

Article

Significant Wave Height Retrieval in Tropical Cyclone Conditions Using CYGNSS Data

Xiangyang Han ^{1,2,3} , Xianwei Wang ^{1,2,3,*} , Zhi He ^{1,2,3}  and Jinhua Wu ⁴ 

¹ School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China; hanxy35@mail2.sysu.edu.cn (X.H.); hezh8@mail.sysu.edu.cn (Z.H.)

² Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

³ Guangdong Provincial Engineering Research Center for Public Security and Disasters, Guangzhou 510275, China

⁴ China Railway Siyuan Survey and Design Group Co., Ltd., Wuhan 430063, China; wjh368903@gmail.com

* Correspondence: wangxw8@mail.sysu.edu.cn

Abstract: The retrieval of global significant wave height (SWH) data is crucial for maritime navigation, aquaculture safety, and oceanographic research. Leveraging the high temporal resolution and spatial coverage of Cyclone Global Navigation Satellite System (CYGNSS) data, machine learning models have shown promise in SWH retrieval. However, existing models struggle with accuracy under high-SWH conditions and discard a significant number of such observations due to low quality, which limits their effectiveness in global SWH retrieval, particularly for monitoring tropical cyclone (TC) events. To address this, this study proposes a daily global SWH retrieval framework through the enhanced eXtreme Gradient Boosting model (XGBoost-SC), which incorporates Cumulative Distribution Function (CDF) matching to introduce prior distribution information and reduce errors for SWH values exceeding 3 m. An enhanced loss function is employed to improve accuracy and mitigate the distribution bias in low-SWH retrieval induced by CDF matching. The results were tested over one million sample points and validated against the European Centre for Medium-Range Weather Forecasts (ECMWF) SWH product. With the help of CDF matching, XGBoost-SC outperformed all models, significantly reducing RMSE and bias while improving the retrieval capability for high SWHs. For SWH values between 3–6 m, the RMSE and bias were 0.94 m and −0.44 m, and for values above 6 m, they were 2.79 m and −2.0 m. The enhanced performance of XGBoost-SC for large SWHs was further confirmed in TC conditions over the Western North Pacific and in the Western Atlantic Ocean. This study provides a reference for large-scale SWH retrieval, particularly under TC conditions.

Keywords: Cyclone Global Navigation Satellite System (CYGNSS); Global Navigation Satellite System-reflectometry (GNSS-R); tropical cyclone; significant wave height



Citation: Han, X.; Wang, X.; He, Z.; Wu, J. Significant Wave Height Retrieval in Tropical Cyclone Conditions Using CYGNSS Data. *Remote Sens.* **2024**, *16*, 4782. <https://doi.org/10.3390/rs16244782>

Academic Editor: Yuriy Kuleshov

Received: 8 November 2024

Revised: 18 December 2024

Accepted: 18 December 2024

Published: 22 December 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

The global retrieval of significant wave height (SWH) data has profound significance for maritime navigation, aquaculture safety, and oceanographic research. Buoy systems and satellite radar altimeters [1,2] are constrained by their monitoring scope or temporal resolution, restricting their ability to rapidly and effectively measure the global SWH.

In 1993, Martín-Neira first proposed an ocean radar altimeter technology to retrieve the SWH using Global Navigation Satellite System-reflectometry (GNSS-R) [3]. Subsequently, various shipborne and shore-based SWH experiments were conducted. However, these techniques were limited in spatial coverage and temporal resolution and couldn't achieve rapid global observation. The successful launch of satellite missions such as TechDemoSat-1 (TDS-1) [4], Cyclone GNSS (CYGNSS) [5], and BuFeng-1 A/B [6] provided opportunities to address these limitations. Spaceborne GNSS-R, particularly CYGNSS, has enabled rapid and efficient global observations of SWH. Additionally, it has been applied to retrieve

other geophysical parameters, such as soil moisture [7], sea surface wind speed [8], oil spill detection [9], and sea surface height (SSH) [10].

Inspired by radar altimetry, Clarizia [11] proposed an approach to estimate SWH using the relationship between SWH and the leading-edge slope (LES) of the Integrated Delay Waveform (IDW). Similarly, Peng and Jin [12] developed an empirical function relating the square root of the signal-to-noise ratio (SNR) to SWH for global estimation. Subsequently, Yang and Jin [13] established a polynomial relationship between SWH, LES, and the delay-Doppler map average (DDMA) to retrieve the SWH. In 2022, Bu and Yu [14] proposed four variables to construct a relation function with SWH, i.e., the LES of a normalized integrated delay waveform (NIDW), LES of a normalized central delay waveform (NCDW), trailing-edge slope (TES) of NCDW, and leading-edge waveform and summation (LEWS) of a NCDW. Among these variables, the NCDW LES achieved the best retrieval performance with a root mean square error (RMSE) of 0.50 m and MAPE of 20.2%. Subsequently, they used DDMA and LES to construct two empirical models and applied the minimum variance estimator (MVE) to combine these models to enhance prediction performance [15]. This study achieved a promising retrieval result with an RMSE of 0.43 m and an MAPE of 25.5%. However, due to strict quality control measures, the study lacked sufficient sample points for large-scale SWH observations and did not account for SWH values exceeding 6 m. In summary, the retrieval accuracy of a single geophysical model function (GMF) is limited. Characterizing the relationship between SWH and complex electromagnetic scattering, especially in high-SWH ranges, is particularly challenging [16].

With the continuous development of artificial intelligence technology, machine learning (ML) algorithms have been widely applied in remote sensing information processing. In recent years, researchers have applied ML to spaceborne GNSS-R for SWH retrieval. It is important to note that CYGNSS does not directly measure sea surface SWH but instead constructs the retrieval model based on the nonlinear relationship between SWH and retrieval parameters. Wang et al. [17] used an artificial neural network (ANN) to construct the relationship between 10 variables and SWH from the European Center for Medium-range Weather Forecasts (ECMWF). It reduced the errors in high SWH values by using an oversampling technique and a soft data filter. Consequently, the study achieved an RMSE of 0.45 m, but the filter resulted in the loss of 16.0% of the observations. Wang et al. [18] demonstrated the superior performance of the Bagging Tree (BT) model for retrieving SWH, achieving an RMSE of 0.50 m, outperforming various other ML models. Bu et al. [19] proposed a deep convolutional neural network (DCNN) model for retrieving swells and wind waves. Compared with the BT model, their model improved the RMSE for swell and wind waves by 26.74% and 27.41%, respectively. Qiao et al. [20] proposed a transformer-based network called WaveTransNet and achieved an RMSE of 0.443 m. However, these papers neither effectively address the errors in high SWH values nor have the capability for global daily retrieval.

Current GNSS-R SWH retrieval models face challenges in accurately capturing high SWH values, mainly due to the scarcity of high-value samples and the complex reflection conditions [18,19]. To the best of our knowledge, no existing studies have specifically addressed the retrieval accuracy issue with high SWHs, particularly during tropical cyclones (TC). In GNSS-R wind speed retrievals, the Cumulative Distribution Function (CDF) matching has been successfully applied to mitigate retrieval bias. Clarizia and Ruf [21] used the CDF of the observable and the ground-truth reference winds to construct a wind speed retrieval model. Reynolds et al. [22] applied the CDF matching after the proposed ANN retrieval model, achieving the lowest RMSE at the time among the literature for wind speed retrievals from CYGNSS. However, no research has applied this method to SWH retrieval.

This paper aims to develop a robust global retrieval model and improve the accuracy of spaceborne GNSS-R with high SWHs (>3 m), especially in surging wave and TC conditions. XGBoost is chosen as the fundamental model [23] and is enhanced by the CDF matching and an improved loss function. The CDF matching is applied to escalate underestimates of

high SWHs (>3 m), while the improved loss function is applied to reduce the errors caused by the CDF matching with moderate and low SWHs (0–3 m). Furthermore, in addition to DDM characteristics, this study incorporates additional auxiliary variables (receiver-related and geometry-related) to assist the ML model in establishing the relationship between SWH and the retrieval parameters.

2. Data

2.1. Datasets

The GNSS-R datasets used in this paper are from the CYGNSS (L1B) v3.1 product in 2022. CYGNSS consisted of eight microsatellites, each carrying a four-channel bistatic radar receiver capable of receiving GPS reflections from 40°S to 40°N (Figure 1a). It provides Earth surface data with a high spatiotemporal resolution at around 25 km and 0.5 s [24].

