MDPI

Article

Machine Learning Applied to a Dual-Polarized Sentinel-1 Image for Wind Retrieval of Tropical Cyclones

Yuyi Hu [1], Weizeng Shao [1,*], Wei Shen [2], Yuhang Zhou [1] and Xingwei Jiang [3]

[1] College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China; d210200046@st.shou.edu.cn (Y.H.); shenwei@wti.ac.cn (Y.Z.)
[2] China Waterborne Transport Research Institute, Ministry of Transport, Beijing 100088, China; m220200672@st.shou.edu.cn
[3] National Satellite Ocean Application Service, Ministry of Natural Resources, Beijing 100081, China; xwjiang@mail.nsoas.org.cn
* Correspondence: wzshao@shou.edu.cn; Tel.: +86-21-61900326

Abstract: In this work, three types of machine learning algorithms are applied for synthetic aperture radar (SAR) wind retrieval in tropical cyclones (TCs), and the optimal method is confirmed. In total, 30 Sentinel-1 (S-1) images in dual-polarization (vertical–vertical [VV] and vertical–horizontal [VH] were collected during the period from 2016 to 2021, which were acquired in interferometric-wide and extra-wide modes with pixels of 10 m and 40 m, respectively. More than 100,000 sub-scenes with a spatial coverage of 3 km are extracted from these images. The dependences of variables estimated from sub-scenes, i.e., VV-polarized and VH-polarized normalized radar cross-section (NRCS), as well as the azimuthal wave cutoff wavelength, on wind speeds from the stepped-frequency microwave radiometer (SFMR) and the soil moisture active passive (SMAP) radiometer are studied, showing the linear relations between wind speed and these three parameters; however, the saturation of VV-polarized NRCS and the azimuthal wave cutoff wavelength is observed. This is the foundation of selecting input variables in machine learning algorithms. Two-thirds of the collocated dataset (20 images) are used for training the process using three machine learning algorithms, i.e., eXtreme Gradient Boosting (XGBoost), Multi-layer Perceptron, and K-Nearest Neighbor, and the coefficients are fitted after training completion through 20 images collocated with SFMR and SMAP data. Another 10 images are taken for validation up to 70 m/s, yielding a 2.53 m/s root mean square error (RMSE) with a 0.96 correlation and 0.12 scatter index (SI) using XGBoost. The result is better than the >5 m/s error achieved using the existing cross-polarized geophysical model function and the other two machine learning algorithms; moreover, the comparison between wind retrievals using XGBoost and Level-2 CyclObs products shows about 4 m/s RMSE and 0.18 SI. This suggests that the machine learning algorithm XGBoost is an effective method for inverting the TC wind field utilizing SAR measurements in dual-polarization.

1. Introduction

Tropical cyclones (TCs) associated with extreme waves and heavy rainfall are huge ocean disasters that cannot be ignored in coastal waters; moreover, TCs are an important way of heat exchange in mid-to-high latitude regions [1]. The on-scene observations in TCs are difficult due to the danger and extreme sea states. Instead, numeric modeling [2] is popularly employed for the hindcasting research on TCs, i.e., their structure, track, intensity, and strong wind-generated waves [3,4]. In addition, spaceborne satellites carrying optical and microwave sensors can be used for TC monitoring in near real-time, i.e., scatterometers [5], polarimetric microwave radiometers [6], and altimeters [7]; however, these remote-sensed products do not meet the requirements of research on the atmospheric and marine dynamics at the submeso- and meso-scales.

Synthetic aperture radar (SAR) has the capability of sea surface dynamics monitoring and target detection in all weather conditions, i.e., winds [8,9], waves [10,11], oceanic fronts [12], oil spills [13] and ships [14]. Although basic TC parameters, i.e., the geographic location of its eye, morphology [15], and intensity [16], are easily obtained on co-polarized (vertical–vertical [VV] and horizontal–horizontal [HH]) SAR, the detailed structure of a wind profile is difficult to retrieve from co-polarized SAR due to the saturation problem in strong winds [17,18]. It is revealed that the SAR backscattering signal in cross-polarization (vertical–horizontal [VH] and horizontal–vertical [HV]) does not easily encounter saturation with increasing wind speed [19,20]; therefore, TC wind information is usually retrieved from the cross-polarized SAR image. Optionally, the TC wind field is also retrieved using the SAR-measured wave cutoff wavelength in the azimuth/flight direction [21,22] or the SAR-derived wave parameters [23].

A geophysical model function (GMF) was developed for SAR wind retrieval [24,25], which relates the normalized radar cross section (NRCS) with a wind vector and incidence angle. As for cross-polarized SAR, it is found that the dependence of the wind direction on NRCS is quite weak; therefore, the cross-polarized GMF is simplified to the function of wind speed at a fixed incidence angle [26]. Several GMFs in cross-polarization are adopted for C-band SARs, i.e., RADARSAT-2 (R-2) [27], Sentinel-1 (S-1) [28], and Gaofen-3 (GF-3) [29]. These GMFs are able to be applied for SAR wind retrieval under extreme weather conditions up to 70 m/s; however, the accuracy (i.e., ~5 m/s error) [30] is worse than the 2 m/s root mean square error (RMSE) of wind speed achieved using the co-polarized GMF (i.e., latest version CMOD7 of the CMOD family) [31] under regular atmospheric conditions [32]. In order to enlarge the spatial coverage, SAR typically uses a wide scanning mode to monitor a TC, i.e., interferometric-wide (IW) and extra-wide (EW) for S-1. Specifically, several sub-swaths are paired into an IW and EW image. Due to different noise-equivalent sigma zero (NESZ), the discontinuity of the wind retrieval at the edge of sub-swaths is the main issue and this problem is quite serious for cross-polarized SAR with weak sea-surface backscattering roughness. Recently, a few studies were conducted for TC wind retrieval using an empirical algorithm combining the co- and cross-polarized SAR measurements [33,34]; however, this advanced algorithm needs a priori information from external information on wind, such as from the European Centre for Medium-Range Weather Forecasts (ECMWF).

It is recognized that TC wind-retrieval algorithms have two advantages: comparable accuracy and independence of prior information, which is difficult for conventional algorithms. Machine learning refers to the general term for algorithms that identify patterns from data and use them for simulation, classification, and clustering. It is commonly recognized that machine learning could effectively explore the inherent correlations between variables and then make accurate predictions. With a large number of satellite products, machine learning is possibly aiming at research on satellite oceanography [35], which has been applied to the development of SAR wave-retrieval algorithms [36,37]. Since the S-1 mission started in 2016, a campaign for TC observation [38] has been prompted during the annual hurricane season. Utilizing abundant dual-polarized images taken of TCs, SAR wind retrieval using machine learning is anticipated to improve, especially for reducing the discontinuity of the retrieval at the edge of sub-swaths, which can be directly implemented without external data.

The purpose of this work is to study the applicability of three machine learning methods for TC wind speed retrieval from the SAR images without external data. The remaining part of our work is organized as follows. Section 2 gives the descriptions of S-1 images collocated with auxiliary data, including along-track observations from the stepped-frequency microwave radiometer (SFMR) and the product from the soil moisture active passive (SMAP) radiometer. The methodology of the three machine learning algorithms and the fitted results of the training process—using 20 images collocated with SFMR and SMAP data—are introduced in Section 3. The derivation of the TC wind-retrieval algorithm using machine learning algorithms and validation of the wind retrievals against SFMR

observations and SMAP products are exhibited in Section 4. Finally, Section 5 summarizes the conclusions.

2. Datasets

During the Satellite Hurricane Observation Campaign (SHOC), a total of 30 S-1 images acquired in IW- and EW-mode were collected in 2016–2022. These images are processed as dual-polarized (VV and VH) for China's and U.S. coastal waters. The TCs' tracks and maximum wind speeds—from the National Hurricane Center (NHC) of the National Oceanic and Atmospheric Administration (NOAA)—overlaid on the geographic coverage of collected images and represented by black rectangles is presented in Figure 1, i.e., (a) at China's coastal waters and (b) the U.S. coastal waters. In particular, SFMR observations from NOAA hurricane aircraft are available for 20 images, which are valuable sources for the development of SAR wind- [39,40] and rain- [41] retrieval algorithms in TCs. As an example, Figure 2a,b illustrate the VV-and VH-polarized NRCS map of an IW image over TC Hermine on 1 September 2016 at 23:45 UTC, in which the track of the SFMR is highlighted by a red line. Similarly, the calibration maps of an EW image over TC Michael on 10 September 2018 at 23:44 UTC are exhibited in Figure 3.
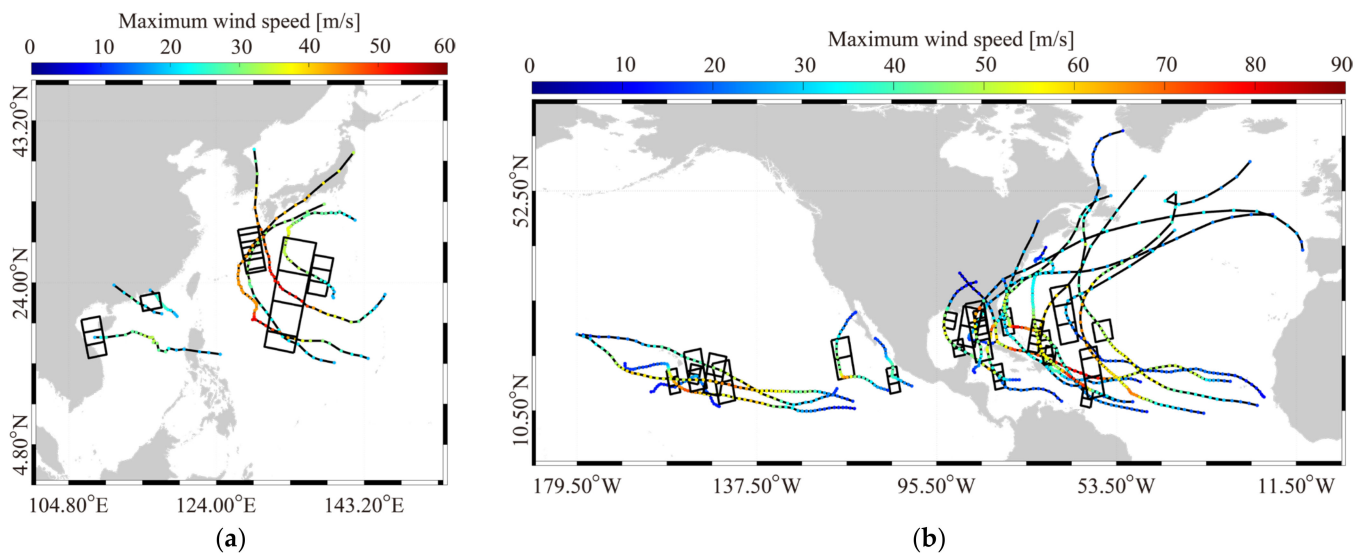

Figure 1. Tropical cyclone (TC) tracks and maximum wind speeds from National Hurricane Center (NHC) of National Oceanic and Atmospheric Administration (NOAA) overlaying the geographic coverage of collected Sentinel-1 (S-1) synthetic aperture radar (SAR) images, represented by black rectangles. The images are located at (**a**) China's coastal waters and (**b**) U.S. coastal waters.

