




Article

A Technique for SAR Significant Wave Height Retrieval Using Azimuthal Cut-Off Wavelength Based on Machine Learning

Shaijie Leng ¹, Mengyu Hao ¹, Weizeng Shao ^{1,*} , Armando Marino ² and Xingwei Jiang ³

¹ College of Oceanography and Ecological Science, Shanghai Ocean University, Shanghai 201306, China; m230200644@st.shou.edu.cn (S.L.); 2169113@st.shou.edu.cn (M.H.)

² Natural Sciences, University of Stirling, Stirling FK9 4LA, UK; armando.marino@stir.ac.uk

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: This study introduces a new machine learning-based algorithm for the retrieving significant wave height (SWH) using synthetic aperture radar (SAR) images. This algorithm is based on the azimuthal cut-off wavelength and was developed in quad-polarized stripmap (QPS) mode in coastal waters. The collected images are collocated with a wave simulation from the numeric model, called WAVEWATCH-III (WW3), and the current speed from the HYbrid Coordinate Ocean Model (HYCOM). The sea surface wind is retrieved from the image at the vertical-vertical polarization channel, using the geophysical model function (GMF) CSARMOD-GF. The results of the algorithm were validated against the measurements obtained from the Haiyang-2B (HY-2B) scatterometer, yielding a root mean squared error (RMSE) of 1.99 m/s with a 0.82 correlation (COR) and 0.27 scatter index of wind speed. It was found that the SWH depends on the wind speed and azimuthal cut-off wavelength. However, the current speed has less of an influence on azimuthal cut-off wavelength. Following this rationale, four widely known machine learning methods were employed that take the SAR-derived azimuthal cut-off wavelength, wind speed, and radar incidence angle as inputs and then output the SWH. The validation result shows that the SAR-derived SWH by eXtreme Gradient Boosting (XGBoost) against the HY-2B altimeter products has a 0.34 m RMSE with a 0.97 COR and a 0.07 bias, which is better than the results obtained using an existing algorithm (i.e., a 1.10 m RMSE with a 0.77 COR and a 0.44 bias) and the other three machine learning methods (i.e., a >0.58 m RMSE with a <0.95 COR), i.e., convolutional neural networks (CNNs), Support Vector Regression (SVR) and the ridge regression model (RR). As a result, XGBoost is a highly efficient approach for GF-3 wave retrieval at the regular sea state.

Keywords: significant wave height; GF-3; machine learning; azimuthal cut-off wavelength; WW3



Citation: Leng, S.; Hao, M.; Shao, W.; Marino, A.; Jiang, X. A Technique for SAR Significant Wave Height Retrieval Using Azimuthal Cut-Off Wavelength Based on Machine Learning. *Remote Sens.* **2024**, *16*, 1644. <https://doi.org/10.3390/rs16091644>

Academic Editor: Chung-Ru Ho

Received: 23 April 2024

Revised: 1 May 2024

Accepted: 2 May 2024

Published: 5 May 2024



Copyright: © 2024 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

It is widely recognized that the sea surface wave plays a crucial role in marine dynamics, especially in climate change. The wave has been traditionally measured using an electro-magnetic current meter and an acoustic Doppler current meter. However, these observations have significant limitations, especially regarding marine research over large spatial coverage. Thanks to advancements in computer science and oceanography theory, numeric wave models such as WAVEWATCH-III (WW3) [1] and Simulation Wave Nearshore (SWAN) [2] have been developed. However, these models typically have a coarse spatial resolution of 10 km, limiting the hindcasting wave application for regional studies on marine science, especially in coastal waters.

Monitoring the sea surface in near real-time has become a crucial area of research thanks to the increasing popularity of remote-sensing technology for earth observation since the 1970s. Currently, the remote-sensed products are officially released, including sea

surface wind from a scatterometer [2], significant wave height (SWH) from an altimeter [3], and wave spectrum from Surface Wave Investigation and Monitoring (SWIM) onboard Chinese–French Oceanography SATellite (CFOSAT) [4,5]. However, monitoring nearshore waves with high precision remains a challenge for the remote sensing community, although two high-frequency phased-array radars [6] can detect swell and current in coastal areas. Synthetic aperture radar (SAR), an advanced sensor with acceptable spatial resolution (i.e., 1 m for TerraSAR-X [TS-X] and TanDEM-X [TD-X]) [7], can significantly retrieve atmospheric–marine dynamics, especially in tropical cyclones [8,9], and detect targets on the sea surface [10].

The geophysical model function (GMF) in vertical–vertical (VV) polarization, initially designed for wind retrieval from a scatterometer [11], can also be used for SAR wind retrieval. The co-polarized (VV and horizontal–horizontal (HH)) GMFs have been further improved through SAR measurements, i.e., C-SARMOD for Sentinel-1 (S-1) [12], C-SARMOD2 [13] for RADARSAT-2 (R-2), and CSARMOD-GF for Gaofen-3 (GF-3) [14]. Other studies have been conducted for wind retrieval using the SAR-derived azimuthal cut-off wavelength [15,16] and theoretical backscattering model [17]. However, due to the saturation of the co-polarized SAR backscattering signal at the regular sea state [18] and at a strong wind speed of >25 m/s (i.e., cyclonic wind profile), GMF usually inverts cross-polarized vertical–horizontal and horizontal–vertical images [19,20] using a machine learning method [21]. Several algorithms have been developed for co-polarized SAR wave retrieval based on a sea surface mapping mechanism, following the introduction of the Max-Planck Institute Algorithm (MPI) [22] and several co-polarized SAR wave retrieval algorithms based on a sea surface mapping mechanism. These algorithms include a semiparametric retrieval algorithm (SPRA) [23], partition rescaling and shift algorithm [24], parameterized first-guess spectrum method (PFSM) [25], and fully polarimetric technique [26,27]. Additionally, empirical models [28,29] and machine learning techniques [30,31] have been implemented for retrieving wave parameters from SAR images directly, without calculating the complex model transfer functions (MTFs) of the mapping mechanism.

In sea surface imaging, the motion between the sea surface and the SAR satellite can make waves at a length smaller than the specific value in the azimuth direction undetected [32]. This phenomenon is called the azimuthal cut-off wavelength caused by unique velocity bunching. SAR wind [33] and wave retrieval algorithms [29] have been developed based on azimuthal cut-off wavelength. Let us consider the ENVISAT-ASAR image in the Agulhas Current region as an example [34]; the Doppler shift frequency represented by the Doppler centroid anomaly (DCA) in the range direction [35,36] is correlated with backscattering roughness of upper ocean dynamics, such as wind, wave, and current speed. However, previous studies [37,38] have reported significant differences between theoretical-based and SAR-measured cut-off wavelengths observed through ENVISAT-ASAR and GF-3 imageries acquired in wave mode. As a result, it is worth investigating the influence of the current on the azimuthal cut-off wavelength.

In this work, 3000 GF-3 images were collected in China's coastal waters. Our goal was to investigate the dependence of SWH on upper oceanic dynamics, such as the sea surface wind speed and azimuthal cut-off wavelength. After analyzing the data, the SWH retrieval algorithm for SAR was developed based on machine learning. The remainder of this study is organized as follows: Section 2 describes the dataset, including SAR images and auxiliary data. Section 3 presents the SAR wave retrieval algorithm methodology, particularly emphasizing studying the dependence of upper oceanic dynamics on the SWH. The applicability of the SWH retrieval algorithm is confirmed in Section 4, and the conclusion is summarized in Section 5.

2. Datasets

This section describes GF-3 SAR images taken in the China Seas. Additionally, the auxiliary data are presented, i.e., European Centre for Medium-Range Weather Forecasts reanalysis (ERA-5) wind, HYCOM current, and the products from Haiyang-2B (HY-2B).

2.1. GF-3 Image

The GF-3, which operates in 12 imaging modes [19], has released the data since August 2016. It is a part of the Dragon Programme, a collaboration project between the large-scale scientific and technological cooperation project between the Chinese Ministry of Science and Technology and the European Space Agency in Earth observation. The calibration method for GF-3 processed as a Level-1A (L-1A) production is presented in the following equation:

$$\sigma^0 = \text{PI}^2 \left(\frac{M}{32,767} \right)^2 - N \text{ [dB]} \quad (1)$$

where σ^0 is normalized radar cross section (NRCS); PI is the pixel intensity of SAR raw data; and M and N are the external calibration constant and the offset factor for a specific imaging mode, respectively, stored in the annotated file. In total, 3000 images acquired in quad-polarized stripmap (QPS) mode with swath coverage of 100 km and a standard pixel of 25 m were collected from 2019 to 2022. Figure 1 shows a quick look at a VV-polarized image after establishing calibration, in which the wind streaks are visible. This image was taken at 9:44 UTC on 29 September 2021, and the incidence angle ranged from 39.68° to 40.85° , following the descending direction. The frame of all images is illustrated in Figure 2, in which the black and blue rectangles represent the spatial coverage of images.

