Article

A Novel Rain Identification and Rain Intensity Classification Method for the CFOSAT Scatterometer

Meixuan Quan [1], Jie Zhang [1,2,*] and Rui Zhang [1]

[1] College of Oceanography and Space Informatics, China University of Petroleum, Qingdao 266580, China; s21160022@s.upc.edu.cn (M.Q.); z21160067@s.upc.edu.cn (R.Z.)

[2] First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China

[*] Correspondence: zhangjie@upc.edu.cn

Abstract: The China–France oceanography satellite scatterometer (CSCAT) is a rotating fan-beam scanning observation scatterometer operating in the Ku-band, and its product quality is affected by rain contamination. The multiple azimuthal NRCS measurements provided by CSCAT L2A, the retrieved wind speed and wind direction provided by CSCAT L2B, as well as the rain data provided by GPM, are used to construct a new rain identification and rain intensity classification model for CSCAT. The EXtreme Gradient Boosting (XGBoost) model, optimized by the Dung Beetle Optimizer (DBO) algorithm, is developed and evaluated. The performance of the DBO-XGBoost exceeds that of the CSCAT rain flag in terms of rain identification ability. Also, compared with XGBoost without parameter optimization, K-nearest Neighbor with K = 5 (KNN5) and K-nearest Neighbor with K = 3 (KNN3), the performance of DBO-XGBoost is better. Its rain identification achieves an accuracy of about 90% and a precision of about 80%, which enhances the quality control of rain. DBO-XGBoost has also shown good results in the classification of rain intensity. This ability is not available in traditional rain flags. In the global regional and local regional tests, most of the accuracy and precision in rain intensity classification have reached more than 80%. This technology makes full use of the rich observed information of CSCAT, realizes rain identification, and can also classify the rain intensity so as to further evaluate the degree of rain contamination of CSCAT products.

Keywords: rain identification; rain intensity classification; CFOSAT scatterometer; XGBoost; DBO

1. Introduction

The China–France oceanography satellite (CFOSAT) was successfully launched on 29 October 2018, through a collaborative effort between the China National Space Administration (CNSA) and the National Center for Space Studies of France (CNES) at the Jiuquan Satellite Launch Center in China. The satellite is equipped with two microwave sensors: the CFOSAT SCATterometer (CSCAT) of China and the Surface Wave Investigation and Monitoring (SWIM) of France. The CSCAT employs a rotating fan beam, which is different from the fixed fan beam and rotating scanning pencil beam of the traditional scatterometer. The innovative rotating fan beam facilitates the simultaneous acquisition of multiple azimuthal measurements of the sea surface, resulting in a large number of independent samples of backscatter coefficients obtained concurrently. The low-speed scanning of antennas enhances the redundancy and reliability of azimuth backscatter measurements. The CSCAT is currently the highest original spatial resolution scatterometer in the world, which provides the possibility to develop high-quality sea surface wind products [1]. However, studies have found that rain has a serious impact on the echo signals received by scatterometers [2–8]. The CSCAT, as a Ku-band scatterometer, will be significantly influenced by rain, thereby limiting the accuracy of its wind products [9]. Therefore, it is necessary to identify the rain-contaminated data in order to improve the quality of CSCAT L2B wind products.

In the past few decades, researchers have developed several quality control (QC) methods for scatterometers. The brightness temperature (TB) provided by the radiometer exhibits high sensitivity to rain, thus making an important contribution to rain identification. The Multidimensional Histogram (MUDH) is the earliest technique for identifying rain using brightness temperature (TB) and other rain characteristics, which was employed to generate a rain flag for SeaWinds on QuikSCAT [10]. The combination of TB and wind speed standard deviation has been applied to OSCAT to develop a rain flag [11]. However, the use of the radiometer TB to identify rain is significantly limited due to the lack of radiometers synchronized with scatterometers on many satellites. Subsequently, maximum likelihood estimation (MLE), which identifies low-quality wind by quantifying the deviation between the measured normalized radar cross-section (NRCS) and the NRCS calculated using the geophysical model function (GMF), has been demonstrated to significantly flag rain-contaminated data [12,13]. The empirically normalized objective function (ENOF) follows a similar principle to MLE, but a weighted approach is used instead of the NRCS measurement error variance of MLE for error quantification [14]. Notably, the ENOF is primarily applicable for rain detection at low wind speeds, while exhibiting an underreporting rate ranging from 65% to 75% during high wind speeds and tropical cyclones.

During the early years, MLE was widely used in wind QC and could reject most of the rain-contaminated data. However, MLE rejects some good wind to ensure the effectiveness of its QC. Subsequently, some rain-flag technologies complementary to MLE are developed to improve QC. Singularity Exponent (SE) is based on the spatial derivatives between wind vector cells (WVCs) [15,16], identifying poor quality wind data from spatial heterogeneity caused by rain [17], and can therefore be complementary to MLE to flag more low quality data. J_{OSS} reduces the false alarm rate (FAR) of rain-contaminated data by using the background wind speed provided in the 2-D variational ambiguity removal (2-DVAR) process [18]. The Bayesian algorithm (P rain flag) provides a posterior rain probability for each measurement in WVC with a low false alarm rate (FAR) and a low missing report rate (MRR) [19].

In recent years, with the rise of artificial intelligence technology, machine learning models have been found to have great potential in the rain. Two neural network (NN) models are used for rain detection and rain rate inversion on five sets of data samples from different regions observed by OCSAT [20]. However, the model cannot independently utilize scatterometer parameters for rain identification and relies on external collocated data sources, including parameters derived from numerical weather prediction (NWP) models: total precipable water (TPW), ground relative humidity (RH), wind speed (WS), and wind direction (WD). The HY2RRM model for rain identification of HY-2A data was developed based on the K-nearest Neighbor (KNN) [21], while employing the same set of rain-sensitive parameters as the MUDH rain flag. Meanwhile, MUDH has also been transplanted to HY-2A data. The experimental results demonstrate that the effect of the KNN model surpasses that of MUDH technology.

At present, there are few studies on the rain flag of CSCAT [22,23]. The J_{OSS} rain flag has been proven to reduce the FAR of rain identification on CSCAT. However, the MLE/Joss-based rain flagging technique has a large MRR [12,18], and the quality control effect of wind products needs to be improved. The current CSCAT rain flag lacks specifications regarding further processing methods for MRR. CFOSAT is not equipped with a radiometer and cannot provide TB synchronized with CSCAT. Consequently, the MUDH-like method cannot be used for the CSCAT rain flag. The direct collocation of rain data is considered the optimal quality control method, but obtaining reliable rain data in the same spatiotemporal domain as the scatterometer often encounters significant challenges.

CSCAT requires the development of more effective quality control methods in order to improve the quality of the CSCAT L2B wind products. The rich observation information in CSCAT provides sufficient data for machine learning to identify rain. EXtreme Gradient Boosting (XGBoost) has high-dimensional data processing and feature selection capabilities to make full use of the rich information of CSCAT. In this paper, a rain identification model

based on the Dung Beetle Optimizer algorithm [24] optimized for XGBoost (DBO-XGBoost) is constructed. The model independently realizes rain identification and rain intensity classification using only the own information of CSCAT, without relying on external wind data or radiometer TB data. This approach enables more timely and efficient rejection of rain data, thereby enhancing the product quality of CSCAT. This paper is organized as follows: Section 2 describes the collocated dataset and the methods of constructing the model. Section 3 analyzes the rain identification and rain intensity classification performance of the model. Section 4 discusses the effect of the model under different sea conditions and the influence of different input information on rain identification. The conclusions are presented in Section 5.

2. Materials and Methods

2.1. Data

In this study, the CSCAT L2A and L2B data and the Global Precipitation Measurement (GPM) Dual-frequency Precipitation Radar (DPR) Ku-band rain data are used to collocate a dataset for rain identification and rain intensity classification. All data are described in detail below.

2.1.1. CSCAT Data

CSCAT is operating at Ku-band (13.256 GHz) and employing medium incidence angles ranging from 28° to 51°. It is the first rotating range-gated fan beam scatterometer, using a rotating 1.2 m slotted-waveguide antenna featuring two fan beams positioned at a 180-degree separation angle, effectively sweeping the surface in a conical manner. One beam is horizontally polarized (H-pol), while the other is vertically polarized (V-pol). Consequently, it can provide measured NRCS at both HH and VV polarizations. Its ground swath width is over 1000 km, so it can achieve global coverage of wind measurements in about 3 days. The two conically scanning fan beams of CSCAT can respectively acquire 2–8 effective observations for a WVC, contingent upon its cross-track position, surpassing the capabilities of fixed fan-beam and pencil-beam scatterometers significantly. Both CSCAT L2A and L2B provide two resolution products, 12.5×12.5 km and 25×25 km. This paper utilizes the 12.5×12.5 km L2A and L2B products, obtained freely from the China Ocean Satellite Data Service Center (COSDSC) (https://osdds.nsoas.org.cn/, Accessed on 20 October 2022), which provides HH- and VV-polarized NRCS measurements from multi-directional observations for the period spanning 1 June 2020 and 30 June 2020. Based on the data quality flags provided by the product, WVCs containing sea ice, land, and those with negative SNR due to significant noise contamination were rejected. It is worth noting that we need to develop a new method to identify rain so the rain-contaminated data can be retained. The characteristic parameters used for rain identification and rain intensity classification include NRCS, azimuth angle, incidence angle, time, longitude, and latitude provided by CSCAT L2A, as well as the retrieved wind speed and wind direction using GMF without taking rain into account provided by CSCAT L2B.

2.1.2. The ECMWF ERA5 Data

ERA5 is the fifth-generation atmospheric reanalysis conducted by the European Centre for Medium-Range Weather Forecasts (ECMWF) to comprehensively analyze global climate and weather patterns. By employing advanced modeling techniques and incorporating the latest data assimilation system, it effectively integrates a wide range of observations into model estimates. The ERA5 hourly dataset encompasses single-level data from 1 January 1950 to near real-time, providing estimations for various atmospheric, ocean-wave, and land-surface variables. It includes the 10 m U-component of neutral wind (U10) and the 10 m V-component of neutral wind (V10), from which the 10 m wind speed (SPD10) and wind direction (WD) can be obtained. The gridding process ensures that the wind data is presented on a regular latitude-longitude grid with a resolution of 0.25 degrees. The data are freely available at the Climate Data Store (https://cds.climate.copernicus.eu/cdsapp#!/

search?type=dataset, Accessed on 11 March 2023). In this paper, the ERA5 global reanalysis containing U10 and V10 is used as a reference to assess the quality of CSCAT L2B wind products. The ERA5 data is interpolated to a resolution of 12.5×12.5 km, which aligns with the spatial resolution of CSCAT data.