The passive L-band radiometer SMAP operating at a sun-synchronous orbit is a National Aeronautics and Space Administration (NASA) mission, active since 2015. Because severe winds can be retrieved from the SMAP-measured brightness temperature, SMAP wind products are beneficial for the analysis of TC intensity and size [42]. Validating SMAP products against SFMR observations for 20 TCs in 2015 and 2016, the standard deviation is ~3 m/s at wind speed greater than 25 m/s [43]. Figure 4a,b show the 0.5° gridded SMAP wind maps over TC Hermine on 1 September 2016 at 23:00 UTC and over TC Michael on 10 September 2018 at 23:00 UTC, respectively.
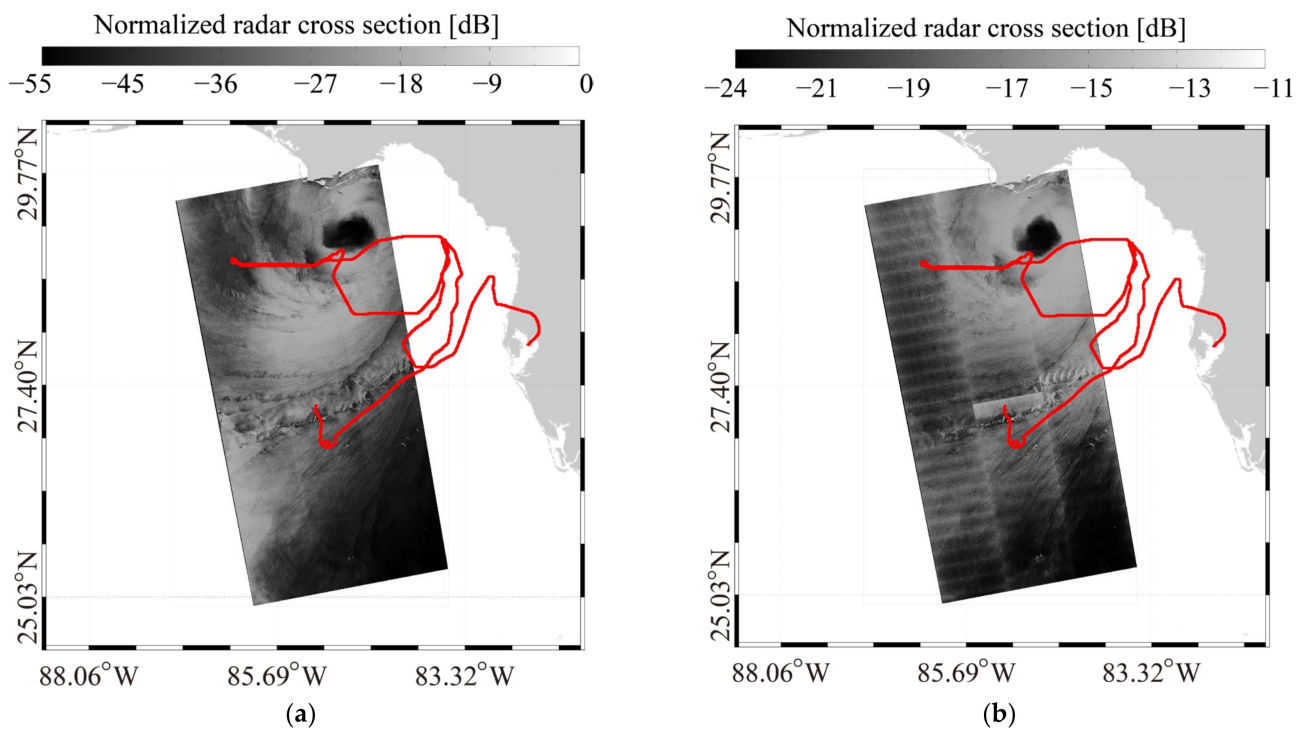
Figure 2. The normalized radar cross-section (NRCS) maps of an S-1 interferometric wide (IW) image over TC Hermine on 1 September 2016 at 23:45 UTC: (**a**) vertical–vertical (VV) and (**b**) vertical–horizontal (VH) polarization. The available measurement from stepped-frequency microwave radiometer (SFMR) is highlighted by red line.
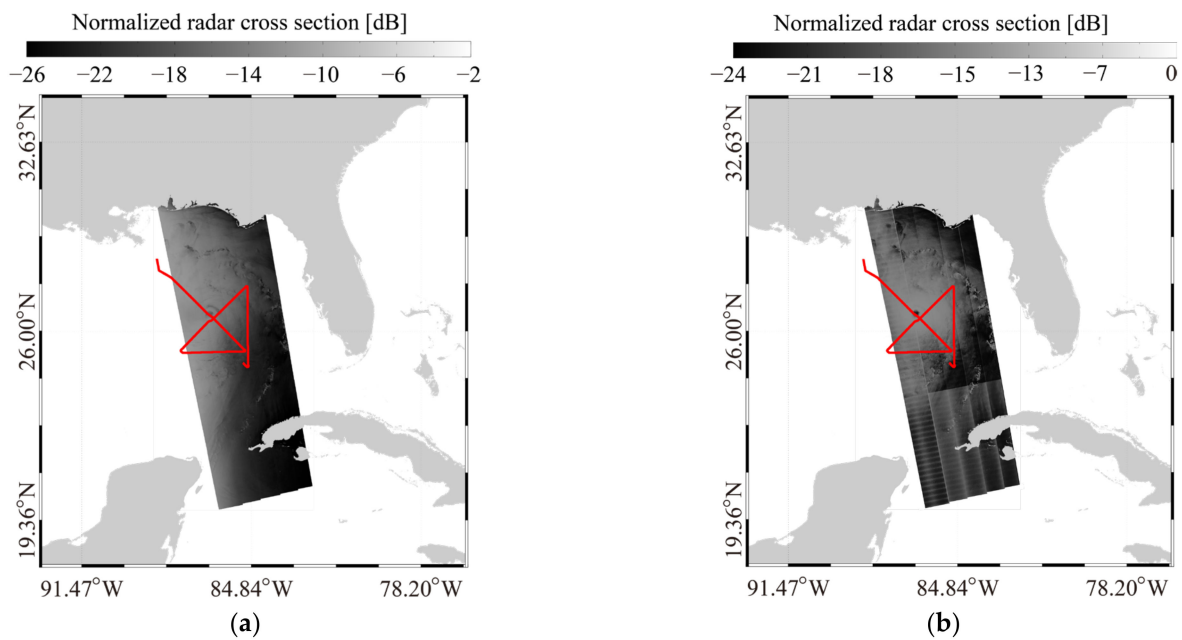
Figure 3. The NRCS maps of an S-1 extra-wide (EW) image over TC Michael on 10 September 2018 at 23:44 UTC: (**a**) VV and (**b**) VH polarization. The red lines represent the tracks of SFMR.
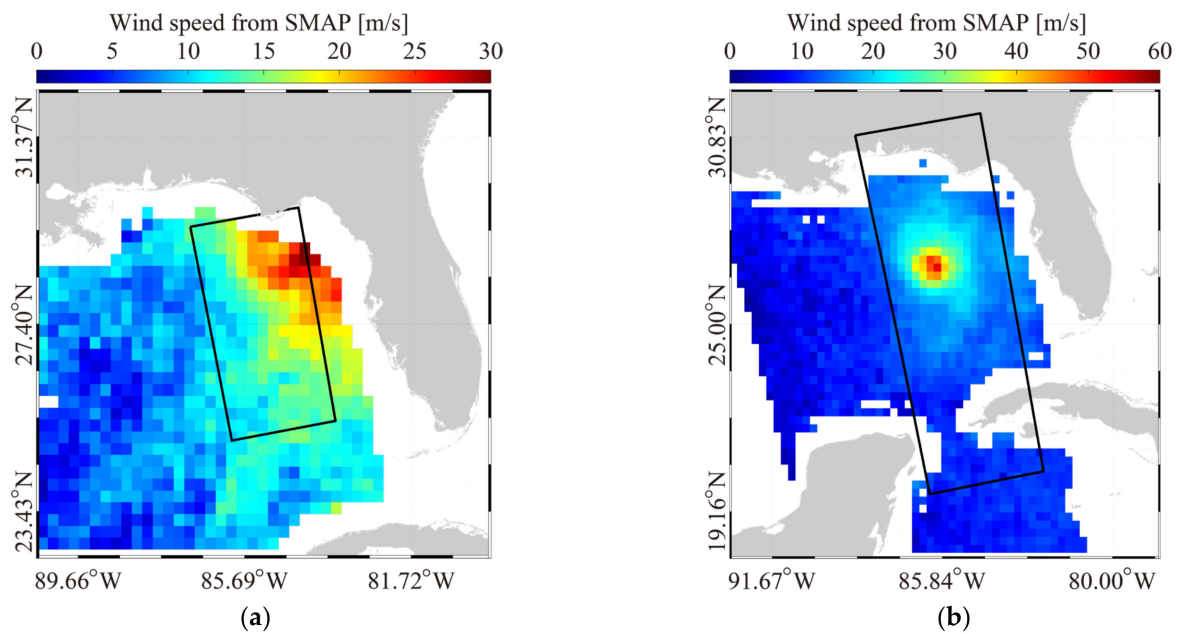
Figure 4. The wind speed maps from soil moisture active passive (SMAP) radiometer over (**a**) TC Hermine on 1 September 2016 at 23:00 UTC, and (**b**) TC Michael on 10 September 2018 at 23:00 UTC. The black rectangles represent the spatial coverages of S-1 IW and EW images in Figures 2 and 3, respectively.

