applied sciences

MDPI

Article

A Quality Control Method for High Frequency Radar Data Based on Machine Learning Neural Networks

Chunye Zhou [1,2], Chunlei Wei [3,*], Fan Yang [3] and Jun Wei [1,2,*]

[1] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), School of Atmospheric Sciences, Sun Yat-sen University, Zhuhai 519082, China; zhouchy28@mail2.sysu.edu.cn
[2] Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, Key Laboratory of Tropical Atmosphere-Ocean System, Ministry of Education, Guangzhou 519082, China
[3] Zhuhai Marine Environmental Monitoring Central Station of the State Oceanic Administration, Zhuhai 519000, China; yangf@scs.mnr.gov.cn
* Correspondence: weicl@scs.mnr.gov.cn (C.W.); weijun5@mail.sysu.edu.cn (J.W.)

Abstract: We propose a quality control method based on machine learning neural networks to enhance the quality of high-frequency (HF) radar data. Unlike traditional quality control methods that rely on radar signals as indicators and involve extensive data manipulation in specialized software, our approach employs a Bi-LSTM neural network model. This method aims to improve data quality and streamline the quality control process. Through a series of analyses, we demonstrate the feasibility of using machine learning techniques to enhance radar data quality.

Keywords: Bi-LSTM; HF radar; quality control; anomaly detection; F-score; recall

1. Introduction

Hainan Island, nestled in the northern part of the South China Sea, enjoys a warm tropical climate. This climate bestows upon the island a rich biodiversity and stunning natural landscapes. Economically speaking, the island stands out as a hotspot for both tourism and fishing. However, the tropical climate brings its own set of challenges, including seasonal storms and the long-term effects of climate change. Given Hainan's unique status in China and the various challenges it faces, gaining insights regarding and monitoring its marine currents is crucial for ensuring the ecological and economic stability of Hainan and its surrounding maritime regions.

High-Frequency Radar (HFR) has emerged as a revolutionary tool in marine observation in recent years. By harnessing high-frequency electromagnetic waves, it not only captures subtle changes on the ocean's surface but also, through cutting-edge inversion techniques, accurately deciphers activities of winds, waves, and crucial ocean currents [1–4]. The application of this technology offers researchers extensive, real-time marine data, and provides essential informational support and solutions in the realms of global climate change and marine safety.

HFR technology has demonstrated high accuracy in measuring winds, waves, and currents on the ocean surface. This achievement has been widely recognized by the academic community [5–10]. However, like other measuring tools, HFR faces inherent challenges in its measurements. These include inaccuracies due to equipment errors, electromagnetic interference, or data loss during transmission [11,12]. In previous studies, the quality control efforts for HFR were mainly centered on radial velocity data. These data reveals the speed component of the ocean's surface from the radar to a specific point. To obtain a complete ocean current vector from this data, radial velocity information is needed from at least two different radar locations. These data are then combined to produce a two-dimensional vector, representing the flow direction and speed of the ocean's surface at a specific point.

check for updates

Citation: Zhou, C.; Wei, C.; Yang, F.; Wei, J. A Quality Control Method for High Frequency Radar Data Based on Machine Learning Neural Networks. *Appl. Sci.* **2023**, *13*, 11826. https://doi.org/10.3390/app132111826

Academic Editor: Alexandre Carvalho

Received: 15 September 2023
Revised: 22 October 2023
Accepted: 27 October 2023
Published: 29 October 2023

Copyright: © 2023 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/).

To form a continuous ocean current vector field, researchers typically employ mathematical methods to merge radial velocity data from multiple radar stations onto a grid. One commonly used method relies on inverse interpolation techniques, utilizing the distance and angular differences between radar stations to estimate velocity values for each grid point. In addition, there is a method based on the least squares principle that optimally fits data gathered from different radar stations by optimizing the model. Despite utilizing these advanced mathematical approaches, the merging process can still produce anomalous vector data. These anomalies can stem from mismatches between radar data, limitations of the merging algorithm, the incomplete or subpar quality control of radial velocities, and other factors. Issues with the quality of radial velocities, such as noise, discontinuity, or other unstable elements, might be partially addressed in the initial quality control phase but can be magnified or lead to new anomalies during the merging process.

In terms of radar data quality control, researchers have adopted various methods. Traditional approaches mainly focus on the quality control of radial velocity data obtained from high-frequency radars. The accuracy of these data largely hinges on the quality of electromagnetic signals and inherent characteristics of radial velocities. Past research has concentrated on the following areas:

Signal-to-noise ratio control: Cosoli [13] and colleagues advocated for using the signal-to-noise ratio as a pivotal quality metric. They discovered that, when the ratio exceeds a certain threshold, the data accuracy notably improves. Based on this, they devised an algorithm to filter and correct radar data according to the signal-to-noise ratio.

Spatial analysis: Roarty [14] and his team proposed a quality control method based on spatial characteristics. They examined the latitude, longitude, average radial direction, and speed of the radial data. Using this information, they assessed the data quality. Moreover, they compared the measured radial velocity with theoretical values, serving as a benchmark for quality control.

Parameter monitoring: Haines's team [15] recognized that radars output parameters beyond just radial velocities. They utilized these parameters for quality control, believing that these parameters provide added evidence for the quality of radial velocity data.

Real-time monitoring: Lorente's group [16] introduced a practical approach. They installed real-time monitoring equipment on buoys within the radar area, enabling instant diagnostics of non-speed parameters. This offered real-time quality feedback for radar data, assisting researchers in adjusting their quality control strategies.

Although these methods significantly improved the quality of radar data, anomalies might still emerge during the vector field synthesis. Thus, we present a new machine learning method specifically tailored for the quality control of these synthesized vector fields.

In this study, we employed machine learning techniques for the quality control of the synthesized radar ocean-current vector velocity data. We used the Bi-LSTM model to analyze the time-domain data and utilized its predictive residuals for anomaly detection. Compared to traditional methods, machine learning offers a more automated, efficient, and real-time solution. Conventional approaches often involve multiple steps, such as manual filtering, threshold setting, and geographical validation, which are not only time-consuming but may also introduce human errors. In contrast, deep learning models can automatically extract data features, substantially reducing the need for manual intervention and providing more consistent and reliable results. By applying machine learning for data anomaly detection, we demonstrated its potential in streamlining processes and enhancing data quality. The results of this research provide valuable guidance for improving radar data quality.

In this section, we have discussed the high-frequency radar technology and its applications and challenges in marine observations. In response to these challenges, we introduce a novel method utilizing machine learning for radar data quality control. In the sections that follow, we will describe in detail the data and methods used, then present our research outcomes, and wrap up with a summary of the entire study.

2. Materials and Methods

2.1. High-Frequency Radar

In our study, we employed the OSMAR-S100 high-frequency compact ground-wave radar system developed by Wuhan University [17]. The system was manufactured by Beijing Highlander Digital Technology Company Limited in Beijing, China. Compared to large line-of-sight radar systems, this compact high-frequency radar boasts multiple advantages, including reduced equipment size, decreased physical space requirements (with antenna aperture no larger than several-hundred meters), lower energy consumption, and the convenience of antenna equipment installation and maintenance. The OSMAR-S100 HF radar operates as a networked radar with a working frequency range from 13 to 16 MHz. As depicted in Figure 1, a total of 19 radar units were installed, distributed across cities like Haikou, Sanya, Ledong, Changjiang, and Wenchang. The synthesized detection area (107.5° E–111.5° E, 17.5° N–21.5° N) spans over 40,000 square kilometers.