2.1.3. The Ku-Band GPM-DPR Data

The GPM satellite, launched on 27 February 2014 by the National Aeronautics and Space Administration (NASA) and the Japanese Aerospace Exploration Agency (JAXA), is equipped with two advanced instruments, including the Dual-frequency Precipitation Radar (DPR) and the GPM Microwave Imager (GMI). The project, which bears resemblance to the Tropical Rain Measuring Mission (TRMM), has been significantly expanded to encompass not only the tropical zone but also mid- to high-latitude areas. It operates on a $65°$ inclined, non-sun-synchronous orbit at an altitude of approximately 407 km, with the primary objective of accurately and precisely quantifying global precipitation. The DPR consists of two precipitation radars, the KuPR on the Ku-band (13.6 GHz) and the KaPR on the Ka-band (35.5 GHz). It provides three different rain rate estimation products based on Ku-band observations, Ka-band observations, and simultaneous use of Ku and Ka-band observations. Toyoshima et al. [25] compared the Ka-band and Ku-band precipitation detection capabilities and found that KaPR has no obvious advantage in rain detectability because the non-Rayleigh scattering effect of KaPR may have partly offset the sensitivity advantage of KaPR relative to KuPR. Lasser et al. [26] evaluated the GPM DPR precipitation estimation based on 15 rain events observed on WegenerNet in 4 years. The results showed that the probability of detection is greater than 0.70 for Ku and DPR but only about 0.50 for Ka. The above experimental findings demonstrate that the ability of KuPR to detect precipitation is superior to that of KaPR. Considering that CSCAT operates in the Ku band, the KuPR data is used to provide the rain information, which is obtained freely from the Globe Portal System (G-Portal) 2AKu standard product (ver. 7) operated by the Japanese Aerospace Exploration Agency (JAXA), including precipitation, latitude, longitude, and time. We get the data from the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) (https://disc.gsfc.nasa.gov/, Accessed on 4 May 2023).

2.1.4. Collocated Dataset

The spatio-temporal collocation of data is a crucial step in establishing the target dataset. Our purpose is to achieve CSCAT and KuPR spatio-temporally collocating to obtain the CSCAT-KuPR dataset containing CSCAT measured information and KuPR rain information. Due to the difference in geographical location and observed time between the two satellites, as well as the temporal and spatial variability of rain, it is imperative to impose reasonable constraints on the spatiotemporal collocation threshold to ensure the validity of the data. The CSCAT measurements are collocated with the ERA5 wind data, ensuring a temporal separation of less than 30 min and a spatial separation of less than 12.5 km. When a CSCAT WVC collocates multiple ERA5 points, the wind data of the ERA5 closest to the CSCAT is selected. The CSCAT-ERA5 collocating data are collocated with the GPM data with temporal and spatial intervals less than 30 min and 6.25 km, respectively. Due to the different resolutions of CSCAT and KuPR, a CSCAT WVC may collocate multiple KuPR rain data sets. We take the average of multiple KuPRs as the rain observed by CSCAT. The formula is as follows:

$$R_{CSCAT} = \frac{\sum_{i=1}^{M} R'_{i(KuPR)}}{M}$$

(1)

where M is the number of KuPRs collocated by a CSCAT WVC.

Figure 1 shows the geographical distribution of the CSCAT-ERA5-KuPR dataset during the period from 1 June 2020 to 30 June 2020, and it can be seen that the data are distributed in the global seas. Due to the short revisit period of the satellite, some data overlap in spatial position.
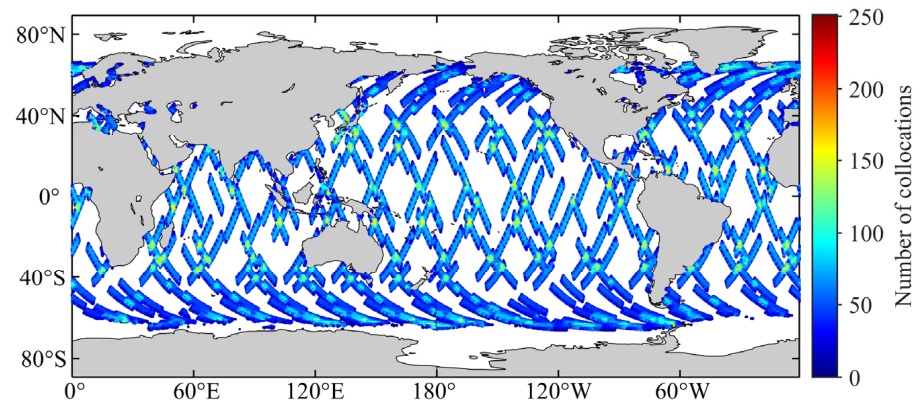
Figure 1. Global geographic distribution of the CSCAT-ERA5-KuPR collocating dataset during the period from 1 June 2020 to 30 June 2020. The colors represent the number of data points matched at the same location within a $1° \times 1°$ bin.

Due to the varying number of effective observations of CSCAT in different WVCs, it is necessary to select and utilize data with consistent observation counts. There are two groups accounting for 100%, six groups accounting for 99.89%, and ten groups accounting for 54.24% of effective WVC data in our dataset. In order to achieve an adequate number of WVC, some of the observed information will inevitably be sacrificed. Conversely, in order to maximize the retention of effective and comprehensively measured information, some of the available WVC will be forfeited. This procedure yields approximately 458,000 data points, of which about 76,400 are contaminated by rain (rain rate > 0.004 mm/h). This dataset is randomly divided into two subsections: 80% for training and 20% for testing. The probability density function (PDF) of retrieval wind speed provided by CSCAT L2B and the data volume of the CSCAT-KuPR dataset are shown in Figure 2. The average wind speed of the PDF is observed to be lower under rain-free conditions, while it is higher during rain events, indicating an overall increase in retrieved wind speed during rain compared to rain-free periods. This characteristic holds significant importance in identifying rain, indicating the necessity of using L2B retrieval wind data.
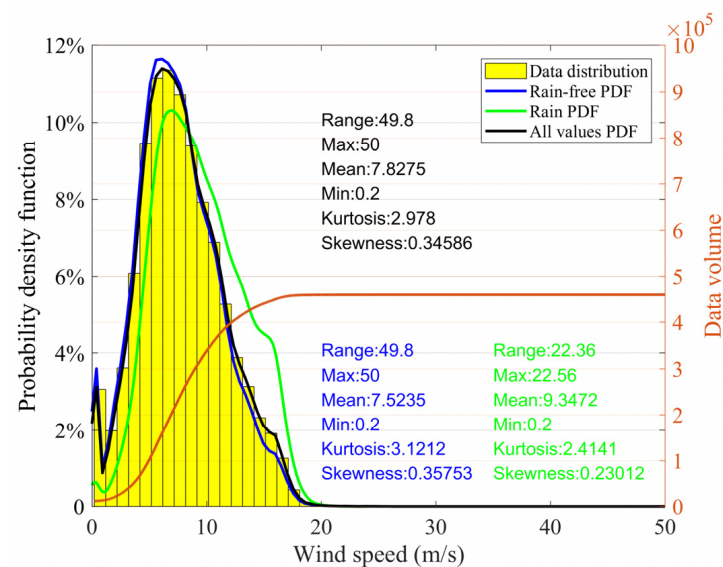

Figure 2. Probability density function of CSCAT L2B wind speed and the data volume of the CSCAT-KuPR dataset. The yellow bar chart represents the wind speed distribution of the total data. The black line represents total data, and the green and blue lines represent rain-contaminated and rain-free data, respectively. Orange is the cumulative amount of data that increases with wind speed.

2.2. Model

A novel rain flag based on the DBO-XGBoost is proposed. XGBoost is a machine learning model used for rain identification and classification of rain intensity. DBO is used to optimize the internal parameters of the XGB model. The model is introduced and constructed as described below.

2.2.1. XGBoost Model

XGBoost is an efficient machine learning technique, particularly well-suited for handling large-scale datasets and high-dimensional features [27,28]. It is an optimized version of the Gradient Boosting Tree, which combines multiple weak classifiers to build a powerful ensemble model. XGBoost reduces overfitting problems with regularization boosting technology and becomes an excellent model in machine learning to solve regression and classification problems.

The expression for XGBoost is as follows:

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{2}$$

where i is the ith sample, k is the kth tree, and $\hat{y}_i^{(t)}$ is the predicted value of the ith sample x_i in the tth tree. The objective function expression is as follows:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} \left(y_i - \hat{y}_i^{(t)} \right)^2 + \sum_{k=1}^{t} \Omega(f_i) \tag{3}$$

where the complexity of the tree Ω is composed of the number of leaf nodes and the l_2 norm of the weight vector of leaf nodes. The second-order Taylor expansion and regularization term expansion of the objective function are carried out to obtain the weight of each leaf node and the optimal objective function.

As an improvement of the decision tree model, XGBoost has some adjustable parameters that can affect the performance, complexity, and training speed of the model. The parameters include: (1) Number of estimators: the number of decisions specified during model training; (2) max_depth: the maximum number of levels that the decision tree may reach in the training process; (3) learning_rate: a large learning rate gives greater weight to the contribution of each tree in the set, which can speed up the training time but may lead to overfitting. In different application scenarios, the performance and generalization ability of the model can be optimized by changing these parameters. Manual parameter adjustment or the grid search method will consume a lot of time. Therefore, it is necessary to find a method to determine the optimal combination of parameters.

2.2.2. Dung Beetle Optimizer Algorithm

DBO, proposed by Xue and Shen, is a novel swarm intelligence optimization algorithm [24]. It updates the position and optimizes the parameters by simulating the ball-rolling, dancing, foraging, stealing, and reproduction behaviors of dung beetles. Among the four individual behaviors, only the rolling ball behavior exhibits superior global search capability throughout each iteration of the algorithm. The foraging behavior conducts a search in proximity to its own position based on dynamically adjusting upper and lower bounds, which gradually decrease over iterations, leading the foraging behavior to transition from global exploration to local exploitation. Reproductive behavior and stealing behavior are local searches based on the dynamic upper and lower bounds near the best individual.

The four behaviors are executed by distinct subsets of dung beetle populations. There is no fixed proportion of the number of dung beetles in each behavior. In this paper, we set the size of the dung beetle population to 30, in which the number of rolling balls, breeding, foraging, and stealing dung beetles is 6, 7, 7, and 10, respectively. It is imperative to ensure that each individual dung beetle within the population exhibits a specialized division of

labor and engages exclusively in one specific behavioral role. The diversified location update strategy of the DBO algorithm enables a more comprehensive exploration of the search space, thereby effectively addressing complex search and optimization problems encountered in practical applications.

2.2.3. XGBoost Parameter Optimization Based on DBO

This paper proposes a method using DBO to optimize XGBoost parameters, thereby improving model performance. The fitness of the DBO population serves as an index that represents the optimal performance of XGBoost. In order to optimize the parameters in XGBoost, it is necessary to define a value range that will serve as the upper and lower bounds for dung beetle activity. The population fitness of DBO is considered the objective function, and the calculated minimum fitness represents the optimal solution. In other words, within the given range, DBO iteratively searches for the optimal fitness of the population by adjusting parameters.