3. Machine Learning Algorithms

In this section, three types of machine learning algorithms, i.e., eXtreme Gradient Boosting (XGBoost), Multi-layer Perceptron (MLP), and K-Nearest Neighbor (KNN), are briefly introduced. The performance of the training processes, using two-thirds of the collocated dataset, is also presented.

3.1. XGBoost

The eXtreme Gradient Boosting (XGBoost) is an efficient ensemble learning algorithm, which is an open-source gradient lifting framework for machine learning with high efficiency, speed, and precision. The features of XGBoost include a model generated through gradient lifting, the usage of decision trees as weak classifiers, and a final strong classifier through continuous iteration, so as to improve the accuracy of each decision tree. Compared with traditional GBDT, XGBoost utilizes the second-order gradient information and L2 regularization to prevent overfitting and adopts some strategies to speed up the computation.

In principle, XGBoost is a boosting model, that is, each iteration optimizes the submodels in the current step only. It mainly consists of two parts: gradient lifting and regularization. Gradient lifting is an iterative method in which each iterative step adds a weak learner to the model by minimizing the loss function, and regularization is a technique for avoiding over-fitting. The main objective function of XGBoost is expressed:

$$\text{Obj} = \sum_{i=1}^{N} \left[g_i f_m(x_i) + \frac{1}{2} h_i f_m^2(x_i) \right] + \Omega(f_m) \tag{1}$$

where

$$g_i = \frac{\partial L}{\partial F_{m-1}(x_i)} \tag{2}$$

$$h_i = \frac{\partial^2 L}{\partial^2 F_{m-1}(x_i)} \tag{3}$$

in which $f_m(x_i)$ is the sub-model of the current step; and $F_{m-1}(x_i)$ is the first $m-1$ sub-model in the training process. The regular term $\Omega(f_m)$ represents the complexity of the sub-model f that is used to control overfitting, and L stands for loss function. As $F_{m-1}(x)$ is determined, g_i and h_i can be easily calculated for each sample point i.

To prevent overfitting, XGBoost sets the tree-based complexity as the regular term:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2 \tag{4}$$

in which T is the number of leaf nodes in tree f; ω is the vector formed by the output regression values of all leaf nodes; $\|\omega\|^2$ is the square of the L2 norm (modulus length) of this vector; and both γ and λ are hyperparameters. As a regression tree, the more leaf nodes and the larger the output regression value, the higher the complexity of the tree.

Overall, XGBoost is a powerful technique with good scalability, generality, and interpretability. Although it requires some domain knowledge and algorithmic understanding to optimize the model, it is still one of the most popular algorithms in the field of machine learning. Figure 5 shows the scheme of the XGBoost algorithm.
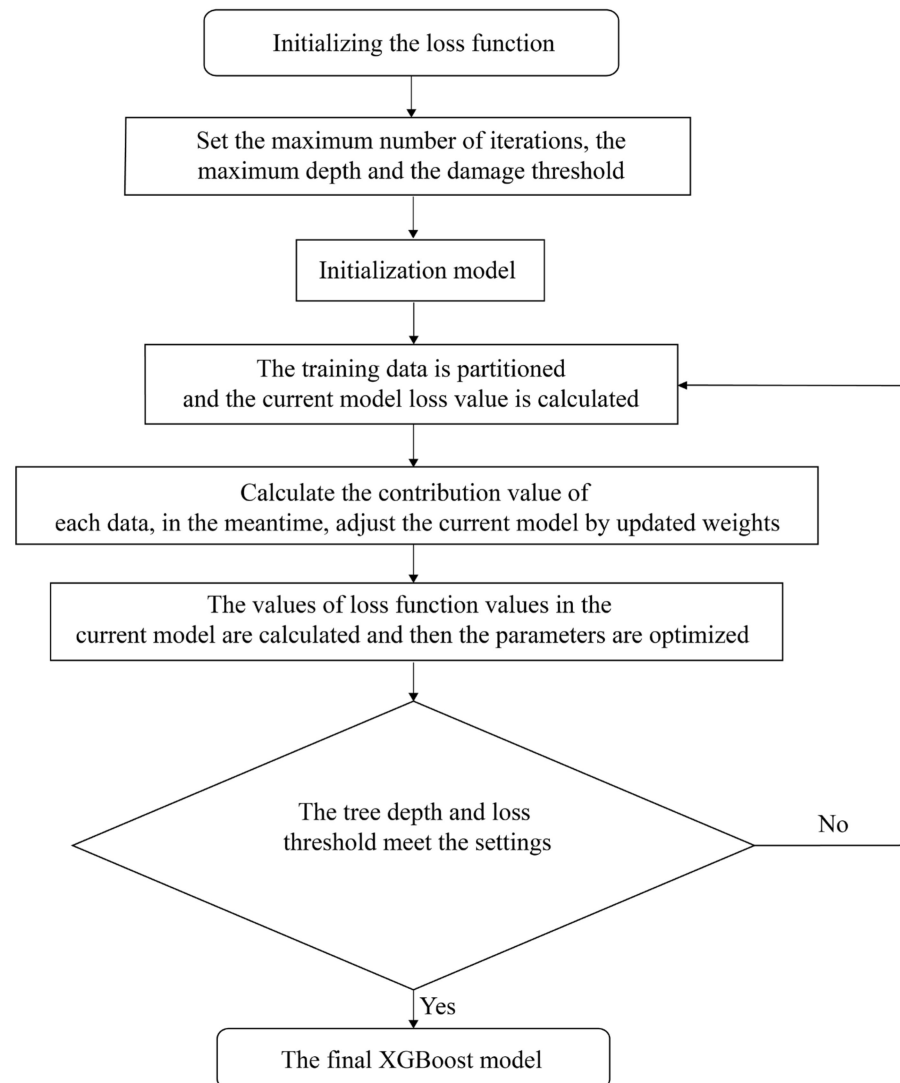
Figure 5. The scheme of eXtreme Gradient Boosting (XGBoost).

3.2. MLP

MLP is a feedforward neural network model that is widely used in various machine learning and deep learning tasks. In the literature, the MLP model is designed based on the logic structure between neurons, which is an adaptive nonlinear system that can automatically learn the features of input data and establish a mapping relationship between input and output.

In the MLP algorithm, each neural unit receives multiple inputs and a biased term and uses a nonlinear activation function to compute the output. The entire network consists of a series of computation layers containing multiple neurons. The input layer is responsible for receiving the input data and the output layer is responsible for the prediction results of the model. The middle layer, which is also known as the hidden layer, is used to synthesize the features of the input data enabling higher-level feature extraction. The weight parameters of the whole model necessitate it to be optimized through a backpropagation algorithm to minimize the error between the predicted results and the true values of the model.
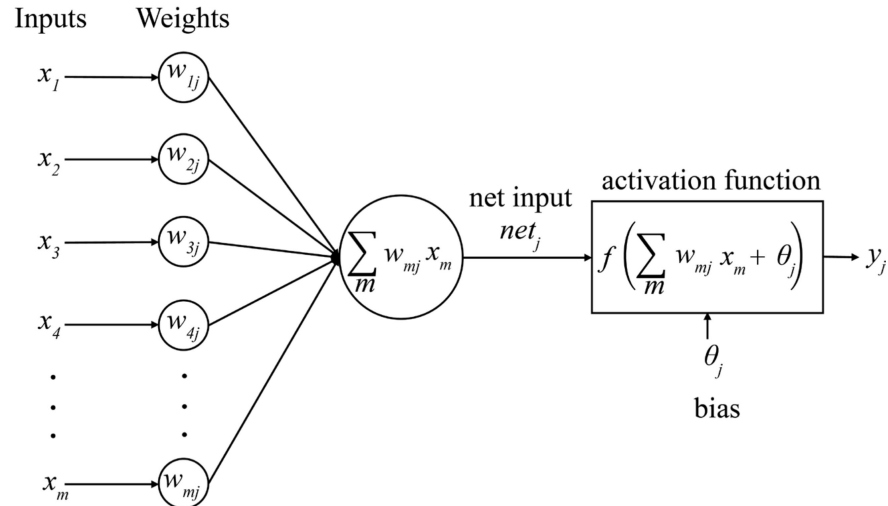
The basic network structure is shown in Figure 6. The training data are divided into different batches and a set of weights w and bias θ are defined using these batch cycle training models; the activation function f used in this model is related to the neurons and final outputs, stated as follows:

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}, \tag{5}$$

where

$$z = \sum_m w_{mj} + \theta_j, \tag{6}$$

in which w_{mj} is the m-th input variable in the neural network and its corresponding j-th weight θ_j is the bias corresponding to the j-th neuron.

Figure 6. The basic network structure of Multi-layer Perceptron (MLP).

3.3. KNN

KNN is a parameterless machine learning algorithm, which relies on comparing a new data point to all the data points in the training dataset and then predicting the label or value of the new data point based on the k-most-similar adjacent data points. In fact, KNN is popularly used in classification problems and regression problems. In the regression of the KNN scheme, given a training sample set and a new data point, the algorithm will predict the numerical result of the new data point by calculating the value of the nearest k points around the sample points, which, using the average or weighted averages, predicts the numerical result of the new sample.