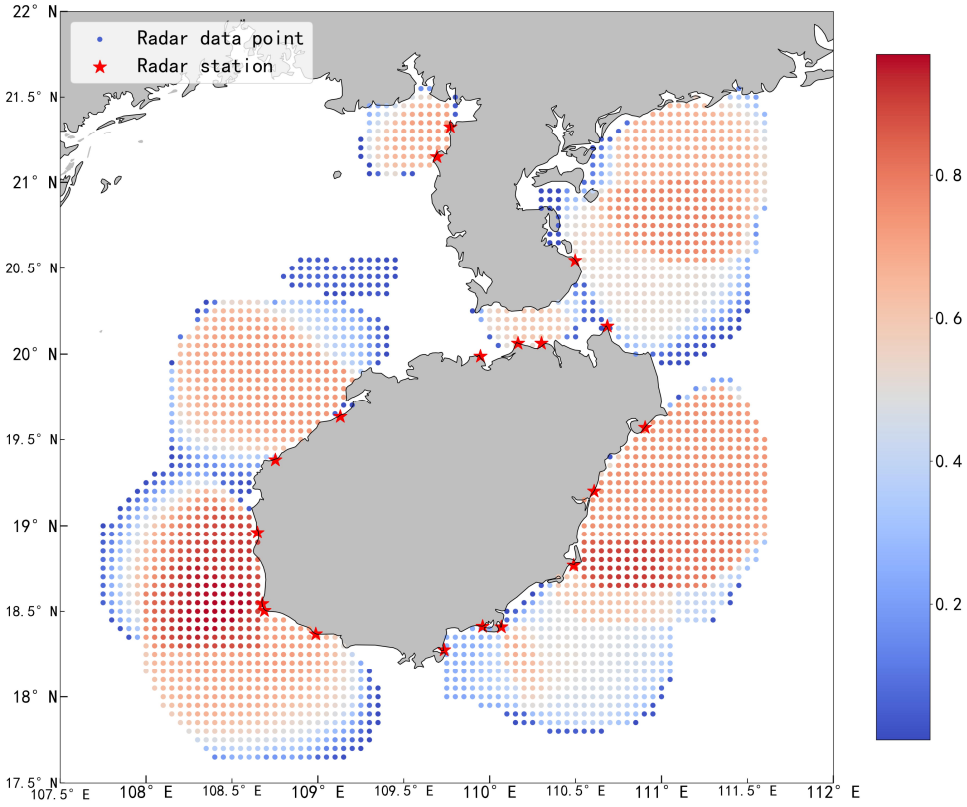


Figure 1. Data coverage of the HF radar network system. Circular markers indicate the locations where radar data are collected, with colors representing radar acquisition rates (indicated by the color bar). Red stars mark single radar station locations.

The data used in this research were collected from 1 January 2022, to 31 October 2022, with a time resolution of 12 min and a spatial resolution of 5 km. Due to factors such as equipment maintenance and varying environmental disturbances at radar stations, the data collection rate differed across locations. In subsequent processing, we selected points with a data collection rate exceeding 80% to ensure data reliability. For the purpose of training and testing our machine learning models, the data were averaged into hourly values.

2.2. Prediction-Based Anomaly Detector

When dealing with one-dimensional time series data, there are two common methods for anomaly detection. The first is window-based anomaly event sequence detection, where the time series data are divided into multiple windows, and then the anomaly score for each time subsequence within each window is calculated [18,19].

The second method is prediction-based anomaly detection. Since anomalies are events that deviate from expectations, a natural approach is to compare the predicted data with the actual data to perform anomaly detection [20,21]. In our study, we employed the prediction-based anomaly detection method as illustrated in the figure below (Figure 2).

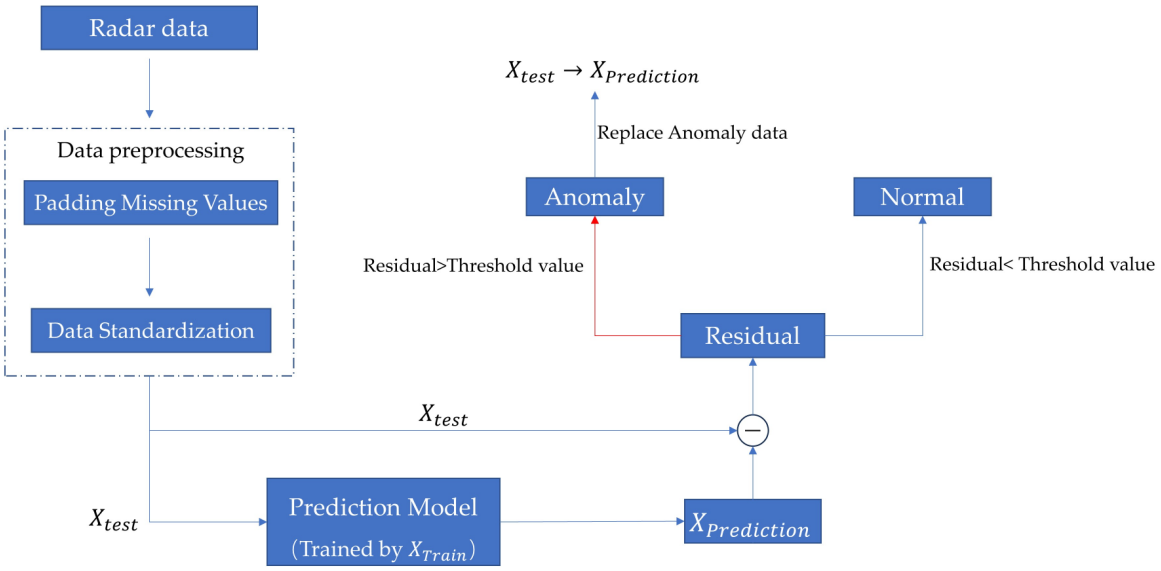


Figure 2. Model anomaly detection process flowchart.

2.2.1. Data Preprocessing

Before the radar data are fed into a predictive model, they undergo essential data preprocessing steps. In the real-world operating environment of radar systems, various interference factors often lead to the omission of data. In this study, to align radar data with the requirements of the model, two critical processes were applied: the imputation of missing data and standardization of data.

To maintain the continuity of the time series data, a thorough timestamp check was conducted on radar time series. This ensures that the time intervals between data points within each segment of the time series remain consistent. If, within a segment, the time interval between adjacent data points exceeds a predefined standard time interval (typically 1 h in this study), it indicates the presence of missing data. In such cases, a linear interpolation technique is employed to reconstruct the missing data. Moreover, recognizing that ocean-current data often exhibit periodicity, in this study, we refrained from imputing missing data for segments with time gaps exceeding 6 h.

To enhance the training effectiveness of neural network models, it is customary to standardize input data. In this research, the z-score standardization method was utilized. This method involves transforming the data to have a mean of zero and a standard deviation of one, making them suitable for input into the neural network model. These preprocessing steps ensure that radar data are effectively prepared for model training and utilization.

2.2.2. Bi LSTM Neural Network

Introduced by Hochreiter and Schmidhuber in 1997 [22], the LSTM neural network model enhances the memory mechanism inherent in Recurrent Neural Networks (RNNs). Traditional RNNs often encounter the "vanishing gradients" problem, leading to challenges in learning long-term dependencies. LSTM addresses this by integrating "gate" units. These gate units control the processes of reading, writing, and erasing information in memory cells, thus enabling the LSTM to capture and learn from long-term dependencies in data sequences more effectively.

Various LSTM variants have been developed over the years. However, research [23] indicates that these adaptations offer no significant improvement over the standard LSTM

model. In light of this, our study utilizes the conventional LSTM model, as depicted in Figure 3. The model is characterized as follows:

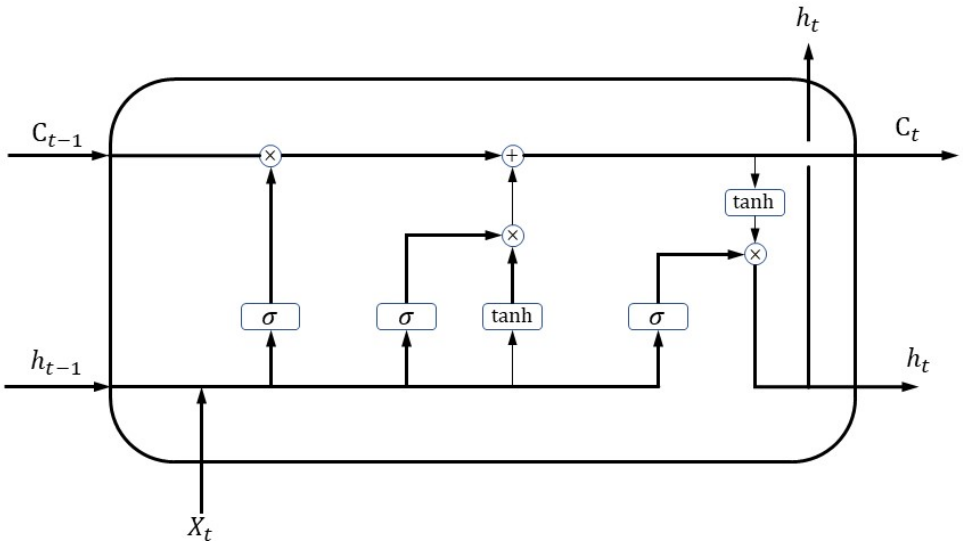


Figure 3. Diagrammatic illustration of the LSTM cell architecture.