DBO-XGBoost is used to classify the data as rain-contaminated or rain-free for the CSCAT-KuPR dataset. The model is trained using a set of physical parameters derived from CSCAT data, including NRCS, azimuth angle, incidence angle, time, longitude, latitude, wind speed, and wind direction as input. The GPM rain data is used as the target variable. Rain-contaminated data is labeled as 1, while rain-free data is labeled as 0. Additionally, the model is also used for rain intensity classification into four levels, with labels ranging from 0 to 3.

The evaluation metrics of the DBO-XGBoost include accuracy, receiver operating characteristic (ROC), precision, recall, etc. ROC is more effective in evaluating the model's performance when the proportion of rain-contaminated data is significantly smaller compared to rain-free data. It is defined as the distribution curve of the true-positive rate (TPR) and false-positive rate (FPR) of the classification model.

$$\begin{cases} \text{TPR} = \frac{\text{TP}}{\text{TP}+\text{FN}} \\ \text{FPR} = \frac{\text{FP}}{\text{FP}+\text{TN}} \end{cases} \tag{4}$$

where TP is a positive sample predicted by the model as a positive class, TN is a negative sample predicted by the model as a negative class, FP is a negative sample predicted by the model as a positive class, and FN is a positive sample predicted by the model as a negative class. The model's performance is considered better when the ROC curve approaches the (0.0, 1.0) point earlier and completely encompasses the other ROC curve at the lower right. When the two curves intersect, it is difficult to visually discern the superior performance between the models. In such cases, it is imperative to compute thea under the curve (AUC) at the lower right section of both ROC curves. A larger AUC value indicates a more favorable model performance. Ultimately, our objective is to achieve an optimal AUC for the ROC. However, the optimal fitness of the population is the minimum value of the objective function, so the objective function is set to the following:

$$\text{Fitness} = -\text{AUC} \tag{5}$$

The process of optimizing XGBoost parameters based on DBO is shown in Figure 3. The training set is inputted into XGBoost, followed by substituting the objective function of XGBoost and the parameters to be optimized into the DBO algorithm with a specified number of iterations. Each dung beetle in the population completes the task of optimizing the target parameters and subsequently calculates and updates the optimal fitness of the population. Consequently, the optimal parameters and the global optimal solution of the objective function are obtained. Constructing a novel DBO-XGBoost model by optimizing the parameters of XGBoost and evaluating its performance using a testing set.
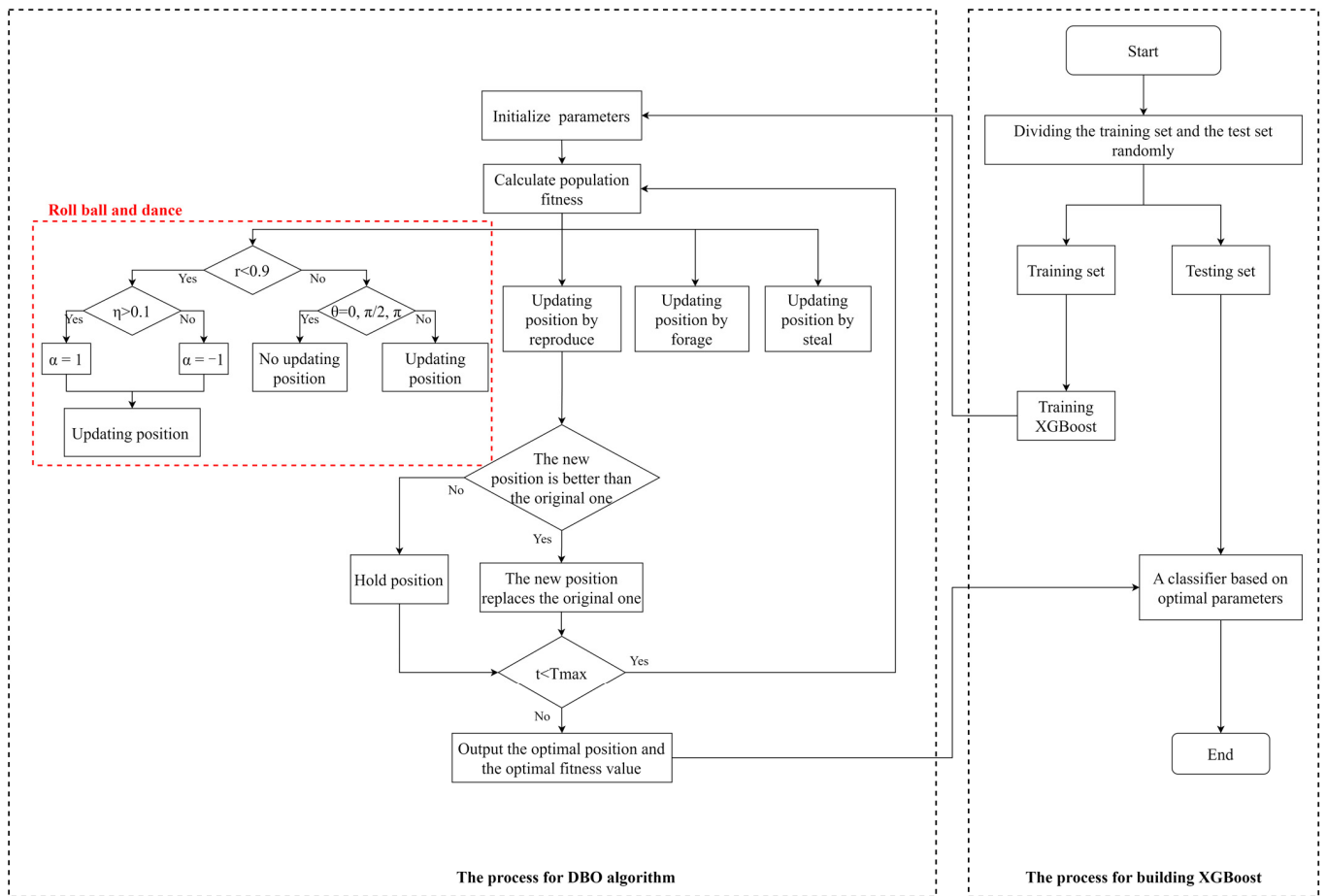
Figure 3. Proposed DBO-based approach for XGBoost.

The parameter value range and the optimal parameters of the XGBoost model obtained by DBO optimization are shown in Table 1. The range of parameters has been preliminarily tested. For the number of estimators, the performance of the XGBoost is obviously bad when the value is less than 100, while the performance changes little when the number is more than 300. Moreover, excessively increasing this value will result in prolonged model training duration, thereby significantly reducing the model's efficiency. The result of the optimization is 300. For max_depth, if this parameter is set too low, it may lead to underfitting of the model, resulting in suboptimal performance. Conversely, setting a higher value increases the complexity of the tree structure and can result in overfitting. The range considered for exploration was from 10 to 60. The result of the optimization is 23. The learning_rate controls the step size of each iteration when the weights are updated. The small learning rate suppresses the contribution of each tree and can obtain a more accurate model, but it also slows down the learning speed. The selected search range is 0.05 to 0.3. The result of the optimization is 0.06.

Table 1. Optimal parameter for XGBoost by DBO.

Parameters	Range	Optimal Value
Number of estimators	[100, 500]	300
Max_depth	[10, 60]	23
Learning_rate	[0.05, 0.3]	0.06

Figure 4 shows the convergence curve of the DBO algorithm to the AUC. It has a strong exploration ability in the early stage of the iteration and can fully search for a promising

spatial range and gradually converge in the later stage of the iteration, so as to ensure that the searched optimal solution is global optimal rather than local optimal.
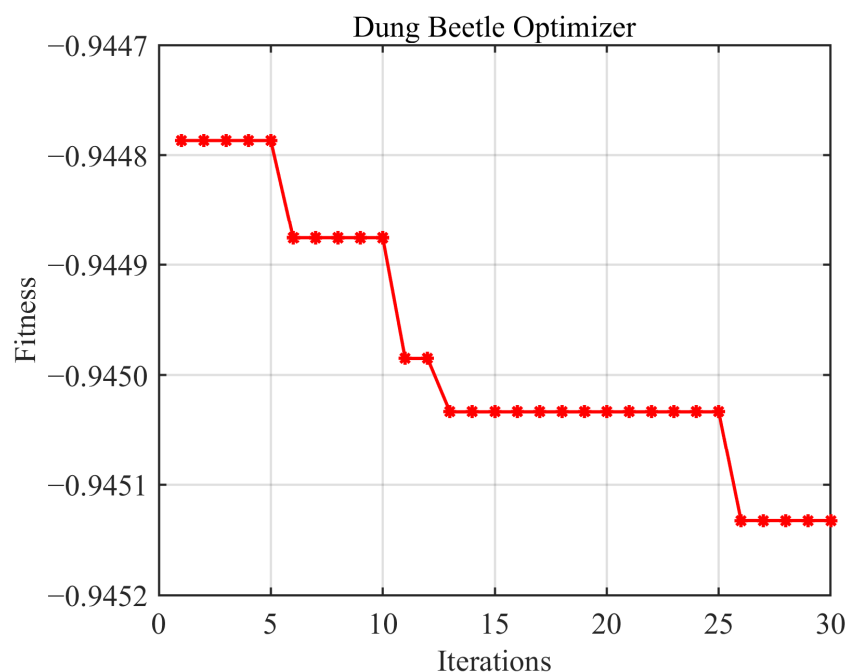

Figure 4. The population fitness curve of the DBO algorithm iteration.

3. Results

3.1. Evaluation of the DBO-XGBoost Model in Rain Identification

In this study, we assessed the performance of DBO-XGBoost and conducted a comparative analysis with K-Nearest Neighborhood (KNN), XGBoost, and CSCAT L2B products in terms of their effectiveness in rain flagging. The XGBoost is the model using the default internal parameters, with the number of estimators set to 100, max_depth set to 6, and learning_rate set to 0.3. KNN is a classical classification algorithm [29]. KNN is a classical classification algorithm, where the underlying principle is that if a majority of the K samples in the feature space surrounding the target point belong to a specific category, then it can be inferred that the sample also belongs to this category. The value of K serves as a hyperparameter in the KNN algorithm, determining the number of nearest neighbors considered for accurate classification. A smaller value of K increases model complexity, resulting in reduced training error but weakened generalization ability. Conversely, a larger value of K reduces model complexity, leading to increased training error but improved generalization ability. Therefore, selecting an appropriate value for K plays a pivotal role in the model. The KNN model has been demonstrated to exhibit favorable performance in the rain identification of HY-2A [21]. The performance of KNN with K = 3 (KNN3) and KNN with K = 5 (KNN5) is compared to that of DBO-XGBoost models. The input features and targets of all comparative models are consistent with DBO-XGBoost.