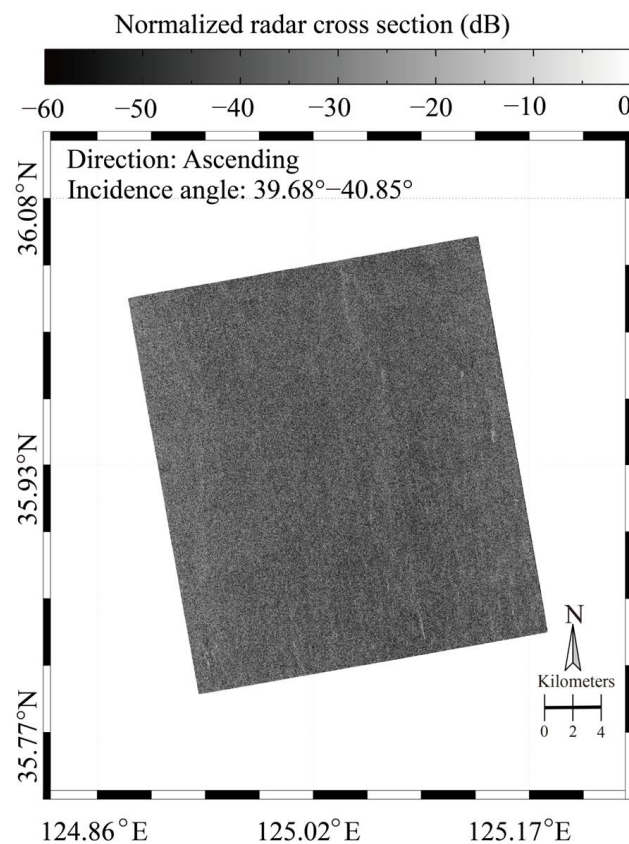


Figure 1. A quick look at a Gaofen-3 (GF-3) synthetic aperture radar (SAR) in vertical–vertical (VV) polarization after calibration, which was taken at 09:44 UTC on 29 September 2021.

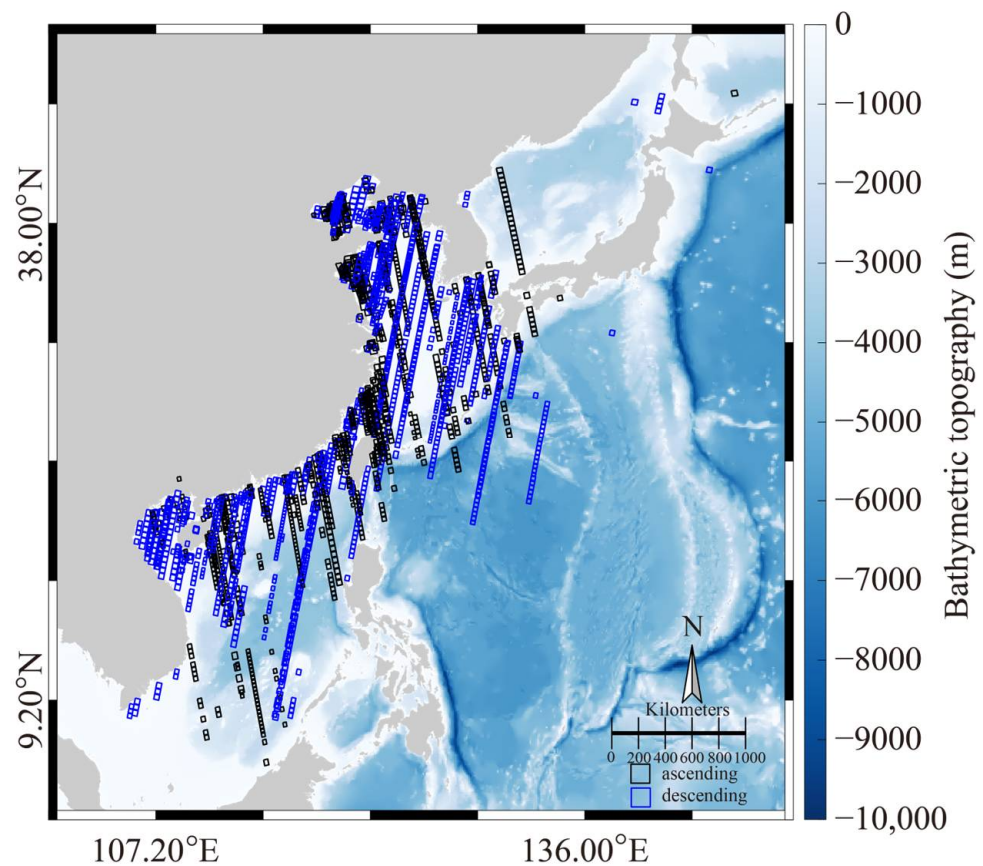


Figure 2. Frame of all images. Black and blue rectangles represent the spatial coverage of images.

2.2. Auxiliary Data

GF-3 satellite image retrieval algorithms for wind and wave [27] have been developed. One such algorithm is the co-polarized GMF CSARMOD-GF, specifically adapted for GF-3 to address calibration issues [14]. The fundamental principle behind the GMF CSARMOD-GF is to establish the relationship between the NRCS and a wind vector, stated as follows:

$$\sigma^0 = B_0(1 + B_1 \cos \varphi + B_2 \cos 2\varphi) \quad (2)$$

where B_0 , B_1 , and B_2 are the functions of wind speed U_{10} at 10 m above the sea surface and incidence angle, θ ; and φ is the wind direction relative to flight orientation. In GMF, when two unknown variables are present, prior information on wind direction from a 0.25° gridded ERA-5 is directly employed. The image is divided into sub-scenes with 215×215 pixels (~ 5 km) in the pretreatment process. Figure 3 depicts a two-dimensional SAR spectrum at a spatial scale between 800 m and 3 km extracted from the image in Figure 1. The red line represents wind direction with 180° ambiguity. Then, the true wind direction is obtained by referring to the ERA-5 wind field (Figure 4) at 10:00 UTC on 29 September 2021. In this figure, the black rectangle represents the spatial coverage of the image in Figure 1.

Validating the SAR-derived wind and wave is crucial for the result's accuracy. To validate this, operational products from HY-2B are used, including the wind from a scatterometer with a spatial resolution of 12.5 km and the SWH with a spatial resolution of 10 km, following the footprint of the altimeter. For instance, Figure 5a,b show the wind and wave maps from the HY-2B scatterometer and altimeter on 25 October 2020. Moreover, the distances between SAR retrievals from the sub-scenes on the image and the measurements from HY-2B are within 3 km. The time differences are each less than 1.5 h. The WW3 model simulates the waves with 0.05° grids at intervals of 0.5 h, treating ERA-5 wind and

water depth from the General Bathymetric Chart of the Oceans (GEBCO) as forcing fields. Our previous studies have shown the reliability of this model, showing the model settings of WW3 and validation of hindcasted simulation in our earlier studies [1]. The numeric circulation models can be used for current simulation, i.e., the Princeton Ocean Model [39]. Its upgraded version is called the Stony Brook Parallel Ocean Model [40], Finite-Volume Community Ocean Model [41], and HYbrid Coordinate Ocean Model (HYCOM) [42]. Fortunately, the current data from the HYCOM reanalysis system are publicly available to investigators worldwide, making them an accessible and reliable resource. The spatial resolution of the HYCOM current is 0.08° grid at intervals of 3 h each day. Figure 6a,b present the examples of HYCOM's current map at 9:00 UTC on 29 September 2021, and the WW3-simulated SWH map at 10:00 UTC on 29 September 2021, in which the black rectangle represents the spatial coverage of the image in Figure 1.

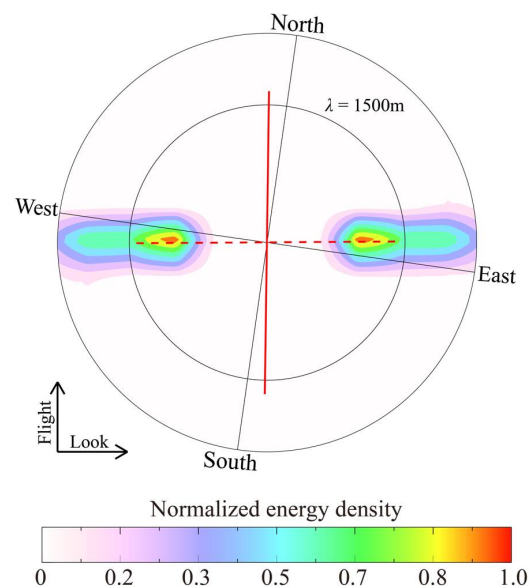


Figure 3. Two-dimensional SAR spectrum at a spatial scale between 800 m and 3 km extracted from the image in Figure 1, in which the red line represents wind direction with 180° ambiguity.

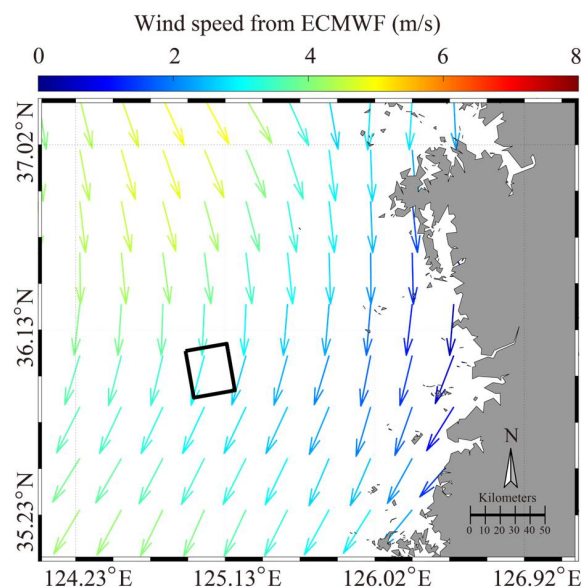


Figure 4. European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA-5) wind at 10:00 UTC on 29 September 2021, in which the black rectangle represents the spatial coverage of the image in Figure 1.