Inputs and gates: At a given time step t, the LSTM unit accepts the output from the previous time step, h_{t-1}, and the current input vector, x_t. These inputs interact with three crucial gate units:

Forget gate (f_t): Determines the extent of information in the memory cell c_t to be discarded.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

Input gate (i_t): Decides the magnitude of new information to be written into the memory cell.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

Output gate (o_t): Establishes the output h_t for the current time step.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{3}$$

$$h_t = o_t \cdot \tan h(C_t) \tag{4}$$

Memory cell update: The memory cell's state is updated based on the outputs from the forget and input gates.

$$\widetilde{C}_t = \tan h(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \tag{6}$$

Key Parameters:

W_f, W_i, W_o, and W_c: Weights for the respective gates and cell update.

b_f, b_i, b_o, and b_c: Biases for the respective gates and cell update.

σ: The sigmoid activation function, responsible for introducing nonlinearity.

The above formulas and terminologies present a comprehensive picture of the LSTM's underlying computations, highlighting its ability to capture and learn long-term dependencies in data sequences.

A vast body of research has proven that the Long Short-Term Memory (LSTM) neural network model is effective for analyzing time series data. The unique architecture of LSTMs enables them to effectively capture long-term dependencies within time series data.

Building upon this foundation, we further adopted the Bidirectional LSTM (Bi-LSTM) model. The concept of the Bi-LSTM neural network is derived from bidirectional recurrent neural networks (RNNs) [24]. Unlike the traditional unidirectional LSTM, the Bi-LSTM incorporates both forward and backward LSTM layers (as illustrated in Figure 4), endowing it with bidirectional data processing capabilities.

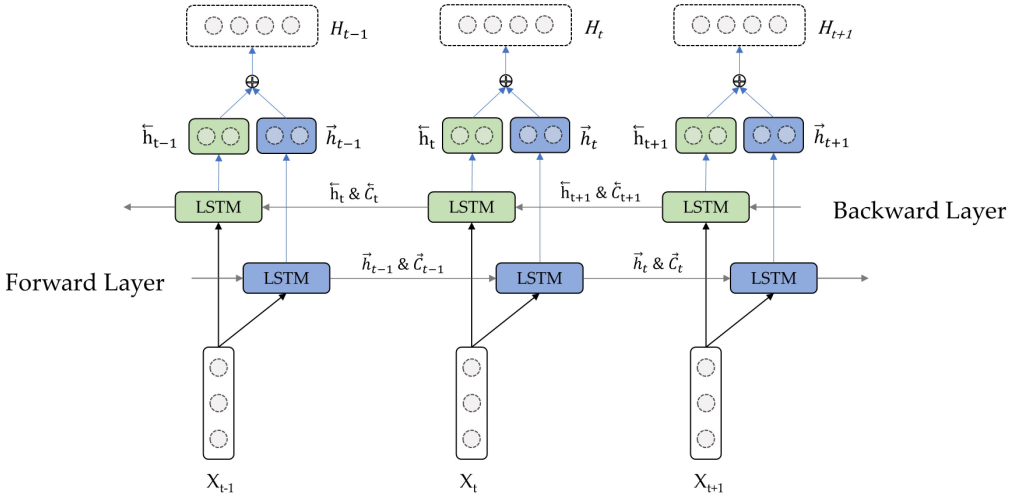


Figure 4. Bidirectional LSTM neural network architecture diagram.

The bidirectionality of Bi-LSTM offers a holistic perspective on ocean-current behaviors, considering both historical and anticipated data. This approach is especially pertinent to the complex patterns of ocean currents, ensuring that detected anomalies are substantial and not fleeting disturbances. The bidirectional structure provides a rich contextual framework for our predictive model, enhancing the precision and reliability of anomaly detection. Numerous studies have affirmed the utility of Bi-LSTM in scenarios requiring consideration of contextual information, such as classification [25,26], speech recognition [27,28], and time series forecasting [29,30].

Neural network structures typically consist of three main components: the input layer, the hidden layers, and the output layer. In this research, the Bi-LSTM layer consists of 512 neurons and is followed by the stacking of multiple fully connected layers, introducing non-linear transformations and dimensionality changes to the output. During the model training process, to ensure that the predictive model can learn the features of normal ocean current data effectively, the training data undergo meticulous curation, excluding anomalous data points. This ensures that the model exhibits a certain level of sensitivity when dealing with abnormal data, thereby ensuring efficiency and accuracy.

2.2.3. Anomaly Detection

This section introduces the anomaly detection process utilized in our study and the associated evaluation metrics. As shown in Figure 2, when the data are fed into the trained model, they produce an output as a time series. The difference between this predicted output and the original input forms the 'residual time series'. Ideally, instances of anomalies should be evident as significant deviations in this residual series, while normal data points are expected to have minor variations from the predicted values. To identify anomalies, we establish a threshold for the residual time series. Data points with residuals below this threshold are classified as normal, while those with residuals exceeding the threshold are identified as anomalies.

In our study, the threshold is determined through an evaluation of anomaly detection performance on the test dataset. For imbalanced datasets such as radar ocean-current data, the performance of anomaly detection algorithms is typically assessed using the Precision–Recall Curve (PRC) and the F-score. The vertical axis of the PRC represents

precision, while the horizontal axis represents recall. Precision, Recall, and the F-score are defined as follows:

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

Precision: The ratio of true positives (correctly identified anomalies) to the sum of true positives and false positives (false alarms), reflecting the algorithm's accuracy in anomaly detection.

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{8}$$

Recall: The ratio of true positives (correctly identified anomalies) to the sum of true positives and false negatives (missed anomalies), indicating the algorithm's capability to detect all anomalies.

$$\text{F} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

F-score: The harmonic mean of precision and recall. It offers a comprehensive measure of the algorithm's performance, factoring in both false positives and false negatives.

In our preceding discussion, we introduced three evaluation metrics for anomaly detection. For the quality control of high-frequency radar data, both precision and recall are of paramount importance. Precision represents the model's ability to correctly label actual anomalies, ensuring the reliability of our anomaly identifications. Meanwhile, recall assesses the model's thoroughness in capturing all genuine anomalies, emphasizing the completeness of the identification process. However, precision and recall often exhibit opposing trends as the detection threshold varies. Over-emphasizing precision may minimize false alarms but at the risk of potentially overlooking some genuine anomalies. Conversely, prioritizing recall could lead to a higher likelihood of labeling normal data as anomalous. To harmonize these two essential metrics and ensure optimal quality control for our radar data, we adopt the F-score. This comprehensive measure takes into account both precision and recall, ensuring that our quality control neither misses genuine anomalies nor raises excessive false alarms.

3. Results

As previously mentioned, the threshold is determined by comparing the anomaly detection performance metrics on the test set. The test set consists of full time series data from 15 different locations, totaling over 58,000 h. These data have been meticulously labeled to distinguish various anomaly types. Their respective quantities and categories are depicted in the accompanying Figure 5.

We categorized anomalies into two main classes:

Single anomalies: These are isolated incidents where data at a specific time point deviate significantly from normal values, often due to momentary equipment faults or external interference.

Continuous anomalies: These anomalies manifest over multiple consecutive time points, possibly arising from equipment issues or external events. They can be further divided into:

Short-term: Spanning two consecutive time points, indicating brief yet noticeable disturbances.

Long-term: Persisting for more than two consecutive time points. These anomalies are less frequent, as they are more likely to have been removed in earlier quality control procedures.

As illustrated in Figure 5, subplot (a) displays the different types of anomalies within a specific time frame. The marked points on this plot provide a clear distinction between the different anomaly classifications. Meanwhile, subplot (b) demonstrates the frequency of each anomaly type. Single anomalies occurred 947 times, making them the most commonly observed type. Short-term continuous anomalies were noted at 194 distinct time points, while long-term continuous anomalies were observed across 102 different time points.

