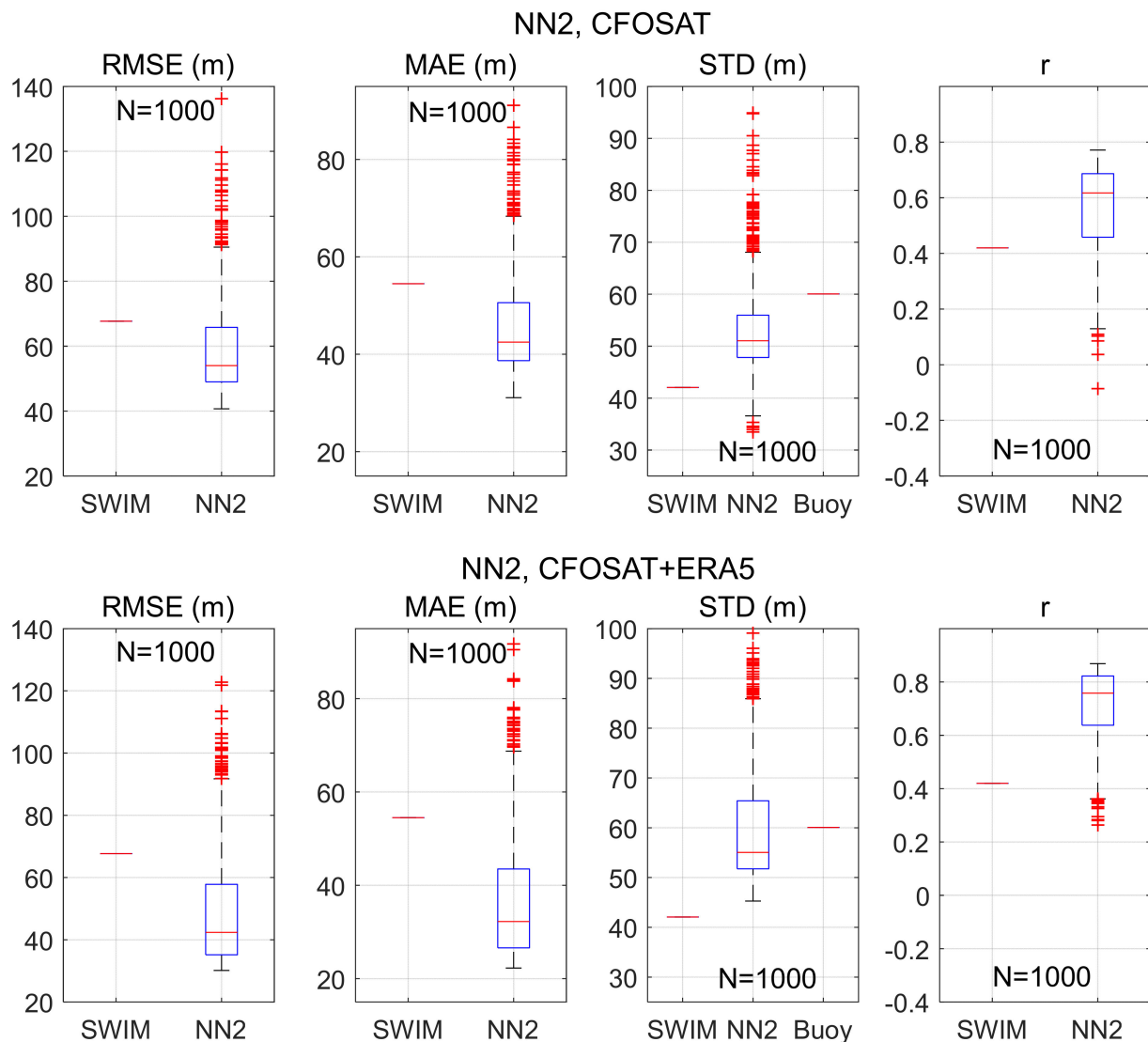


Figure 13. Boxplots of the RMSE, MAE, STD, and r of the SWIM dominant wavelengths and the validation results of the NN2 models at buoy B. The upper panels show the self-calibration results with CFOSAT parameters, and the lower panels show external-source calibration in which both CFOSAT and ERA5 parameters are included. The NN2 models are established by the Monte Carlo experiments based on the data at buoy A. Points in red “+” are drawn as outliers if they are larger than the maximum ($Q3 + 1.5 \times (Q3 - Q1)$) or smaller than the minimum ($Q1 - 1.5 \times (Q3 - Q1)$), where $Q1$ and $Q3$ are the 25th and 75th percentiles, respectively.