In order to evaluate the performance of the model more comprehensively, the classification evaluation metric is used to systematically score the model. The selected model evaluation indicators are as follows: (1) Accuracy: the proportion of correctly classified data in the total dataset; (2) Precision: the proportion of accurately predicted rain-contaminated data among all predicted rain-contaminated data; (3) False alarm rate (FAR): the proportion of data predicted as rain-contaminated but actually rain-free that accounted for the entirety of the rain-free dataset; (4) Missing report rate (MRR): rain data is predicted as the proportion of no rain to all rain data; (5) Rejection rate: the proportion of rain-contaminated data identified by the model to the total data; (6) Actual rain: the proportion

of actual rain-contaminated data in the total data. The formulas for the evaluation metric are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{FAR} = \frac{FP}{FP + TN} \tag{8}$$

$$\text{MRR} = \frac{FN}{TP + FN} \tag{9}$$

where TP is a positive sample predicted by the model as a positive class; TN is a negative sample predicted by the model as a negative class; FP is a negative sample predicted by the model as a positive class; and FN is a positive sample predicted by the model as a negative class.

Accuracy is the most commonly used evaluation metric. However, given the infrequent occurrence of rain events, even if a substantial amount of rain-contaminated data is misclassified as rain-free data, accurate partitioning of the proportion of rain-free data will still yield a model with high accuracy. Nevertheless, our primary objective remains to maximize the identification of rain-contaminated data. Therefore, a superior rain identification model should have the ability to classify rain and no rain with high precision, low FAR, and low MRR. In practical applications, these requirements are often not fully realized. Increasing the precision of reporting each rain event may lead to a higher likelihood of MRR, whereas aiming for comprehensive reporting may result in increased FAR and reduced precision.

Table 2 shows the evaluation of DBO-XGBoost, XGBoost, KNN5, KNN3 models, and the CSCAT rain flag. Among all the models, DBO-XGBoost exhibits the highest accuracy and precision. Compared with KNN5 and KNN3, XGBoost reduces the FAR while obviously increasing the MRR. Although the precision of rain identification has improved, the accuracy of the model has been reduced, and the overall performance has deteriorated. DBO-XGBoost found a balance between the FAR and the MRR of rain identification. Compared with XGBoost, the DBO-XGBoost model exhibits a slight increase in the FAR while significantly reducing the MRR, thereby enhancing its overall performance. Compared with KNN3 and KNN5, the DBO-XGBoost model has a lower FAR. The ROC curve and AUC of DBO-XGBoost, XGBoost, KNN5, KNN3, and CSCAT rain flags are shown in Figure 5. All machine learning models performed better than the CSCAT rain flag. Among all the curves, the ROC curve of the DBO-XGBoost classifier is the closest to the coordinate point (0.0, 1.0), and the area under the curve is the largest. The AUC of the DBO-XGBoost model is 3.93% higher than XGBoost, 3.08% higher than KNN5, 4.84% higher than KNN3, and 35.63% higher than the CSCAT rain flag. This shows that the performance of the DBO-XGBoost is better compared to the XGBoost, KNN5, KNN3, and CSCAT rain flags. Overall, DBO-XGBoost has the best comprehensive performance and excellent rain identification ability.

Table 2. Evaluation of DBO-XGBoost, XGBoost, KNN5, KNN3 models, and CSCAT rain flag.

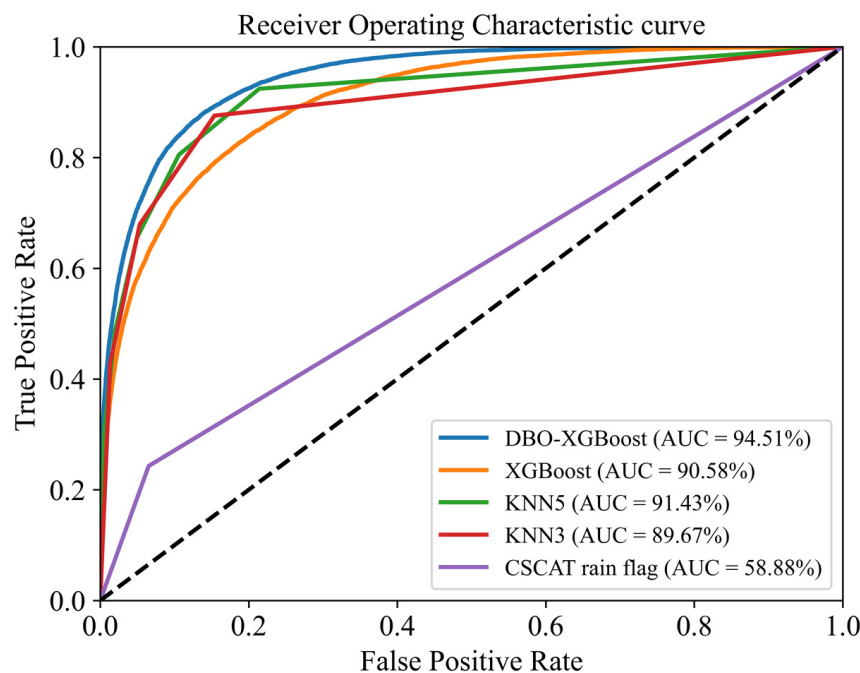| Model	Accuracy	Precision	FAR	MRR	Reject Rate	Actual Rain
DBO-XGBoost	91.03%	80.65%	2.92%	39.20%	12.57%	
XGBoost	88.89%	79.76%	2.21%	55.25%	9.36%	
KNN5	90.14%	72.68%	4.93%	34.54%	15.02%	16.68%
KNN3	90.21%	71.86%	5.33%	32.05%	15.77%	
CSCAT Rain Flag	86.72%	28.73%	6.53%	75.71%	8.26%	

Figure 5. ROC curve and AUC of DBO-XGBoost, XGBoost, KNN5, KNN3, and the CSCAT rain flag.
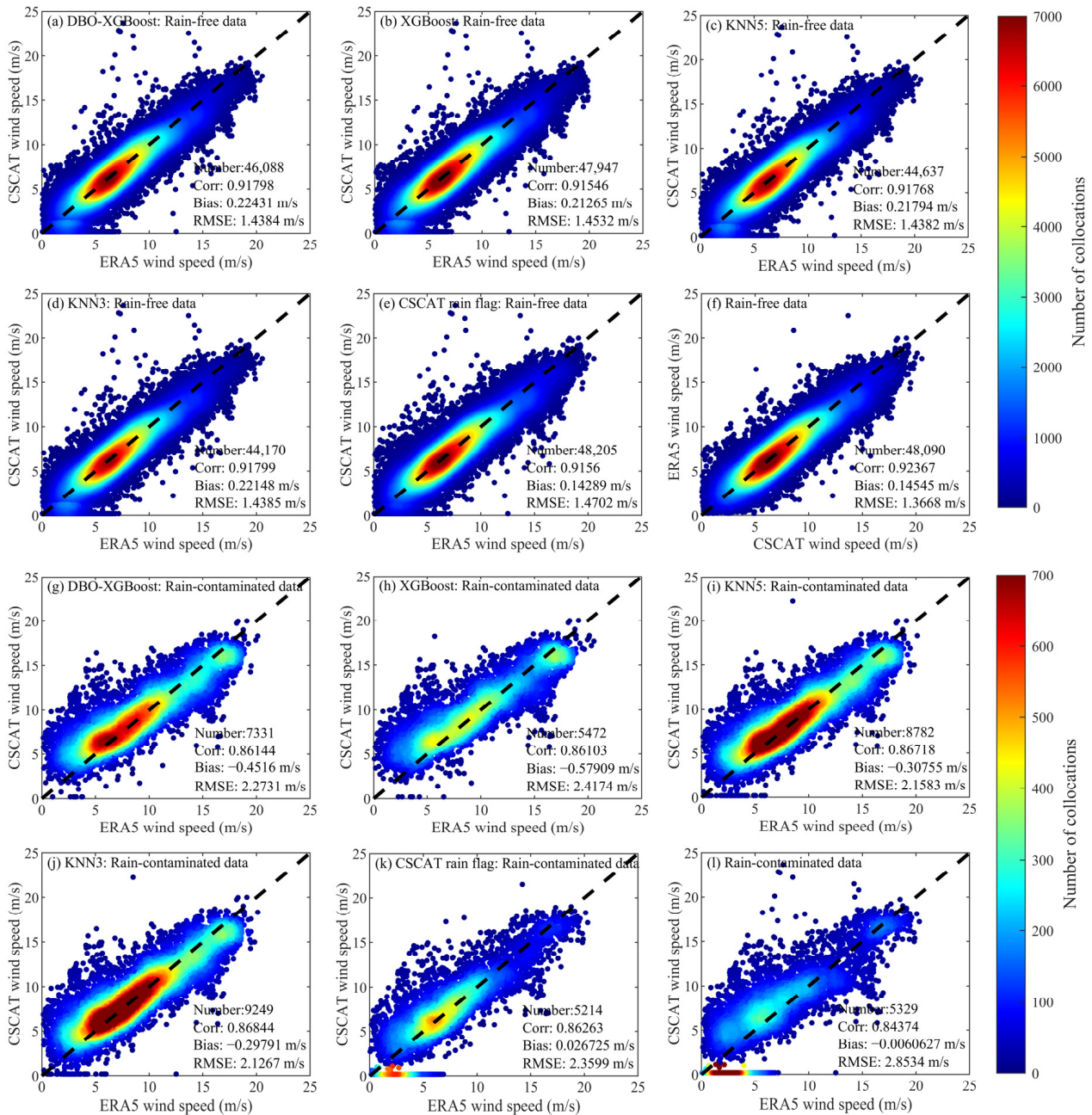
Figure 6 shows the retrieved wind speed scatter point density of rain-free data and rain-contaminated data flagged by DBO-XGBoost, XGBoost, KNN5, KNN3, and the CSCAT rain flag, which indirectly reflects the accuracy of the model in identifying rain. The root mean square error (RMSE) between CSCAT L2B wind speed and ERA5 wind speed without rain-contaminated data is calculated, as is the RMSE between CSCAT L2B wind speed and ERA5 wind speed flagged as rain-contaminated data by the models. The rain-free data (Figure 6a–f) and rain-contaminated data (Figure 6g–l) were flagged in the following four machine learning models: the CSCAT rain flag and the GPM KuPR collocation. The corr of all data, including rain-free data and rain-contaminated data, is 0.90752; the bias is 0.13155 m/s; and the RMSE is 1.5793 m/s. Due to the limited panel, this data is not shown in the figure. It can be seen that the RMSE of the wind field is reduced by all methods, and the effect of the machine learning model is better than that of the CSCAT rain flag. The effect of DBO-XGBoost is significantly improved compared to XGBoost and is comparable to the results of KNN5 and KNN3. The RMSE of the retrieved wind speed for data flagged as rain-contaminated is considerably high, surpassing 2 m/s in all cases. This indicates that filtered rain data significantly contributes to deviations in the retrieved wind speed. From the overall trend, the data affected by rain is overestimated in the low wind speed region and underestimated in the high wind speed region when the wind speed is retrieved.