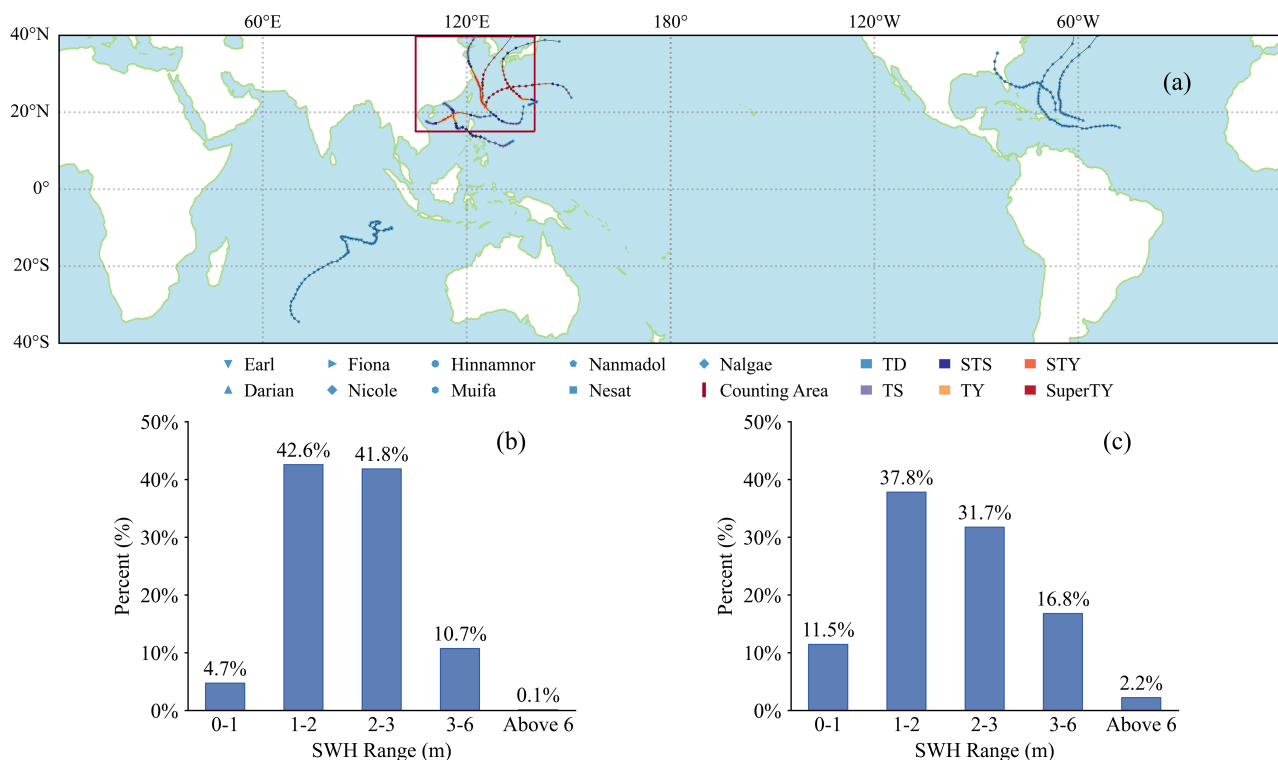


Figure 1. Distribution of the dataset (a), with the blue area indicating the coverage of CYGNSS data over the sea and the green area representing coastal seas, defined as regions 50 km away from the coastline. Histogram of SWH distribution for the overall dataset (b). Histogram of SWH distribution for the five typhoons in the testing dataset (c). ‘TD’, ‘TS’, ‘STS’, ‘TY’, ‘STY’, and ‘SuperTY’ denote tropical depression, tropical storm, severe tropical storm, tropical cyclone, severe typhoon, and super typhoon, respectively. The red box represents the area of statistics shown in (c).

The reference SWH is provided by fifth-generation reanalysis data from the ECMWF (ERA-5) with a grid resolution of $0.5^{\circ} \times 0.5^{\circ}$ and a temporal resolution of 1 h. The ECMWF reanalysis data have been widely used in constructing geophysical parameters models, which have provided global atmospheric, land, and oceanic geophysical parameters from 1950 to the present [25]. There are numerous swell-related parameters provided by the ECMWF. In this paper, the significant height of combined wind waves and swell is selected as the truth value of SWH which is closer to the real sea state [14,18].

To build a reliable dataset and obtain accurate SWH truth values for specular points, the matchup of the CYGNSS L1 product with the ECMWF SWH data was performed by spatial bilinear and temporal linear interpolation. A geographically partitioned random down-sampling model is applied to minimize computational resource expenditure and cap the total volume of data at nearly 5 million samples. Additionally, to ensure the limited

samples more effectively cover the areas prone to TC impacts [26], the down-sampling weights are set to 1:6:1 for the equatorial region, 5°N to 30°N, and 30°N to 40°N, with the same approach applied in the Southern Hemisphere.

All the datasets were segmented for training the model from 1 January to 31 May 2022, validating the model from 1 June to 31 August and testing the model from 1 September to 31 December. The partitioning ratio for these three datasets was set at 3:1:1, and the temporal segment helps mitigate the impact of data leakage caused by temporal correlation between datasets on model accuracy. The average minimum SWH and median values in the dataset are 0.17 m and 2.04 m, respectively. The maximum daily SWH fluctuates significantly over time, with its minimum value being 4.11 m and its maximum value reaching 13.45 m (Figure 2).

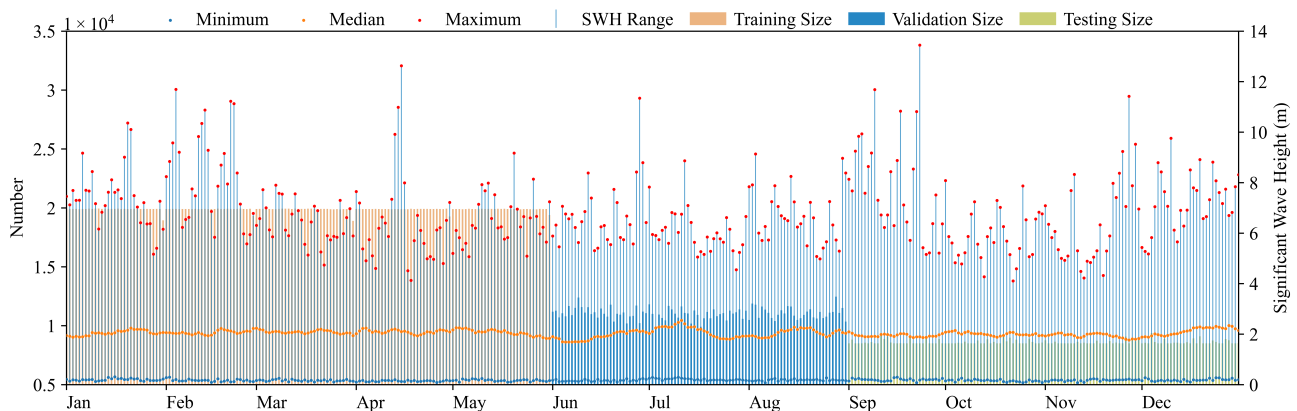


Figure 2. Daily data volume, maximum, minimum, and median values for the training, validation, and test datasets.

As a satellite with the scientific motivation to observe TC [5], it is regrettable that CYGNSS cannot accurately obtain high SWH values [18,19], especially during TCs. There were 13 typhoons that occurred in 2022 over the Western North Pacific (WNP) ocean [27]. Five typhoons were selected to test our model, i.e., Hinnamnor, Muifa, Nanmadol, Nesat, and Nalgae (Figure 1a, Table 1). During these TC periods, 19% of the SWHs were above 3 m and the highest value reached 10.83 m (Figure 1c). In contrast, only 10.8% of the SWHs were above 3 m in the global test data (Figure 1b). Therefore, it is crucial to enhance the retrieval accuracy of high SWHs, especially in TC conditions.

Table 1. Detailed information of selected five typhoons.

Typhoons	Start Time	End Time	Max SWH	Data Volume
Hinnamnor	1 September 2022	6 September 2022	9.93 m	1839
Muifa	10 September 2022	14 September 2022	6.14 m	1442
Nanmadol	15 September 2022	19 September 2022	10.83 m	1110
Nesat	16 October 2022	18 October 2022	7.30 m	611
Nalgae	30 October 2022	2 November 2022	7.09 m	1432

2.2. Feature Selection

This paper utilizes features related to the delay-Doppler map (DDM) and other auxiliary variables provided by the CYGNSS L1B v3.1 product. DDM is a 17-discrete time delay and 11-discrete Doppler frequency 2-D power diagram, with a delay bandwidth of 0.25 chip units and a Doppler frequency bandwidth of 500 Hz. When its specular point (SP) is located on the sea surface, its image presents a classic horseshoe shape [28]. Different sea surface SWHs cause changes in sea surface roughness, which affects the DDM power values (Figure 3). When the sea is calm, good specular reflection conditions are formed, resulting in higher power values in the red box in Figure 3a. As the SWH increases, the reflection conditions worsen, leading to weaker power values in the same region (Figure 3b,c).

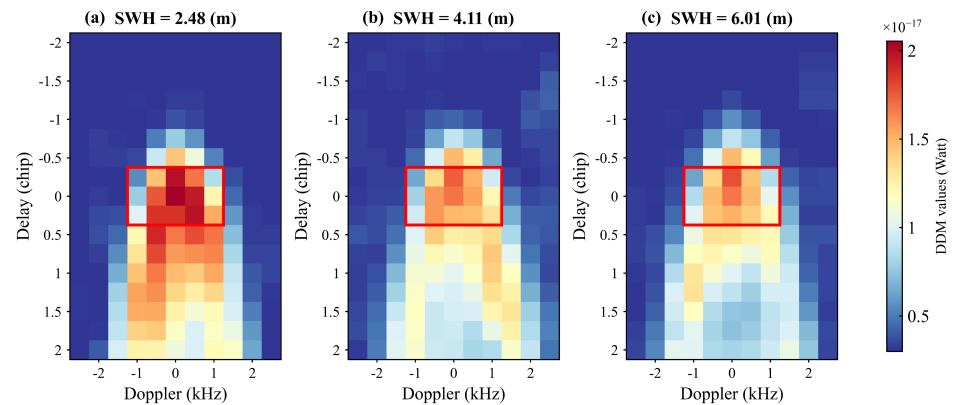


Figure 3. The CYGNSS L1 delay-Doppler map (DDM) under three SWH conditions, (a) SWH = 2.48 m, (b) SWH = 4.11 m, (c) SWH = 6.01 m.

Compared with previous studies [14,15,18], additional waveform features of integral delay waveform (IDW) are incorporated to better characterize the DDM, including LES, TES, LEWS, and trailing-edge waveform summation (TEWS). Other DDM-related features are also considered, such as DDMA, SNR, Scatter Area, Noise Floor, NBRCS, and DDM maximum bin (DDM_peak).

DDMA is the average of $5 \text{ Doppler} \times 3 \text{ delay}$ bin box analog power in DDM, which is shown in Figure 3 with the red line. LES, TES, and NBRCS are also calculated from the $5 \text{ Doppler} \times 3 \text{ delay}$ bin box [11]. Detailed calculation methods for LEWS and TEWS are described in [14,29]. As indispensable features of DDM, SNR [12] and DDM_peak [30] also have the potential to retrieve SWH at the global scale.

Although many feature variables are available for retrieving SWH, the relationship between these variables and SWH becomes increasingly challenging as SWH increases (Figure 4). This complexity also requires the inclusion of auxiliary variables. Two types of auxiliary variables (receiver-related and geometry-related) are incorporated to improve the retrieval accuracy of the ML model. The detailed information of selected feature variables is shown in Table 2.