It is worth noting that the improved indicators of different models for the verification sample at buoy B show a very discrete pattern. For the statistics of 1000 groups of model samples, there are many outliers with poor verification effect, indicating the random instability generated by the model in migration (Figure 13). For example, according to the RMSE of 1000 calibration models that both CFOSAT and ERA5 parameters are input, there are 36 models are zoned as outliers with red “+”, so they are classified as abnormal models, and the remaining 964 models are classified as normal models.

By tracing the source of the two types of models, their improvement effect on the sample at buoy A can be reanalyzed. For example, the lower-left panel in Figure 14 shows

that the median declines in RMSE of 36 abnormal models and 964 normal models based on CFOSAT + ERA5 parameters are 52.52 m (cyan blue arrow) and 62.64 m (orange arrow), respectively. The improvement range of the SWIM wavelength of the normal model is better. By comparing the error indexes of normal models and abnormal models, we can find that except for MEA of the self-calibration model based on CFOSAT parameters (upper middle panel in Figure 14), the improvement of other indexes of the normal models is better than that of abnormal models. This reveals that models with better local calibration performances are more suitable and stable to calibrate the SWIM wavelength for neighboring locations and thus have higher promotion and application ability.

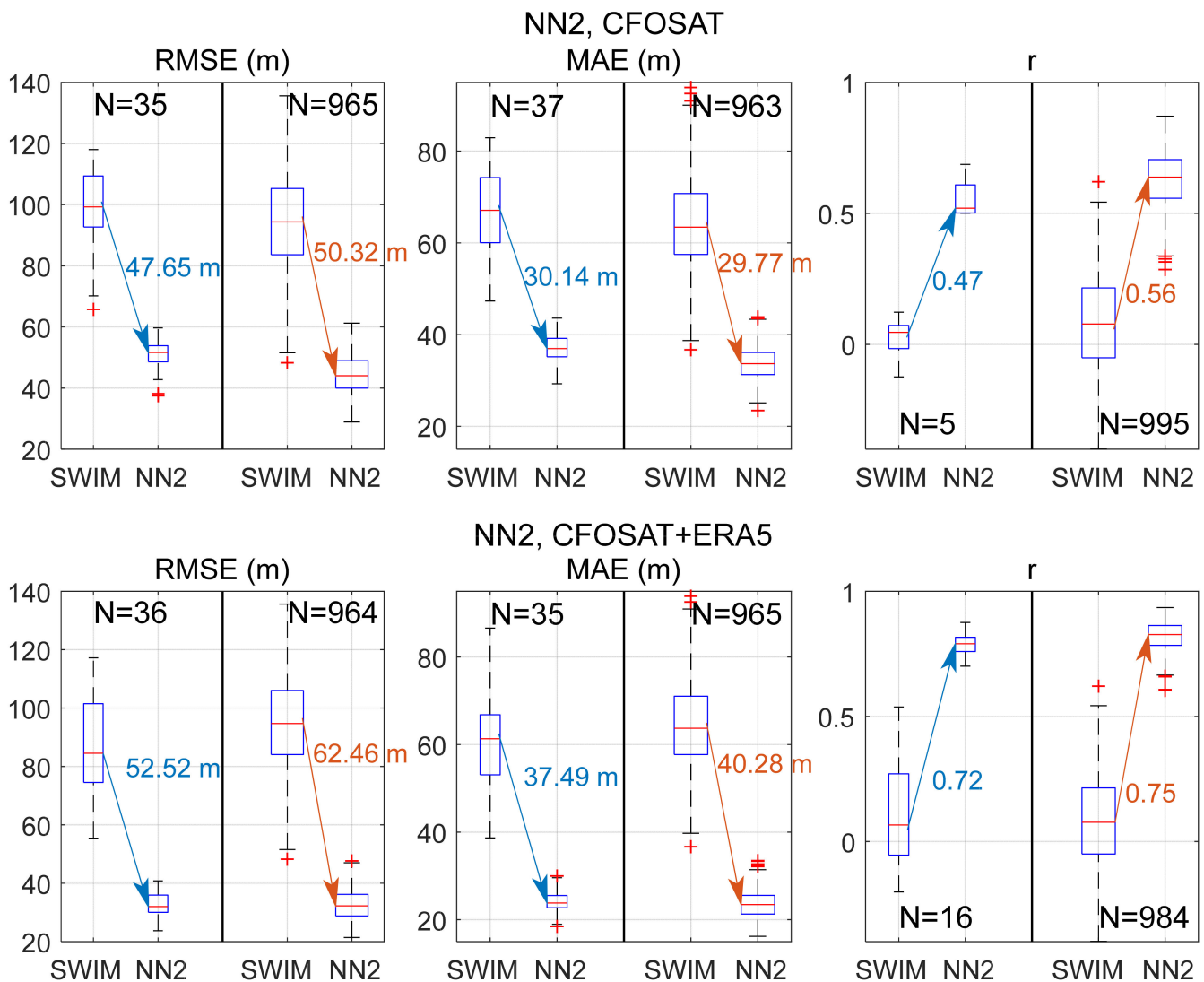


Figure 14. Boxplot comparison of validation results between the abnormal (left in each subplot) and normal (right in each subplot) NN2 models at buoy A. The abnormal and normal models are corresponding to the outlier samples (in red “+”) and remaining samples in Figure 13, respectively. The upper panels show the self-calibration results with CFOSAT parameters, and the lower panels show external-source calibration in which both CFOSAT and ERA5 parameters are included.

5. Discussion

5.1. Self-Calibration Ability of the Dominant Wavelength of CFOSAT

Satellite remote sensing is an effective means to realize large-area wave observations, e.g., [1,13,14]. Traditional satellites only use altimeters to measure wave height, due to the new type of off-nadir spectrometer SWIM carried by CFOSAT, it is possible to directly

obtain the wavelength information [4–7]. However, in certain local sea areas, such as the SSCS, the results of satellite measurements are susceptible to multiple islands or reefs due to the complexity of terrain and environment [10–12]. Therefore, the wavelength of satellite measurements can easily deviate from field observations (Figure 2).