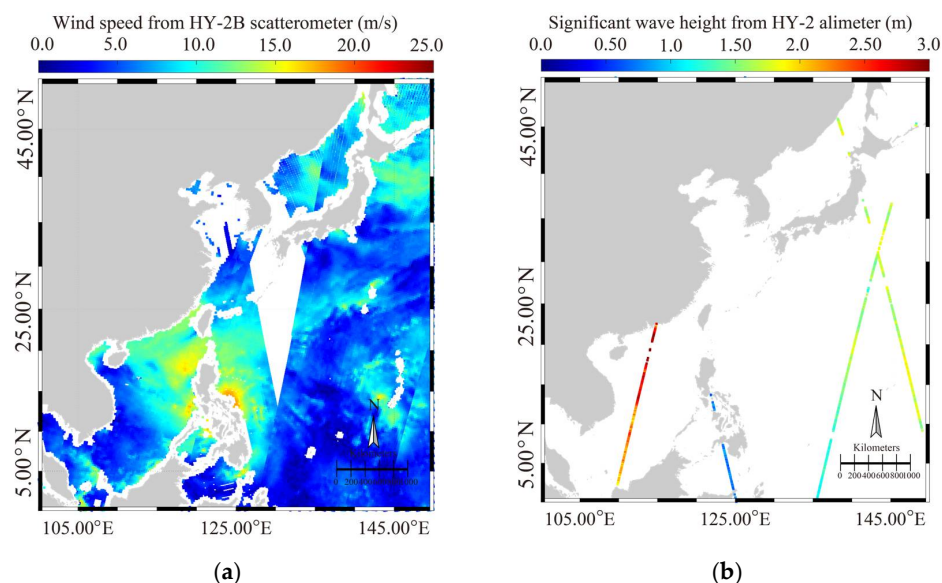


Figure 5. Maps from the HY-2B scatterometer and altimeter on 25 October 2020: (a) wind and (b) significant wave height (SWH).

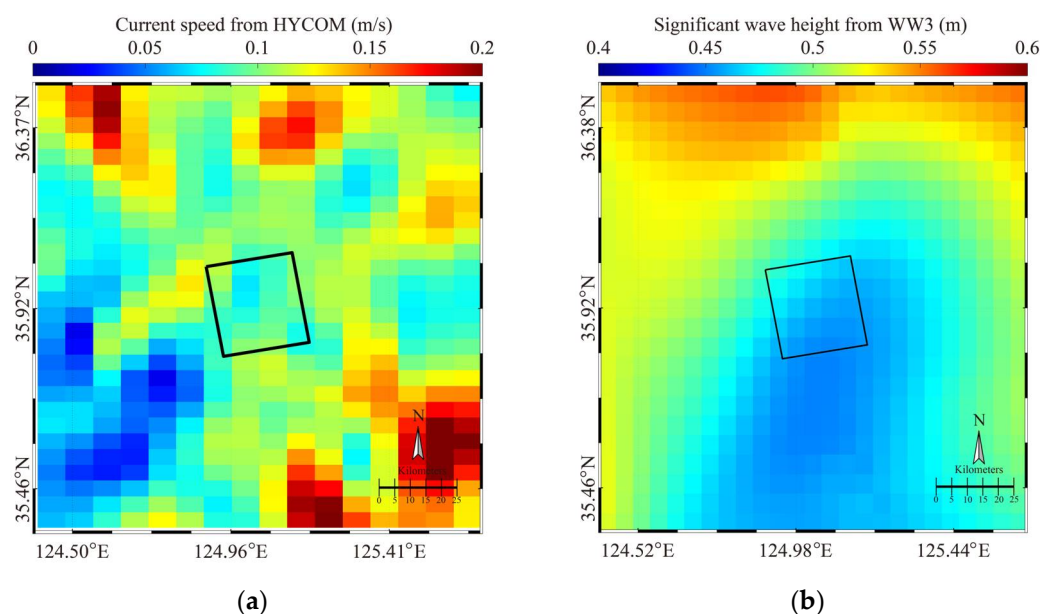


Figure 6. (a) Current map at 9:00 UTC on 29 September 2021, from Hybrid Coordinate Ocean Model (HYCOM), and (b) the WW3-simulated SWH map at 10:00 UTC on 29 September 2021, in which the black rectangle represents the spatial coverage of the image in Figure 1.

3. Methodology

This section presents the wave retrieval algorithms from VV-polarized SAR images, presenting the theoretical calculation of azimuthal cut-off wavelength associated with wave spectrum. The SWH retrieval algorithms based on machine learning are proposed after studying the dependence of upper oceanic dynamics on the azimuthal cut-off wavelength. The general processing flow diagram of the methodology is presented in Figure 7. Firstly, the wind speed, azimuthal cut-off wavelength and radar incidence angle are obtained from the GF-3 images, which re-located with the SWH simulated by the WW3. Secondly, those matchups are treated as the dataset for developing the SWH retrieval algorithms based on machine learning models. In addition, the measurements from the HY-2B are used for validating the SAR-derived wind speeds and SWHs.

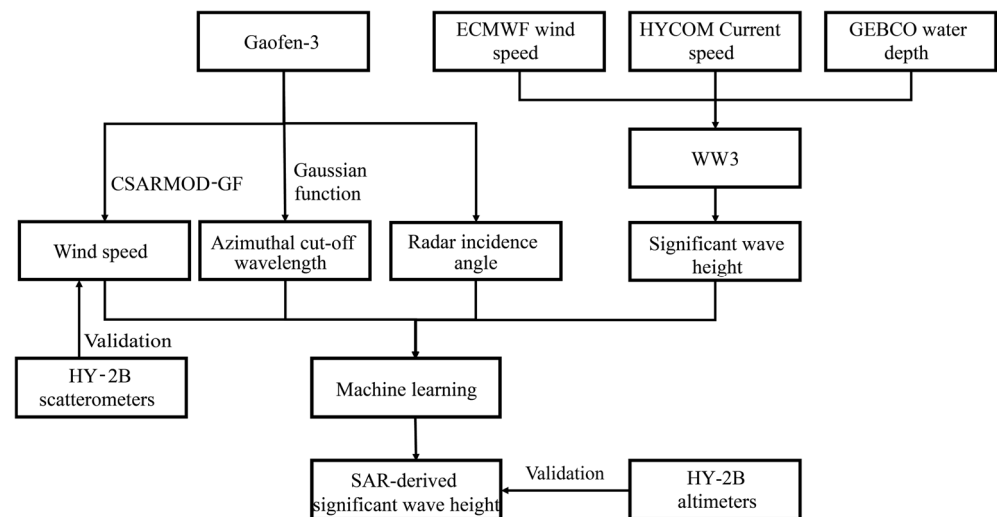


Figure 7. The general processing flow diagram.

3.1. SAR Wave Retrieval Algorithm

As mentioned earlier, SAR wave retrieval can be performed using a theoretical-based scheme, an empirical model, or an intelligent algorithm. The theoretical-based scheme, also known as PFSM, combines the benefits of MPI and SPRA algorithms to invert wave spectrum from VV-polarized GF-3 image. The critical aspect of the PFSM algorithm is the separation of the SAR intensity spectrum in the dimension of wave number, and the threshold k_s [43] is estimated by

$$k_s = \left[\frac{2.87 g V^2}{R^2 U_{10}^4 \cos^2 \phi (\sin^2 \phi \sin^2 \theta + \cos^2 \phi)} \right]^{0.33} \quad (3)$$

where R is the slant distance, V is the flight velocity, g is gravitational acceleration, ϕ is the wave propagation direction relative to the radar look orientation, θ is the radar incidence angle, and U_{10} is the wind speed. Here, the SAR-derived wind speed, U_{10} , is derived from (3). Figure 8a displays the SAR-derived wind map corresponding to the image in Figure 1. Moreover, a statistical analysis was conducted through more than 1000 matchups between SAR retrievals and products of the HY-2B scatterometer, as depicted in Figure 8b. Our findings show a 1.99 m/s root mean squared error (RMSE), a 0.82 correlation (COR), and a 0.27 scatter index (SI) of wind speed, indicating the reliability of that SAR-derived wind for this study.

The portion of an SAR intensity spectrum at the wave number, k , greater than threshold, k_s , corresponds to the wind wave, obtaining the dominant velocity, c_p , and propagation wave direction, ϕ_p . Afterward, the discretized values, i.e., $0.8c_p \leq c_p \leq 1.2c_p$ at intervals of 0.1 m/s and $(\phi_p - 20^\circ) \leq \phi_p \leq (\phi_p + 20^\circ)$ at intervals of 1° , were used. Then, SAR-derived wind speeds are used in the wave spectrum, called the Elfouhaily model [44]. Those wave spectra and mapping MTFs stimulate the intensity spectra, and the best-fit first-guess wave spectrum corresponds to the minimum difference between the simulated and SAR intensity spectra. The wind wave spectrum is inverted by minimizing the cost function, J , expressed as follows:

$$J = \int [T_k - \bar{T}_k]^2 dk + \mu \int \left\{ \frac{[F_k - \bar{F}_k]}{[C - \bar{F}_k]} \right\}^2 dk \quad (4)$$

where T_k is the inverted wave spectrum; \bar{T}_k is the best-fit first-guess wave spectrum; F_k is the SAR intensity spectrum; \bar{F}_k is the mapped intensity spectrum; μ indicates the weight coefficient; and C is constant as 0.001, ensuring the computational convergence. The process involves inserting the swell spectrum by reversely solving the SAR intensity spectrum at a wave number, k , smaller than the threshold, k_s . Lastly, the wave spectrum is a composite of

the wind wave and the swell. After applying FFT-2, the two-dimensional SAR intensity spectrum of the sub-scene in Figure 3a at a spatial scale between 60 m and 1 km is shown in Figure 9a. The one-dimensional wave spectrum derived from SAR is presented in Figure 9b. The SAR-derived result is duplicated with the true wave spectrum because of the 180-degree ambiguity of the SAR intensity spectrum.