KNN mainly consists of the following steps:

1. Determine the number of neighbors, *k*, and the new data points;
2. For a new data point, the distance between it and every data point in the training set is calculated. It is necessary to figure out the distance measurement method used in the KNN algorithm, which depends on the type of data and the relationship between the data. Typically, Euclidean distance is used to measure the distance *d* between two data points. The formulation is described as follows:

$$d = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2}, \tag{7}$$

3. Select the *k* data point closest to the logarithmic data point. The adjustment of *k* has an important impact on the prediction result of the model. When the value of *k* is small, the model becomes more complex and can fit the training data well but it may overfit; in contrast, if the value of *k* is large, the model becomes simpler;
4. Predict the output of the new data points based on the category (or value) of the nearest neighbor.

The flowchart of the KNN is illustrated in Figure 7.
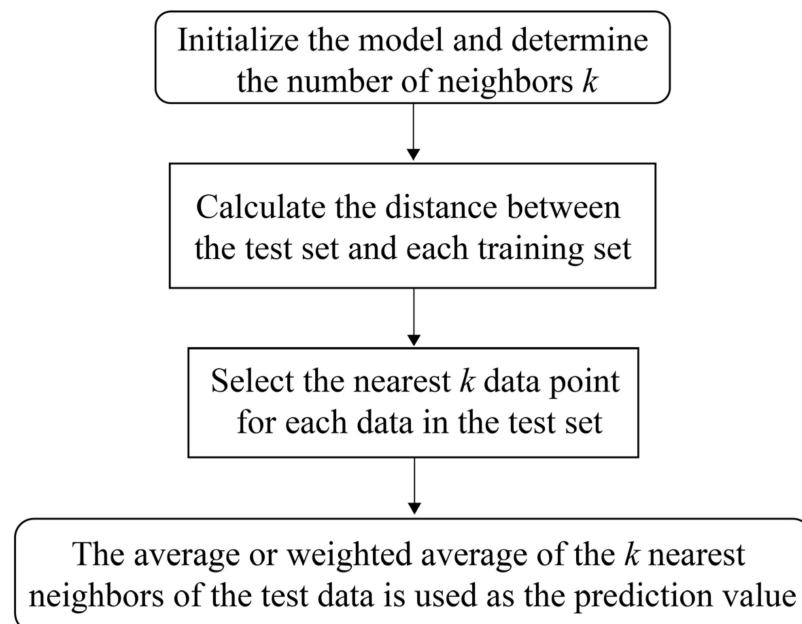
Figure 7. The flowchart of K-Nearest Neighbor (KNN).

3.4. Advanced Wind Dataset

At present, the French Research Institute for Exploitation of the Oceans (IFREMER) team provides an advanced SAR algorithm to retrieve winds in TCs, which combines the measurements at VV- and VH-polarization channels with a priori information [33,44], such as the ECWMF (its spatial resolution is 0.125° with a time step of 3 h). In this method, the prior information on a wind vector is used and then the retrieval wind is obtained by minimizing the cost function *J*, which is generated using SAR-measured NRCS in VH and VV polarization:

$$J(u,v) = \sum_{\text{PP} \in [\text{VV,VH}]} \left[\frac{\sigma_0^{\text{PP}} - \text{GMF}^{\text{PP}}(\theta, \phi, U_{10})}{\Delta\sigma_0^{\text{PP}}}\right] + \left[\frac{u^{\text{priori}} - u}{\Delta u}\right]^2 + \left[\frac{v^{\text{priori}} - v}{\Delta v}\right] \tag{8}$$

in which (*u*,*v*) represent the space of solution in the geographical referential from −80 to 80 m/s; u^{priori} and v^{priori} are the a priori information given by the ECMWF model; σ_0^{PP} and $\Delta\sigma_0^{\text{PP}}$ represent the NRCS measurements in co- (pp = VV) and cross- (pp = VH) polarization and their associated errors; $\text{GMF}^{\text{PP}}(\theta, \phi, U_{10})$ stands for the GMF defined

for each polarization with respect to the incidence angle θ, the wind direction relative to the azimuth direction ϕ, and wind speed U_{10}. The TC wind product denoted as Level-2 CyclObs is also used in our work.

4. Results

In this section, the TC wind-retrieval algorithm based on three machine learning algorithms is adopted through 20 images. Then, the validation of retrieved wind speeds from 10 images is exhibited.

4.1. TC Wind-Retrieval Algorithm

The key factor for applying machine learning is searching for the variables as the input in the training process. According to backscattering theory, NRCS is determined by wind and this is the foundation of the development of the SAR wind-retrieval algorithm [45]. Under cyclonic conditions, the wave parameters are exponentially related to wind considering the fetch and duration-limited feature inside a TC [46]. on the other hand, the azimuthal cutoff wavelength caused by velocity bunching is determined by the growth of the wave. In this sense, TC wind has a positive relationship with the azimuthal cutoff wavelength [47].

A few sub-scenes extracted from the S-1 image have 256×256 pixels and 128×128 pixels with a spatial coverage of 3 km for the EW and IW images, respectively. Because the S-1 IW and EW images are composited by several sub-swathes, the pretreatment for denoise is necessary; therefore, the sub-scenes are smoothed using a 3×3 Gaussian filter. The azimuthal cutoff wavelength λ_c is conveniently calculated by minimizing the standard error of a Gaussian function $G(k_x)$ with respect to wave number k_x in the azimuthal direction, expressed:

$$G(k_x) = \exp\left\{-\pi\left(\frac{k_x}{k_c}\right)^2\right\},\qquad(9)$$

where

$$\lambda_c = \frac{2\pi}{k_c}\qquad(10)$$

Note that velocity bunching is independent of polarization; thus, the azimuthal cutoff wavelength in VV-polarization is only used here. In total, three variables are estimated from the sub-scenes, i.e., VV-polarized σ_0^{VV} and VH-polarized NRCS σ_0^{VH}, and the VV-polarized cutoff wavelength λ_c. There are more than 10,000 samples collocated with wind speed U_{10} from SFMR and SMAP, which is treated as the training dataset. Figure 8 shows the relations between U_{10} and three variables, i.e., (a) σ_0^{VV}, (b) σ_0^{VH}, and (c) λ_c. It is not surprising that the saturation of σ_0^{VV} and λ_c is observed at high winds, this is because wind speed and SWH are explicitly related, considering the fetch and duration-limited features inside a TC [48]; however, this behavior is not found in relation with respect to σ_0^{VH}.

In the training procedure, the input parameters include σ_0^{VV}, σ_0^{VH}, λ_c, and incidence angle θ and the output is U_{10}. Figure 9 shows the performance of the training process, i.e., (a) XGBoost and (b) MLP; however, the KNN has no such process. It is observed that the three machine learning algorithms will eventually converge and the RMSE achieved using XGBoost is less than 2 m/s, which is better than the 4 m/s RMSE achieved using MPL.