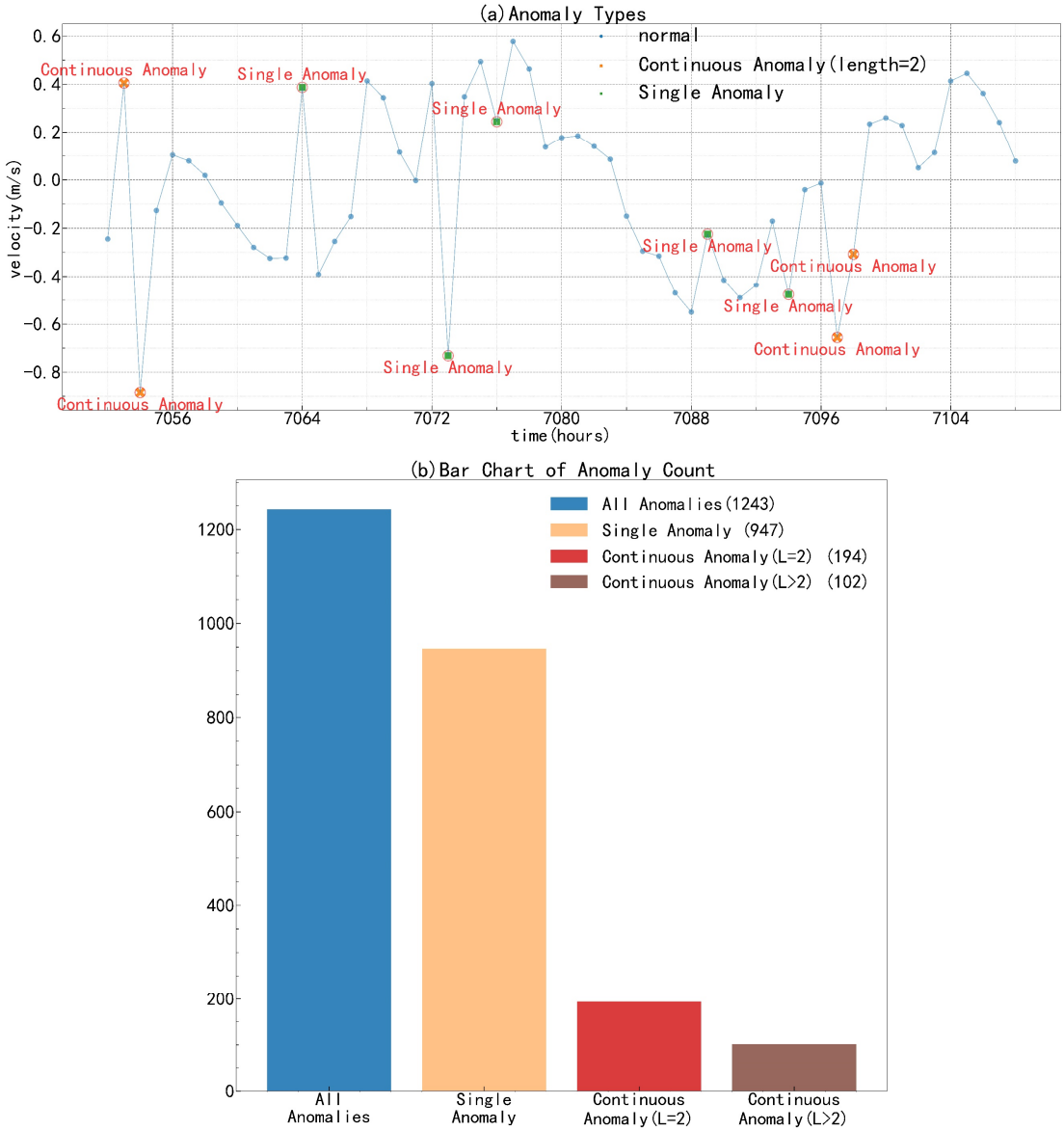


Figure 5. (**a**) Schematic representation of anomaly types and (**b**) bar chart of anomaly type counts.

Despite the chosen locations being in areas with high data coverage, evident anomalies were observed in these core regions. This underscores the central role of anomaly detection in data analysis. Particularly in the edge regions probed by the radar station, which are low data coverage areas, the data not only face significant continuity issues but are also more susceptible to various interferences and synthetic problems, exacerbating the anomaly conditions. Hence, the precise detection and handling of these anomalies become particularly vital to ensure data quality and accuracy, laying a solid foundation for subsequent data analysis.

In the ensuing discussions, we will delve into the anomaly detection methods and results for each of these three anomaly types.

3.1. Single Anomaly

As illustrated by Figure 5b, single anomalies are the most prevalent type of anomaly. In the actual operation of radar systems, due to the combined effects of various random error factors, such as short-term signal interference, single anomalies tend to occur more frequently. On the other hand, long-duration anomalies, because of their more evident

abnormal features, are typically filtered out before the synthesizing of radar ocean current vector data. Consequently, single anomalies are the primary focus of this study.

Figure 6a illustrates the input sequence structure of the prediction model. When the model attempts to predict the time point X_t, it considers data spanning six hours around that point. This design ensures that the model can fully capture the temporal context information adjacent to X_t, thereby enhancing the prediction accuracy. More detailed explanations regarding the choice of the input sequence length will be provided in the sensitivity analysis experiment at the end of the Section 3.

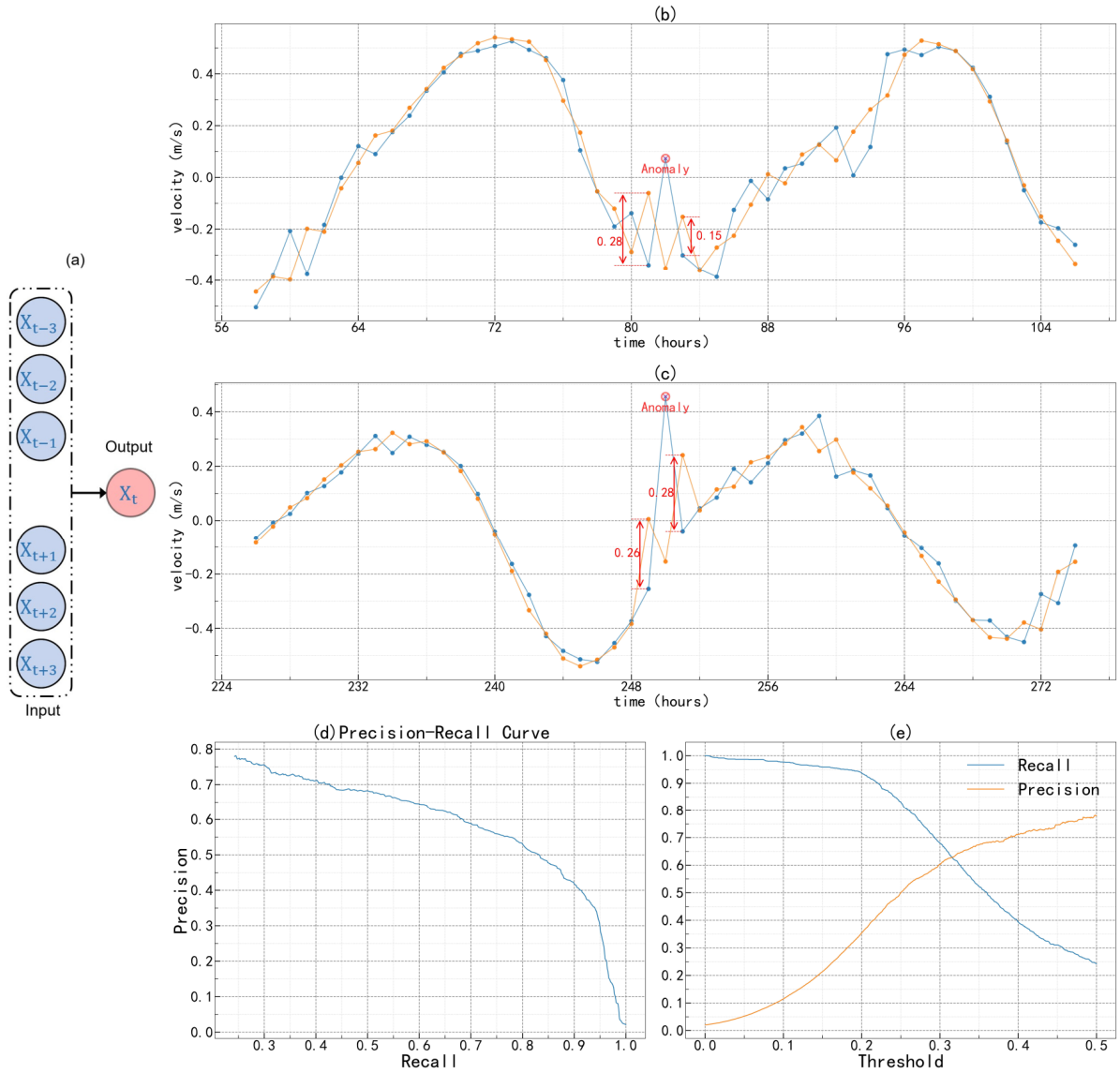


Figure 6. (**a**) Schematic representation of model input sequence, (**b**,**c**) comparison between model predicted output (orange curve) and radar data (blue curve) for single-point anomalies, (**d**) model PR curve plot, (**e**) trend graph of recall and precision as threshold varies.