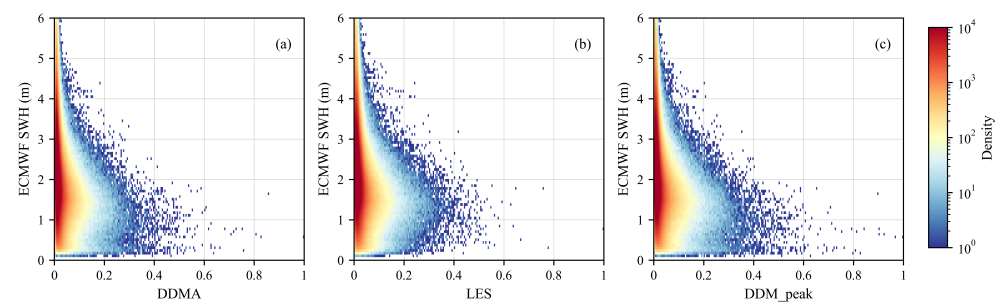


Figure 4. The relationships between the feature variables and SWH. (a) DDMA, (b) LES, (c) DDM_peak.

2.3. Data Preprocessing

The following quality control methods are performed to screen the CYGNSS L1B data.

1. The quality control (QC) flag excludes the observation with specular point less than 25 km from land and other invalid or large errors in DDMs.
2. All the DDM observables and auxiliary parameters (e.g., LES, TES, and sp_rx_gain) must be positive, and all the Nan values are discarded.
3. The nano star tracker attitude status (NST_att_status) should be “OK” (i.e., NST_att_status = 0).
4. The range-corrected gain (RCG) figure of merit (FOM) for the DDM should be larger than or equal to 10.

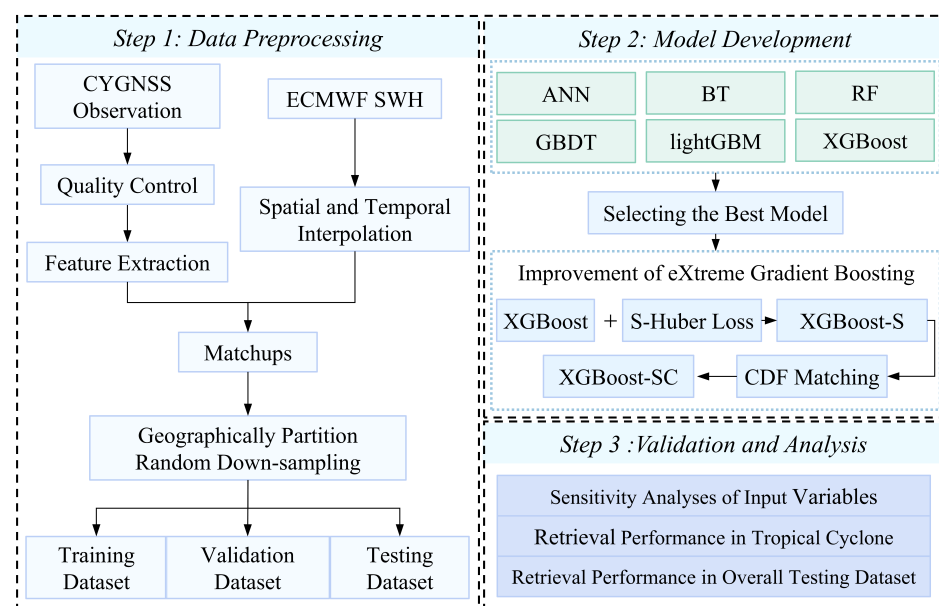
Table 2. List of input variables for ML model.

	Variables	Description
DDM-related	LES	Leading-edge slope
	TES	Trailing-edge slope
	LEWS	Leading-edge waveform summation
	TEWS	Trailing-edge waveform summation
	DDMA	DDM average
	SNR	DDM signal-to-noise ratio
	NBRCS	Normalized bistatic radar cross section
	Scatter Area	Scattering area of NBRCS
	Noise Floor	DDM noise floor
	DDM_peak	The max scatter power of DDM
Receiver-related	gps_erip	GPS effective isotropic radiated power
Geometry-related	sp_rx_gain	Specular point receiver antenna gain
	sp_lon	Specular point longitude
	sp_lat	Specular point latitude
	sp_inc_angle	Specular point incidence angle
	sp_az_body	Specular point azimuth angle
	RCG	Range-corrected gain

3. Methodology

Inspired by the effectiveness of ensemble learning in constructing geophysical parameter models [18], the eXtreme Gradient Boosting (XGBoost) [23] is selected as the base model after comparing with the BT, Artificial Neural Networks (ANN), Random Forests (RF), Gradient Boosting Decision Trees (GBDT), and Light Gradient Boosting Machine (lightGBM) models.

The overall structure of this paper is illustrated in Figure 5. First, we constructed a dataset using ERA-5 SWH data and the CYGNSS L1B product. This dataset was then used to test various machine learning methods and parameter settings, from which the optimal model and input parameters were selected. Subsequently, to address the issue of lower retrieval accuracy for high SWH values, CDF matching was employed to reduce model errors for SWH values above 3 m (XGBoost-SC). Concurrently, the S-Huber loss function (XGBoost-S) was applied to enhance the retrieval accuracy for SWH values in the range of 0 to 3 m.

**Figure 5.** The main structure of this paper.

3.1. XGBoost Model

XGBoost is a highly efficient and scalable ML algorithm known for its performance and speed in data modeling. It has been widely applied in regression and classification tasks [23]. Its high efficiency in optimizing the objective function enhances its robustness and versatility across various applications.

The equation for the prediction of the XGBoost model is given in (1).

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (1)$$

where $\hat{y}_i$ represents the predicted values, K is the number of decision trees, x_i denotes the sample input features, and F is the set of all possible decision trees. The objective function of XGBoost contains two parts: the training loss and regularization term in (2):

$$Obj = \sum_i^N Loss(y_i, \hat{y}_i) + \sum_i^K \Omega(f_k) \quad (2)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (3)$$

where $Loss(y_i, \hat{y}_i)$ is the loss function to measure the residual between $\hat{y}_i$ and y_i (reference value) during each iteration of XGBoost, $\Omega(f_k)$ is a regularization parameter that aids in managing the model's complexity, thereby preventing overfitting and ensuring better generalization to unseen data, T is the number of the leaf node, w is the score of each leaf nodes, γ and λ are the hyper-parameters to control the model structure, and N represents the number of samples.

3.2. XGBoost-SC Model

The fundamental structure of the enhanced XGBoost-SC model is illustrated in Figure 6.

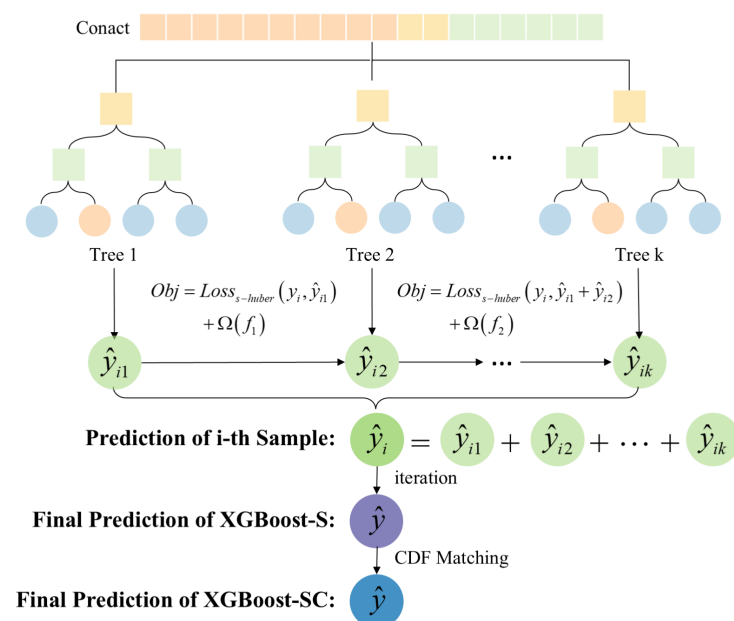


Figure 6. The structure of the XGBoost-SC model.

The accuracy and generalization ability of XGBoost are influenced by the distribution of the sample data. While the Mean Squared Error (MSE, (4)) is commonly used as a loss function for regression tasks in XGBoost, it tends to overly magnify the importance of larger errors. This can lead to a focus on minimizing larger errors at the expense of smaller ones,

especially when handling imbalanced datasets. As a result, the model's robustness may be compromised.

$$Loss_{MSE}(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \quad (4)$$

Unlike the MSE, the Mean Absolute Error (MAE) loss function exhibits superior robustness due to its low sensitivity to outliers. Unfortunately, the MAE is not differentiable at zero, so it cannot be used as a loss function. The Huber loss effectively combines MSE and MAE, mitigating the limitations inherent to both approaches, being mathematically expressed in (5):

$$Loss_{Huber}(y_i, \hat{y}_i) = \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2, & |y_i - \hat{y}_i| \leq \delta \\ \delta(|y_i - \hat{y}_i| - \frac{1}{2}\delta), & \text{otherwise} \end{cases} \quad (5)$$

Here, δ is a threshold that determines the transition between the quadratic and linear loss. When the error in $(-\delta, \delta)$, the Huber loss equals the MSE. When the error in $(-\infty, \delta]$ or $[\delta, \infty)$, it equates to the MAE. To smooth the Huber loss near the zero interval and stabilize the calculation of gradient, it is smoothed into (6).

$$Loss_{Huber}(y_i, \hat{y}_i) = \delta^2 \left(\sqrt{1 + \left(\frac{y_i - \hat{y}_i}{\delta} \right)^2} - 1 \right) \quad (6)$$

Nevertheless, the convergence rate of the Huber loss is slower than the MSE's. In this paper, the S-Huber loss is employed to expedite model convergence by introducing a squared term into the fundamental structure of the Huber loss, as described in (7).

$$Loss_{S-Huber}(y_i, \hat{y}_i) = \delta^2 \left(\sqrt{1 + \left(\frac{y_i - \hat{y}_i}{\delta} \right)^2} - 1 \right) + h(y_i - \hat{y}_i)^2 \quad (7)$$

where h represents the weight coefficient of the square term. As a crucial step in model development, optimizing the hyperparameters (such as h , δ , etc.) of the proposed XGBoost model is conducted through grid search.

As shown in Figure 7, when the model encounters large errors, the S-Huber loss function behaves more like the MAE to promote rapid convergence. For smaller errors, it leans toward the MSE loss, which has a smoother gradient, enhancing accuracy with the low-value SWHs (0–3 m).