4.2. Validation

The inverted wind maps from the IW image over TC Maria on 21 September 2017 at 22:45 UTC using the three retrieval algorithms are shown in Figure 10, i.e., (a) XGBoost, (b) MLP, and (c) KNN. In the meantime, winds are also retrieved from the VH-polarized image using the GMF S-1 IW Mode Wind Speed Retrieval Model after Noise Removal (S1IW.NR) [40], as presented in Figure 10d. Similarly, the inverted wind maps from the IW image TC Larry 7 September 2021 at 21:48 UTC using XGBoost, MLP, KNN, and the GMF S-1 EW Mode Wind Speed Retrieval Model after Noise Removal (S1EW.NR) [39,49], are illustrated in Figure 11a–d, respectively. It is surprising that the cyclonic structure using KNN
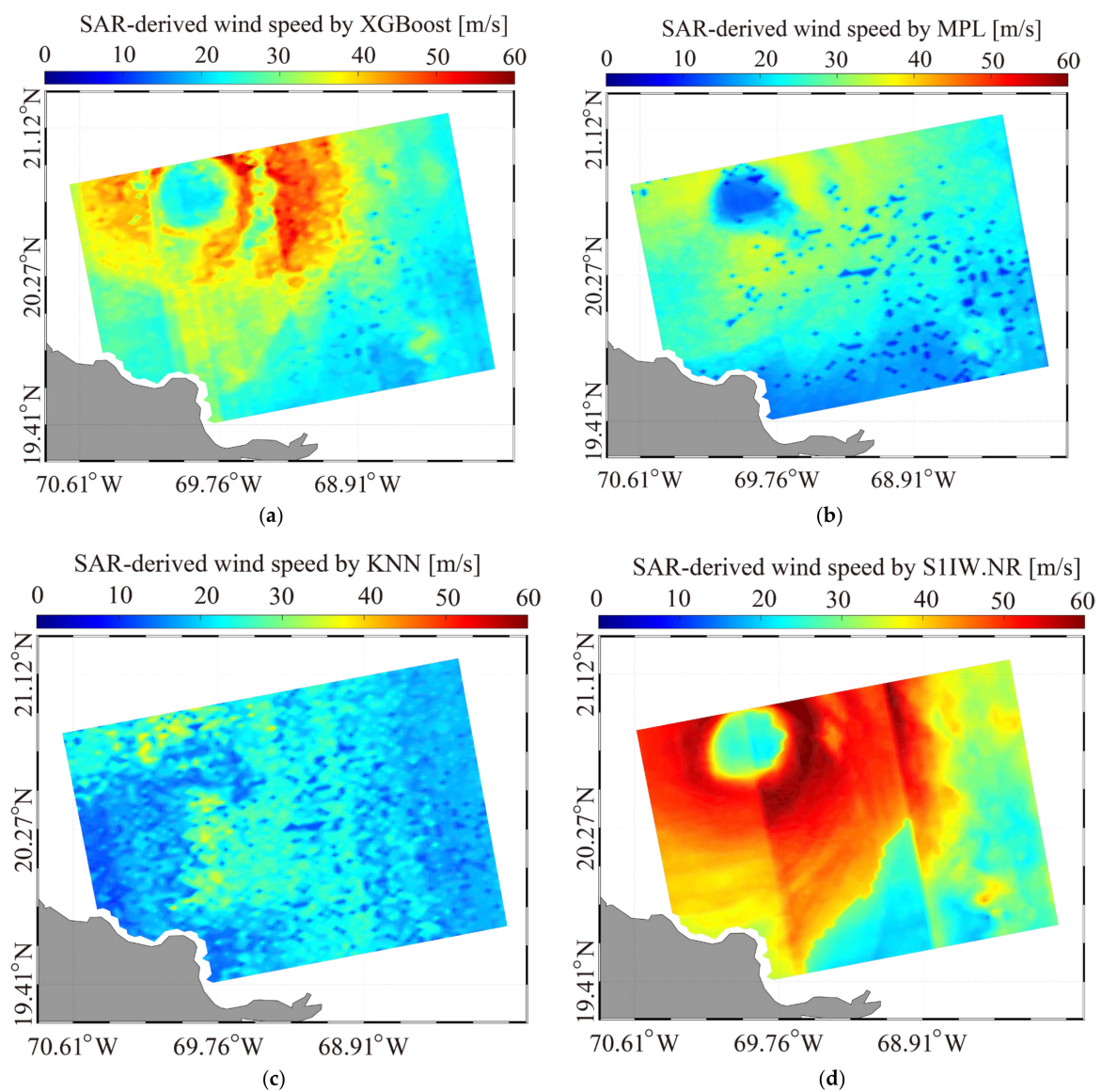
is unclear because there is no iterative regression process; furthermore, the discontinuity of retrievals at the edge of sub-swaths seriously occurs using VH-polarized GMFs.
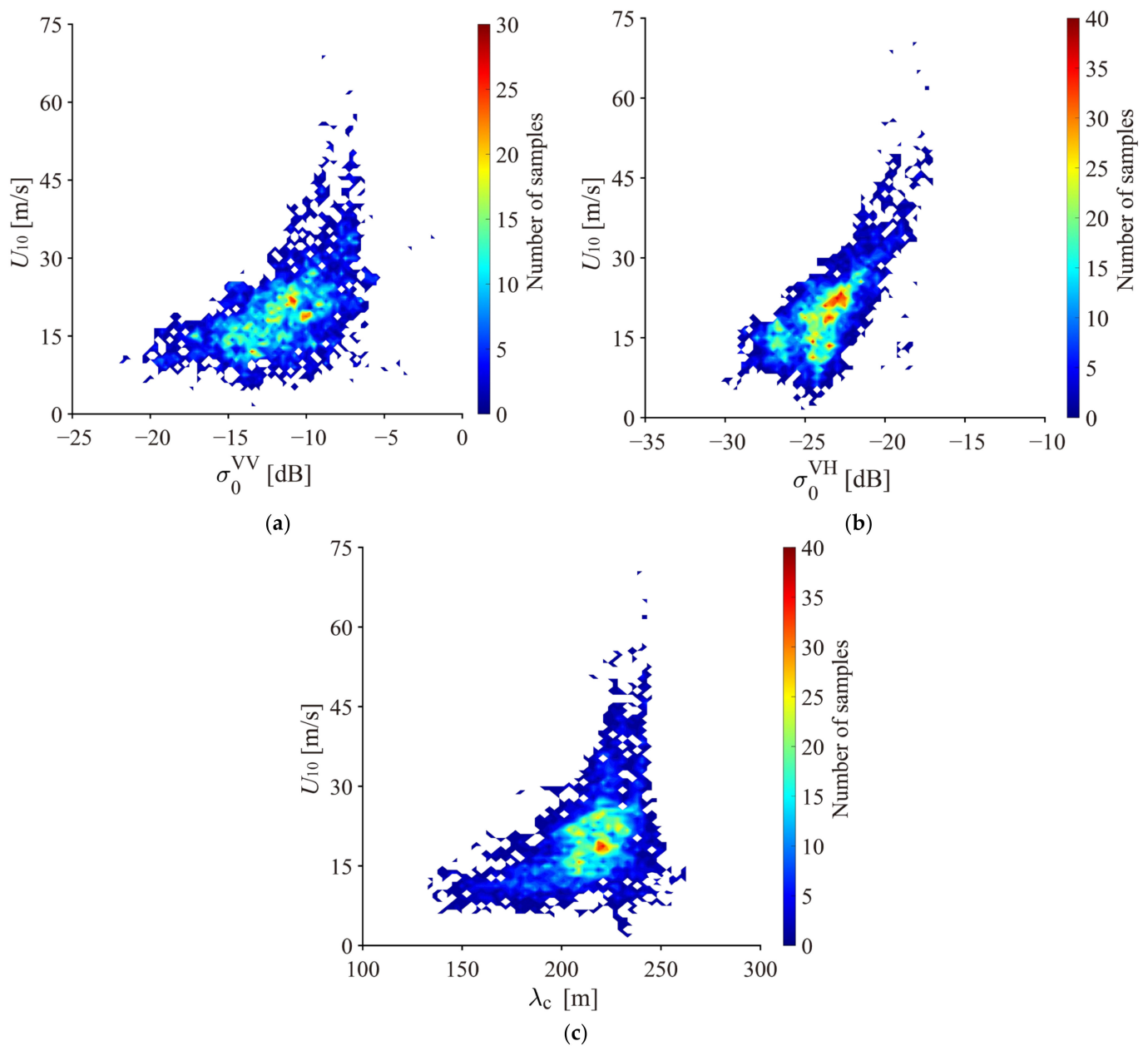

Figure 8. Wind speed versus three variables: (**a**) VV-polarized NRCS; (**b**) VH-polarized NRCS; and (**c**) VV-polarized cutoff wavelength.
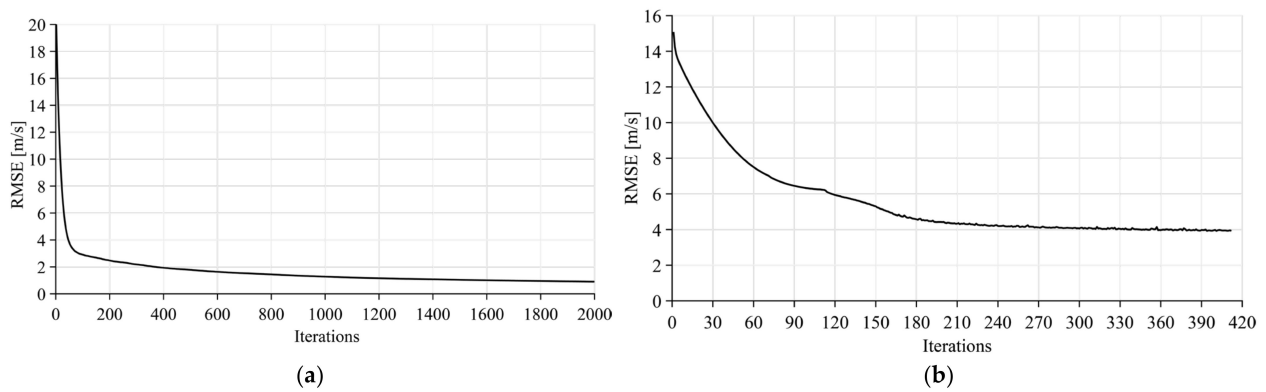
Figure 9. The performance of the training process: (**a**) XGBoost and (**b**) MLP.

Figure 10. The inverted wind maps from IW images over TC Maria on 21 September 2017 at 22:45 UTC using three machine learning algorithms, i.e., (**a**) XGBoost, (**b**) MLP, (**c**) KNN; and (**d**) the VH-polarized geophysical model function (GMF) denoted as S-1 IW Mode Wind Speed Retrieval Model after Noise Removal (S1IW.NR).
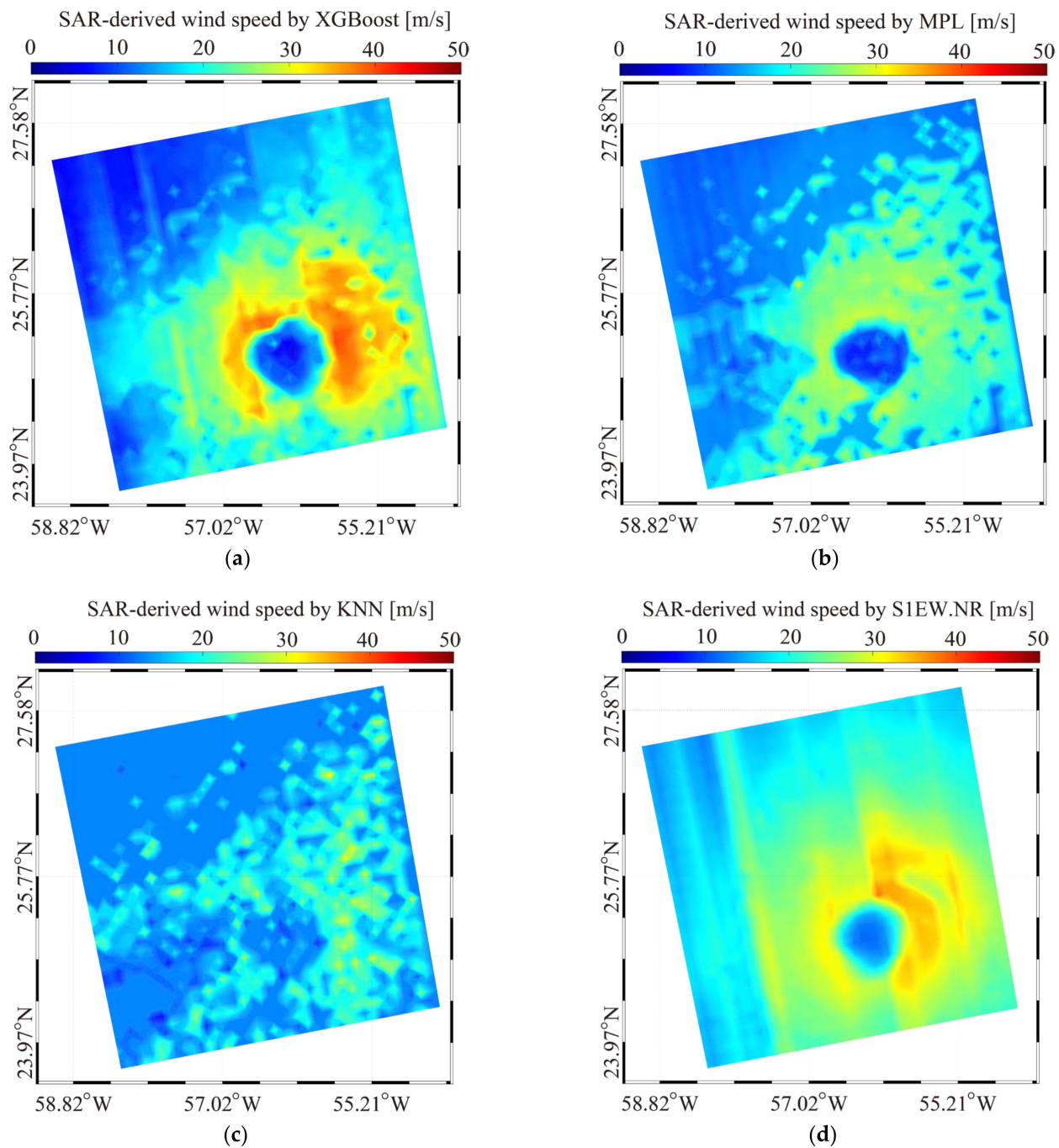
Figure 11. The inverted wind maps from EW images over TC Larry 7 September 2021 at 21:48 UTC using three machine learning algorithms, i.e., (**a**) XGBoost, (**b**) MLP, (**c**) KNN; and (**d**) the VH-polarized GMF denoted as S-1 EW Mode Wind Speed Retrieval Model after Noise Removal (S1EW.NR).