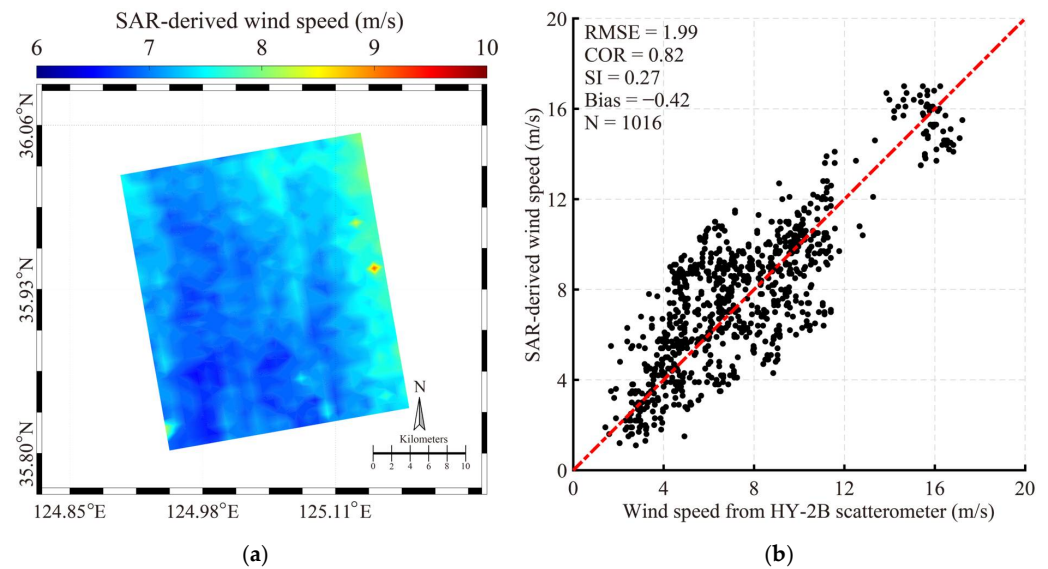


Figure 8. (a) SAR-derived wind map corresponding to the image in Figure 1, and (b) a comparison between SAR retrievals and wind speeds of the HY-2B scatterometer.

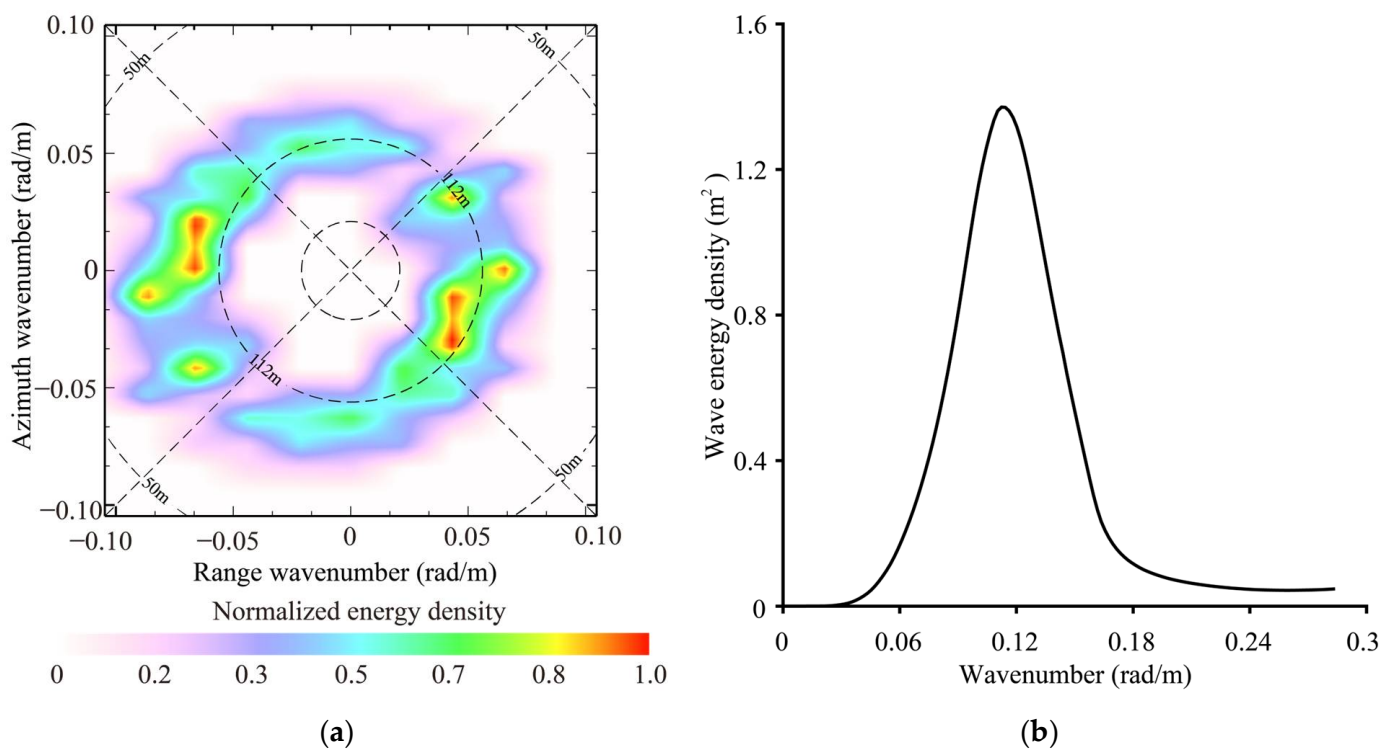


Figure 9. (a) Two-dimensional SAR intensity spectrum of the sub-scene in Figure 3a at a spatial scale between 60 m and 1 km. (b) The one-dimensional SAR-derived wave spectrum.

3.2. Dependence of Upper Oceanic Dynamics on the Azimuthal Cut-Off Wavelength

In practice, the azimuthal cut-off wavelength, λ , is estimated by fitting it to the normalized one-dimensional spectrum based on a Gaussian function, $G(k_x)$, as described in the following equation:

$$G(k_x) = \exp\left\{\pi\left(\frac{k_x}{k_c}\right)^2\right\} \quad (5)$$

where k_x is the wavenumber in the azimuthal direction, and $k_c = 2\pi/\lambda$ [45]. Note that velocity bunching represented by azimuthal cut-off wavelength is independent of polarization [38]. As a result, the azimuthal cut-off wavelength in VV-polarization was only used.

Figure 10a,b depict the relationships between the SWH and two variables: wind speed for a 1 m/s bin and azimuthal cut-off wavelength for a 1 m bin. Unsurprisingly, wind speed is correlated with the SWH (COR = 0.56) because the sea state is determined by wind stress. As a result, the NRCS was included in the empirical wave retrieval algorithm (i.e., CWAVE). Additionally, the azimuthal cut-off wavelength is correlated with the SWH (COR = 0.50). This behavior is more apparent in the tropical cyclones following the fetch- and duration-limited features [46]. Similarly, Figure 10c shows the dependence of current speed on azimuthal cut-off wavelength, indicates that the COR (=0.11) is quite weak because the current has less influence on azimuthal cut-off wavelength.