One of the main results of this study is that the wavelength measurement deviation of CFOSAT can be corrected effectively by other parameters measured by the satellite. Both the results of 1000 calibration experiments with random sampling and the results of specific sampling methods with the earlier 75% data for calibration followed by 25% data for validation show that the model established by CFOSAT parameters can effectively reduce the RMSE and MAE of the dominant wavelength of satellite measurement, improve r and make the STD closer to the observed values (Figures 7, 11 and 12). This result is further confirmed by verification against independent buoy samples (Figure 13).

The reason why this calibration can be effective is that several critical parameters (Table 1 and Figure 6) obtained by the satellite provide relatively accurate information to establish the relationship between satellite measurements and the accurate wavelength value. Previous studies have retrieved wave period parameters from altimetry wave height and wind speed parameters based on analytical equations with artificial neural networks, e.g., [2,3]. However, in this study, we use NN models rather than analytical equations to directly describe this correlation, to introduce more information (parameters) into the correlation. For example, if the nine factors in Table 1 are introduced to structure a fixed equation, the equation will be complicated to establish. Using a neural network model can effectively solve this problem. For example, the recent research of Wang et al. [20,30,38] used NN models to calibrate and reconstruct the wave parameters (including wave period) of CFOSAT and achieved good results.

The dominant wavelength data measured by satellites can play an important role in engineering applications and scientific research. For disaster early warning, military activities, and other applications that require high timeliness of data, satellites must synchronously calibrate the data with other parameters measured by themselves. The results of this study confirm that CFOSAT has the potential to complete such synchronization calibration of the dominant wavelength (Figures 7 and 12). In addition, for applications requiring high data quality, such as scientific research using satellite data, it is necessary to further improve the wavelength data by using external-source parameters. The results show that supplementary input of ERA5 data can indeed further optimize the dominant wavelength data (Figures 8, 9 and 12), mainly through providing relatively accurate wavelength (wave period) information. However, due to the difficulty of synchronizing the data from other sources with satellite data, such as ERA5, which needs to be reanalyzed through model calculation [39], the timeliness of the generated data inevitably lags.