As shown in Figure 6b,c, the values in the residual time series at the timestamps corresponding to anomalies are significantly high, confirming our model's anomaly detection capabilities. Nevertheless, it is worth mentioning that the normal data around these single-point anomalies also exhibit significant discrepancies in the difference series. One reason for this is that, when predicting these normal points, the model used data from the single-point anomalies in constructing its input sequence. This led to pronounced

discrepancies even in what should have been normal predictions. With a lower threshold, this can mistakenly classify these points as anomalies, compromising the overall accuracy.

Figure 6d depicts the relationship between the precision and recall. The thresholds start from 0 and increase in increments of 0.01 m/s up to 0.5 m/s. These thresholds are applied to the residual time series generated by the model and the test set to obtain the variation curves of precision and recall, further aiming to determine the optimal threshold. As previously discussed in Figure 6b,c, Figure 6e provides a broader perspective from the entire dataset, showcasing the misclassifications at lower thresholds: despite the high recall in this range, the precision is low. When the threshold exceeds 0.2 m/s, there is a notable decline in recall with only a slight increase in precision. Constrained by these factors, the F-score reaches its maximum value of 0.627.

These prediction biases were expected given the nature of the training data. During previous model training, only normal data were used as input. However, when applying the model in real-world scenarios, the inclusion of a significant amount of anomalous data in the dataset becomes part of the prediction model's input sequence, affecting its capability to accurately predict regular data.

Addressing the aforementioned issue and enhancing the anomaly detection performance necessitates improvements to our model. It should be capable of discerning the most relevant input information while minimizing the influence of anomalous data on its outputs.

To enhance the model's ability to recognize anomalous data and its robustness, we decided to draw inspiration from data augmentation techniques in deep learning. We adopted an innovative strategy: injecting simulated anomalous information into the normal data. This strategy aims to optimize the model's generalization capabilities. Experience tells us that solely relying on the original training data may not be sufficient to achieve optimal results in real-world anomaly detection tasks. However, if the model can learn and adapt to simulated anomalous data during the training phase, its ability to recognize anomalies and stability in real scenarios will be significantly improved.

For simulating anomalous information, we opted for Gaussian distribution, also termed normal distribution, a common probability distribution in statistics. This distribution simulates the overall effect of multiple small random error factors in the real world. According to the Central Limit Theorem, these effects exhibit a shape that is approximately normally distributed, making Gaussian distribution an ideal choice for our purposes.

During the anomaly injection process, we fine-tuned the Gaussian distribution's relevant parameters to ensure that the produced anomalies clustered mainly within a range akin to observed real-world anomalies. Specifically, from 15 different locations, we randomly selected 10% of the data and introduced anomalies at the 't' time point. To emulate various sudden changes potentially present in actual data, we then chose another distinct 10% data subset, introducing anomalies at either the 't − 1' or 't + 1' time points.

Figure 7a,b compares the predictions of the original and improved models at two different time points. As observed from the charts, the improved model can effectively identify anomalous data by producing large residuals. Simultaneously, it excels at making accurate predictions for normal data, with almost negligible residuals.

Figure 7c presents the PR curves of both models. As illustrated in the chart, the PR curve of the improved model consistently envelops that of the original model, indicating that, at equivalent recall levels, the improved model achieves greater precision.

Figure 7d displays the precision and recall of the new model at different thresholds. Viewing the entire dataset further validates the observations made from Figure 7a,b, where the model significantly reduced residuals for normal data even in the presence of anomalous data. This is reflected in the graph where the model retains high precision even at lower thresholds.

All sub-figures in Figure 7 validate the efficacy of our model's improvements. The F-score of the new model has increased from 0.63 to 0.79, indicating, roughly, a 16% improvement over the old model. By injecting simulated anomalous data into the training set,

we have not only enhanced the model's generalization capabilities but also its robustness against anomalies.

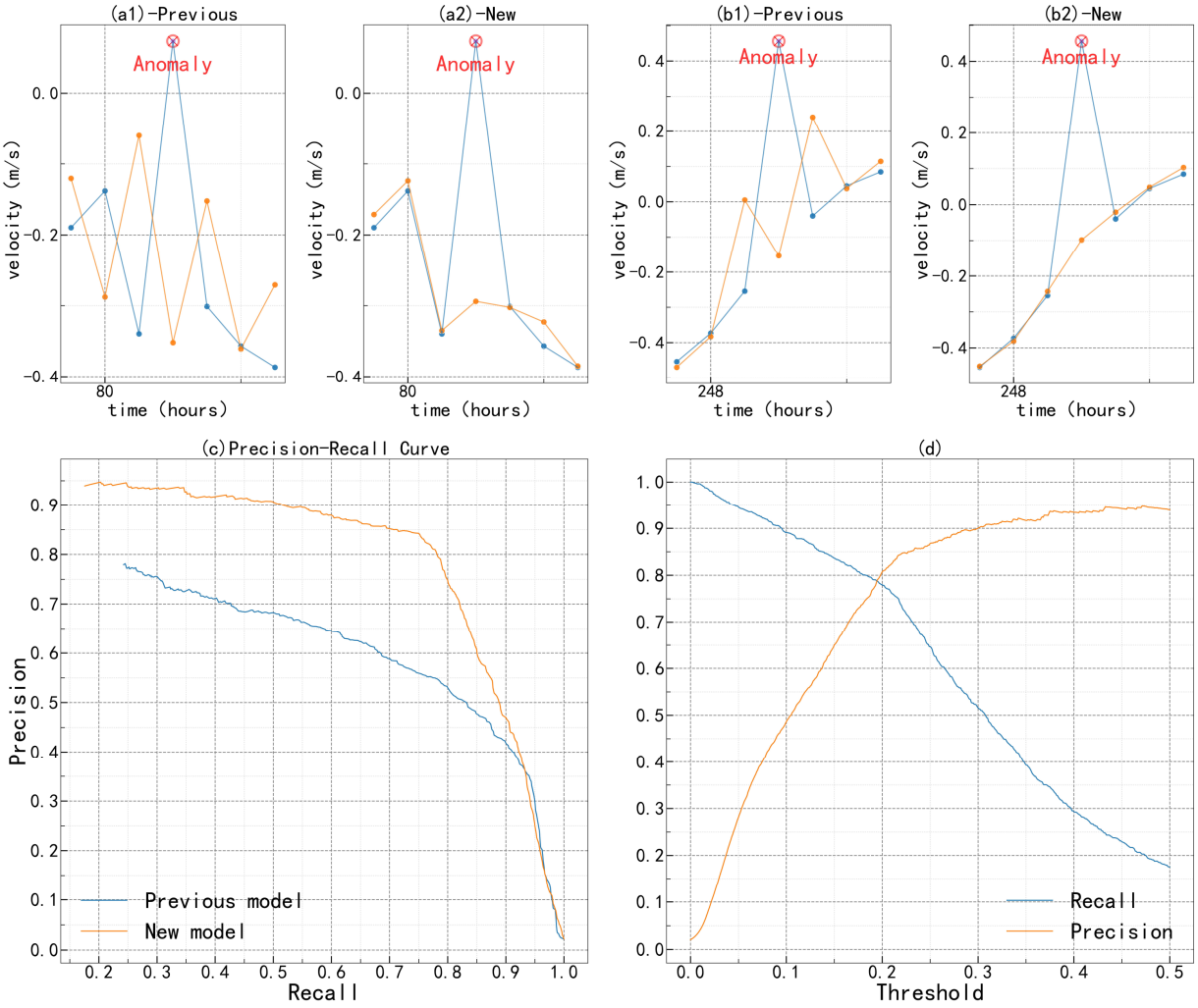


Figure 7. (**a**,**b**) Comparison of predictions before and after model improvement, with the orange curve representing the predicted model data and the blue curve representing the original radar data. (**c**) PR (precision–recall) curve before and after model improvement, (**d**) recall and precision graph as threshold varies after improvement.

3.2. Continuous Anomalies

Continuous anomalies represent another focal area of our study. Compared to single-point anomalies, these anomalies present a greater detection challenge due to their persistent nature in time series data. As shown in Figure 8a,b, when the predictive model encounters continuous anomalies, it struggles to generate predictions that significantly deviate from the anomalous data. This observation can be explained in two ways. Firstly, while the Bi-LSTM model excels at capturing long-term dependencies in time series data, when faced with consecutive anomalous data points, the model may over-rely on previous "memories" to interpret the current data point, leading to reduced sensitivity to continuous anomalies. Secondly, although the bidirectional nature of Bi-LSTM allows it to capture both past and future context in an input sequence, this mechanism may lack sufficient discriminative power when faced with a large number of consecutive anomalies.