3.2. Evaluation of the DBO-XGBoost Model in Rain Intensity Classification

In the experiment on rain identification, DBO-XGBoost, KNN5, and KNN3 all had good performance. In contrast, the XGBoost model performs worse but is still better than the CSCAT rain flag. The multi-classification ability of the machine learning model can further classify the rain intensity of the rain-contaminated data, which is an ability that the CSCAT rain flag does not have. Such classification can augment our capability to evaluate the degree of rain contamination in CSCAT data and facilitate subsequent product processing. According to the standard of the China Meteorological Administration, we classify rain intensity into four levels: Light rain ranges from 0.004 to 0.41 mm/h, heavy rain ranges from 0.41 to 2.08 mm/h, torrential rain ranges from 2.08 to 4.16 mm/h, and the rain rate of heavy downpour is above 4.16 mm/h. The experimental procedure for this subsection is as follows: Firstly, the training set used for constructing the rain identification model is also utilized in building the rain intensity classification model with the same hyperparameter.

Then, the real labeled rain-contaminated data derived from GPM collocation is used to train the model to ensure that it is realistic. However, as the rain intensity classification model was developed after the rain identification experiment, its objective is to further categorize the identified rain into distinct intensities. Therefore, during the model testing phase, we specifically selected the dataset that was previously identified as rain contamination to evaluate the performance of the rain intensity classification model. Figure 7 shows the process of rain intensity classification.

Figure 6. The scatter point density of rain-free data and rain-contaminated data flagged by DBO-XGBoost, XGBoost, KNN5, KNN3, and the CSCAT rain flag for retrieved wind speed.
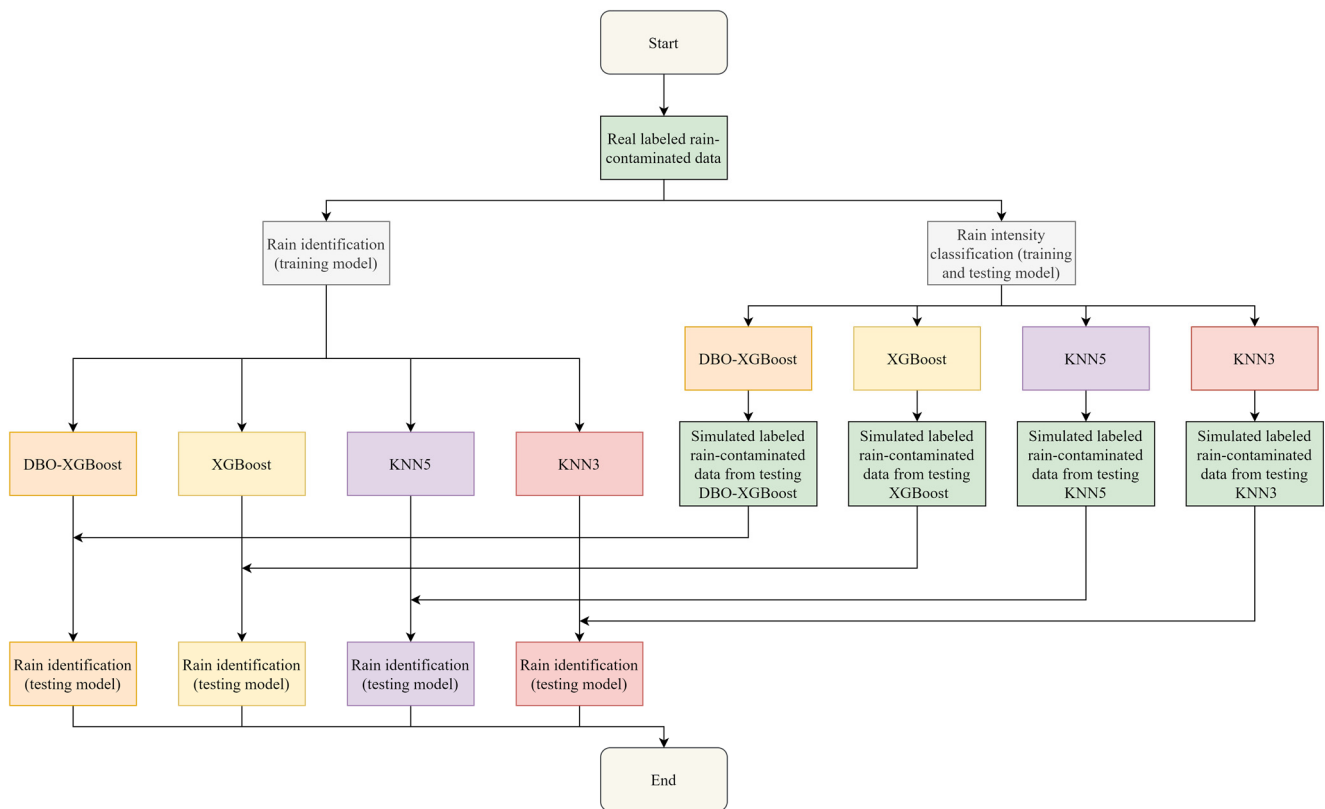
Figure 7. The process of rain intensity classification by using DBO-XGBoost, XGBoost, KNN5, and KNN3.

The evaluation of the DBO-XGBoost, XGBoost, KNN5, and KNN3 in the rain intensity classification is shown in Table 3. The rain intensity classification problem is addressed in the presence of rain, thereby obviating the need to consider FAR and MRR. Furthermore, accuracy is defined as the proportion of accurately predicted rain-contaminated data among all actual rain-level data. The evaluation metrics include accuracy, precision, and the comparison between rejection rate and proportion of actual rain.

The accuracy and precision of the XGBoost, KNN5, and KNN3 rain intensity classifications are only a few, reaching more than 80%. The performance of these models in accurately categorizing rain intensity is suboptimal, with more than half of the classification accuracy and rate falling below 70%. Rain intensity classification requires high classification ability of the model because there exists a certain correlation between the accuracy of classification across each level. For instance, the classification accuracy of KNN5 for rain is merely 32.01%, indicating that approximately 70% of rain instances are misclassified into other three levels instead of being correctly classified, thereby resulting in a low precision for those levels. The KNN5 model has demonstrated commendable performance in rain identification. However, it falls short of meeting the requirements for rain intensity classification. Obviously, the DBO-XGBoost model performs best among the four machine learning models, and its accuracy and precision for light rain, torrential rain, and heavy downpour levels have reached more than 80%, or even more than 90%. For the four machine learning models, the classification precision of heavy rain levels is lower than that of other levels. The precision of KNN3 is only 19.72%, but DBO-XGBoost in this term can still be close to 80%, which proves that DBO-XGBoost has an excellent classification ability across all rain intensities.

Table 3. The evaluation of the DBO-XGBoost, XGBoost, KNN5, and KNN3 in rain intensity classification.

Model	Rain Intensity	Accuracy	Precision	Reject Rate	Actual Rain
DBO-XGBoost	Light rain	89.08%	93.88%	52.30%	55.11%
	Heavy rain	90.28%	79.92%	35.35%	31.29%
	Torrential rain	82.88%	93.36%	6.79%	7.65%
	Heavy downpour	88.63%	94.70%	5.57%	5.95%
XGBoost	Light rain	78.34%	79.57%	53.03%	55.11%
	Heavy rain	69.04%	57.08%	37.04%	31.29%
	Torrential rain	35.54%	74.78%	4.02%	7.65%
	Heavy downpour	62.38%	74.41%	5.92%	5.95
KNN5	Light rain	78.71%	81.95%	58.86%	55.11%
	Heavy rain	65.14%	53.09%	33.51%	31.29%
	Torrential rain	32.01%	54.55%	3.75%	7.65%
	Heavy downpour	52.03%	67.23%	3.87%	5.95%
KNN3	Light rain	81.74%	97.46%	77.67%	55.11%
	Heavy rain	70.40%	19.72%	20.10%	31.29%
	Torrential rain	40.00%	39.29%	1.14%	7.65%
	Heavy downpour	53.12%	31.48%	1.10%	5.95%

During the process of rain identification, some rain-free data are mistakenly classified as rain-contaminated, which is expected to be classified as light rain rather than higher rain levels in the experiment of rain-level classification. This approach aims to minimize misclassification errors. In the results of the DBO-XGBoost model, the proportion of rain-free data being classified in the light rain category is 78%, and the proportion of no rain data being classified in the moderate rain category is 20.39%. Additionally, a mere 0.36% and 0.27% of no rain data are respectively misclassified as stormy rain and heavy stormy rain. This shows that DBO-XGBoost not only exhibits high accuracy in classifying rain intensity but also effectively reduces the misclassification observed in previous classification methods.