5.2. Critical Parameters for the Calibration of Dominant Wavelength of CFOSAT

Research in different fields has shown that the selection of appropriate input parameters has an important influence on the effectiveness of NN models [20,21]. The input of irrelevant parameters may not improve the NN model but may even worsen the model [23]. The results of this paper also show that using critical parameters to build the NN model (NN2) requires fewer computing resources and is more accurate than using all parameters to build the NN model (NN1) and even adding kernel principal components of noncritical parameters (NN3) (Figures 7, 9 and 11). This is probably because avoiding input of information that is not closely related to wavelength calibration can avoid the misleading effect of those factors on NN, thus allowing NN to focus more on the relationship between closely related parameters and wavelength in the modeling process and achieve better results.

In this study, by carrying out systematic modeling experiments, the critical parameters of calibrating the CFOSAT dominant wavelength by the NN model were extracted, including the distance of the off-nadir sampling point from the calibration point ($dists$), nadir significant wave height (SWH_n), nadir wind speed ($Wind_n$), and the dominant wavelength ($wavelens$) itself observed by the satellite (Figure 6). When these four parameters were

used for the calibration, the dominant wavelength would become much more closed to the reference values observed by buoys, with the indexes of RMSE, MAE, r , and STD significantly improved (e.g., Figures 7, 10e and 12c). The identification of the four parameters as the critical parameters is consistent with existing knowledge. For example, by providing accurate wave height information, a relatively accurate scale of wave sizes can be indicated to the calibration model. Moreover, the wind is closely related to the development degree of the wave [9], so providing accurate wind speed information can make the model realize the real development degree of the wave and then adjust the proportion relationship between wave height and wavelength. Previous studies have also pointed out that wind speed and significant wave height are also parameters that need to be input for the wave period retrieval of altimeter data [2,3].

It is also worth noting that the contribution of the wave period (T_{m_e}) is more than 3/4, much larger than the total influence of the other four critical parameters (Figure 8). Therefore, if external data are to be used, the most effective way is to provide a promising reference value of the wavelength (or the wave period according to the dispersion relationship), which is more meaningful than supplementing other parameter information.

6. Conclusions

The off-nadir spectrometer SWIM (Surface Investigation and Monitoring) of CFOSAT (China France Oceanography Satellite) can obtain the directional spectrum in the global ocean scale, to provide wavelength data. Although the data quality of the satellite has been widely recognized in most open oceans, there will inevitably be a shortage in the data quality in some sea areas, such as the southern South China Sea (SSCS), because the environmental conditions may be inconsistent with the open sea area. In this paper, the practicability of calibrating CFOSAT wavelength data against in situ observations in the SSCS by using artificial neural network (NN) models is systematically analyzed. To avoid the uncertainty caused by insufficient data, the Monte Carlo method was used to conduct random sampling, and a large number of repeated experiments were carried out. For each group of sampling data, NN model construction and mean impact value (MIV) analysis are carried out. The results show that the NN model based on the parameters measured by CFOSAT can greatly reduce the error of the SWIM dominant wavelength (e.g., Figures 7 and 10–13).

This study mainly contributes to two aspects. Firstly, the feasibility of self-calibration of the dominant wavelength by using CFOSAT measurements is verified. In addition, the critical parameters to realize the calibration are obtained. Besides the dominant wavelength itself, parameters such as the distance of the sampling point from the calibration point, the nadir significant wave height, and the nadir wind speed are also important for the calibration.

This study may provide reference and technical support for the application of CFOSAT data in local sea areas. In future research, it is still necessary to further calibrate and verify the wave data observed by CFOSAT using multichannel data in more sea areas to further optimize the satellite algorithm. Improving the quality of data, such as the dominant wavelength of satellite measurement, is important for providing better service of satellite data in engineering applications and scientific research.