Three machine learning algorithms and cross-polarized GMFs are implemented for 10 images. The wind retrievals are validated against the SMAP and SFMR products up to 70 m/s. The following formula is used to calculate the scatter index (SI):

$$\text{SI} = \frac{1}{\overline{X}}\sqrt{\frac{\sum_{i=1}^{n}\left[\left(X_i - \overline{X}\right) - \left(Y_i - \overline{Y}\right)\right]^2}{n}} \qquad (11)$$

It is found that the RMSE of wind speed is 2.53 m/s with a 0.96 correlation (COR) and 0.12 SI using XGBoost (Figure 12a), which is better than the results using other algorithms, i.e., a 5.98 m/s RMSR with a 0.75 COR and a 0.29 SI using MPL (Figure 12b), a 6.41 m/s RMSR with a 0.69 COR and a 0.31 SI using KNN (Figure 12c), and an 8.76 m/s RMSE with a 0.57 COR and a 0.40 SI using cross-polarized GMFs (Figure 12d); therefore, it is believed that XGBoost is the optimal algorithm based on machine learning for inverting TC wind from a dual-polarized S-1 image.
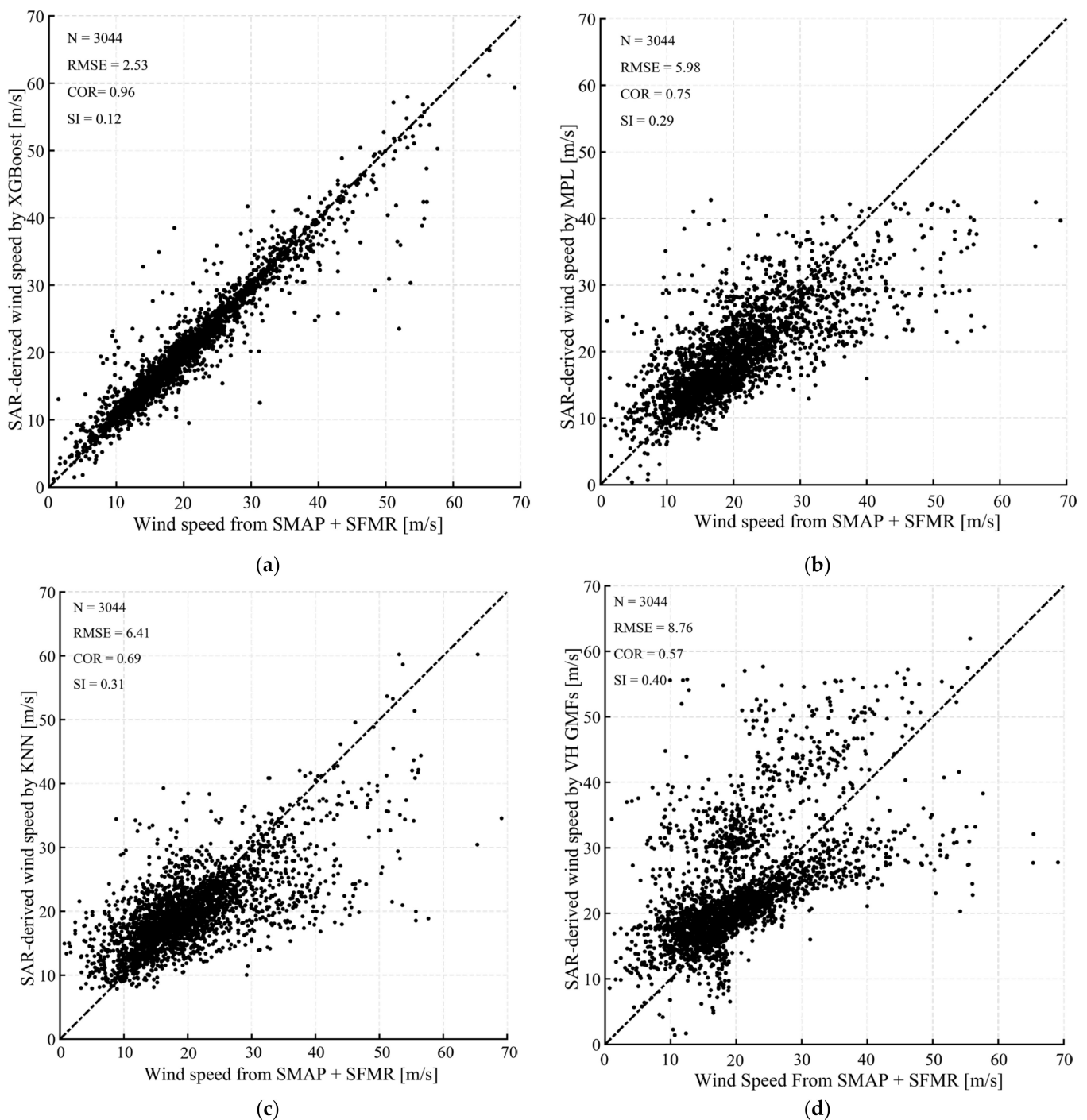
Figure 12. Comparisons between the combinations of SMAP and SFMR and SAR-derived wind speeds using four algorithms: (**a**) XGBoost, (**b**) MLP, (**c**) KNN, and (**d**) VH-polarized GMFs.