Figures 8–11 show the retrieved wind speed scatter point density of different rain intensities classified by DBO-XGBoost, XGBoost, KNN5, and KNN3, respectively, which indirectly reflects the performance of the model in identifying rain. It can be seen that higher rain intensity is associated with worse wind quality. Comparing the wind speed RMSE of DBO-XGBoost and XGBoost in light rain, it is found that the XGBoost model exhibits a larger RMSE in comparison to the DBO-XGBoost model when discerning the impact of light rain on wind speed. This discrepancy arises due to the enhanced capability of DBO-XGBoost in identifying a greater volume of light rain data. To be specific, due to the low accuracy and precision of light rain detected by XGBoost, a portion of higher-intensity rain is classified into the category of light rain, leading to an increased RMSE. Therefore, the effect of the wind speed on the same rain intensity between models is not suitable for direct comparison, necessitating a comprehensive analysis in conjunction with the evaluation metrics of the models.
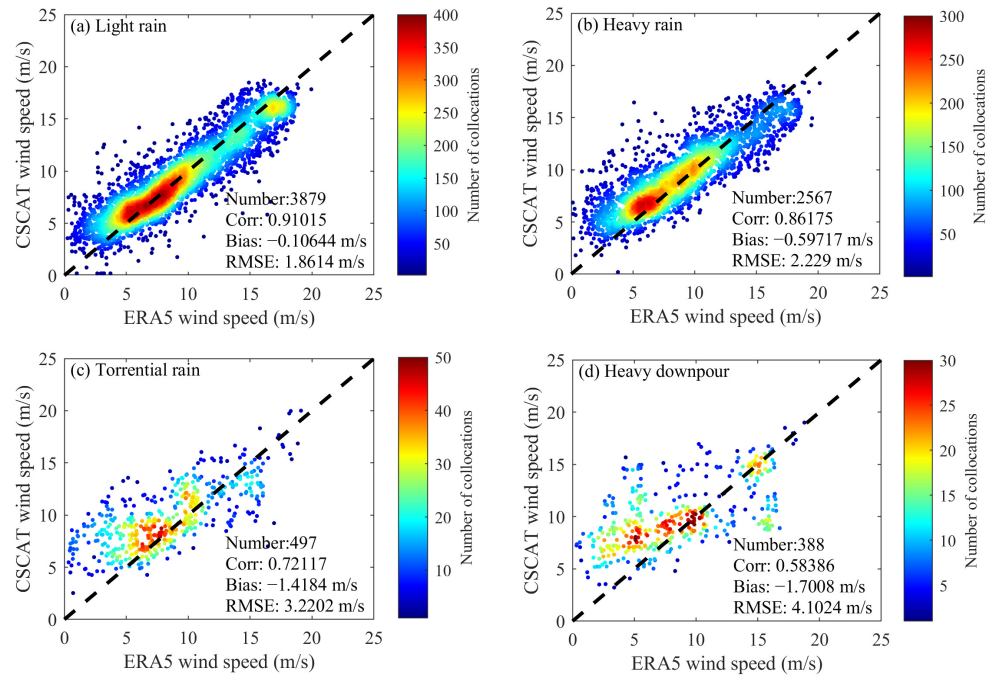
Figure 8. The scatter point density of different rain intensities classified by DBO-XGBoost for retrieved wind speed compared with ERA5 wind speed.
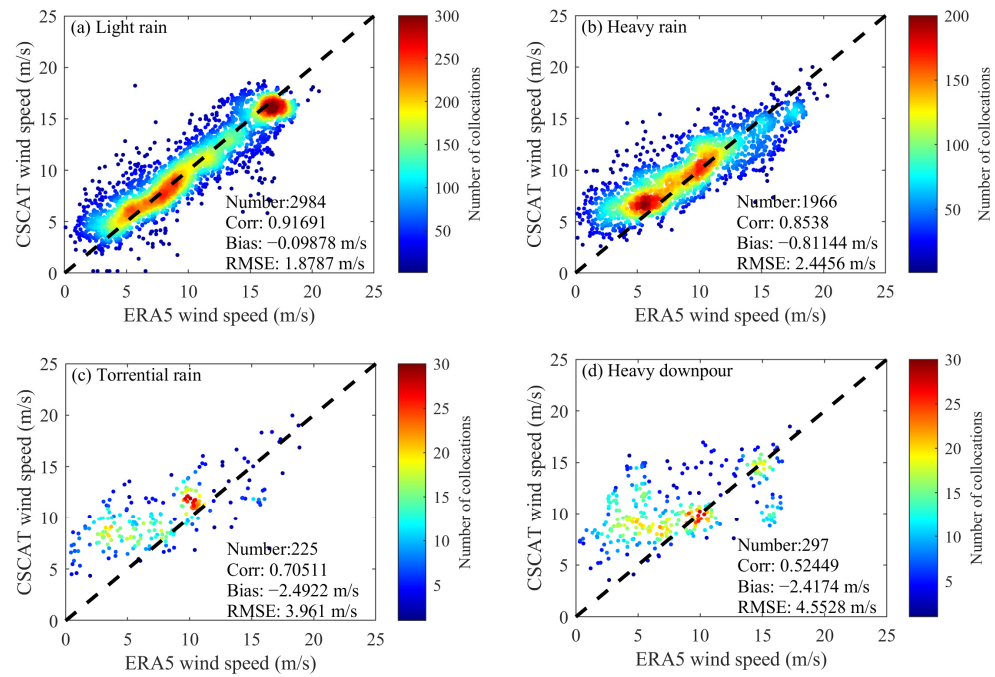
Figure 9. The scatter point density of different rain intensities classified by XGBoost for retrieved wind speed compared with ERA5 wind speed.
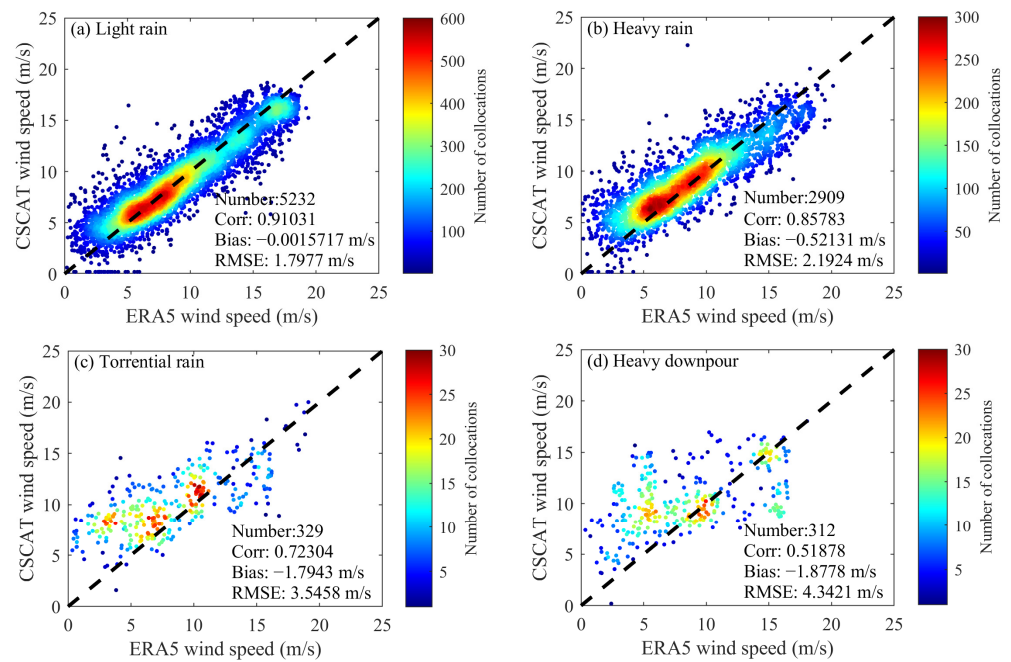
Figure 10. The scatter point density of different rain intensities classified by KNN5 for retrieved wind speed compared with ERA5 wind speed.
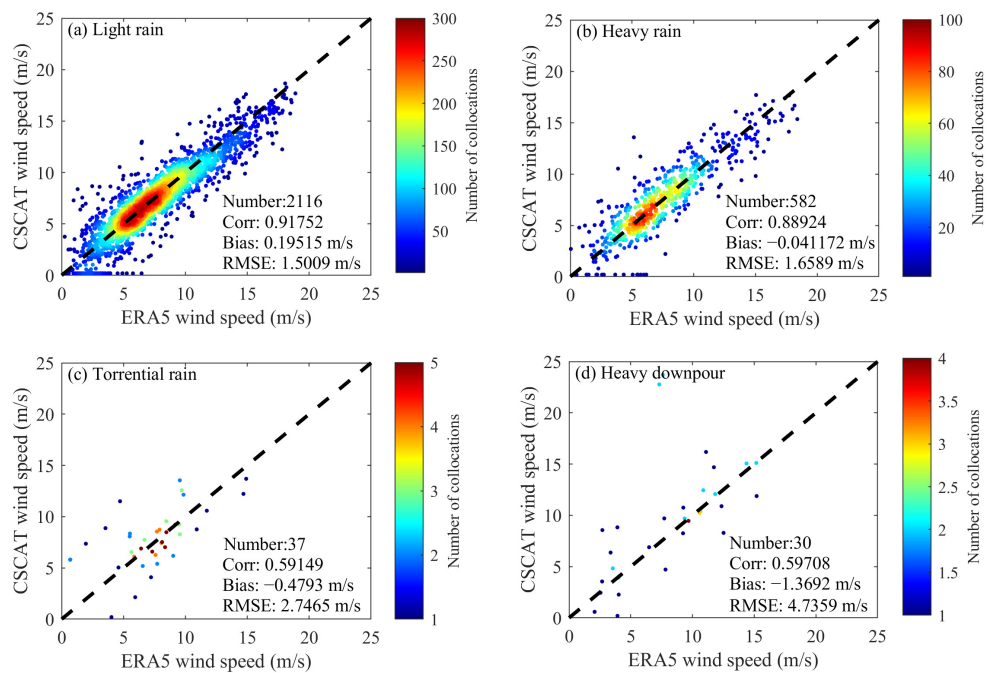

Figure 11. The scatter point density of different rain intensities classified by KNN3 for retrieved wind speed compared with ERA5 wind speed.

4. Discussion

The DBO-XGBoost model is constructed to realize rain identification and rain intensity classification in Section 3. However, the application ability of DBO-XGBoost is not clear under different background conditions. In this section, we use the data obtained under different background conditions to test the ability of DBO-XGBoost and determine its applicability and conditions of use.

4.1. Analysis of Model Performance under Two Different Sea Conditions

Currently, research findings indicate that variations in background wind speeds can influence the effect of rain on NRCS [9,30]. The impact of rain on NRCS is more pronounced at lower wind speeds but diminishes as wind speed increases. When the wind speed reaches approximately 10–15 m/s, the rain-induced effect on VV polarized NRCS becomes practically imperceptible, and the effect on HH-polarized NRCS is significantly mitigated compared to low wind speeds. When the wind speed reaches approximately 15–20 m/s, the rain-induced effect on HH-polarized NRCS disappears. As an important feature of rain identification, if the NRCS variation between rain and no-rain conditions is minimal, it will pose challenges in accurately identifying rain. Therefore, we suspect that the accuracy of rain identification will be reduced when the background wind speed is high. On the contrary, it will be more conducive to rain identification when the background wind speed is low.