Author Contributions: Conceptualization, B.L. and J.L. (Junmin Li); methodology, B.L. and J.L. (Junliang Liu); validation, B.L., P.S. and Y.L.; formal analysis, B.L.; investigation, J.L. (Junliang Liu), B.L. and W.C.; resources, S.T.; data curation, S.T.; writing—original draft preparation, B.L.; writing—review and editing, J.L. (Junmin Li); supervision, J.L. (Junmin Li); project administration, J.L. (Junmin Li); funding acquisition, J.L. (Junmin Li), B.L. and P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2019ZD0302), Hainan Provincial Natural Science Foundation of China (421QN380), CAS Key Laboratory of Science and Technology on Operational Oceanography (OOST2021-02), Guangzhou Science and

Technology Project (202102020464), National Key Research and Development Program of China (2021YFC3100501), National Natural Science Foundation of China (41776005), the CAS Key Technology Talent Program (173059000000160006), the Science and Technology Projects of Guangdong Province (2021B1212050023), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA13030304).

Data Availability Statement: The CFOSAT wave data are downloaded from <https://osdds.nsoas.org.cn> (accessed on 31 December 2021). The buoy data are available from the corresponding author by request.

Acknowledgments: We acknowledge the support of the CFOSAT team, CNES, and NSOAS in providing the data. The calculation in this study is supported by the High-Performance Computing Division in the South China Sea Institute of Oceanology.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Shanas, P.R.; Kumar, V.S.; Hithin, N.K. Comparison of gridded multi-mission and along-track mono-mission satellite altimetry wave heights with in situ near-shore buoy data. *Ocean Eng.* **2014**, *83*, 24–35. [CrossRef]
- Gommenginger, C.P.; Srokosz, M.A.; Challenor, P.G.; Cotton, P.D. Measuring ocean wave period with satellite altimeters: A simple empirical model, *Geophys. Res. Lett.* **2003**, *30*, 2150. [CrossRef]
- Quilfen, Y.; Chapron, B.; Collard, F.; Serre, M. Calibration/Validation of an Altimeter Wave Period Model and Application to TOPEX/Poseidon and Jason-1 Altimeters. *Mar. Geod.* **2004**, *27*, 535–549. [CrossRef]
- Hauser, D.; Tourain, C.; Hermozo, L.; Alraddawi, D.; Tran, N.T. New observations from the SWIM radar on-board CFOSAT: Instrument validation and ocean wave measurement assessment. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5–26. [CrossRef]
- Hauser, D.; Tourain, C.; Lachiver, J.-M. CFOSAT: A New Mission in Orbit to Observe Simultaneously Wind and Waves at the Ocean Surface. *Space Res. Today* **2019**, *206*, 15–21. [CrossRef]
- Hauser, D.; Tourain, C.; Hermozo, L. Report on the SWIM cal/val at the End of the Vervification Phase. 2019. Available online: https://www.aviso.altimetry.fr/fileadmin/user_upload/SWIM_CalvalReport_compressed.pdf (accessed on 31 December 2021).