Figure 8c illustrates the comparison of the recall rates for continuous and single-point anomalies at different thresholds. As the threshold increases, the recall rate for single-point anomalies decreases slowly, while that for continuous anomalies drops rapidly. Combining

the specific examples from Figure 8a,b, it becomes clear that the model, when predicting certain continuous anomalies, cannot generate sufficiently large residuals to distinguish them from normal data. Figure 8d shows that, even when the model's F-score is at its peak, the recall rate for continuous anomalies remains significantly lower than that for single-point anomalies. The findings from all sub-figures in Figure 8 emphasize the challenges that continuous anomalies pose to the predictive model.

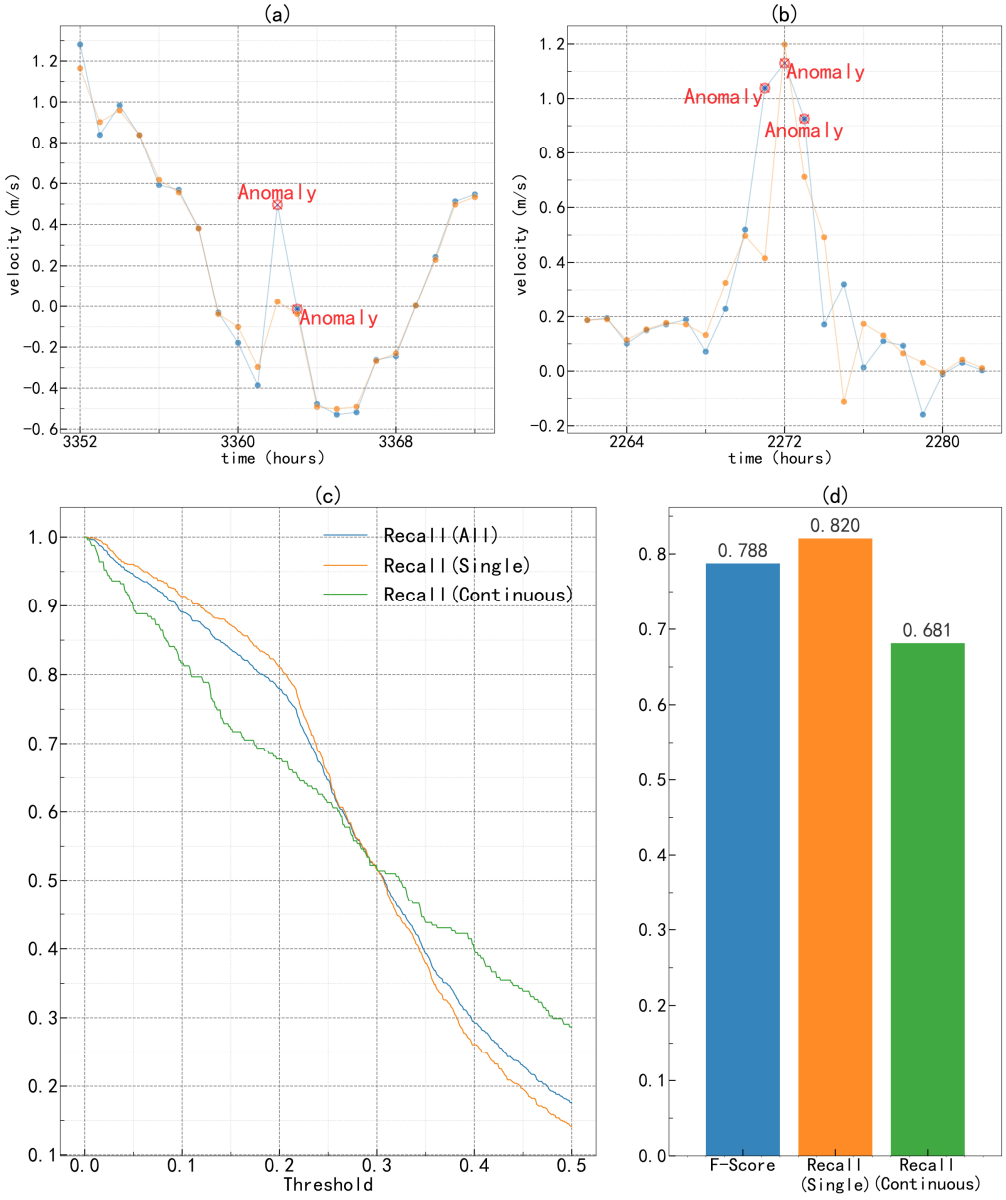


Figure 8. (**a**,**b**) Comparison between model predicted output (orange curve) and radar data (blue curve) for continuous anomalies, (**c**) illustration of recall for different anomaly types as threshold varies, (**d**) bar chart of recall for various anomaly types when model achieves maximum F-score.

In summary, while the model's performance in detecting single-point anomalies is commendable, there is room for optimization in the detection of continuous anomalies. Thus, the next steps in research should focus on further enhancing the model's performance in this area.

This iterative process presents a practical solution for addressing continuous anomalies, effectively transforming them into a sequence of single-point anomaly detections. This approach streamlines the detection process and has the potential to enhance the model's performance in handling continuous anomalies.

In order to ensure rationality, during the iterative forecasting process, the data being replaced should ideally consist of as many abnormal data points as possible. As a result, the selected threshold should guarantee a high level of accuracy. In this study, a threshold corresponding to an accuracy of 90% or higher was used. As shown in Figure 7d, the minimum threshold value is 0.3 m/s. The threshold values range from 0.3 to 0.5 m/s in increments of 0.01 m/s, and the number of iterations ranges from one to five times. The specific results are illustrated in Figure 9.

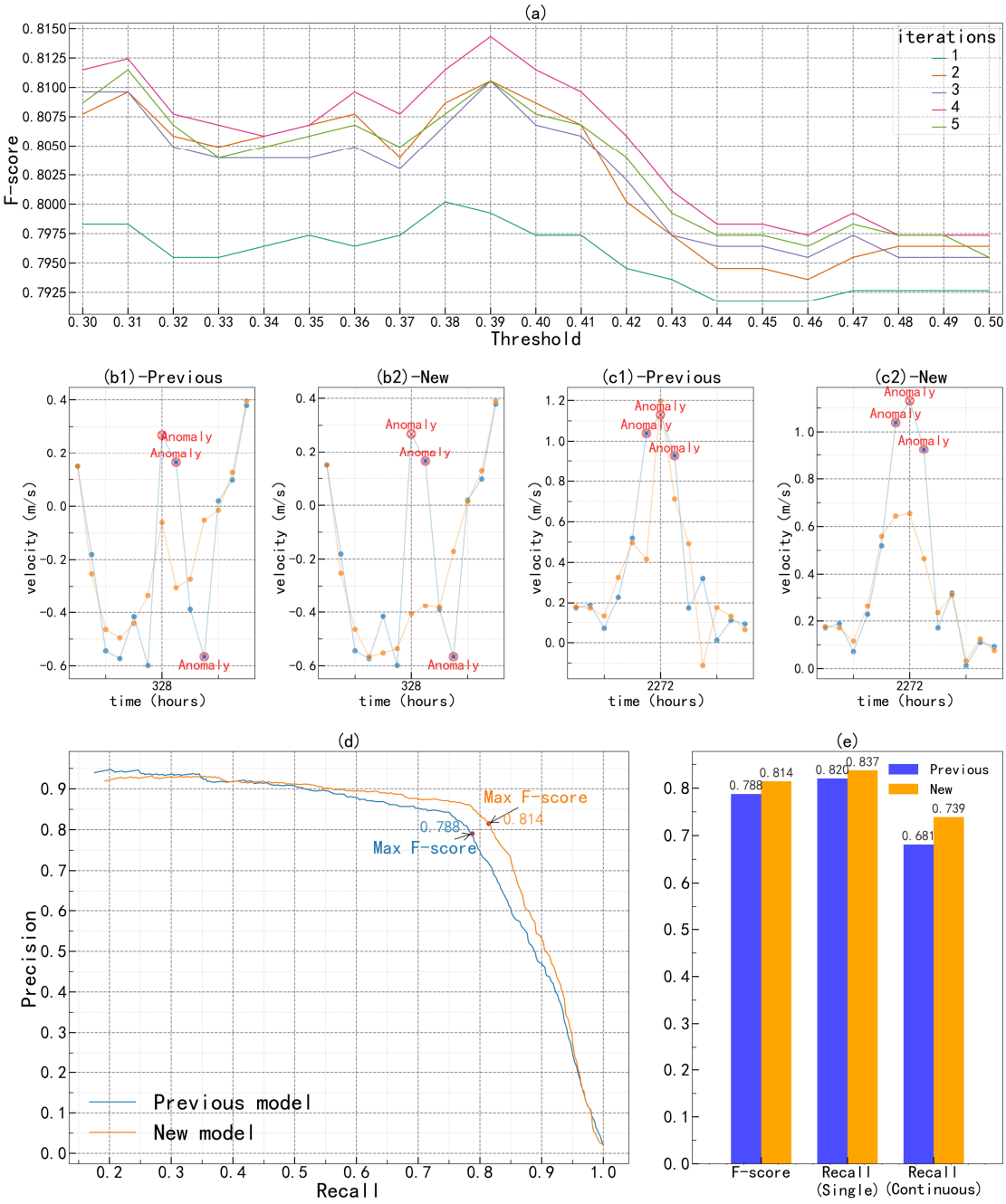


Figure 9. (**a**) F-score graphs under different thresholds, with different line colors indicating iteration counts. (**b**,**c**) Comparison graphs of predictions before and after model improvement, with the orange curve representing the predicted model data and the blue curve representing the radar data. (**d**) PR (precision–recall) curve graphs before and after model improvement, (**e**) bar chart comparing recall rates for different anomaly types before and after model improvement.