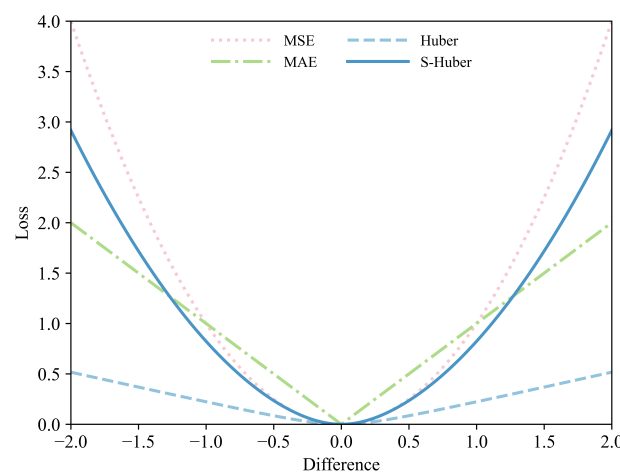


Figure 7. Comparison of different loss function where δ in Huber loss is 0.3 and the δ and h in S-Huber loss are 0.3 and 0.6, respectively.

In the final step, CDF matching is employed to reduce the errors of XGBoost-S in the high-SWH range (>3 m). It assumes that the Probability Density Function (PDF) of the training dataset is almost equal to the testing dataset's. An interpolated function ($F_{cdf}(x, y)$) is constructed to map the XGBoost-S predictions of the testing dataset ($\hat{y}_{test}$) to their corresponding CDF ($CDF_{\hat{y}_{test}}$).

$$F_{cdf} = f(CDF_{\hat{y}_{test}}, \hat{y}_{test}) \quad (8)$$

This function will be used to obtain the interpolated reference values (y_{train_CDF}) of the training dataset, which conform to the CDF of the model retrieval results ($\hat{y}_{test}$).

$$y_{train_CDF} = F_{cdf}(CDF_{y_{train}}) \quad (9)$$

Next, a polynomial function ($F_{corr}(x, y)$) is employed to fit the difference between the interpolated reference values and the real reference SWH of the training data. The coefficients of the function are optimized through the least squares method. The order of it, setting to 5, is ascertained through empirical analysis.

$$F_{corr} = f(y_{train_CDF}, y_{train}) \quad (10)$$

Finally, the model predictions of XGBoost-SC are derived using this polynomial function ($F_{corr}(x, y)$) along with the XGBoost-S predictions of the testing dataset, ensuring that they follow the probability distribution of the training dataset.

3.3. Experiment Designs and Model Evaluation

The effectiveness of the retrieval parameters and models was assessed through four comparison experiments involving ANN, RF, BT, GBDT, lightGBM, and XGBoost using the global testing data from 1 September to 31 December 2022. Four settings of the parameters were used for each model as follows:

1. *Setting 1*: DDM-related variables, 10 parameters.
2. *Setting 2*: DDM-related variables and receiver-related variables, 12 parameters.
3. *Setting 3*: DDM-related variables and geometry-related variables, 15 parameters.
4. *Setting 4*: DDM-related variables, receiver-related variables, and satellite geometry-related variables, 17 parameters.

Four evaluation metrics were computed, i.e., the bias, RMSE, correlation coefficient (R), and Mean Absolute Percentage Error (MAPE).

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (11)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (13)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (14)$$

where n is the volume of the testing dataset, y_i and $\hat{y}_i$ are the reference values and estimated SWH values by ML models, and $\bar{y}$ is the average of y_i .

4. Results

4.1. SWH Retrievals in the Overall Testing Dataset

XGBoost achieves the best performance with the lowest RMSE and the highest R-value across the four experiment settings compared with the ANN, RF, BT, GBDT and lightGBM models (Table 3), and also allows users to customize the loss function for various applications. Therefore, XGBoost is chosen as the fundamental model, and Setting 4 is applied for further improvement.

Table 3. Retrieval performance of ML models under four experiment settings.

Methods	Setting 1		Setting 2		Setting 3		Setting 4	
	RMSE (m)	R	RMSE (m)	R	RMSE (m)	R	RMSE (m)	R
ANN	0.65	0.46	0.68	0.51	0.58	0.60	0.55	0.65
RF	0.67	0.39	0.64	0.48	0.55	0.66	0.54	0.67
BT	0.66	0.43	0.64	0.50	0.55	0.66	0.54	0.67
GBDT	0.64	0.47	0.62	0.51	0.57	0.62	0.56	0.64
lightGBM	0.63	0.49	0.61	0.55	0.53	0.68	0.53	0.69
XGBoost	0.63	0.50	0.60	0.56	0.52	0.70	0.51	0.71

Using Setting 4 as the input variables, the improved S-Huber loss function was applied to XGBoost, and the results (XGBoost-S) were further filtered using CDF matching to obtain the XGBoost-SC results. We also used DDMA and LES to construct two empirical models and applied the MVE to combine them [15], enhancing prediction performance and enabling a comparison between the empirical models and the ML-based models. XGBoost-S performs better and has a smaller RMSE, less bias, and higher R-values than the other models in the global ocean (Table 4). The improved S-Huber loss function and CDF matching algorithm together greatly reduce the RMSE and bias and have the smallest RMSE for SWH above 3 m. Meanwhile, XGBoost-S achieves the best performance with the smallest RMSE of 0.51 and bias of 0.02.

Table 4. The retrieval performance in Setting 4 (S represents S-Huber loss and C represents CDF matching).

Methods	All Testing Dataset				0–3 m		3–6 m		Above 6 m	
	RMSE (m)	Bias (m)	R	MAPE (%)	RMSE (m)	Bias (m)	RMSE (m)	Bias (m)	RMSE (m)	Bias (m)
MVE *	0.70	0.05	0.27	32.6	0.55	0.20	1.42	−1.31	4.73	−4.64
ANN	0.55	0.05	0.65	23.9	0.47	0.14	1.01	−0.81	3.45	−3.23
RF	0.54	0.06	0.67	21.0	0.45	0.15	0.99	−0.76	3.55	−3.24
BT	0.54	0.06	0.67	20.8	0.45	0.14	0.99	−0.77	3.56	−3.25
GBDT	0.56	0.06	0.64	24.3	0.46	0.16	1.06	−0.90	3.80	−3.61
lightGBM	0.53	0.05	0.69	21.7	0.44	0.14	0.99	−0.81	3.55	−3.30
XGBoost	0.51	0.06	0.71	20.2	0.43	0.14	0.94	−0.73	3.26	−2.96
XGBoost-S	0.51	0.02	0.71	19.5	0.42	0.11	0.98	−0.77	3.34	−3.05
XGBoost-SC	0.56	0.06	0.63	20.7	0.50	0.12	0.94	−0.44	2.79	−2.00

* The MVE was conducted using DDMA and LES.

The density scatter plots in Figure 8a,b show that the SWHs retrieved by the XGBoost-S and XGBoost-SC models are well comparable to the ECMWF data within the 0–3 m range. As the SWH increases, the XGBoost-S model tends to underestimate values above 3 m, while XGBoost-SC mitigates this underestimation with CDF matching.

Furthermore, the histograms of SWH in Figure 9a–c show that the XGBoost-SC model has better agreement with the ECMWF data than the XGBoost-S model in each value range. In particular, the proportions of SWH above 3 m are 9.2%, 4.7%, and 10.3% for the ECMWF, XGBoost-S, and XGBoost-SC models, respectively.

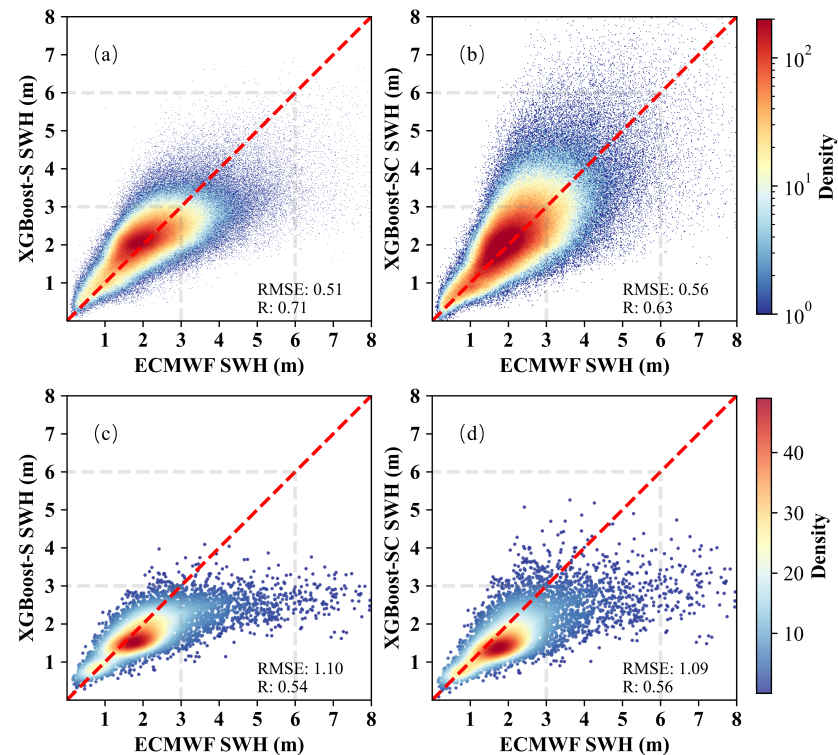


Figure 8. Scatter density plots of SWH between the ECMWF and the XGBoost-S and XGBoost-SC models in the overall testing dataset (a,b) and in the WNP during the five TCs from September to November in 2022 (c,d). The red line is the 1:1 line. (S represents S-Huber loss, and C represents the CDF matching).