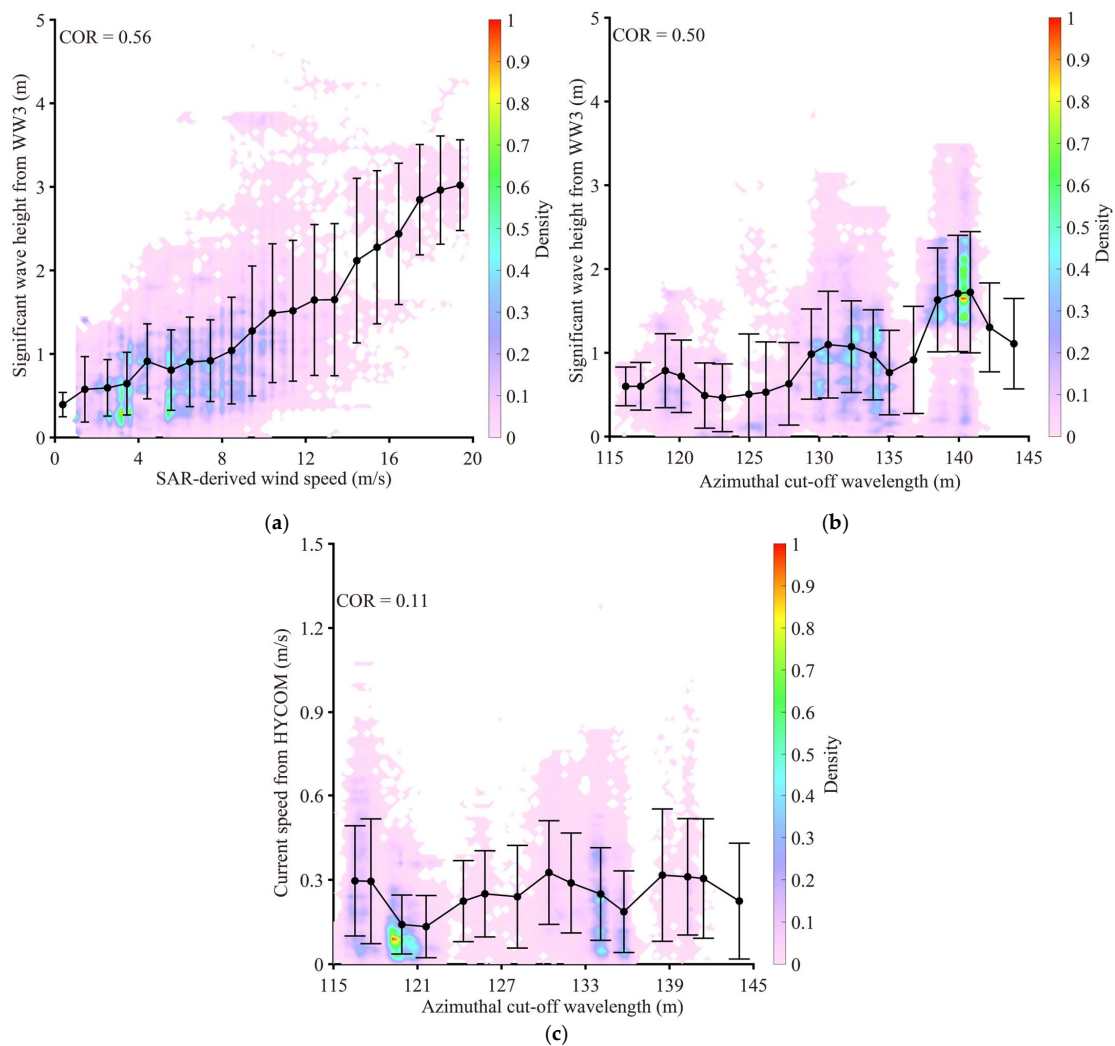


Figure 10. Relation between SWH and two variables: (a) wind speed for a 1 m/s bin and (b) azimuthal cut-off wavelength for a 1 m bin. (c) Relation between azimuthal cut-off wavelength and current speed for a 0.1 m/s bin.

3.3. Development of SWH Retrieval Algorithm

Advanced techniques such as machine learning are becoming increasingly integral to the development of SAR wind technology in the era of artificial intelligence [47] and wave retrieval [29,31] algorithms. In a recent study [9], three machine learning schemes were applied for wind retrieval in tropical cyclones. In this study, the four machine learning methods, i.e., eXtreme Gradient Boosting (XGBoost), convolutional neural networks (CNN), Support Vector Regression (SVR), the ridge regression model (RR), were used to develop the new SWH retrieval algorithm. Detailed information on the four methods is described as follows.

In principle, XGBoost uses the multiple decision trees generated by gradient lifting. Then, a strong classifier is employed to improve each decision tree's accuracy in the iterations. Moreover, the second-order gradient information and regularization can avoid over-fitting effectively and offer better computing efficiency. Hu et al. [9] presented the details of the XGBoost process, and the hyperparameters used in the XGBoost model are listed in Table 1.

Table 1. The Hyperparameters used in XGBoost.

Hyperparameter	Value
max_depth	50
n_estimators	300
Gamma	0.1
Subsample	0.9
min_child_weights	3
reg_lambda	1
reg_alpha	0.001
Gamma	0.1

As one of the representative algorithms of deep learning, the CNN is a class of feed-forward neural networks that contains convolutional computation and a deep structure. The CNN network model is constructed by stacking multiple convolutional layers, pooling layers, and fully connected layers to learn the features of the input data and ultimately make regression predictions. The convolutional layer extracts features from the input data through convolutional operations, and the pooling layer conducts a dimensionality reduction to extract important feature information. Finally, the fully connected layer integrates the features obtained by integrating the convolutional and pooling layers for the final regression prediction.

SVR is a machine learning algorithm for solving regression problems. Firstly, the kernel function is used to obtain the high-dimensional spatial characteristics of the original data, and then the loss function is used to evaluate the difference between the predicted value and the real value. In order to avoid the overfitting problem, the model is adjusted by the gradient descent optimization algorithm and the regularization parameter. At the same time, the boundary parameters are set to define a tolerable range. Therefore, by setting different parameters, SVR can be adapted to various scientific problems.

In general, RR is a linear model that deals with regression problems. Compared with the traditional linear regression model, the regularization term and the complexity of the model can be optimized by adjusting the regularization parameter. A larger regularization parameter reduces the complexity of the model and avoids the overfitting problem. In addition, when the features of input data are multicollinear, the ridge regression model can improve the stability and reliability of the model.

Our work has datasets consisting of 40,000 matchups and 2000 images available in the training process using four machine learning methods. As reported in a recent study [48], the SAR features under 15 polarization modes are used in the development of GF-3 wave retrieval based on XGBoost and are quite complex to implement. In contrast, the azimuthal cut-off wavelength caused by velocity bunching is independent of the imaging mode

and polarization. In our work, the inputs of this model include the azimuthal cut-off wavelength, SAR-derived wind speed by GMF CSARMOD-GF, and radar incidence angle. The output is the SWH. Figure 11 exhibited the training process performance, indicating that the XGBoost eventually converges and that approximately 0.30 m RMSE and the CNN eventually converges and that approximately 0.58 m RMSE is achieved as the iteration approaches 200. However, the SVR and RR have no iteration process. Figure 11c shows the SHAP values highlighted by the feature value. It was found that the wind speed and azimuthal cut-off wavelength have the greatest impact on the XGBoost model.

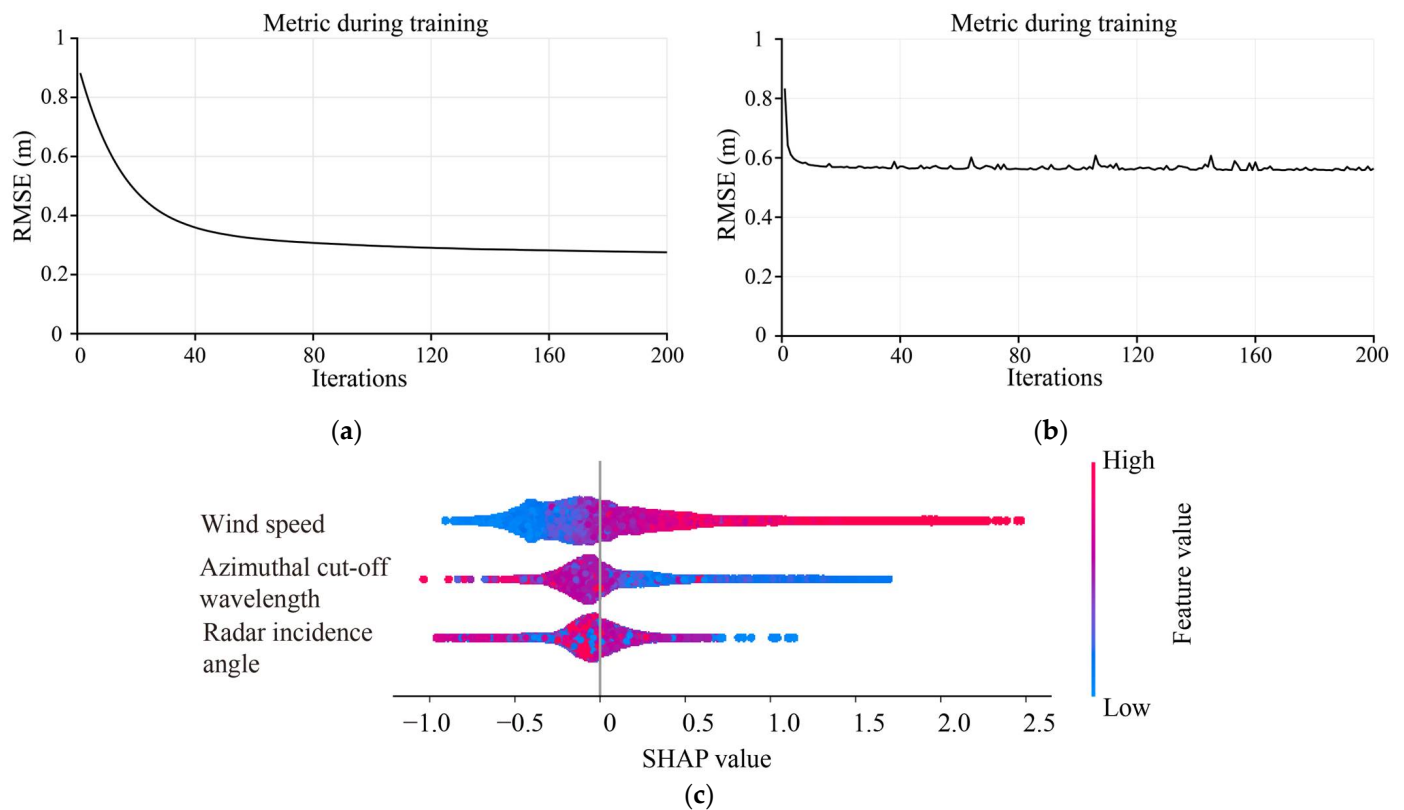


Figure 11. Performance of the training process: (a) eXtreme Gradient Boosting (XGBoost), (b) convolutional neural networks (CNN), and (c) the SHAP value map.

4. Results and Discussion

This section proposes an algorithm applied for 1000 images, presents the validation against HY-2B measurements, and discusses the error analysis.