Figure 9a clearly illustrates how the iterative prediction method significantly enhances the efficiency of the anomaly detection. The trend in F-score variation is influenced by several factors, including the type and quantity of anomalous data, the selected threshold, and the number of iterations.

As is shown in Figure 5b, anomalies with a length of 2 predominate within the realm of continuous anomalies. Figure 9b,c further emphasizes the exemplary performance of the iterative prediction model in addressing consecutive anomalies with lengths of 2 and 3. Combining these insights, even though the model encounters challenges when detecting longer consecutive anomalies due to constraints related to the input sequence length, its outstanding performance in detecting consecutive anomalies with lengths of 2 and 3 remains evident in Figure 9d,e.

In Figure 9d, the overall curve shifts upward and to the right, indicating an improvement in the model compared to the previous version. In Figure 9e, it can be observed that the iterative prediction model effectively enhances the model's anomaly detection performance, with improved recall rates for both single-point anomalies and consecutive anomalies.

3.3. Sensitivity Study

In the concluding part of this section, we will examine variations in the model's anomaly detection performance under different input sequence lengths, considering two key aspects.

Firstly, as discussed in Section 2 regarding missing data imputation, we only impute missing data when the gap between sequences has fewer than six standard time intervals. However, when processing radar sea-current data, we face a particularly severe problem: significant data are missing, especially in the peripheral regions covered by the radar (as depicted in Figure 1). Even after applying missing value imputation, the data inevitably get fragmented into numerous time series segments. Due to the Bi-LSTM model requiring a certain amount of data both before and after the data point being predicted, the data at the beginning and end of each segment are undetectable. Consequently, as the input sequence length increases, the amount of data genuinely available for detection diminishes. This results in certain anomalous data being undetected, leading to the 'escape' phenomenon, as illustrated in Figure 10a,b.

In the second aspect, we compared the performance metrics of anomaly detection for the model under different input sequence lengths. To account for shorter input sequence lengths, we introduced anomalies only at the points under examination, as opposed to the method described in Section 3.1, where anomalies were added to both the points under examination and their neighboring points. The specific results are shown in Figure 10c.

In an overall assessment, longer input sequences lead to the presence of more undetectable data. As the amount of detectable data diminishes, some anomalous data can slip through the detection process, thereby undermining the overall quality control effectiveness. Conversely, overly short input sequence lengths make it challenging for the model to address consecutive anomalies, resulting in suboptimal quality control. Considering these factors, we chose an input sequence length of 7 to strike a balance.

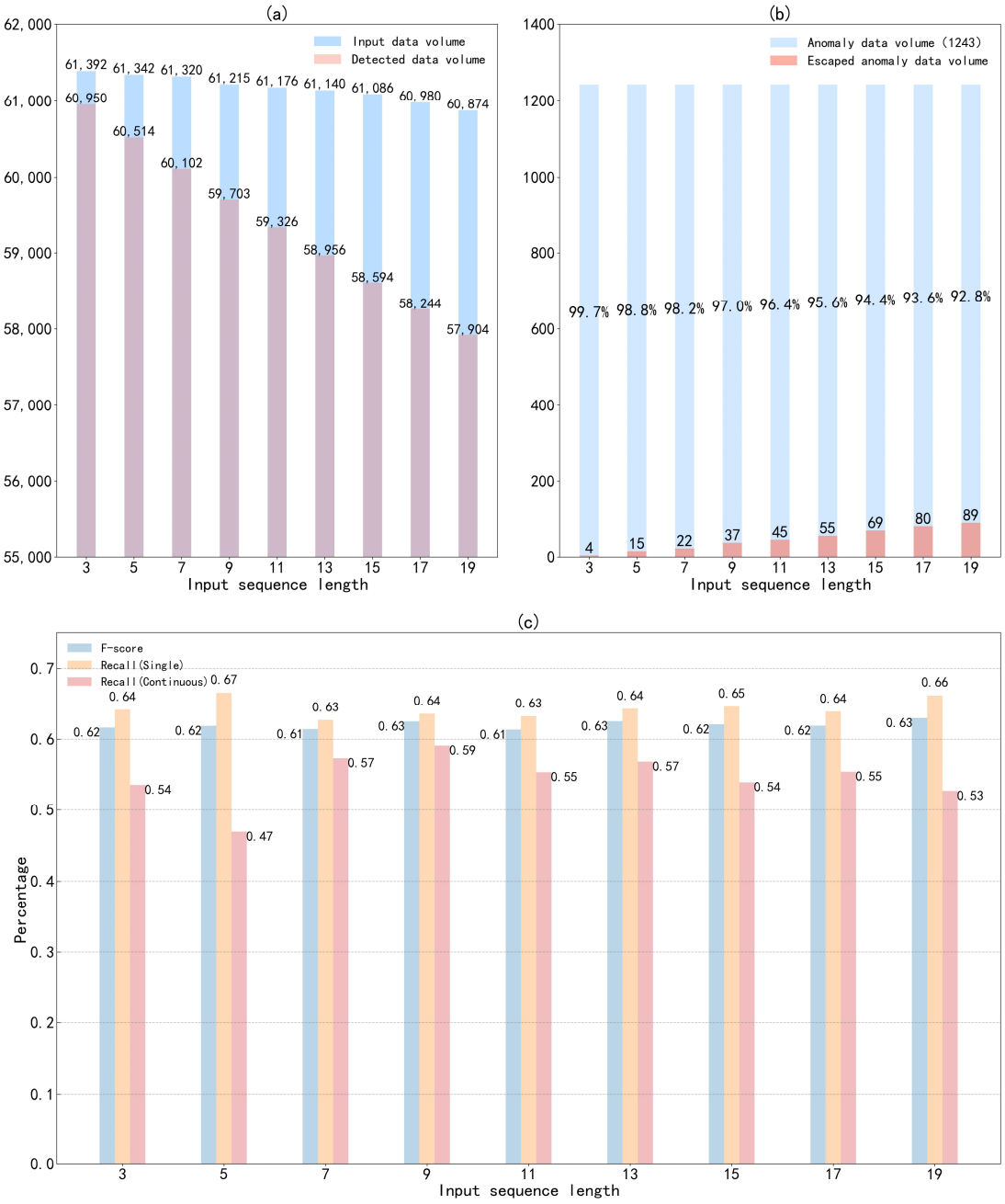


Figure 10. (**a**,**b**) illustrate the variations in data detection volume and the amount of undetected anomalous data under different input sequence lengths, and (**c**) depicts the changes in anomaly detection performance under varying input sequence lengths.

4. Conclusions

Compared to earlier research on the quality control of radar data, the deep learning approach employed in our study primarily focuses on generated vector sea-current charts rather than radial velocities. This allows for a more effective integration with previous quality control efforts on radial velocities. Not only can we address anomalies resulting from synthetic algorithms, but our study also helps to rectify anomalies that persisted due to incomplete or suboptimal previous quality control. Moreover, our method does not rely on radar echo-related signals for quality assessment. Instead, it capitalizes on the continuity of time series and the strengths of deep learning, significantly simplifying the quality control process and improving the data quality.

In Section 2, we provided a comprehensive introduction to the Bi-LSTM neural network model, clarified the data preprocessing process, explained our reasons for choosing specific anomaly detection metrics, and detailed the architecture of the anomaly detection system. In Section 3, we began by discussing the various types of anomalies commonly found in radar sea current data and their respective distribution percentages. Subsequently, we optimized the model's input and detection process specifically for different anomaly types. In the concluding section of Section 3, we conducted a sensitivity analysis experiment designed to elucidate the effects of input sequence length variations on model performance.