- Tison, C.; Hauser, D. SWIM Products Users Guide: Product Description and Algorithm Theoretical Baseline Description. 2018. Available online: https://www.aviso.altimetry.fr/fileadmin/documents/data/tools/SWIM_ProductUserGuide.pdf (accessed on 31 December 2021).
- Liang, G.; Yang, J.; Wang, J. Accuracy evaluation of CFOSAT SWIM L2 products based on NDBC buoy and Jason-3 altimeter data. *Remote Sens.* **2021**, *13*, 887. [CrossRef]
- Su, H.; Wei, C.; Jiang, S.; Li, P.; Zhai, F. Revisiting the seasonal wave height variability in the South China Sea with merged satellite altimetry observations. *Acta Oceanol. Sin.* **2017**, *36*, 38–50. [CrossRef]
- Sun, J.; Guan, C.; Liu, B. Ocean wave diffraction in near-shore regions observed by Synthetic Aperture Radar. *Chin. J. Oceanol. Limnol.* **2006**, *24*, 48–56. [CrossRef]
- Lentz, S.J.; Churchill, J.H.; Davis, K.A.; Farrar, J.T. Surface gravity wave transformation across a platform coral reef in the Red Sea. *J. Geophys. Res. Oceans* **2016**, *121*, 693–705. [CrossRef]
- Sun, Z.; Zhang, H.; Xu, D.; Liu, X.; Ding, J. Assessment of wave power in the South China Sea based on 26-year high-resolution hindcast data. *Energy* **2020**, *197*, 117218. [CrossRef]
- Zieger, S.; Vinoth, J.; Young, I.R. Joint calibration of multiplatform altimeter measurements of wind speed and wave height over the past 20 years. *J. Atmos. Oceanic Technol.* **2009**, *26*, 2549–2564. [CrossRef]
- Abdalla, S.; Janssen, P.A.E.M.; Bidlot, J.R. Jason-2 OGDR Wind and Wave Products: Monitoring, Validation and Assimilation. *Mar. Geod.* **2010**, *33*, 239–255. [CrossRef]
- Albuquerque, J.; Antolínez, J.A.A.; Rueda, A.; Méndez, F.J.; Coco, G. Directional correction of modeled sea and swell wave heights using satellite altimeter data. *Ocean Model.* **2018**, *131*, 103–114. [CrossRef]
- Yang, J.; Zhang, J. Validation of Sentinel-3A/3B Satellite Altimetry Wave Heights with Buoy and Jason-3 Data. *Sensors* **2019**, *19*, 2914. [CrossRef] [PubMed]
- Zamani, A.; Azimian, A.; Heemink, A.; Solomatine, D. Wave height prediction at the Caspian Sea using a data-driven model and ensemble-based data assimilation methods. *J. Hydroinform.* **2009**, *11*, 154–164. [CrossRef]
- Peres, D.; Iuppa, C.; Cavallaro, L.; Cancelliere, A.; Foti, E. Significant wave height record extension by neural networks and reanalysis wind data. *Ocean Model.* **2015**, *94*, 128–140. [CrossRef]
- Wu, K.; Li, X.; Huang, B. Retrieval of ocean wave heights from spaceborne SAR in the Arctic Ocean with a neural network. *J. Geophys. Res. Oceans* **2021**, *162*, e2020JC016946. [CrossRef]
- Wang, J.K.; Aouf, L.; Dalphiné, A.; Zhang, Y.G.; Liu, J.Q. The wide swath significant wave height: An innovative reconstruction of significant wave heights from CFOSAT's SWIM and scatterometer using deep learning. *Geophys. Res. Lett.* **2020**, *48*, e2020GL091276. [CrossRef]