4.1. Validation

The machine learning-based algorithm for SWH retrieval was implemented for the other 1000 images. Afterward, the retrievals from SAR were compared with the measurements obtained from the HY-2B altimeter and the simulations generated by the WW3 model. The HY-2B swath covering the SAR scenes is collected, and the temporal difference between them is less than 1.5 h, revealing over 600 matchups for validation and error analysis. Figure 12a shows the retrieval result from several along-track images at 9:44 UTC on 29 September 2021, where the gaps are caused by an invalid azimuthal cut-off wavelength. The color circles represent the footprints of the HY-2B altimeter, and the black rectangles represent the spatial coverage of the image corresponding to Figure 1. Figure 12b shows that the retrieval results are consistent with those from HY-2B footprints concerning latitude. Therefore, the case study at SWH up to 1 m preliminarily confirmed that SWH could be practically inverted from the SAR image.

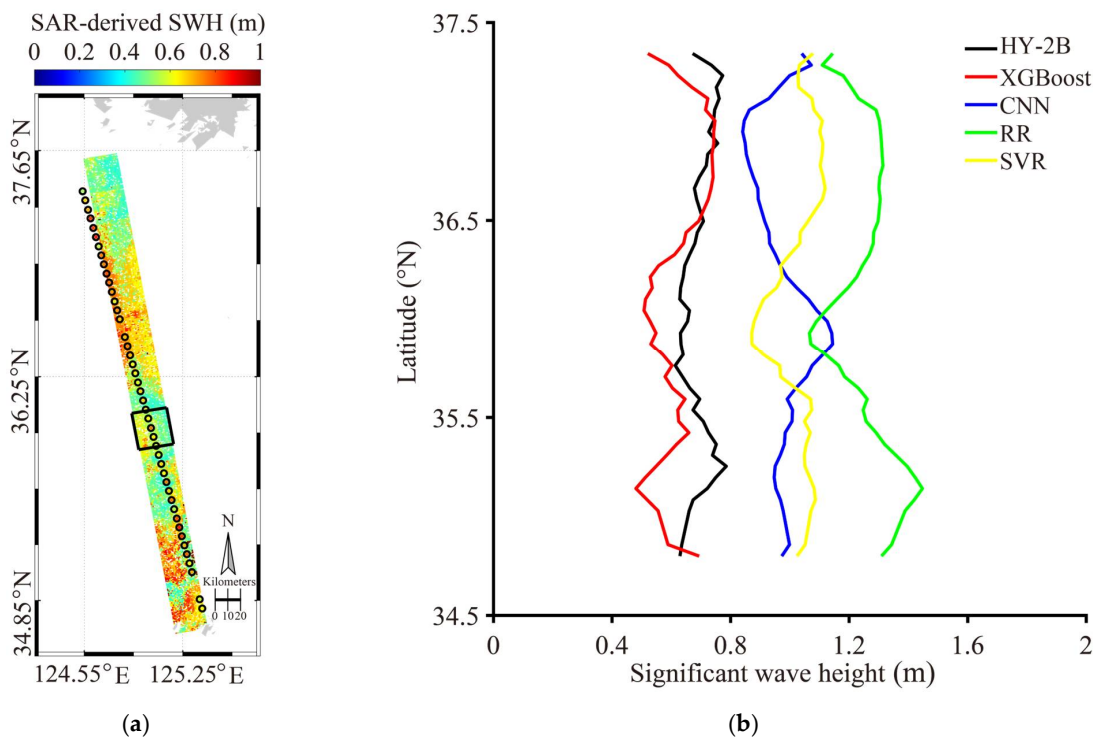


Figure 12. (a) Retrieval results along the track corresponding to the image in Figure 1; (b) the retrieval results and HY-2B footprints with respect to latitude. The color circles represent the footprints of the HY-2B altimeter, and the black rectangles represent the spatial coverage of the image corresponding to Figure 1.

Based on Figure 13a, the statistical analysis conducted at SWH goes up to 5 m. The results show that the RMSE of SWH determined by the XGBoost model is 0.34 m and has a 0.97 COR and a 0.07 m bias, which is an improvement over the PFSM algorithm (Figure 13b), achieving a 1.10 m RMSE with a 0.77 COR and a 0.44 bias. Moreover, this result is also better than those achieved by other machine learning methods, i.e., a RMSE of 0.58 m and a COR of 0.94 by CNN (Figure 13c), a RMSE of 0.77 m, and a COR of 0.91 (Figure 13d) by RR and a RMSE of 0.62 m and a COR of 0.95 by SVR (Figure 13e). It was concluded that the XGBoost has the best performance at the SWH retrieval from GF-3 images. Additionally, this algorithm has the advantage of high computational efficiency without requiring the calculation of complex MTFs. However, it should be noted that the PFSM algorithm seems to overestimate the values at a weak sea state (SWH < 0.5 m) due to the absence of the wind term in MTFs.

4.2. Discussion

Figure 14 exhibits the variations in the bias (SAR retrievals by the XGBoost minus HY-2B measurements) regarding SAR-derived azimuthal cut-off wavelength, SAR-derived wind speed, and SWH measured by the HY-2B altimeter. The results show that the bias within -0.5 m varies with the SAR-derived azimuthal cut-off wavelength. When the SAR-derived wind speed increases, the bias rate does not show much of a change. According to the findings, the bias tends to decrease when the HY-2B altimeter measures the SWH and is less than 2 m. However, it increases as the SWH reaches 5 m. When the SWH ranges from 3 m to 4 m, the bias varies at 0.30 m, which aligns with the outcome of the training process. Based on this, it is concluded that XGBoost is a reliable method for SAR wave retrieval. The limitation of the proposed algorithm is a lack of samples in the training and validation dataset at an extreme sea state (>7 m). This indicates that, in the case of extreme sea state, the trained XGBoost is not suitable for SAR wave retrievals. In the future, this issue can be solved using images acquired during a tropical cyclone.

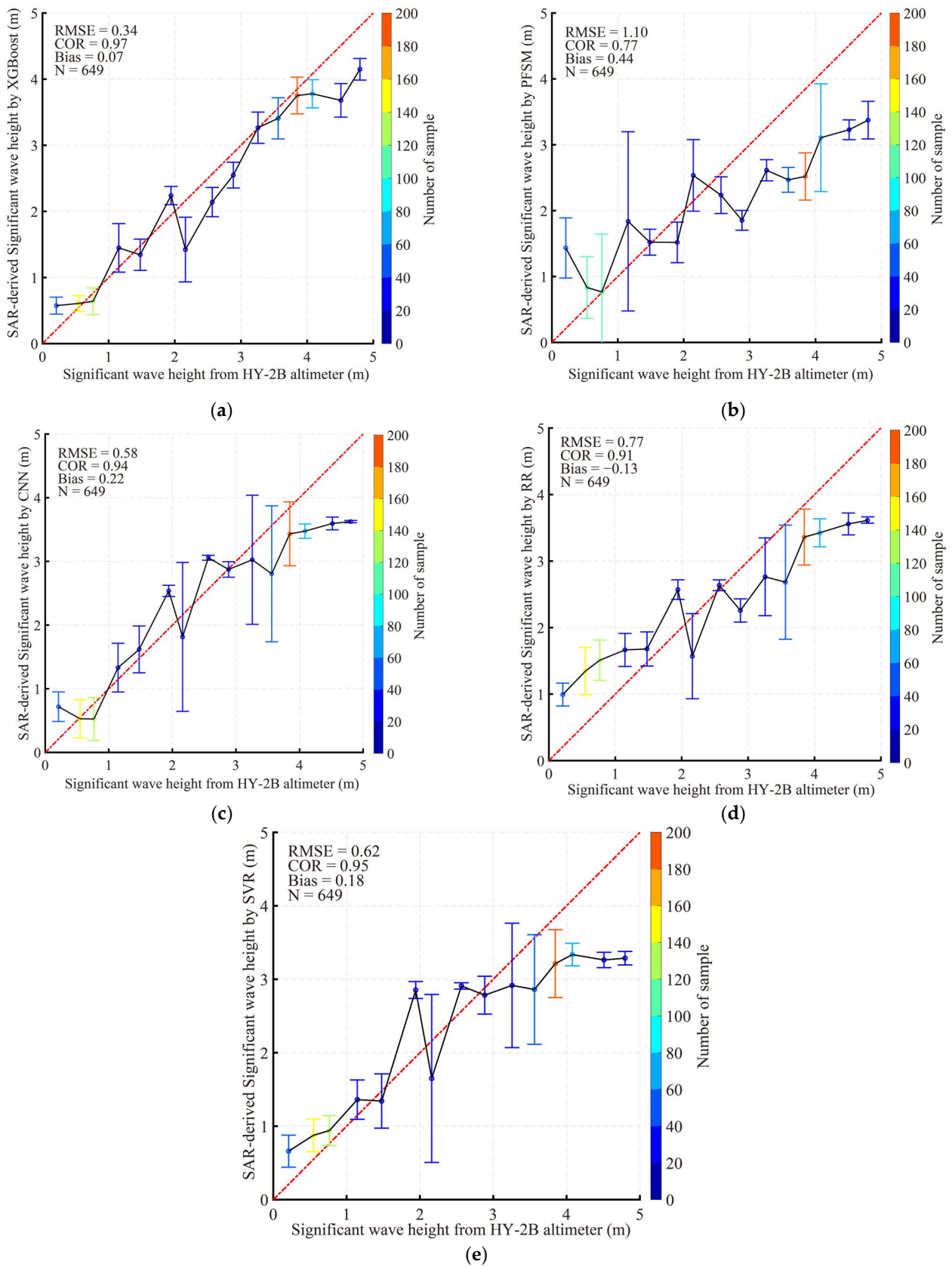


Figure 13. Validation of SAR retrievals by (a) the XGBoost, (b) parameterized first-guess spectrum method (PFSM), (c) the CNN, (d) the RR, and (e) the SVR against the measurements from the HY-2B altimeter.