In our study, the quantitative results, as detailed in Figure 9d,e, display the final PR curve of the model and the F-score, which stands as the core evaluation metric in our anomaly detection research. For datasets like high-frequency radar sea-current data, imbalances are prevalent where anomalies (or 'positive instances') are significantly outnumbered by normal observations. Such imbalanced datasets, characterized by large volumes of data and a high complexity, can pose challenges to conventional machine learning algorithms. In this context, our proposed method achieved an F-score of 0.814, signifying its strong performance in both precision and recall, and thus effectively tackling the anomalies.

Our method has provided substantial assistance for the quality control of radar and offered insights into potential areas of exploration. However, there remains potential for further optimization in certain areas. In subsequent research, we aim to enhance the model's sensitivity to specific anomaly patterns and consider increasing the model's complexity to capture more information.

Author Contributions: Conceptualization, J.W.; Data curation, F.Y.; Formal analysis, C.Z.; Funding acquisition, J.W.; Investigation, C.Z.; Methodology, C.Z.; Resources, F.Y.; Supervision, C.W.; Validation, C.Z.; Visualization, C.Z.; Writing—original draft, C.Z.; Writing—review& editing, C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Basic Research and Development Project of China (2022YFF0802000), the project supported by Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) No. SML2020SP009 and the National Natural Science Foundation of China (41976007).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the fact that parts of radar data on oceanic state information is classified.

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

References

1. Teague, C.C.; Vesecky, J.F.; Fernandez, D.M. HF Radar Instruments, Past to Present. *Oceanography* **1997**, *10*, 40–44. [CrossRef]
2. Paduan, J.D.; Washburn, L. High-Frequency Radar Observations of Ocean Surface Currents. *Ann. Rev. Mar. Sci.* **2013**, *5*, 115–136. [CrossRef] [PubMed]
3. Crombie, D.D. Doppler Spectrum of Sea Echo at 13.56 Mc./s. *Nature* **1955**, *175*, 681–682. [CrossRef]
4. Lipa, B.; Barrick, D. Least-Squares Methods for the Extraction of Surface Currents from CODAR Crossed-Loop Data: Application at ARSLOE. *IEEE J. Ocean. Eng.* **1983**, *8*, 226–253. [CrossRef]
5. Liu, Y.; Weisberg, R.H.; Merz, C.R.; Lichtenwalner, S.; Kirkpatrick, G.J. HF Radar Performance in a Low-Energy Environment: CODAR SeaSonde Experience on the West Florida Shelf. *J. Atmos. Ocean Technol.* **2010**, *27*, 1689–1710. [CrossRef]
6. Liu, Y.; Weisberg, R.H.; Merz, C.R. Assessment of CODAR SeaSonde and WERA HF Radars in Mapping Surface Currents on the West Florida Shelf. *J. Atmos. Ocean Technol.* **2014**, *31*, 1363–1382. [CrossRef]
7. Liu, Y.; Merz, C.R.; Weisberg, R.H.; O'Loughlin, B.K.; Subramanian, V. Data Return Aspects of CODAR and WERA HighFrequency Radars in Mapping Currents. In *Observing the Oceans in Real Time*; Springer: Cham, Switzerland, 2018; pp. 227–240. [CrossRef]
8. Sun, Y.; Chen, C.; Beardsley, R.C.; Ullman, D.; Butman, B.; Lin, H. Surface Circulation in Block Island Sound and Adjacent Coastal and Shelf Regions: A FVCOM-CODAR Comparison. *Prog. Oceanogr.* **2016**, *143*, 26–45. [CrossRef]
9. Mau, J.-C.; Wang, D.-P.; Ullman, D.S.; Codiga, D.L. Comparison of Observed (HF Radar, ADCP) and Model Barotropic Tidal Currents in the New York Bight and Block Island Sound. *Estuar. Coast. Shelf. Sci.* **2007**, *72*, 129–137. [CrossRef]

10. Lipphardt, B.; Kirwan, A.; Grosch, C.; Lewis, J.; Paduan, J. Blending HF Radar and Model Velocities in Monterey Bay through Normal Mode Analysis. *J. Geophys. Res.* **2000**, *105*, 3425–3450. [CrossRef]
11. Roarty, H.; Cook, T.; Hazard, L.; George, D.; Harlan, J.; Cosoli, S.; Wyatt, L.; Alvarez Fanjul, E.; Terrill, E.; Otero, M.; et al. The Global High Frequency Radar Network. *Front. Mar. Sci.* **2019**, *6*, 164. [CrossRef]
12. Fujii, S.; Heron, M.L.; Kim, K.; Lai, J.-W.; Lee, S.-H.; Wu, X.; Wu, X.; Wyatt, L.R.; Yang, W.-C. An Overview of Developments and Applications of Oceanographic Radar Networks in Asia and Oceania Countries. *Ocean. Sci. J.* **2013**, *48*, 69–97. [CrossRef]
13. Cosoli, S.; Bolzon, G.; Mazzoldi, A. A Real-Time and Offline Quality Control Methodology for SeaSonde High-Frequency Radar Currents. *J. Atmos. Ocean Technol.* **2012**, *29*, 1313–1328. [CrossRef]
14. Roarty, H.; Smith, M.; Kerfoot, J.; Kohut, J.T.; Glenn, S. Automated Quality Control of High Frequency Radar Data. In Proceedings of the 2012 Oceans, Hampton Roads, VA, USA, 14–19 October 2012; pp. 1–7.
15. Haines, S.; Seim, H.; Muglia, M. Implementing Quality Control of High-Frequency Radar Estimates and Application to Gulf Stream Surface Currents. *J. Atmos. Ocean Technol.* **2017**, *34*, 1207–1224. [CrossRef]
16. Lorente, P.; Piedracoba, S.; Soto-Navarro, J.; Alvarez-Fanjul, E. Evaluating the Surface Circulation in the Ebro Delta (Northeastern Spain) with Quality-Controlled High-Frequency Radar Measurements. *Ocean Sci.* **2015**, *11*, 921–935. [CrossRef]
17. Tian, Y.; Wen, B.; Li, Z.; Hou, Y.; Tian, Z.; Huang, W. Fully Digital Multi-Frequency Compact High-Frequency Radar System for Sea Surface Remote Sensing. *IET Radar Sonar Navig.* **2019**, *13*, 1359–1365. [CrossRef]
18. Keogh, E.; Lin, J.; Fu, A. HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence. In Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05), Houston, TX, USA, 27–30 November 2005; p. 8.
19. Wei, L.; Kumar, N.; Lolla, V.N.; Keogh, E.J.; Lonardi, S.; Ratanamahatana, C. Assumption-Free Anomaly Detection in Time Series. In Proceedings of the International Conference on Statistical and Scientific Database Management, Santa Barbara, CA, USA, 27–29 June 2005; pp. 237–240.
20. Yaacob, A.H.; Tan, I.K.T.; Chien, S.F.; Tan, H.K. ARIMA Based Network Anomaly Detection. In Proceedings of the 2010 Second International Conference on Communication Software and Networks, Singapore, 26–28 February 2010; pp. 205–209.
21. Jones, M.; Nikovski, D.; Imamura, M.; Hirata, T. Exemplar Learning for Extremely Efficient Anomaly Detection in Real-Valued Time Series. *Data Min. Knowl. Discov.* **2016**, *30*, 1427–1454. [CrossRef]
22. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural. Comput.* **1997**, *9*, 1735–1780. [CrossRef]
23. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural. Netw. Learn Syst.* **2017**, *28*, 2222–2232. [CrossRef]
24. Schuster, M.; Paliwal, K.K. Bidirectional Recurrent Neural Networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
25. Graves, A.; Schmidhuber, J. Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures. *Neural Netw.* **2005**, *18*, 602–610. [CrossRef]
26. Shrestha, A.; Li, H.; Kernec, J.L.; Fioranelli, F. Continuous Human Activity Classification From FMCW Radar with Bi-LSTM Networks. *IEEE Sens. J.* **2020**, *20*, 13607–13619. [CrossRef]
27. Graves, A.; Jaitly, N.; Mohamed, A. Hybrid Speech Recognition with Deep Bidirectional LSTM. In Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, Czech Republic, 8–12 December 2013; pp. 273–278.
28. Zhang, H.; Huang, H.; Han, H. A Novel Heterogeneous Parallel Convolution Bi-LSTM for Speech Emotion Recognition. *Appl. Sci.* **2021**, *11*, 9897. [CrossRef]
29. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.
30. Cui, Z.; Ke, R.; Pu, Z.; Wang, Y. Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Forecasting Network-Wide Traffic State with Missing Values. *Transp. Res. Part C Emerg. Technol.* **2020**, *118*, 102674. [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.