The advanced CyclObs wind product was also used to evaluate the SAR-derived wind speed using the XGBoost method. The wind map from the CyclObs wind product over TC Maria on 21 September 2017 at 22:45 UTC is shown in Figure 13a, in which the black rectangle represents the spatial coverage seen in Figure 11. It is found that the XGBoost retrieval result is similar to the CyclObs wind product, i.e., cyclonic structure and maximum wind speeds. Moreover, the CyclObs products of 10 images are validated against the collocated matchups from SMAP and SFMR, showing a 4.48 m/s RMSE with a 0.22 SI, as presented in Figure 13b. This is better than the result in Figure 12d because the algorithm generating the CyclObs products is more advanced than previous VH-polarized algorithms. Figure 14 shows the comparison between the CyclObs wind products and the SAR-derived wind speeds using the herein algorithm, in which the wind speed is grouped into a 0.07 m/s bin. The RMSE of wind speed is about 4 m/s with a 0.92 correlation and 0.18 SI. Therefore, we think the XGBoost is the optimal algorithm for inverting TC wind from a dual-polarized S-1 image without external data.
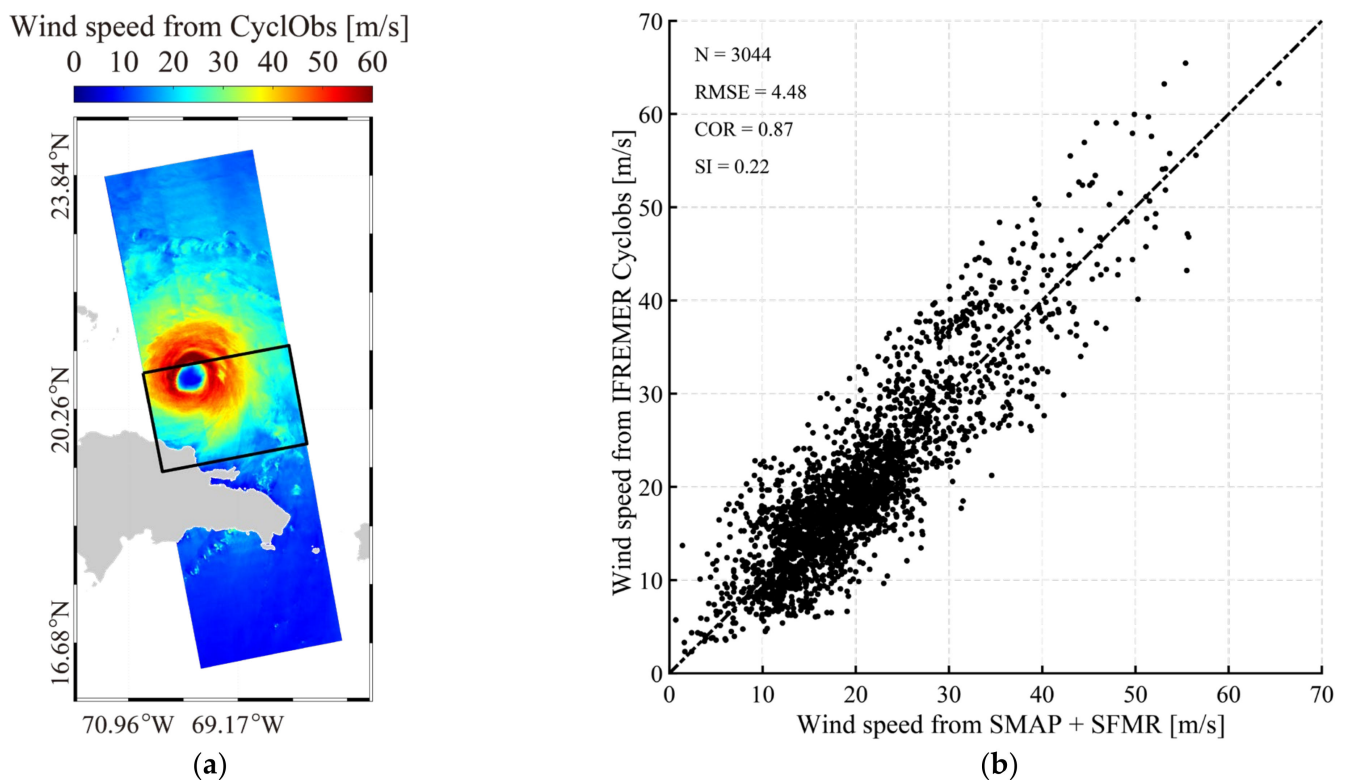
Figure 13. (**a**) The wind maps from CyclObs wind product over TC Maria on 21 September 2017 at 22:45 UTC, in which black rectangle represents the position of Figure 11; (**b**) Comparisons between the combination of SMAP and SFMR and CyclObs wind product.
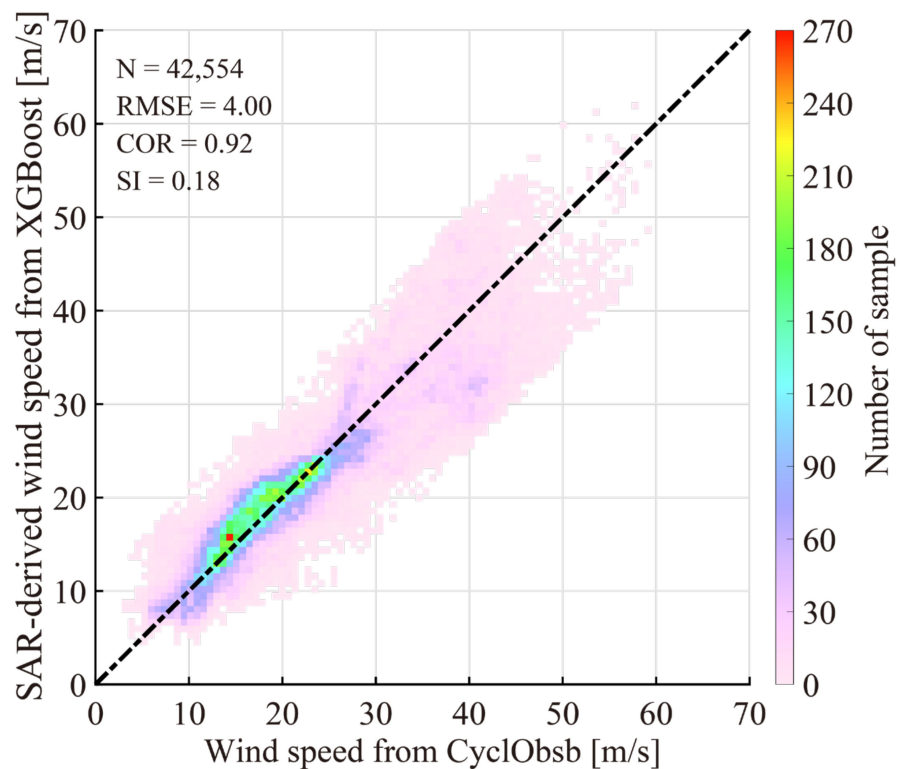
Figure 14. Validation of SAR-derived wind speed using XGBoost against the CyclObs wind product, in which the wind speed is grouped into a 0.07 m/s bin.

5. Conclusions

Accurate retrieval of TC winds and waves remains a hot topic for the SAR community. Although several GMFs aim to cross-polarize SAR wind retrieval in TCs, discontinuity in the retrieval caused by different NESZs at the edge of sub-swaths is apparently observed. In addition, the accuracy of cross-polarized SAR wind retrieval is relatively lower than that in co-polarization. With a background in artificial intelligence, in this study, the machine learning method has been applied and aimed to solve this problem through abundant data.

In total, 30 S-1 IW/EW images were collected during SHOC in the past 5-year mission of S-1, which were collocated with SMAP products. In particular, SFMR observations aboard NOAA hurricane aircraft are available for 20 images. Theoretically, SAR roughness and the azimuthal cutoff wave length are related to sea-surface wind speed. This is also demonstrated by analyzing the collocated dataset. Three algorithms based on machine learning, i.e., XGBoost, MLP, and KNN were trained through the matchups with U_{10} from SFMR and SMAP, in which σ_0^{VV}, σ_0^{VH}, λ_c, and θ are set as inputs in order to fit the coefficients. Three machine learning algorithms and VH-polarized GMFs are implemented for the other 10 images collocated with the SFMR observations and SMAP products. It is found that the accuracy of inversion using XGBoost has good performance up to 70 m/s, indicating a 2.98 m/s RMSE of wind speed with a 0.94 COR and 0.14 SI, which is significantly better than the results (> 5 m/s RMSE) from the other two algorithms and the VH-polarized GMFs. It is necessary to figure out if the cyclonic structure of the inverted wind using KNN is unclear, in which case there is no iterative regression process. Again, the discontinuity of retrievals at the edge of sub-swaths is improved using XGBoost, which is inevitably achieved using VH-polarized GMFs. Moreover, the retrieval results from XGBoost have good performance with the CyclObs wind product up to 70 m/s, indicating about 4 m/s RMSE of wind speed with a 0.92 COR and 0.18 SI. Collectively, the XGBoost is the optimal algorithm for inverting TC wind from dual-polarized S-1 images without external data.

In the near future, more SAR images at the C-band will be collected during the hurricane season [49,50], and more tropical cyclonic parameters, i.e., maximum wind speed

and the radius of maximum wind speed will be considered in the training process of machine learning methods.

Author Contributions: Conceptualization, W.S. (Weizeng Shao) and Y.H.; methodology, W.S. (Wei Shen) and Y.Z.; validation, Y.H.; formal analysis, W.S. (Weizeng Shao); investigation, W.S. (Weizeng Shao) and W.S. (Wei Shen); resources, W.S. (Weizeng Shao); writing (original draft preparation), W.S. (Wei Shen) and X.J.; writing (review and editing), W.S. (Weizeng Shao) and W.S. (Wei Shen); visualization, Y.H.; funding acquisition, W.S. (Weizeng Shao) and X.J. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Due to the nature of this study, participants did not agree to their data being shared publicly; therefore, supporting data are unavailable.

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

References

1. Liu, Y.P.; Tang, D.L.; Evgeny, M. Chlorophyll concentration response to the typhoon wind-pump induced upper ocean processes considering air-sea heat exchange. *Remote Sens.* **2019**, *11*, 1825. [CrossRef]
2. Zahir, R.K.; Quadir, D.A.; Mannan, C.M.; Ahasan, M.N.; Haque, M.S. Simulation of structure, track and landfall of tropical cyclone Bijli using WRF-ARW model. *J. Bangladesh Acad. Sci.* **2015**, *39*, 157–167.
3. Sheng, Y.X.; Shao, W.Z.; Li, S.Q.; Zhang, Y.M.; Yang, H.W.; Zuo, J.C. Evaluation of typhoon waves simulated by WaveWatch-III model in shallow waters around Zhoushan islands. *J. Ocean. Univ. China* **2019**, *18*, 365–375. [CrossRef]
4. Hu, Y.; Shao, W.Z.; Wei, Y.L.; Zuo, J.C. Analysis of typhoon-induced waves along typhoon tracks in the western North Pacific Ocean, 1998–2017. *J. Mar. Sci. Eng.* **2020**, *8*, 521. [CrossRef]
5. Shao, W.Z.; Jiang, T.; Jiang, X.W.; Zhang, Y.G.; Zhou, W. Evaluation of sea surface winds and waves retrieved from the Chinese HY-2B data. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2021**, *14*, 9624–9635. [CrossRef]
6. Gaiser, P.W.; Germain, K.M.; Twarog, E.M.; Poe, G.A.; Chang, P.S. The Windsat spaceborne polarimetric microwave radiometer: Sensor description and early orbit performance. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 2347–2361. [CrossRef]
7. Lagerloef, G.S.; Mitchum, G.T.; Lukas, R.B.; Niiler, P.P. Tropical pacific near-surface currents estimated from altimeter, wind, and drifter data. *J. Geophys. Res.* **1999**, *104*, 23313–23326. [CrossRef]
8. Shao, W.Z.; Hu, Y.Y.; Jiang, X.W.; Zhang, Y.G. Wave retrieval from quad-polarized Chinese Gaofen-3 SAR image using an improved tilt modulation transfer function. *Geo-Spat. Inf. Sci.* **2023**. [CrossRef]
9. Shao, W.Z.; Nunziata, F.; Zhang, Y.G.; Corcione, V.; Migliaccio, M. Wind speed retrieval from the Gaofen-3 synthetic aperture radar for VV and HH-polarization using a re-tuned algorithm. *Eur. J. Remote Sens.* **2021**, *54*, 1318–1337. [CrossRef]
10. Sheng, Y.X.; Shao, W.Z.; Zhu, S.; Sun, J.; Yuan, X.Z.; Li, S.Q.; Shi, J.; Zuo, J.C. Validation of significant wave height retrieval from co-polarization Chinese Gaofen-3 SAR imagery using an improved algorithm. *Acta Oceanol. Sin.* **2018**, *37*, 1–10. [CrossRef]
11. Shao, W.Z.; Jiang, X.W.; Sun, Z.F.; Hu, Y.Y.; Marino, A.; Zhang, Y.G. Evaluation of wave retrieval for Chinese Gaofen-3 synthetic aperture radar. *Geo-Spat. Inf. Sci.* **2022**, *25*, 229–243. [CrossRef]
12. Xie, T.; Perrie, W.; Chen, W. Gulf stream thermal fronts detected by synthetic aperture radar. *Geophys. Res. Lett.* **2010**, *37*, L06601. [CrossRef]
13. Jiang, T.; Shao, W.Z.; Hu, Y.Y.; Zheng, G.; Shen, W. L-band analysis of the effects of oil slicks on sea wave characteristics. *J. Ocean Univ. China* **2023**, *22*, 9–20. [CrossRef]
14. Xu, Q.; Li, Y.Z.; Li, X.F.; Zhang, Z.H.; Cao, Y.N.; Cheng, Y.C. Impact of ships and ocean fronts on coastal sea surface wind measurements from the advanced scatterometer. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2018**, *11*, 2162–2169. [CrossRef]
15. Ni, W.; Stoffelen, A.; Ren, K. Hurricane eye morphology extraction from SAR images by texture analysis. *Front. Earth Sci.* **2022**, *16*, 190–205. [CrossRef]