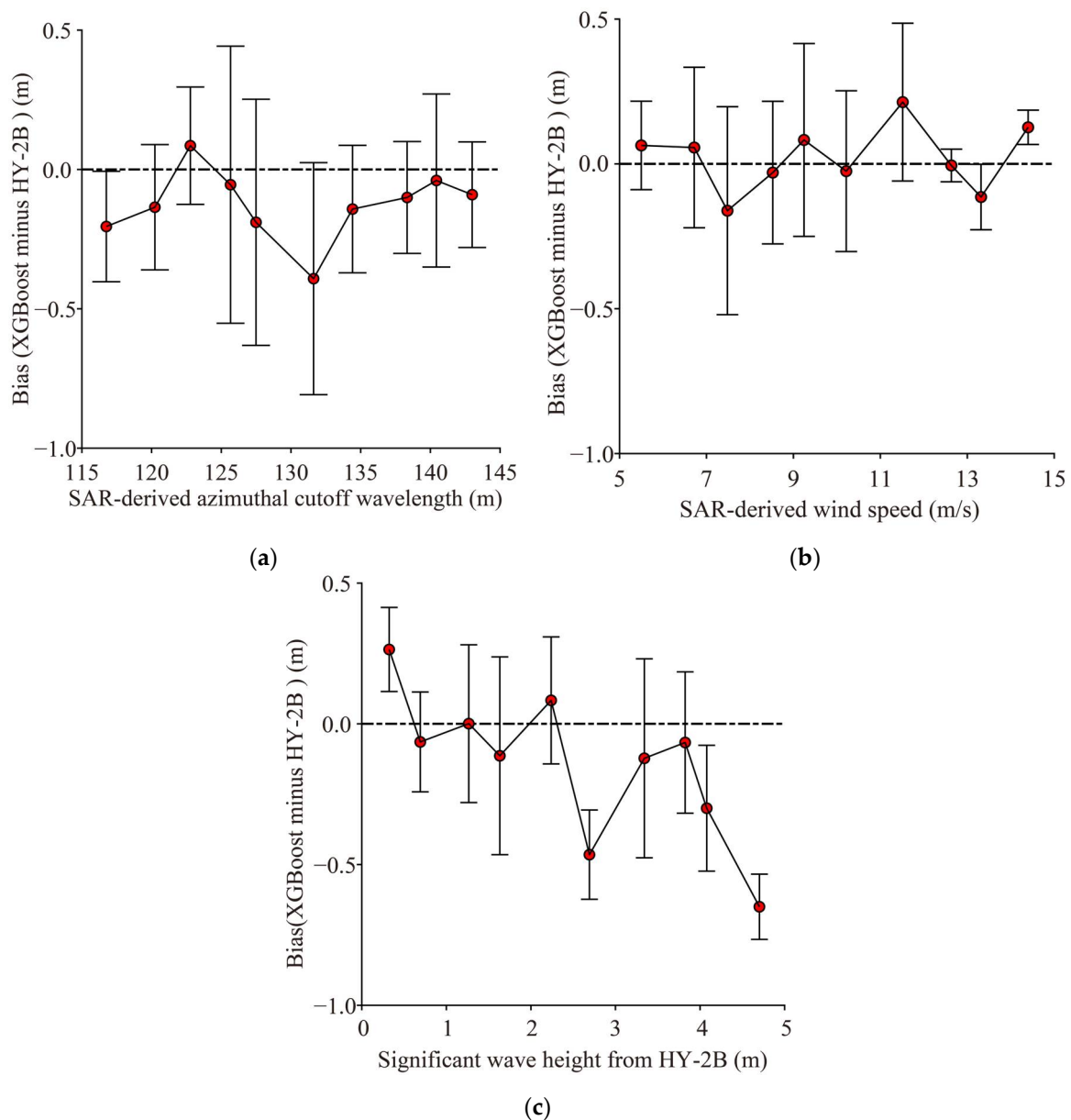


Figure 14. Variations in the bias (SAR retrievals minus HY-2B measurements) with respect to (a) SAR-derived azimuthal cut-off wavelength, (b) SAR-derived wind speed, and (c) SWH measured by the HY-2B altimeter.

5. Conclusions

SAR is well recognized as an effective sensor for sea surface monitoring with an acceptable spatial resolution. Over time, theoretical and empirical algorithms related to wind and wave retrieval from SAR images have been developed and maturely implemented for C-band SAR, i.e., S-1 [12] and GF-3 [17]. Additionally, a DCA-based algorithm empirically inverts SAR current velocity in the radar look direction [34–36]. However, the SAR wave retrieval accuracy necessitates further improvements aimed at marine-time awareness. Under this circumstance, this study aimed to develop a machine learning-based SAR wave retrieval algorithm after analyzing the dependences of the azimuthal cut-off wavelength on sea state, i.e., wave and current.

Based on the information provided, 3000 GF-3 images were acquired in QPS mode and collocated with wave simulations, using the WW3 model and currents provided by HYCOM. The GMF CSARMOD-GF was used to invert the wind. Then, the retrievals were

validated against the measurements from the HY-2B scatterometer. The results show a 1.99 m/s RMSE with a 0.82 COR and a 0.27 SI of wind speed. These images were collocated with a wave simulation, using the WW3 model, in which ERA-5 wind was treated as a forcing field. The dependence of WW3-simulated SWH on SAR-derived wind speed, azimuthal cut-off wavelength, and HYCOM current speed was studied. The linear relation between SWH, azimuthal cut-off wavelength, and wind speed was observed at the regular sea state (COR > 0.5). However, the current speed has less influence on the azimuthal cut-off wavelength (COR = 0.11). Based on this finding, the four well-known machine learning methods (i.e., XGBoost, CNN, SVR, and RR) were applied to develop a SAR wave retrieval algorithm through 2000 images collocated with WW3-simulated SWH. In the training process, XGBoost eventually converges to be about 0.30 m RMSE, and the CNN eventually converges, and that is approximately 0.58 m RMSE as the iteration approaches 200. However, the SVR and RR have no iteration process. The validation of SWH inverted from 1000 images by XGBoost against the measurements from the HY-2B altimeter yields a 0.34 m RMSE with a 0.97 COR. In contrast, the RMSE of SWH is 1.10 m, with a 0.77 COR, using the PFSM algorithm; the RMSE of SWH is 0.58 m, with a 0.94 COR, using the CNN; the RMSE of SWH is 0.77 m, with a 0.91 COR, using the RR; and the RMSE of SWH is 0.62 m, with a 0.95 COR, using the SVR algorithm. The analysis results indicated that the XGBoost has the best performance at the SWH retrieval from GF-3 images.

Since 2018, tropical cyclones have been captured by S-1 during the Satellite Hurricane Observation Campaign [16]. In the literature, the wind speed and SWH are explicitly related following the fetch- and duration-limited features inside a tropical cyclone [46]. Future studies will explore machine learning adopted for SAR wave retrieval during tropical cyclones.

Author Contributions: Conceptualization, W.S. and S.L.; methodology, W.S. and S.L.; software, S.L.; validation, W.S., M.H. and A.M.; formal analysis, W.S., M.H. and X.J.; investigation, S.L. and M.H.; resources, W.S.; data curation, X.J. and A.M.; writing—original draft preparation, W.S. and X.J.; writing—review and editing, W.S.; visualization, S.L. and W.S.; supervision, M.H.; project administration, A.M. and W.S.; funding acquisition, W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly supported by the National Key Research and Development Program of China (2023YFE0102400), the National Natural Science Foundation of China (42076238 and 42376174), and the Natural Science Foundation of Shanghai (23ZR1426900).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: We genuinely appreciate the National Satellite Ocean Application Service for providing the Gaofen-3 images and the Haiyang-2B products through an authorized account at <https://osdds.nsoas.org.cn> (accessed on 1 May 2024). The European Centre for Medium-Range Weather Forecasts data can be accessed via <http://www.ecmwf.int> (accessed on 1 May 2024). The General Bathymetric Chart of the Oceans data can be downloaded from <ftp.edcftp.cr.usgs.gov> (accessed on 1 May 2024). Additionally, sea surface current data from the HYbrid Coordinate Ocean Model are open access and can be obtained from <https://www.hycom.org> (accessed on 1 May 2024).