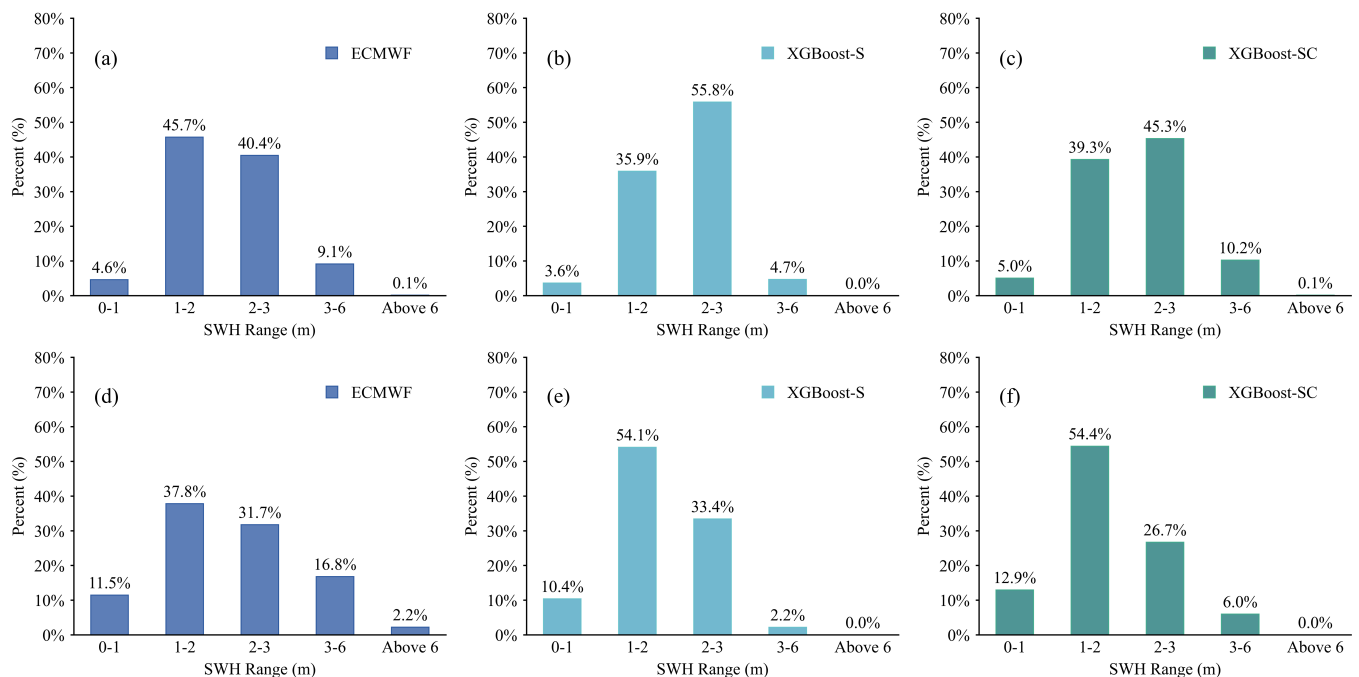


Figure 9. The histograms of SWHs by the ECMWF, XGBoost-S, and XGBoost-SC models in the overall testing dataset (a–c) and in the WNP during the five TCs from September to November in 2022 (d–f).

Figure 10 illustrates the global SWHs by the ECMWF, BT, XGBoost-S, and XGBoost-SC on 24 December 2022. As shown in Figure 10a, the atmospheric circulation caused higher SWHs in the South Indian Ocean, North Atlantic, and North Pacific regions than

in other areas [31]. Most of those high SWHs were also captured by the XGBoost-SC model (Figure 10g), but fewer of them were captured by the BT and XGBoost-S models (Figure 10c,e). The proportions of high SWHs above 3 m were 20.4% by the ECMWF and 12.6% by the XGBoost-SC model, but 7.8% and 7.2% by the BT and XGBoost-S models. The XGBoost-SC model again showed better performance in estimating the high SWHs than the BT and XGBoost-S models.

Figure 11 illustrates the global daily retrieval accuracy of significant wave heights (SWH) above 3 m for BT, XGBoost-S, and XGBoost-SC in the entire testing dataset. From a time series perspective, XGBoost-SC demonstrates higher accuracy in daily high-SWH retrievals, with its RMSE series showing less fluctuation compared with the other two models. The daily retrieval bias for XGBoost-SC remains around -0.6 m, while for the BT and XGBoost-S models, it fluctuates around -0.9 m. This trend aligns with the results in Table 4, further validating XGBoost-SC's superior performance in retrieving high SWHs (>3 m). It highlights the model's enhanced accuracy under challenging oceanic conditions, such as those seen during TCs.

4.2. SWH Retrievals in TC Conditions

The retrieval results of BT, XGBoost-S, and XGBoost-SC during the five typhoons are compared with the SWHs by the ECMWF. The BT model was selected as the baseline to verify the improvement of the models in TC conditions because of its superior performance, as reported in [18].

The XGBoost model has a smaller RMSE, less bias, and higher R-values than the BT model for SWH retrievals in TC conditions, while the XGBoost-SC model further reduces the RMSE and bias of SWHs above 3 m (Table 5). The underestimates of SWHs above 3 m by the XGBoost-S model are greatly mitigated by the XGBoost-SC model (Figure 8c,d). The proportion of SWHs above 3 m is only 2.2% by the XGBoost-S model and increases to 6.0% by the XGBoost-SC model, although it is still much less than that (19.0%) by the ECMWF (Figure 9d–f).

Table 5. Retrieval performance in the WNP during the five TCs in 2022 (S represents S-Huber loss, and C represents CDF matching).

Methods	Data in Selected Typhoons				0–3 m		3–6 m		Above 6 m	
	RMSE (m)	Bias (m)	R	MAPE (%)	RMSE (m)	Bias (m)	RMSE (m)	Bias (m)	RMSE (m)	Bias (m)
BT	1.17	−0.48	0.45	27.9	0.50	−0.10	1.86	−1.68	5.12	−4.97
XGBoost	1.07	−0.42	0.57	26.5	0.60	−0.20	2.27	−2.17	4.73	−4.59
XGBoost-S	1.10	−0.47	0.54	26.5	0.48	−0.13	1.75	−1.58	4.84	−4.71
XGBoost-SC	1.09	−0.50	0.56	27.2	0.55	−0.20	1.69	−1.46	4.56	−4.41

Figure 12 shows that the SWHs retrieved by the BT and XGBoost-S models in the TC centers were much lower than those by the ECMWF. The improved XGBoost-SC model retrieved higher SWHs than the BT and XGBoost-S models, but still lower than those by the ECMWF. During the five TCs from September to November in the WNP, the proportions of SWHs above 3 m by the ECMWF were 50.3%, 16.4%, 39.5%, 62.1%, and 38.2% for the TC Hinnamnor, Muifa, Nanmadol, Nesat, and Nalgae, respectively. The BT model only retrieved 1.7%, 0.8%, 2.4%, 7.8%, and 7.7%, respectively, and the improved XGBoost-SC model retrieved 10.3%, 4.9%, 15.9%, 14.7%, and 13.6%, respectively (Figure 13).

Besides the five tropical cyclones (TCs) in the WNP, Hurricane Earl, Fiona, and Nicole and Severe Tropical Cyclone Dorian were selected to further validate the global applicability of XGBoost-SC (Figure 14). Compared with the TCs in the WNP, Hurricane Earl, Fiona, and Nicole in the western Atlantic Ocean generated higher waves and had a more concentrated impact area. In contrast, Severe Tropical Cyclone Dorian was exceptionally

intense, generating widespread large waves over the South Indian Ocean from 21 December to 25 December 2022.

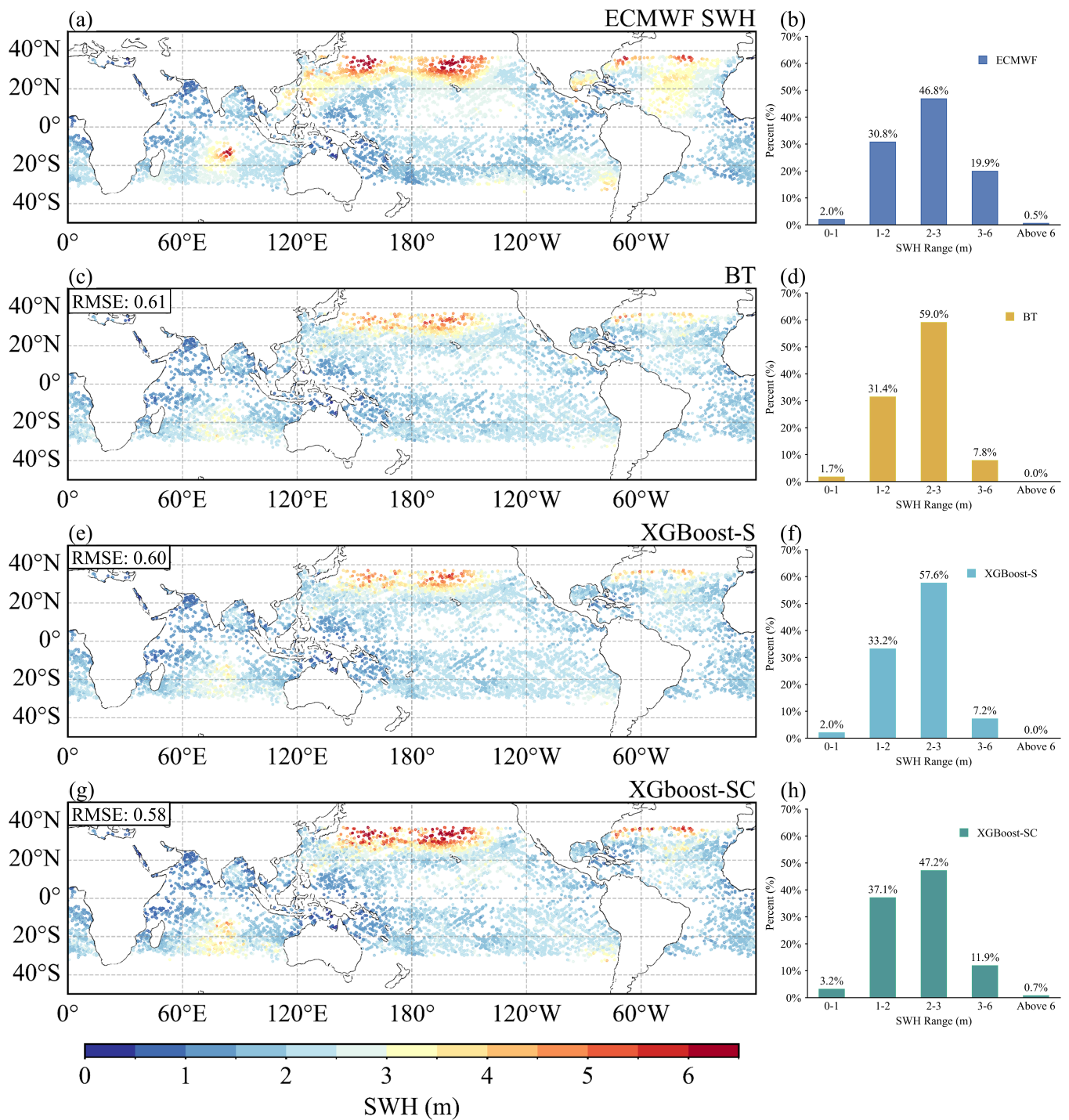


Figure 10. The spatial distribution and histograms of SWHs by the ECMWF (a,b), BT (c,d), XGBoost-S (e,f), and XGBoost-SC (g,h) on 24 December 2022.