16. Reppucci, A.; Lehner, S.; Schulz-Stellenfleth, J.; Brusch, S. Tropical cyclone intensity estimated from wide-swath SAR images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 1639–1649. [CrossRef]

17. Hwang, P.A.; Zhang, B.; Toporkov, J.V.; Perrie, W. Comparison of composite Bragg theory and quad-polarization radar backscatter from RADARSAT-2: With applications to wave breaking and high wind retrieval. *J. Geophys. Res. Oceans* **2010**, *115*, C08019. [CrossRef]

18. Hersbach, H. Comparison of C-Band scatterometer CMOD5.N equivalent neutral winds with ECMWF. *J. Atmos. Ocean. Technol.* **2010**, *27*, 721–736. [CrossRef]

19. Vachon, P.W.; Wolfe, J. C-band cross-polarization wind speed retrieval. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 456–459. [CrossRef]

20. Zhang, B.; Perrie, W. Cross-polarized synthetic aperture radar: A new potential measurement technique for hurricanes. *Bull. Amer. Meteorol. Soc.* **2012**, *93*, 531–541. [CrossRef]

21. Grieco, G.; Lin, W.; Migliaccio, M.; Nirchio, F.; Portabella, M. Dependency of the Sentinel-1 azimuth wavelength cut-off on significant wave height and wind speed. *Int. J. Remote Sens.* **2016**, *37*, 5086–5104. [CrossRef]

22. Migliaccio, M.; Huang, L.Q.; Buono, A. SAR speckle dependence on ocean surface wind field. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5447–5455. [CrossRef]

23. Shao, W.Z.; Hu, Y.Y.; Nunziata, F.; Corcione, V.; Li, X.M. Cyclone wind retrieval based on X-band SAR-derived wave parameter estimation. *J. Atmos. Ocean. Technol.* **2020**, *37*, 1907–1924.

24. Shao, W.Z.; Zhu, S.; Sun, J.; Yuan, X.Z.; Sheng, Y.X.; Zhang, Q.J.; Ji, Q.Y. Evaluation of wind retrieval from co-polarization Gaofen-3 SAR imagery around China seas. *J. Ocean Univ. China* **2019**, *18*, 84–96. [CrossRef]

25. Yao, R.; Shao, W.Z.; Jiang, X.W.; Yu, T. Wind speed retrieval from Chinese Gaofen-3 synthetic aperture radar using an analytical approach in the nearshore waters of China's seas. *Int. J. Remote Sens.* **2022**, *43*, 3028–3048. [CrossRef]

26. Zhu, S.; Shao, W.Z.; Marino, A.; Sun, J.; Yuan, X.Z. Semi-empirical algorithm for wind speed retrieval from Gaofen-3 quad-polarization strip mode SAR data. *J. Ocean Univ. China* **2020**, *19*, 23–35. [CrossRef]

27. Zhang, G.S.; Li, X.F.; Perrie, W.; Hwang, P.; Zhang, B.; Yang, X.F. A hurricane wind speed retrieval model for C-band RADARSAT-2 cross-polarization ScanSAR images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4766–4774. [CrossRef]

28. Zhang, K.; Huang, J.; Mansaray, L.R.; Guo, Q.; Wang, X. Developing a subswath-based wind speed retrieval model for sentinel-1 VH-polarized SAR data over the ocean surface. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1561–1572. [CrossRef]

29. Shao, W.Z.; Yuan, X.Z.; Sheng, Y.X.; Sun, J.; Zhou, W.; Zhang, Q.J. Development of wind speed retrieval from cross-polarization Chinese Gaofen-3 synthetic aperture radar in typhoons. *Sensors* **2018**, *18*, 412. [CrossRef]

30. Gao, Y.; Zhang, J.; Guan, C.L.; Sun, J. Analyzing sea surface wind distribution characteristics of tropical cyclone based on sentinel-1 SAR images. *Remote Sens.* **2021**, *13*, 4501. [CrossRef]

31. Stoffelen, A.; Verspeek, J.A.; Vogelzang, J.; Verhoef, A. The CMOD7 geophysical model function for ASCAT and ERS wind retrievals. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2123–2134. [CrossRef]

32. Yang, X.F.; Li, X.F.; Zheng, Q.A.; Gu, X.F.; Pichel, W.; Li, Z.W. Comparison of ocean-surface winds retrieved from quikscat scatterometer and Radarsat-1 SAR in offshore waters of the U.S. West coast. *IEEE Trans. Geosci. Remote Sens.* **2011**, *8*, 163–167.

33. Mouche, A.A.; Chapron, B.; Zhang, B.; Husson, R. Combined co- and cross-polarized SAR measurements under extreme wind conditions. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6746–6755. [CrossRef]

34. Shao, W.Z.; Lai, Z.Z.; Nunziata, F.; Buono, A.; Jiang, X.W.; Zuo, J.C. Wind field retrieval with rain correction from dual-polarized Sentinel-1 SAR imagery collected during tropical cyclones. *Remote Sens.* **2022**, *14*, 5006. [CrossRef]

35. Jin, Q.; Fan, X.; Liu, J.; Xue, Z.; Jian, H. Estimating tropical cyclone intensity in the South China sea using the XGBoost model and Fengyun satellite images. *Atmosphere* **2020**, *11*, 423. [CrossRef]

36. Shao, W.Z.; Ding, Y.Y.; Li, J.C.; Gou, S.P.; Nunziata, F.; Yuan, X.Z.; Zhao, L.B. Wave retrieval under typhoon conditions using a machine learning method applied to Gaofen-3 SAR imagery. *Can. J. Remote Sens.* **2020**, *45*, 723–732. [CrossRef]

37. Wang, H.; Yang, J.; Lin, M.; Li, W.; Zhu, J.; Ren, L.; Cui, L. Quad-polarimetric SAR sea state retrieval algorithm from Chinese Gaofen-3 wave mode imagettes via deep learning. *Remote Sens. Environ.* **2022**, *273*, 112969. [CrossRef]

38. Zhao, X.B.; Shao, W.Z.; Zhao, L.B.; Gao, Y.; Hu, Y.Y.; Yuan, X.Z. Impact of rain on wave retrieval from Sentinel-1 synthetic aperture radar images in tropical cyclones. *Adv. Space Res.* **2021**, *67*, 3072–3086. [CrossRef]

39. Gao, Y.; Guan, C.L.; Sun, J.; Xie, L. Tropical cyclone wind speed retrieval from dual-polarization Sentinel-1 EW mode products. *J. Atmos. Ocean. Technol.* **2020**, *3*, 1713–1724. [CrossRef]

40. Gao, Y.; Sun, J.; Zhang, J.; Guan, C.L. Extreme wind speeds retrieval using Sentinel-1 IW mode SAR data. *Remote Sens.* **2021**, *13*, 1867. [CrossRef]

41. Zhao, X.B.; Shao, W.Z.; Lai, Z.Z.; Jiang, X.W. Retrieval of rain rates for tropical cyclones from Sentinel-1 synthetic aperture radar images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 3187–3197. [CrossRef]

42. Mayers, D.; Ruf, C. MTrack: Improved center fix of tropical cyclones from SMAP wind observations. *Bull. Am. Meteorol. Soc.* **2020**, *102*, 1–23. [CrossRef]

43. Meissner, T.; Ricciardulli, L.; Wentz, F.J. Capability of the SMAP mission to measure ocean surface winds in storms. *Bull. Amer. Meteorol. Soc.* **2017**, *98*, 1660–1677.

44. Mouche, A.; Chapron, B.; Knaff, J.; Zhao, Y.; Zhang, B.; Combot, C. Copolarized and cross-polarized sar measurements for high-resolution description of major hurricane wind structures: Application to irma category 5 hurricane. *J. Geophys. Res. Oceans* **2019**, *124*, 3905–3922. [CrossRef]

45.　Vachon, P.W.; Dobson, F.W. Validation of wind vector retrieval from ERS-1 SAR images over the ocean. *Glob. Atmos. Ocean. Syst.* **1996**, *5*, 177–187.

46.　Shao, W.Z.; Li, X.F.; Hwang, P.; Zhang, B.; Yang, X.F. Bridging the gap between cyclone wind and wave by C-band SAR measurements. *J. Geophys. Res. Oceans* **2017**, *122*, 6714–6724. [CrossRef]

47.　Corcione, V.; Grieco, G.; Portabella, M.; Nunziata, F.; Migliaccio, M. A novel azimuth cutoff implementation to retrieve sea surface wind speed from SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3331–3340. [CrossRef]

48.　Kudryavtsev, V.N.; Chapron, B.; Myasoedov, A.G.; Collard, F.; Johannessen, J.A. On dual co-polarized SAR measurements of the ocean surface. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 761–765. [CrossRef]

49.　Yuan, X.Z.; Shao, W.Z.; Han, B.; Wang, X.C.; Wang, X.Q.; Gao, Y. Rain-induced characteristics in C- and X-band synthetic aperture radar observations of tropical cyclones. *Remote Sens. Lett.* **2021**, *12*, 573–584. [CrossRef]

50.　Hu, Y.Y.; Shao, W.Z.; Jiang, X.W.; Zhou, W.; Zuo, J.C. Improvement of VV-polarization tilt MTF for Gaofen-3 SAR data of a tropical cyclone. *Remote Sens. Lett.* **2023**, *14*, 461–468. [CrossRef]