The data selected for this study includes regions in the North India (55–95°E, 0–30°S) and the Northwest Pacific (20−40°N, 140−180°E), as there are significant differences in the WVC collocated ERA5 wind speeds between these two regions. Rain rates and wind speeds in the North India and Northwest Pacific are shown in Figure 12. The wind speed in the North India is comparatively higher, ranging between 10–15 m/s, whereas that of the Northwest Pacific exhibits a relatively lower range at 3–8 m/s. To facilitate a comparative analysis of the DBO-XGBoost performance in these two regions, we developed separate models exclusively for the North India and the Northwest Pacific using training data based on each region, which are more targeted than models built using global data.
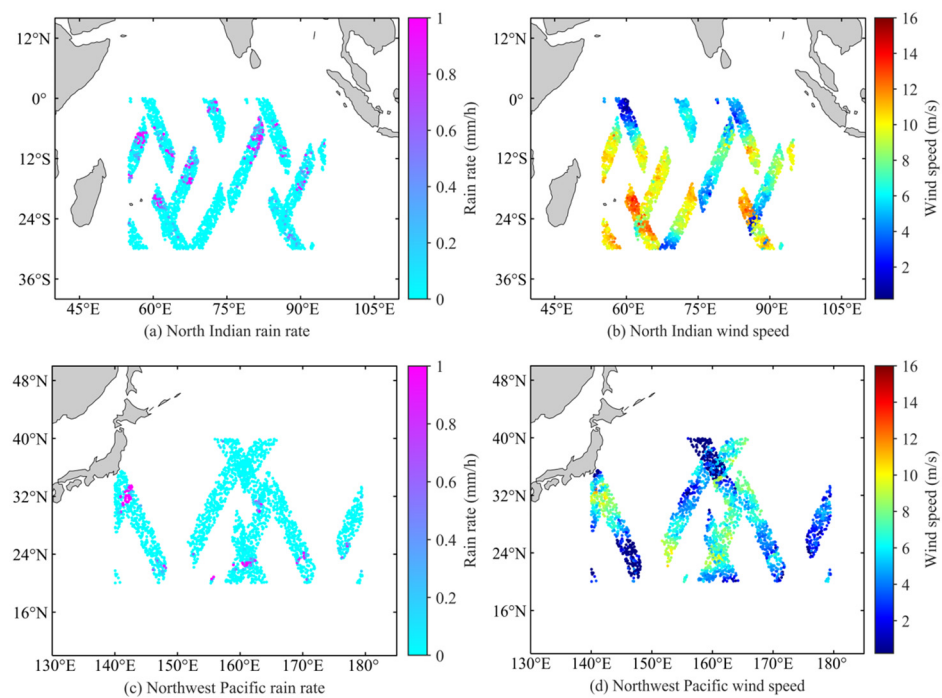

Figure 12. Rain and wind speed information for the North India and Northwest Pacific.

The evaluation of the DBO-XGBoost model, CSCAT rain flag, and local models in the North India and Northwest Pacific is shown in Table 4. The application of the DBO-XGBoost model in the Northwest Pacific outperforms its performance in the North India, which may be attributed to the more pronounced effect of rain on NRCS at lower wind speeds. Under different polarization and incidence angle conditions, the increased effect of rain on NRCS disappears in the wind speed range of 10–20 m/s. Therefore, it is more conducive to the identification of rain under a lower background wind speed [9]. The ROC curve and AUC of the DBO-XGBoost, local model, and CSCAT rain flag in North India and the Northwest Pacific are shown in Figure 13. The application effectiveness

of the constructed rain identification model surpasses that of the CSCAT rain flag across diverse sea conditions in both regions. In general, the global rain identification model has better generalization, and it is better than that of local rain identification models in terms of performance.

Table 4. The evaluation of the DBO-XGBoost model, CSCAT rain flag, and local models in the North Indian and Northwest Pacific.

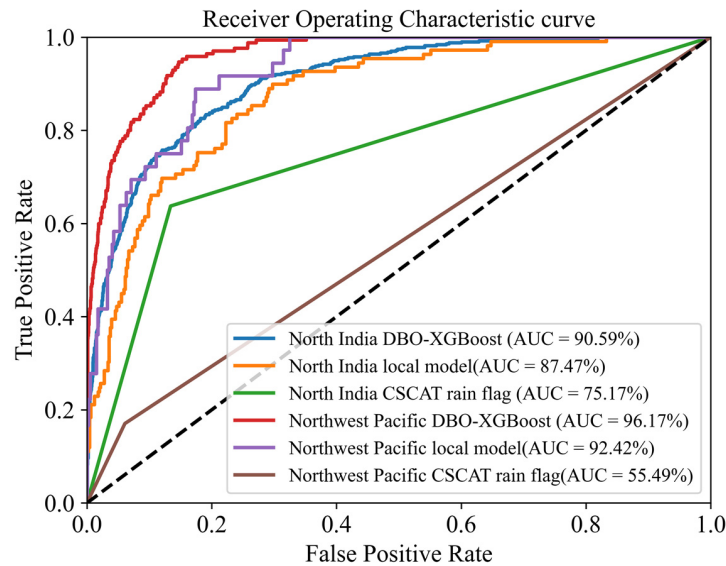| Region	Model	Accuracy	Precision	FAR	MRR	Reject Rate	Actual Rain
North Indian	DBO-XGBoost	88.77%	77.22%	2.48%	56.51%	9.12%	
	CSCAT rain flag	85.33%	63.74%	2.40%	78.16%	5.55%	16.20%
	Local model	85.90%	85.90%	4.13%	60.55%	10.37	
Northwest Pacific	DBO-XGBoost	95.15%	82.83%	0.85%	51.67%	4.57%	
	CSCAT rain flag	81.42%	17.05%	14.64%	64.71%	16.27%	7.86%
	Local model	93.30%	66.67%	1.76%	61.17%	4.85%	

Figure 13. ROC curve and AUC of the DBO-XGBoost, local model, and CSCAT rain flag in North India and the Northwest Pacific.

The evaluation of DBO-XGBoost in rain intensity classification in the North India and Northwest Pacific is shown in Table 5. Due to the use of rain-contaminated data identified by the DBO-XGBoost in the rain identification experiment (simulated rain-contaminated data) for classifying rain intensity, instead of actual rain data, the existence of FAR and MRR led to changes in the proportion of data under different rain levels. The proportion of simulated rain data for the four levels in the North India is 92.17%, 6.05%, 1.42%, and 0.36%, respectively, and for the Northwest Pacific, it is 92.93%, 4.04%, 2.02%, and 1.01%, respectively. The classification of rain intensity in the two regions shows that the accuracy and precision for both torrential rain and heavy downpours are perfect, with a remarkable rate of 100%. However, a misclassification occurs between light rain and heavy rain, with some instances of light rain being erroneously categorized as heavy rain and some instances of heavy rain being erroneously categorized as light rain in the North India. The rain intensity classification in the Northwest Pacific indicates that, except for a light rain accuracy rate of 83.70% and a heavy rain precision rate of 21.05%, all other accuracy and precision achieved 100%. This suggests that the model misclassifies a part of light rain into heavy rain, and the number of light rains accounts for 92.93% of the total rain-contaminated data, leading to the low precision of heavy rain. Overall, there is a common problem in the application of the DBO-XGBoost model in these two regions, wherein there exists a

certain misjudgment in distinguishing between light rain and heavy rain. Despite the limited occurrence of torrential rain and heavy downpours, they are still accurately and comprehensively identified.

Table 5. Evaluation of the DBO-XGBoost model in the North India and Northwest Pacific.

Region	Rain Level	Accuracy	Precision	Reject Rate	Actual Rain
North Indian	Light rain	89.58%	98.31%	83.99%	93.15%
	Heavy rain	76.47%	32.50%	14.23%	5.00%
	Torrential rain	100%	100%	1.42%	1.17%
	Heavy downpour	100%	100%	0.36%	0.68%
Northwest Pacific	Light rain	83.70%	100%	77.78%	95.93%
	Heavy rain	100%	21.05%	19.19%	2.45%
	Torrential rain	100%	100%	2.02%	0.74%
	Heavy downpour	100%	100%	1.01%	0.88%

4.2. Analysis of Model Performance by Using an Orbit of CSCAT Data

The occurrence of rain exhibits both temporal and spatial correlation, resulting in the observation of rain by a WVC of CSCAT being indicative of the presence of rain in the surrounding WVCs [31,32]. This phenomenon is inherent in physics, but it also has a certain randomness. Therefore, we chose an orbit of CSCAT data that has continuous space-time to test the performance of DBO-XGBoost.

The performance of the model constructed in this study is evaluated using an orbit of CSCAT data, identified as CFO_EXPR_SCA_C_L2A_OR_20200601T071229_08806. The wind speeds and rain rates of CSCAT and GPM detection tracks and collocating regions are shown in Figure 14. The red dotted box area is the area where the CSCAT data can collocate the KuPR. The upper and lower figures on the right are the CSCAT wind speed diagram and the KuPR rain diagram of the matching area, respectively. The white area is no data or invalid data, and the gray area is land.
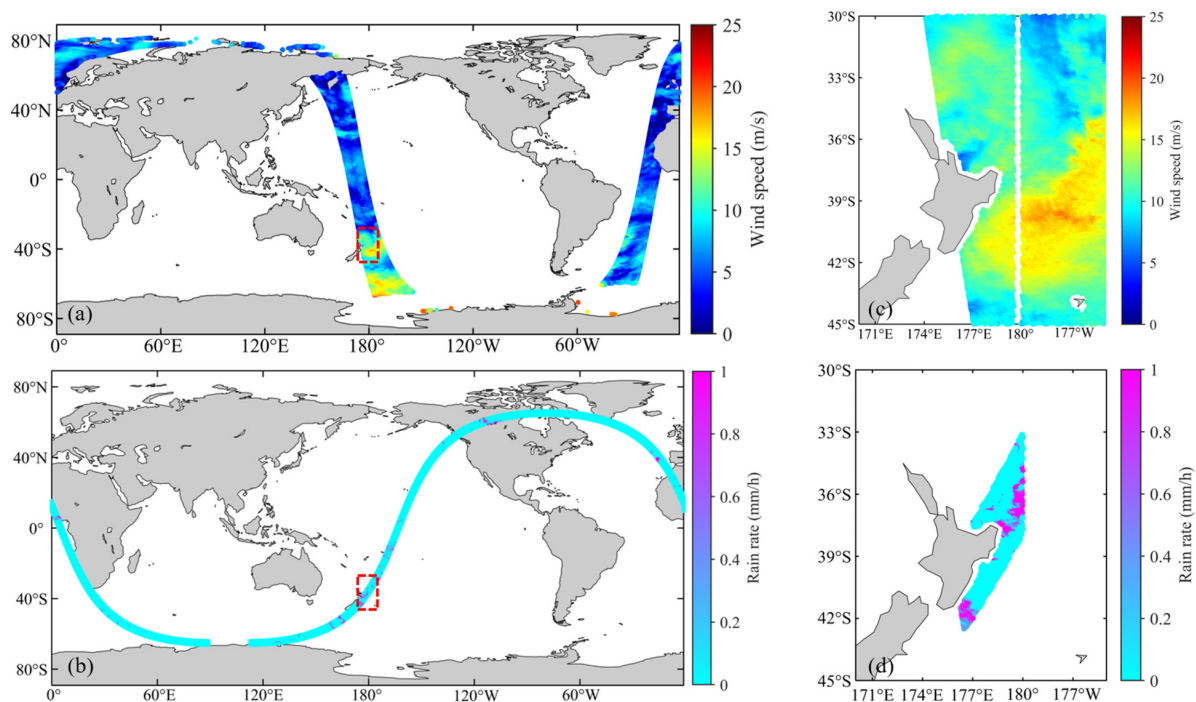

Figure 14. Rain identification by using CFO_EXPR_20200601T071229_08806 data. (**a**) CSCAT wind speed collocated area; (**b**) GPM rain collocated area; (**c**) collocated wind speed information; (**d**) collocated rain information.