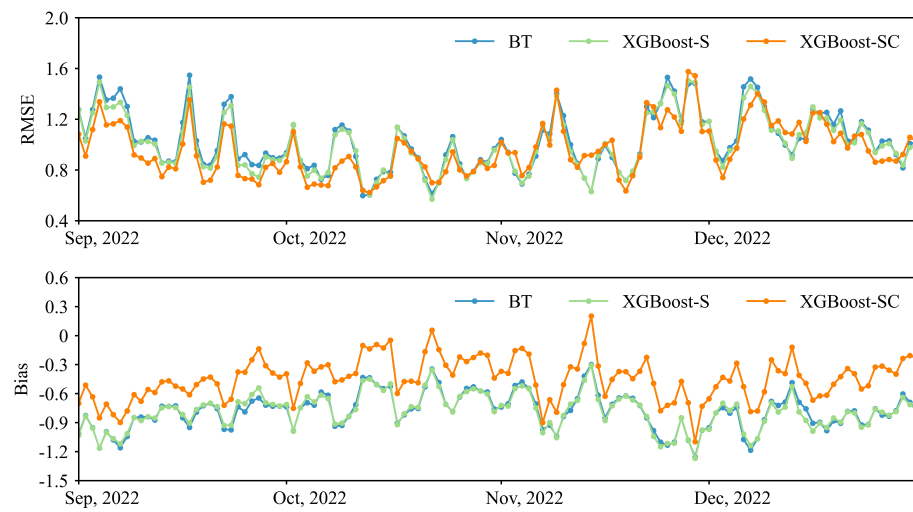


Figure 11. Time series of daily RMSE and bias for SWH retrievals above 3 m by BT, XGBoost-S, and XGBoost-SC.

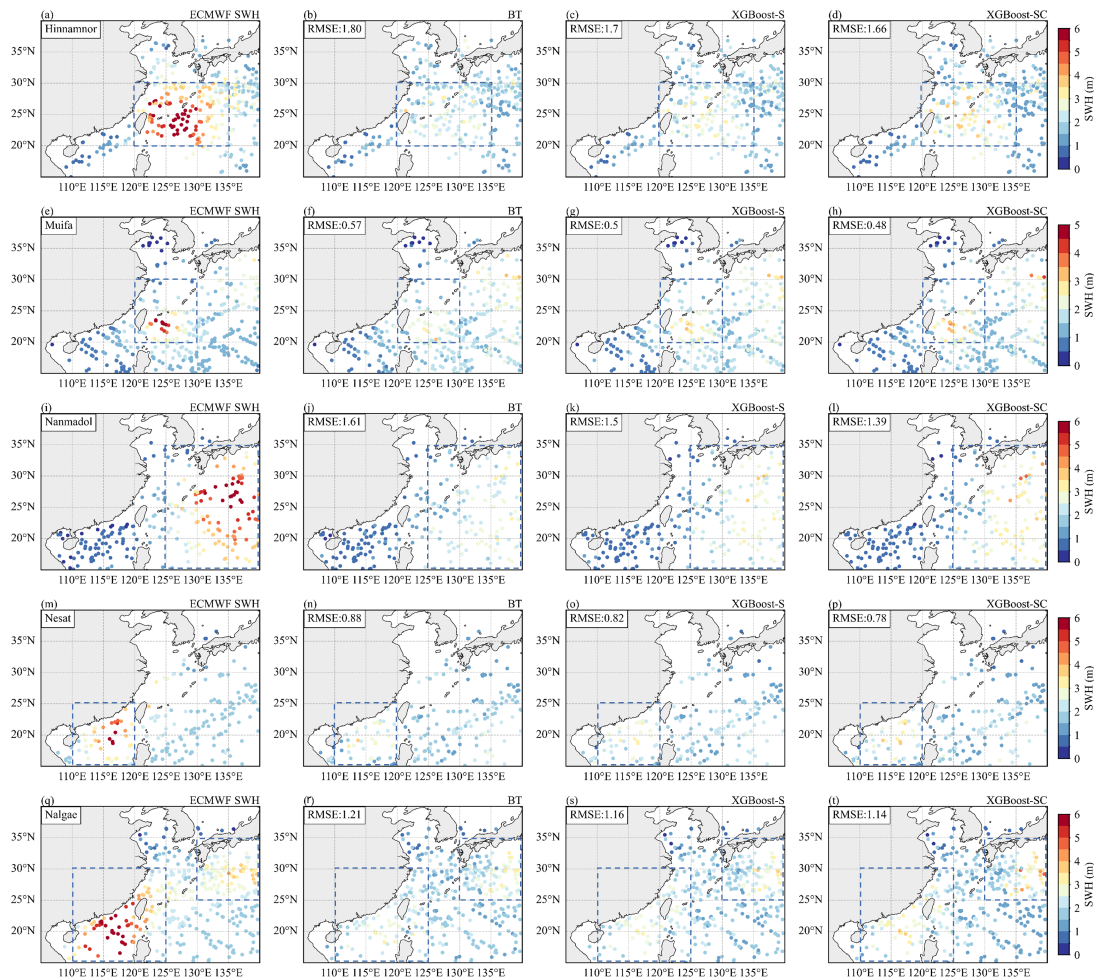


Figure 12. Spatial distributions of SWHs by the ECMWF, BT, XGBoost-S, and XGBoost-SC models in the WNP during the five TCs of Hinnamnor on September 3 (a–d), Muifa on September 11 (e–h), Nanmadol on September 16 (i–l), Nesat on October 17 (m–p), and Nalgae on November 1 (q–t). The blue boxes represent the statistic area.

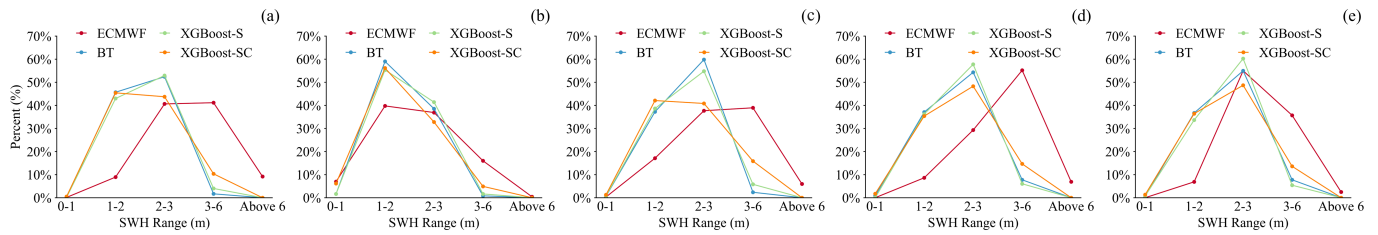


Figure 13. The line chart of SWHs proportion by the ECMWF, BT, XGBoost-S, and XGBoost-SC models in the WNP during the five TCs of Hinnamnor (a), Muifa (b), Nanmadol (c), Nesat (d), and Nalgae (e).