21. Lu, W.; Su, H.; Yang, X.; Yan, X.H. Subsurface temperature estimation from remote sensing data using a clustering-neural network method. *Remote Sens. Environ.* **2019**, *422*, 213–222. [CrossRef]
22. Guenaydin, K. The estimation of monthly mean significant wave heights by using artificial neural network and regression methods. *Ocean Eng.* **2008**, *35*, 1406–1415. [CrossRef]
23. Li, B.; Li, J.; Li, Y.; Zhang, Z.; Shi, P.; Liu, J.; Chen, W. Application of artificial neural network to numerical wave simulation in the coastal region of island. *J. Xiamen Univ. (Nat. Sci.)* **2020**, *59*, 420–427. (In Chinese) [CrossRef]
24. Shiokawa, Y.; Date, Y.; Kikuchi, J. Application of kernel principal component analysis and computational machine learning to exploration of metabolites strongly associated with diet. *Sci. Rep.* **2018**, *8*, 3426. [CrossRef]
25. Dombi, G.W.; Nandi, P.; Saxe, J.M.; Ledgerwood, A.M.; Lucas, C.E. Prediction of Rib Fracture Injury Outcome by an Artificial Neural Network. *J. Trauma* **1995**, *39*, 915–921. [CrossRef] [PubMed]
26. Jiang, J.L.; Su, X.; Zhang, H.; Zhang, X.H.; Yuan, Y.J. A Novel Approach to Active Compounds Identification Based on Support Vector Regression Model and Mean Impact Value. *Chem. Biol. Drug Des.* **2013**, *81*, 650–657. [CrossRef] [PubMed]
27. Yao, Y.; Yang, X.; Lai, S.; Chin, R.J. Predicting tsunami like solitary wave run up over fringing reefs using the multi layer perceptron neural network. *Nat. Hazards* **2021**, *107*, 601–616. [CrossRef]
28. Li, X.; Liu, B.; Zheng, G.; Ren, Y.; Zhang, S.; Liu, Y.; Gao, L.; Zhang, B.; Wang, F. Deep learning-based information mining from ocean remote sensing imagery. *Natl. Sci. Rev.* **2020**, *7*, 1584–1605. [CrossRef]
29. Shapiro, A. Monte Carlo Sampling Methods-ScienceDirect. *Handb. Oper. Res. Manag. Sci.* **2003**, *10*, 353–425.
30. Wang, J.; Aouf, L.; Badulin, S. Retrieval of wave period from altimetry: Deep learning accounting for random wave field dynamics. *Remote Sens. Environ.* **2021**, *265*, 112629. [CrossRef]
31. Ye, H.; Li, J.; Li, B.; Liu, J.; Tang, D.; Chen, W.; Yang, H.; Zhou, F.; Zhang, R.; Wang, S.; et al. Evaluation of CFOSAT Scatterometer Wind Data in Global Oceans. *Remote Sens.* **2021**, *13*, 1926. [CrossRef]
32. Chen, G.; Lin, H. Technical note: Impacts of collocation window on the accuracy of altimeter/buoy wind-speed comparison-a simulation study. *Int. J. Remote Sens.* **2001**, *22*, 35–44. [CrossRef]
33. Caballero, I.; Gómez-Enri, J.; Cipollini, P.; Navarro, G. Validation of High Spatial Resolution Wave Data From Envisat RA-2 Altimeter in the Gulf of Cádiz. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 371–375. [CrossRef]
34. Fett, R.W.; Kevin, M.R. Island barrier effect on sea state as revealed by a numerical wave model and DMSP satellite data. *J. Phys. Oceanogr.* **1976**, *6*, 324–334. [CrossRef]
35. Andréfouët, S.; Ardhuin, F.; Queffelec, P.; Gendre, R.L. Island shadow effects and the wave climate of the Western Tuamotu Archipelago (French Polynesia) inferred from altimetry and numerical model data. *Mar. Pollut. Bull.* **2012**, *65*, 415–424. [CrossRef] [PubMed]
36. Pawka, S.S. Island shadows in wave directional spectra. *J. Geophys. Res.* **1983**, *88*, 2579–2591. [CrossRef]
37. Ponce de León, S.; Soares, C.G. On the sheltering effect of islands in ocean wave models. *J. Geophys. Res. Ocean.* **2005**, *110*, C09020. [CrossRef]
38. Wang, J.K.; Aouf, L.; Dalphin, A.; Li, B.X.; Xu, Y.; Liu, J.Q. Acquisition of the significant wave height from CFOSAT SWIM spectra through a deep neural network and its impact on wave model assimilation. *J. Geophys. Res. Ocean.* **2021**, *126*, e2020JC016885. [CrossRef]
39. Copernicus Climate Change Service (C3S). ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate. Copernicus Climate Change Service Climate Data Store (CDS). 2017. Available online: <https://cds.climate.copernicus.eu/cdsapp#!/home> (accessed on 31 December 2021).
40. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control. Signals Syst.* **1989**, *2*, 301–314. [CrossRef]
41. Deo, M.C.; Naidu, C.S. Real time forecasting using Neural Networks. *Ocean Eng.* **1998**, *26*, 191–203. [CrossRef]
42. Zhang, Z.; Li, C.W.; Qi, Y.; Li, Y.S. Incorporation of artificial neural networks and data assimilation techniques into a third-generation wind-wave model for wave forecasting. *J. Hydroinform.* **2006**, *8*, 65–76. [CrossRef]
43. Londhe, S.N.; Shah, S.; Dixit, P.R.; Nair, T.M.B.; Sirisha, P.; Jain, R. A coupled numerical and artificial neural network model for improving location specific wave forecast. *Appl. Ocean Res.* **2016**, *59*, 483–491. [CrossRef]
44. Beale, M.; Hagan, M.T.; Demuth, H.B. Neural Network Toolbox 7—User’s Guide, 951 pp. *MathWorks Natick Mass.* **2010**, *1*, 77–81.
45. Hagan, M.T.; Menhaj, M.B. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* **1994**, *5*, 989–993. [CrossRef] [PubMed]
46. Scholkopf, B.; Smola, A.; Muller, K.R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **1998**, *10*, 1299–1319. [CrossRef]