The evaluation of DBO-XGBoost and CSCAT rain flags in rain identification using an orbit of CSCAT data is shown in Table 6. The rain identification effect of an orbit of CSCAT data is better than that of the local area in the previous section and has obvious advantages over the CSCAT rain indicator. Although the wind speed ranges from 10–15 m/s, the ability of the model to identify rain is still outstanding, possibly attributed to the analogous spatiotemporal characteristics of the environment, facilitating accurate identification of nearby precipitation events. The results show that the time-space continuous data meets the timeliness and spatial correlation of rain, thereby improving the ability of the DBO-XGBoost model to identify rain. The evaluation of DBO-XGBoost and CSCAT rain flags in rain intensity classification using an orbit of CSCAT data is shown in Table 7. The model has a good effect on the division of light rain and heavy downpour; the precision is more than 95%, and the accuracy is more than 85%. However, the classification effect of heavy rain and torrential rain did not reach the average level of the model, indicating a need for further research in the classification of heavy rain and torrential rain.

Table 6. Evaluation of DBO-XGBoost and CSCAT rain flags in rain identification using CFO_EXPR_20200601T071229_08806.

Model	Accuracy	Precision	FAR	MRR	Reject Rate	Actual Rain
DBO-XGBoost	94.39%	84.49%	3.06%	18.12%	16.40%	16.92%
CSCAT rain flag	59.62%	14.45%	36.07%	65.06%	35.90%	

Table 7. Evaluation of the DBO-XGBoost in rain intensity classification using CFO_EXPR_20200601T071229_08806.

Model	Rain Intensity	Accuracy	Precision	Reject Rate	Actual Rain
DBO-XGBoost	Light rain	91.04%	96.38%	84.06%	45.79%
	Heavy rain	67.27%	39.78%	13.48%	37.36%
	Torrential rain	64.29%	81.82%	1.59%	11.10%
	Heavy downpour	85.71%	100%	0.87%	5.76%

5. Conclusions

The XGBoost optimized by the DBO algorithm is used to construct DBO-XGBoost to realize rain identification and rain intensity classification, so as to realize quality control of CSCAT wind products. A dataset generated by the collocation of CSCAT and GPM is used for constructing DBO-XGBoost. We evaluated the performance of DBO-XGBoost and conducted a comparative analysis with KNN, XGBoost, and CSCAT rain flag. In terms of rain identification, the QCs of all machine learning models are better than those of the CSCAT rain flag. DBO-XGBoost shows better performance than XGBoost and is comparable to KNN5 and KNN3. In the experiment of classifying light rain, heavy rain, torrential rain, and heavy downpour, the DBO-XGBoost demonstrates its excellent performance compared with XGBoost, KNN3, and KNN5.

Furthermore, we evaluate the performance of DBO-XGBoost in rain identification and rain intensity classification under two different sea conditions. The accuracy and precision of DBO-XGBoost in rain identification at low wind speeds are higher than those at high wind speeds, and the FAR and MRR are lower than those at high wind speeds. This is probably due to the diminishing impact of rain on NRCS with increasing wind speed. DBO-XGBoost misclassifies only a part of light rain at low wind speed while misclassifying light rain and heavy rain at high wind speed, indicating that the classification of rain intensity at low wind speed is more accurate. An orbit of CSCAT data is used to evaluate the performance of DBO-XGBoost. The results show that continuous data in time and space is more conducive to rain identification but did not significantly improve the classification of rain intensity. Therefore, the classification of heavy rain and torrential rain still needs further research.

The DBO-XGBoost model developed for rain identification and rain intensity classification has several advantages: (1) The rain-contaminated data can be directly flagged without collocating with other external data, which improves the timeliness and utilization rate of CSCAT data. (2) Compared with the CSCAT rain flag, the DBO-XGBoost model exhibits superior rain identification ability and possesses the capacity to classify rain intensity to evaluate the severity of rain events that the CSCAT rain flag lacks. (3) The machine learning model can simplify the data processing flow, which is more efficient than the traditional rain flag method. In the future, we will consider the correction of rain-contaminated data and make full use of CSCAT-measured information to play to its advantages.

References

1. Liu, J.; Lin, W.; Dong, X.; Lang, S.; Yun, R.; Zhu, D.; Zhang, K.; Sun, C.; Mu, B.; Ma, J. First results from the rotating fan beam scatterometer onboard CFOSAT. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8793–8806. [CrossRef]
2. Melsheimer, C.; Alpers, W.; Gade, M. Simultaneous observations of rain cells over the ocean by the synthetic aperture radar aboard the ERS satellites and by surface-based weather radars. *J. Geophys. Res. Ocean.* **2001**, *106*, 4665–4677. [CrossRef]
3. Stiles, B.W.; Yueh, S.H. Impact of rain on spaceborne Ku-band wind scatterometer data. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1973–1983. [CrossRef]
4. Weissman, D.E.; Bourassa, M.A.; Tongue, J. Effects of rain rate and wind magnitude on SeaWinds scatterometer wind speed errors. *J. Atmos. Ocean. Technol.* **2002**, *19*, 738–746. [CrossRef]
5. Tournadre, J.; Quilfen, Y. Impact of rain cell on scatterometer data: 1. Theory and modeling. *J. Geophys. Res. Ocean.* **2003**, *108*. [CrossRef]
6. Draper, D.W.; Long, D.G. Evaluating the effect of rain on SeaWinds scatterometer measurements. *J. Geophys. Res. Ocean.* **2004**, *109*. [CrossRef]
7. Hilburn, K.; Smith, D.K.; Wentz, F.J. Rain effects on scatterometer systems. In Proceedings of the NASA Ocean Vector Wind Science Team Meeting, Boulder, CO, USA, 18–20 May 2009.
8. Zhang, B.; Alpers, W. The effect of rain on radar backscattering from the ocean. In *Advances in SAR Remote Sensing of Oceans*; CRC Press: Boca Raton, FL, USA, 2018; pp. 317–330.
9. Zhao, X.; Lin, W.; Portabella, M.; Wang, Z.; He, Y. Effects of rain on CFOSAT scatterometer measurements. *Remote Sens. Environ.* **2022**, *274*, 113015. [CrossRef]
10. Huddleston, J.N.; Stiles, B.W. A multidimensional histogram rain-flagging technique for SeaWinds on QuikSCAT. In *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No. 00CH37120)*; IEEE: Honolulu, HI, USA, 2000; pp. 1232–1234.
11. Gohil, B.S.; Sikhakolli, R.; Gangwar, R.K.; Kumar, A.S.K. Oceanic rain flagging using radar backscatter and noise measurements from Oceansat-2 scatterometer. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 2050–2055. [CrossRef]
12. Portabella, M.; Stoffelen, A. Rain detection and quality control of SeaWinds. *J. Atmos. Ocean. Technol.* **2001**, *18*, 1171–1183. [CrossRef]
13. Portabella, M.; Stoffelen, A. Characterization of residual information for SeaWinds quality control. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2747–2759. [CrossRef]

14. Mears, C.; Wentz, F.; Smith, D. *SeaWinds on QuikSCAT Normalized Objective Function Rain Flag*; Version 1.2; Product Description; Remote Sensing Systems: Santa Rosa, CA, USA, 2000.
15. Portabella, M.; Stoffelen, A.; Lin, W.; Turiel, A.; Verhoef, A.; Verspeek, J.; Ballabrera-Poy, J. Rain effects on ASCAT-retrieved winds: Toward an improved quality control. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2495–2506. [CrossRef]
16. Lin, W.; Portabella, M.; Stoffelen, A.; Verhoef, A.; Turiel, A. ASCAT wind quality control near rain. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4165–4177. [CrossRef]
17. Koch, W. Directional analysis of SAR images aiming at wind direction. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 702–710. [CrossRef]
18. Xu, X.; Stoffelen, A. Improved rain screening for ku-band wind scatterometry. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2494–2503. [CrossRef]
19. Zhao, K.; Stoffelen, A.; Verspeek, J.; Verhoef, A.; Zhao, C. Bayesian Algorithm for Rain Detection in Ku-Band Scatterometer Data. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–16. [CrossRef]
20. Ghosh, A.; Varma, A.K.; Shah, S.; Gohil, B.; Pal, P.K. Rain identification and measurement using Oceansat-II scatterometer observations. *Remote Sens. Environ.* **2014**, *142*, 20–32. [CrossRef]
21. Peng, Y.; Xie, X.; Lin, M.; Ran, L.; Yuan, F.; Zhou, Y.; Tang, L. A study of sea surface rain identification based on HY-2A scatterometer. *Remote Sens.* **2021**, *13*, 3475. [CrossRef]
22. Xu, X.; Stoffelen, A.; Lin, W.; Dong, X. Rain false-alarm-rate reduction for CSCAT. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1–5. [CrossRef]
23. Chen, Y.; Cui, Y.; Lin, W.; Liu, J.; Sun, C.; Lang, S. The impacts of assimilating CFOSAT scatterometer winds for Typhoon cases based on real-time rain quality control. *Atmos. Res.* **2023**, *285*, 106621. [CrossRef]
24. Xue, J.; Shen, B. Dung beetle optimizer: A new meta-heuristic algorithm for global optimization. *J. Supercomput.* **2023**, *79*, 7305–7336. [CrossRef]
25. Toyoshima, K.; Masunaga, H.; Furuzawa, F.A. Early evaluation of Ku-and Ka-band sensitivities for the global precipitation measurement (GPM) dual-frequency precipitation radar (DPR). *Sola* **2015**, *11*, 14–17. [CrossRef]
26. Lasser, M.; Foelsche, U. Evaluation of GPM-DPR precipitation estimates with WegenerNet gauge data. *Atmos. Meas. Technol.* **2019**, *12*, 5055–5070. [CrossRef]
27. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T. *Xgboost: Extreme Gradient Boosting*; R Package Version 0.4-2; 2015; Volume 1, pp. 1–4. Available online: https://cran.ms.unimelb.edu.au/web/packages/xgboost/vignettes/xgboost.pdf (accessed on 10 September 2023).
28. 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.
29. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory* **1967**, *13*, 21–27. [CrossRef]
30. Moore, R.; Yu, Y.; Fung, A.; Kaneko, D.; Dome, G.; Werp, R. Preliminary study of rain effects on radar scattering from water surfaces. *IEEE J. Ocean. Eng.* **1979**, *4*, 31–32. [CrossRef]
31. Bell, T.L. A space-time stochastic model of rainfall for satellite remote-sensing studies. *J. Geophys. Res. Atmos.* **1987**, *92*, 9631–9643. [CrossRef]
32. Bacchi, B.; Kottegoda, N.T. Identification and calibration of spatial correlation patterns of rainfall. *J. Hydrol.* **1995**, *165*, 311–348. [CrossRef]