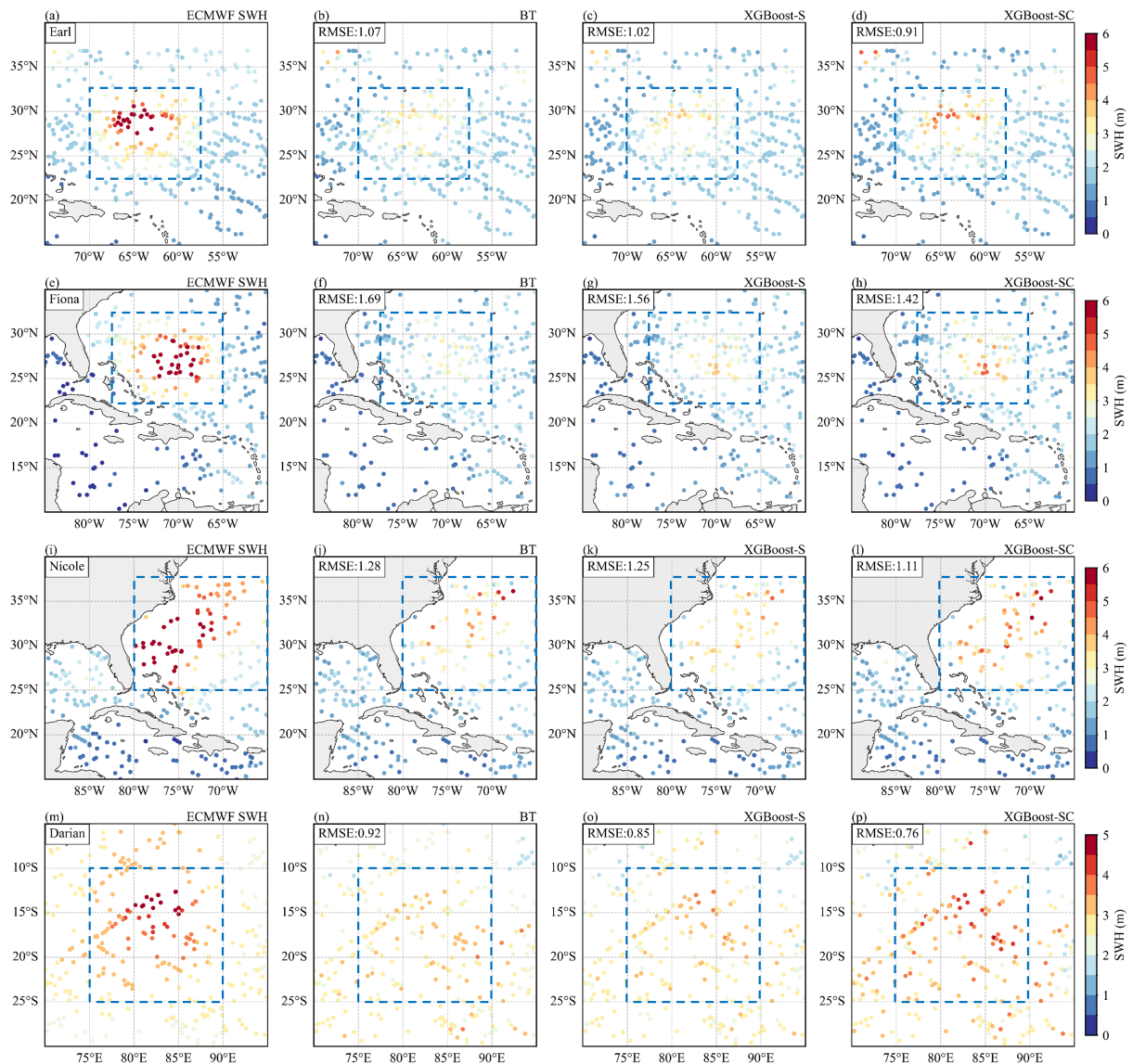


Figure 14. Spatial distributions of SWHs by the ECMWF, BT, XGBoost-S, and XGBoost-SC models during Hurricane Earl on September 8 (a–d), Fiona on September 22 (e–h), and Nicole on November 9 (i–l) and Severe Tropical Cyclone Darian on December 23 (m–p). The blue boxes represent the statistic area.

The improved XGBoost-SC model successfully captured high SWHs around the TC centers, showing results more comparable to those of the ECMWF model, while the BT model significantly underestimated these high SWH values (Figure 14). For Hurricane Earl and Fiona, the ECMWF shows that 50.0% and 63.8% of the SWHs exceeded 3 m. However,

the BT model retrieved only 13.7% and 6.1% for SWHs above 3 m, whereas the improved XGBoost-SC model achieved 26.2% and 24.1%, respectively (Figure 15a,b). The stronger the TC, the better the performance of the XGBoost-SC model, highlighting its versatility for applications in both the WNP and the western Atlantic Ocean.

For Hurricane Nicole, although the BT model retrieved a high proportion of large SWHs, it erroneously estimated higher SWHs in high-latitude regions, in contrast with the reference, which showed Nicole's main impact in the lower-left region of the blue box (Figure 14j). XGBoost-S did not show this misjudgment, but its retrievals did not accurately reflect the high-SWH region. In contrast, XGBoost-SC effectively identified the impact area of Nicole and retrieved higher SWH values (Figure 15c).

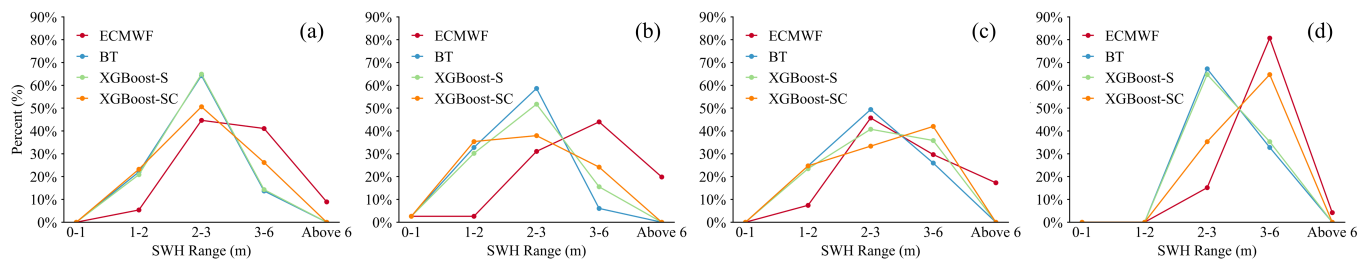


Figure 15. The line chart of SWHs proportionally retrieved by the ECMWF, BT, XGBoost-S, and XGBoost-SC models during Earl (a), Fiona (b), Nicole (c), and Darian (d).

Severe Tropical Cyclone Darian had an even larger area of wave influence, with 84.9% of the SWHs exceeding 3 m, according to the ECMWF. However, the BT model only captured 32.8% for the SWHs above 3 m, while XGBoost-S detected 35.3%. Utilizing the XGBoost-S results along with prior distribution knowledge, XGBoost-SC achieved a more accurate retrieval of 64.7%, accurately reflecting SWH variations across the affected ocean surface (Figure 15d).

5. Discussion

5.1. Sensitivity Analyses of Input Variables

We designed four experiment settings to test the input variables. The RMSE of each model significantly decreased from Setting 1 to Setting 4. For example, in the case of XGBoost, the RMSE reduced from 0.63 to 0.51, and the R increased from 0.50 to 0.71. The addition of geometry-related variables plays a key role in this improvement. Using the importance analysis function of XGBoost, we identified that the latitude and longitude of the specular point contribute most significantly among geometry-related variables. A similar conclusion was drawn by Li et al. [32] in their study on wind speed retrieval using CYGNSS.

5.2. Model Improvements

Previous studies on SWH retrieval models, particularly [18,33], have demonstrated the excellent performance of tree-based GNSS-R SWH retrieval models, which has provided valuable guidance for this research. In recent years, the adoption of Vision Transformers has gained momentum in remote sensing retrieval tasks, providing a robust alternative to traditional deep learning models. Despite the advantages of deeper network structures in enhancing nonlinear modeling, challenges arise when the nonlinear relationships within data are inherently weak, as seen with SWH values surpassing 3 m. In such cases, these weaker relationships may hinder the ability of a Transformer to capture meaningful patterns effectively.

Qiao et al. [20] developed a Transformer-based network called WaveTransNet. It demonstrated high accuracy for SWHs ranging from 0 to 4 m. However, for SWH values between 4 to 8 m, its performance lagged behind that of the XGBoost-SC model, as shown in Table 6. This discrepancy highlights the need for further exploration into the comparative

effectiveness of Transformer-based retrieval models relative to other methodologies, as their superiority may not be universal across varying data ranges.

This study applies CDF matching with the XGBoost model in SWH retrievals from the GNSS-R data with the help of an improved Huber loss function for the first time, as far as we know. We observe the inherent characteristics of the CDF matching [21,22,28], which increase the retrieval accuracy and the proportions for large SWHs above 3 m and increase the RMSE and biases for SWHs below 3 m (Table 4 and Figures 8 and 9). The SWH values below 3 m constitute 89.1% of the dataset and play a critical role in the overall accuracy of the global ocean data. The performance limitations of the XGBoost model, alongside the characteristics of CDF matching and the increased proportion of low-quality observations at higher SWH values, cause overall poorer performance in the metrics relative to previous studies [18–20,33].

Table 6. The global retrieval performance of Tree-based and Transformer-based models (S represents S-Huber loss and C represents CDF matching).

Methods	0–4 m		4–8 m		Above 8 m	
	RMSE (m)	Bias (m)	RMSE (m)	Bias (m)	RMSE (m)	Bias (m)
RF	0.49	0.09	1.74	−1.48	5.76	−5.63
BT	0.49	0.08	1.75	−1.49	5.78	−5.65
GBDT	0.51	0.09	1.90	−1.73	5.92	−5.81
lightGBM	0.48	0.08	1.75	−1.54	5.78	−5.64
XGBoost	0.47	0.08	1.63	−1.39	5.30	−5.17
XGBoost-S	0.46	0.05	1.68	−1.45	5.37	−5.24
XGBoost-SC	0.54	0.08	1.45	−0.81	4.72	−4.33
WaveTransNet [20]	0.43	0.04	1.72	−1.65	—	—

However, our approach demonstrates superior accuracy in the high-SWH range, achieving our objective of improving retrieval accuracy under tropical cyclone (TC) conditions. Meanwhile, the proportions of high SWH values ranged from 16.4% for TC Muifa to 62.1% for TC Nesat in the WNP and 63.8% for TC Fiona in the western Atlantic Ocean. It is critical to map the high SWHs in TC conditions from the GNSS-R data, whose primary mission is to map the TC wind fields and wave field, such as the CYGNSS mission [5,18,19].

5.3. Model Constraints and Future Work

The complex nonlinear relationship between the SWH and the input variables makes modeling quite challenging for ML models with SWHs above 3 meters. Therefore, the higher SWHs tend to have larger errors [17,19]. We improved the Huber loss function and applied it to mitigate the errors induced by CDF matching. The improved XGBoost model, i.e., XGBoost-SC, performs better in retrieving high SWHs above 3 m than the BT, XGBoost, and other models, especially in TC conditions (Figures 12 and 14). In spite of the performance enhancement, XGBoost-SC can only retrieve about one third of the high SWHs above 3 m against those of the ECMWF in TC conditions. It is still quite challenging to accurately retrieve the high SWHs from the GNSS-R data. This is due to the fact that high wave heights hinder the formation of effective reflection zones in the DDM, thereby reducing inversion accuracy [19].

There is still significant work to be done in retrieving global SWHs from the spaceborne GNSS-R data. Firstly, although the SWH data retrieved by the ECMWF have been widely used as reference values to retrieve various oceanic physical parameters, their quality is still worse than the in-situ data's. Therefore, obtaining more reliable global SWH reference values is a crucial but very challenging task.

Secondly, reliable data quality control and down-sampling methods are essential. Balancing data quality and spatial coverage is key for the operational application of CYGNSS in SWH retrievals. In this study, the dataset was constructed with increased sampling

weights in the 5°N to 30°N and 5°S to 30°S regions, and lower quality control standards were applied to enhance coverage in high-SWH areas and to obtain more SWH samples in the high range. However, due to the system characteristics of CYGNSS, this approach may introduce more unstable observations and reduce the overall dataset quality.

Finally, more effective methods for representing the sea state are needed. Comprehensive descriptions of sea conditions using CYGNSS observations are crucial for further improving retrieval accuracy, particularly under TC conditions. Meanwhile, the experimental results show that utilizing the S-Huber loss function allows XGBoost-S to achieve better accuracy for lower SWH values. Additionally, CDF matching significantly improves the accuracy of SWH predictions above 3 m, which leverages historical distribution information. Therefore, identifying key variables to distinguish between high- and low-SWH conditions and developing a segmented SWH retrieval model could further enhance prediction accuracy.