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

References

1. Wei, M.; Shao, W.Z.; Shen, W.; Hu, Y.Y.; Zhang, Y.; Zuo, J.C. Contribution of surface waves to sea surface temperatures in the Arctic Ocean. *J. Ocean Univ. China* **2024**. [CrossRef]
2. Shi, J.; Shao, W.Z.; Shi, S.H.; Hu, Y.Y.; Jiang, T.; Zhang, Y.G. Can sea surface waves be simulated by numerical wave models using the fusion data from remote-sensed winds? *Remote Sens.* **2023**, *15*, 3825. [CrossRef]
3. 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. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 9624–9635. [CrossRef]

4. Xu, Y.; Liu, J.Q.; Xie, L.L.; Sun, C.R.; Liu, J.P.; Xian, D. China-France Oceanography Satellite (CFOSAT) simultaneously observes the typhoon-induced wind and wave fields. *Acta Oceanol. Sin.* **2019**, *38*, 158–161. [[CrossRef](#)]
5. Hao, M.Y.; Shao, W.Z.; Shi, S.H.; Liu, X.; Hu, Y.Y.; Zuo, J.C. Validation of Surface Waves Investigation and Monitoring Data against Simulation by Simulating Waves Nearshore and Wave Retrieval from Gaofen-3 Synthetic Aperture Radar Image. *Remote Sens.* **2023**, *15*, 4402. [[CrossRef](#)]
6. Wang, W.L.; Forget, P.; Guan, C.L. Inversion and assessment of swell wave heights from HF radar spectra in the Iroise Sea. *Ocean Dyn.* **2016**, *66*, 527–538. [[CrossRef](#)]
7. Shao, W.Z.; Li, X.M.; Lehner, S.; Guan, C.L. Development of polarization ratio model for sea surface wind field retrieval from TerraSAR-X HH polarization data. *Int. J. Remote Sens.* **2014**, *35*, 4046–4063. [[CrossRef](#)]
8. Corcione, V.; Grieco, G.; Portabella, M.; Nunziata, F.; Migliaccio, M. A novel azimuth cut-off implementation to retrieve sea surface wind speed from SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3331–3340. [[CrossRef](#)]
9. Hu, Y.Y.; Shao, W.Z.; Shen, W.; Zhou, Y.; Jiang, X.W. Machine learning applied to a dual-polarized Sentinel-1 image for wind retrieval of tropical cyclones. *Remote Sens.* **2023**, *15*, 3948. [[CrossRef](#)]
10. 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](#)]
11. 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 Observ. Remote Sens.* **2017**, *10*, 2123–2134. [[CrossRef](#)]
12. Mouche, A.; Chapron, B. Global C-band Envisat, RADARSAT-2 and Sentinel-1 SAR measurements in copolarization and crosspolarization. *J. Geophys. Res. Oceans* **2015**, *120*, 7195–7207. [[CrossRef](#)]
13. Lu, Y.R.; Zhang, B.; William, P.; Mouche, A.; Wang, H. A C-band geophysical model function for determining coastal wind speed using synthetic aperture radar. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2018**, *11*, 2417–2428. [[CrossRef](#)]
14. 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](#)]
15. 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](#)]
16. 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 Observ. Remote Sens.* **2023**, *16*, 3187–3197. [[CrossRef](#)]
17. 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](#)]
18. Kudryavtsev, V.; Hauser, G.; Caudal, G.; Chapron, B. A semiempirical model of the normalized radar cross-section of the sea surface 1. Background model. *J. Geophys. Res. Oceans* **2003**, *108*, FET 2-1–FET 2-24. [[CrossRef](#)]
19. 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](#)]
20. 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](#)]
21. 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](#)]
22. Hasselmann, K.; Hasselmann, S. On the nonlinear mapping of an ocean wave spectrum into a synthetic aperture radar image spectrum. *J. Geophys. Res.* **1991**, *96*, 10713–10729. [[CrossRef](#)]
23. Mastenbroek, C.; Valk, C.F.de. A semi-parametric algorithm to retrieve ocean wave spectra from synthetic aperture radar. *J. Geophys. Res.* **2000**, *105*, 3497–3516. [[CrossRef](#)]
24. Schulz-Stellenfleth, J.; Lehner, S.; Hoja, D. A parametric scheme for the retrieval of two-dimensional ocean wave spectra from synthetic aperture radar look cross spectra. *J. Geophys. Res.* **2005**, *101*, 297–314. [[CrossRef](#)]
25. 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](#)]
26. Zhang, B.; Perrie, W.; He, Y.J. Validation of RADARSAT-2 fully polarimetric SAR measurements of ocean surface waves. *J. Geophys. Res.* **2010**, *115*, C06031. [[CrossRef](#)]
27. 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](#)]
28. Li, X.M.; Konig, T.; Schulz-Stellenfleth, J.; Lehner, S. Validation and intercomparison of ocean wave spectra inversion schemes using ASAR wave mode data. *Int. J. Remote Sens.* **2010**, *31*, 4969–4993. [[CrossRef](#)]
29. Zhu, S.; Shao, W.Z.; Armando, M.; Shi, J.; Sun, J.; Yuan, X.Z.; Hu, J.C.; Yang, D.K.; Zuo, J.C. Evaluation of Chinese quad-polarization Gaofen-3 SAR wave mode data for significant wave height retrieval. *Can. J. Remote Sens.* **2018**, *44*, 588–600. [[CrossRef](#)]
30. Pleskachevsky, A.; Tings, B.; Wiehle, S.; Imber, J.; Jacobsen, S. Multiparametric sea state fields from synthetic aperture radar for maritime situational awareness. *Remote Sens. Environ.* **2022**, *280*, 113200. [[CrossRef](#)]
31. Wang, H.; Yang, J.S.; Lin, M.S.; Li, W.W.; Zhu, J.H.; Ren, L.; Cui, L.M. Quad-polarimetric SAR sea state retrieval algorithm from Chinese Gaofen-3 wave mode images via deep learning. *Remote Sens. Environ.* **2022**, *273*, 112969. [[CrossRef](#)]

32. Kudryavtsev, V.; Fan, S.; Zhang, B.; Mouche, A.; Chapron, B. On quad-polarized SAR measurements of the ocean surface. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8362–8370. [[CrossRef](#)]
33. Zhao, X.B.; Shao, W.Z.; Hao, M.Y.; Jiang, X.W. Novel approach to wind retrieval from Sentinel-1 SAR in tropical cyclones. *Can. J. Remote Sens.* **2023**, *49*, 2254839. [[CrossRef](#)]
34. Johannessen, J.A.; Chapron, B.; Collard, F.; Kudryavtsev, V.; Mouche, A.; Akimov, D.; Dagestad, K. Direct ocean surface velocity measurements from space: Improved quantitative interpretation of Envisat ASAR observations. *Geophys. Res. Lett.* **2008**, *35*, L22608. [[CrossRef](#)]
35. Hansen, M.W.; Collard, F.; Dagestad, K.F.; Johannessen, J.A.; Chapron, B. Retrieval of sea surface range velocities from Envisat ASAR Doppler centroid measurements. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3582–3592. [[CrossRef](#)]
36. Wang, L.H.; Zhou, Y.X.; Ge, J.Z.; Johannessen, J.A.; Shen, F. Mapping sea surface velocities in the Changjiang coastal zone with advanced synthetic aperture radar. *Acta Oceanol. Sin.* **2014**, *33*, 141–149. [[CrossRef](#)]
37. Stopa, J.E.; Arduin, F.; Chapron, B.; Collard, F. Estimating wave orbital velocity through the azimuth cut-off from space-borne satellites. *J. Geophys. Res.* **2015**, *120*, 761–7634. [[CrossRef](#)]
38. Hao, M.Y.; Shao, W.Z.; Yao, R.; Zhang, Y.G.; Jiang, X.W. Improvement of quad-polarized velocity bunching modulation transfer function by C-band Gaofen-3 SAR. *Remote Sens. Lett.* **2023**, *14*, 968–978. [[CrossRef](#)]
39. Robertson, R.; Padman, L.; Levine, M.D. A correction to the baroclinic pressure gradient Term in the Princeton Ocean Model. *J. Atmos. Ocean. Technol.* **2001**, *18*, 1068–1075. [[CrossRef](#)]
40. Hu, Y.Y.; Shao, W.Z.; Li, J.; Zhang, C.L.; Cheng, L.Q.; Ji, Q.Y. Short-term variations in water temperature of the Antarctic surface layer. *J. Mar. Sci. Eng.* **2022**, *10*, 287. [[CrossRef](#)]
41. Yang, Z.H.; Shao, W.Z.; Ding, Y.; Shi, J.; Ji, Q.Y. Wave simulation by the SWAN model and FVCOM considering the sea-water level around the Zhoushan Islands. *J. Mar. Sci. Eng.* **2020**, *8*, 783. [[CrossRef](#)]
42. Chassignet, E.P.; Smith, L.T.; Halliwell, G.R.; Bleck, R. North Atlantic simulations with the HYbrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity. *J. Phys. Oceanogr.* **2003**, *3*, 2504–2526. [[CrossRef](#)]
43. Sun, J.; Guan, C.L. Parameterized first-guess spectrum method for retrieving directional spectrum of swell-dominated waves and huge waves from SAR images. *Chin. J. Oceanol. Limnol.* **2006**, *24*, 12–20.
44. Elfouhaily, T.; Chapron, B.; Katsaros, K. A unified directional spectrum for long and short wind-driven waves. *J. Geophys. Res. Atmos.* **1997**, *102*, 15781. [[CrossRef](#)]
45. Shao, W.Z.; Zhang, Z.; Li, X.F.; Li, H. Ocean wave parameters retrieval from Sentinel-1 SAR imagery. *Remote Sens.* **2016**, *8*, 707. [[CrossRef](#)]
46. Hwang, P.A. Fetch and duration-limited nature of surface wave growth inside tropical cyclones: With applications to air-sea exchange and remote sensing. *J. Phys. Oceanogr.* **2016**, *46*, 41–56. [[CrossRef](#)]
47. Shao, W.Z.; Zhu, S.; Zhang, X.P.; Gou, S.P.; Zhao, L.B. Intelligent wind retrieval from Chinese Gaofen-3 SAR imagery in quad-polarization. *J. Atmos. Ocean. Technol.* **2019**, *36*, 2121–2137. [[CrossRef](#)]
48. Song, T.R.; Yan, Q.S.; Fan, C.Q.; Meng, J.M.; Wu, Y.Q.; Zhang, J. Significant wave height retrieval using XGBoost from polarimetric Gaofen-3 SAR and feature importance analysis. *Remote Sens.* **2023**, *15*, 149. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.