6. Conclusions

This study focuses on improving the accuracy of high SWH retrieving using space-borne GNSS-R data from CYGNSS. XGBoost is chosen as the fundamental model and is enhanced by CDF matching and an improved loss function. The CDF matching is applied to escalate underestimates with high SWHs (>3 m), while the improved loss function is applied to reduce the errors caused by CDF matching with moderate and low SWHs.

The model was validated by over one million sample points, with the ECMWF SWH as the reference. On the global scale, XGBoost-S achieved the lowest RMSE of 0.51 m and bias of 0.02 m with the help of the S-Huber loss function. However, it tended to underestimate SWH values above 3 m and was mitigated by CDF matching, leading to significantly larger proportions and a reduced RMSE and bias for SWHs above 3 m by XGBoost-SC.

The enhanced performance of XGBoost-SC for large SWHs was further confirmed in TC conditions over the Western North Pacific and in the Western Atlantic Ocean. Nevertheless, it is still quite challenging and needs further improvement for high SWH retrieval using CYGNSS data, particularly in surging wave and TC conditions.

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

Funding: This study was supported financially by the National Key R&D Program of China (2021YFC3001000), the National Natural Science Foundation of China (41871085), the Innovation-Group Project of the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311021004) and the State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences: G2014-02-06. And the APC was funded by the State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences (G2014-02-06).

Data Availability Statement: The data supporting the findings of this study, including ERA-5 significant wave height data from the European Centre for Medium-Range Weather Forecasts (<https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download>, accessed on 17 December 2024) and Cyclone Global Navigation Satellite System observation data from NASA (<https://search.earthdata.nasa.gov/search/granules?p=C2205618435-POCLOUD>, accessed on 17 December 2024), are publicly available.

Acknowledgments: We thank NASA and the European Centre for Medium-Range Weather Forecasts (ECMWF) for publicly providing Cyclone Global Navigation Satellite System (CYGNSS) observation data and significant wave height products.

Conflicts of Interest: Author Jinhua Wu was employed by the company China Railway Siyuan Survey and Design Group Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Yang, J.; Zhang, J.; Jia, Y.; Fan, C.; Cui, W. Validation of Sentinel-3A/3B and Jason-3 Altimeter Wind Speeds and Significant Wave Heights Using Buoy and ASCAT Data. *Remote Sens.* **2020**, *12*, 2079. [\[CrossRef\]](#)
2. Xue, S.; Geng, X.; Yan, X.H.; Xie, T.; Yu, Q. Significant wave height retrieval from Sentinel-1 SAR imagery by convolutional neural network. *J. Oceanogr.* **2020**, *76*, 465–477. [\[CrossRef\]](#)
3. M, M.N. A passive reflectometry and interferometry system (PARIS) application to ocean altimetry. *ESA J.* **1993**, *17*, 331–355.
4. Foti, G.; Gommenginger, C.; Srokosz, M. First Spaceborne GNSS-Reflectometry Observations of Hurricanes From the UK TechDemoSat-1 Mission. *Geophys. Res. Lett.* **2017**, *44*, 12358–12366. [\[CrossRef\]](#)
5. Ruf, C.S.; Atlas, R.; Chang, P.S.; Clarizia, M.P.; Garrison, J.L.; Gleason, S.; Katzberg, S.J.; Jelenak, Z.; Johnson, J.T.; Majumdar, S.J.; et al. New Ocean Winds Satellite Mission to Probe Hurricanes and Tropical Convection. *Bull. Amer. Meteorol. Soc.* **2016**, *97*, 385–395. [\[CrossRef\]](#)
6. Wan, W.; Liu, B.; Guo, Z.; Lu, F.; Niu, X.; Li, H.; Ji, R.; Cheng, J.; Li, W.; Chen, X.; et al. Initial Evaluation of the First Chinese GNSS-R Mission BuFeng-1 A/B for Soil Moisture Estimation. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [\[CrossRef\]](#)
7. Yan, Q.; Huang, W.; Jin, S.; Jia, Y. Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data. *Remote Sens. Environ.* **2020**, *247*, 111944. [\[CrossRef\]](#)
8. Zhang, Y.; Zhao, X.; Yang, S.; Han, Y.; Hong, Z.; Meng, W.; Chen, Z.; Liu, W. Research on Sea Surface Wind Speed FM Based on CYGNSS and HY-2B Microwave Scatterometer. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–15. [\[CrossRef\]](#)
9. Li, C.; Huang, W.; Gleason, S. Dual Antenna Space-Based GNSS-R Ocean Surface Mapping: Oil Slick and Tropical Cyclone Sensing. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2015**, *8*, 425–435. [\[CrossRef\]](#)
10. Zhang, Y.; Lu, Q.; Jin, Q.; Meng, W.; Yang, S.; Huang, S.; Han, Y.; Hong, Z.; Chen, Z.; Liu, W. Global Sea Surface Height Measurement From CYGNSS Based on Machine Learning. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2023**, *16*, 841–852. [\[CrossRef\]](#)
11. Clarizia, M.P.; Ruf, C.S.; Jales, P.; Gommenginger, C. Spaceborne GNSS-R Minimum Variance Wind Speed Estimator. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6829–6843. [\[CrossRef\]](#)
12. Peng, Q.; Jin, S. Significant Wave Height Estimation from Space-Borne Cyclone-GNSS Reflectometry. *Remote Sens.* **2019**, *11*, 584. [\[CrossRef\]](#)
13. Yang, S.; Jin, S.; Jia, Y.; Ye, M. Significant Wave Height Estimation from Joint CYGNSS DDMA and LES Observations. *Sensors* **2021**, *21*, 6123. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Bu, J.; Yu, K. Significant Wave Height Retrieval Method Based on Spaceborne GNSS Reflectometry. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1503705. [\[CrossRef\]](#)
15. Bu, J.; Yu, K. A New Integrated Method of CYGNSS DDMA and LES Measurements for Significant Wave Height Estimation. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1505605. [\[CrossRef\]](#)
16. Ruf, C.S.; Gleason, S.; McKague, D.S. Assessment of CYGNSS Wind Speed Retrieval Uncertainty. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2019**, *12*, 87–97. [\[CrossRef\]](#)
17. Wang, F.; Yang, D.; Yang, L. Retrieval and Assessment of Significant Wave Height from CYGNSS Mission Using Neural Network. *Remote Sens.* **2022**, *14*, 3666. [\[CrossRef\]](#)
18. Wang, C.; Yu, K.; Zhang, K.; Bu, J.; Qu, F. Significant Wave Height Retrieval Based on Multivariable Regression Models Developed With CYGNSS Data. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [\[CrossRef\]](#)
19. Bu, J.; Yu, K.; Ni, J.; Huang, W. Combining ERA5 data and CYGNSS observations for the joint retrieval of global significant wave height of ocean swell and wind wave: A deep convolutional neural network approach. *J. Geod.* **2023**, *97*, 81. [\[CrossRef\]](#)
20. Qiao, X.; Huang, W. WaveTransNet: A Transformer-Based Network for Global Significant Wave Height Retrieval From Spaceborne GNSS-R Data. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–11. [\[CrossRef\]](#)
21. Clarizia, M.P.; Ruf, C.S. Statistical Derivation of Wind Speeds From CYGNSS Data. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3955–3964. [\[CrossRef\]](#)
22. Reynolds, J.; Clarizia, M.P.; Santi, E. Wind Speed Estimation From CYGNSS Using Artificial Neural Networks. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2020**, *13*, 708–716. [\[CrossRef\]](#)
23. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
24. Clarizia, M.P.; Ruf, C.S. Wind Speed Retrieval Algorithm for the Cyclone Global Navigation Satellite System (CYGNSS) Mission. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4419–4432. [\[CrossRef\]](#)
25. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.* **2020**, *146*, 1999–2049. [\[CrossRef\]](#)
26. Knapp, K.R.; Kruk, M.C.; Levinson, D.H.; Diamond, H.J.; Neumann, C.J. The international best track archive for climate stewardship (IBTrACS) Unifying Tropical Cyclone Data. *Bull. Amer. Meteor. Soc.* **2010**, *91*, 363–376. [\[CrossRef\]](#)
27. Huang, X.; Bai, L.; Yu, Z.; Chan, J.C.L.; Yu, H.; Tang, J.; Guo, R.; Wan, R. Tropical cyclone activities in the Western North Pacific in 2022. *Trop. Cyclone Res. Rev.* **2024**, *13*, 125–135. [\[CrossRef\]](#)
28. Liu, X.; Bai, W.; Tan, G.; Huang, F.; Xia, J.; Yin, C.; Sun, Y.; Du, Q.; Meng, X.; Liu, C.; et al. GNSS-R Global Sea Surface Wind Speed Retrieval Based on Deep Learning. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [\[CrossRef\]](#)

29. Yongchao, Z.; Tao, T.; Li, Z.; Qu, X.; Wickert, J. Sensing Sea Ice Based on Doppler Spread Analysis of Spaceborne GNSS-R Data. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2019**, *13*, 1–10. [[CrossRef](#)]
30. Cheng, P.H.; Lin, C.C.H.; Morton, Y.T.J.; Yang, S.C.; Liu, J.Y. A Bagged-Tree Machine Learning Model for High and Low Wind Speed Ocean Wind Retrieval From CYGNSS Measurements. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–10. [[CrossRef](#)]
31. Han, S.; Zhang, H.; Zheng, Y. A global study of temporal and spatial variation of SWH and wind speed and their correlation. *Acta Oceanolog. Sin.* **2014**, *33*, 48–54. [[CrossRef](#)]
32. Li, Z.; Guo, F.; Zhang, X.; Guo, Y.; Zhang, Z. Analysis of factors influencing significant wave height retrieval and performance improvement in spaceborne GNSS-R. *GPS Solut.* **2024**, *28*, 64. [[CrossRef](#)]
33. Wang, Q.; Bu, J.; Ni, J.; Li, L.; Liu, X.; Huang, W. Ocean swell height estimation from spaceborne GNSS-R data using hybrid deep learning model. *GPS Solutions* **2024**, *28*, 1–19. [[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.