*remote sensing*

MDPI

*Article*

# A Multi-Scale Fusion Deep Learning Approach for Wind Field Retrieval Based on Geostationary Satellite Imagery

**Wei Zhang [1,*], Yapeng Wu [1], Kunkun Fan [1], Xiaojiang Song [2], Renbo Pang [2] and Boyu Guoan [2]**

[1] College of Computer Science and Technology, Ocean University of China, Qingdao 266100, China
[2] National Marine Environment Forecasting Center, Beijing 100082, China
* Correspondence: weizhang@ouc.edu.cn

**Abstract:** Wind field retrieval, a crucial component of weather forecasting, has been significantly enhanced by recent advances in deep learning. However, existing approaches that are primarily focused on wind speed retrieval are limited by their inability to achieve real-time, full-coverage retrievals at large scales. To address this problem, we propose a novel multi-scale fusion retrieval (MFR) method, leveraging geostationary observation satellites. At the mesoscale, MFR incorporates a cloud-to-wind transformer model, which employs local self-attention mechanisms to extract detailed wind field features. At large scales, MFR incorporates a multi-encoder coordinate U-net model, which incorporates multiple encoders and utilises coordinate information to fuse meso- to large-scale features, enabling accurate and regionally complete wind field retrievals, while reducing the computational resources required. The MFR method was validated using Level 1 data from the Himawari-8 satellite, covering a geographic range of 0–60°N and 100–160°E, at a resolution of 0.25°. Wind field retrieval was accomplished within seconds using a single graphics processing unit. The mean absolute error of wind speed obtained by the MFR was 0.97 m/s, surpassing the accuracy of the CFOSAT and HY-2B Level 2B wind field products. The mean absolute error for wind direction achieved by the MFR was 23.31°, outperforming CFOSAT Level 2B products and aligning closely with HY-2B Level 2B products. The MFR represents a pioneering approach for generating initial fields for large-scale grid forecasting models.

**Keywords:** wind field retrieval; deep learning; geostationary satellite imagery; multi-scale fusion retrieval

check for updates

Academic Editor: Xinghua Li

Received: 24 December 2024
Revised: 8 February 2025
Accepted: 8 February 2025
Published: 11 February 2025

**Citation:** Zhang, W.; Wu, Y.; Fan, K.; Song, X.; Pang, R.; Guoan, B. A Multi-Scale Fusion Deep Learning Approach for Wind Field Retrieval Based on Geostationary Satellite Imagery. *Remote Sens.* **2025**, *17*, 610. https://doi.org/10.3390/rs17040610

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Wind is a fundamental component of atmospheric dynamics, and is critically important for the global climate system and human activities [1]. Variations in wind patterns correlate directly with the frequency and intensity of extreme weather phenomena, such as storms, typhoons, and tornadoes, presenting threats to human lives and property, as well as to infrastructure and ecological stability [2]. Conversely, wind energy is a sustainable and clean energy source that has undergone rapid global development and widespread adoption. The distribution and dynamics of wind fields play pivotal roles in determining the potential and efficiency of wind energy resources, providing essential guidance for the planning, construction, and operation of wind farms [3,4]. Consequently, wind field research is valuable. Through advanced wind field retrieval techniques, a more accurate understanding of wind dynamics can be achieved, providing essential initial data for forecasting models and thereby improving the accuracy and timeliness of weather predictions.

In recent decades, numerous methodologies for wind field retrieval have emerged. Traditional approaches predominantly rely on empirical relationships, mathematical mod-

els [5,6], and meteorological principles to estimate wind field states from observational data. Waveform matching [7] involves comparing waveform characteristics observed in data with those in established models or reference datasets to identify the most compatible wind field model. Although this method offers a direct estimation of the wind field and is adaptable to various types of observational data, it typically requires a substantial volume of observational data and reference libraries to attain precise outcomes. It may also encounter challenges in effectively mitigating noise within observations [8]. The geophysical model function (GMF) method [9] uses geophysical model functions to interpret remote sensing data and interactions among the atmosphere, oceans, and Earth's surface. Although it is frequently employed in the interpretation and analysis of remote sensing observational data, this method is challenged by its model complexity and dependence on the quality of observational data [10]. The Cloud Motion Winds method [11] involves extracting wind speed and direction data by monitoring cloud movements. This method offers wide coverage and real-time capability, but currently faces limitations in accuracy and is dependent on cloud cover [12,13].

In recent years, the development of neural networks and deep learning technologies has brought new perspectives to wind field retrieval, compensating for the shortcomings of traditional methods. For example, Li et al. proposed a wind speed retrieval model near the coastline based on an artificial neural network [14]. The model utilised global navigation satellite system reflectometry (GNSS-R) from Cyclone GNSS (CYGNSS) Level 1B data for retrieval, resulting in a 24.4% improvement in wind speed root mean square error (RMSE) compared to CYGNSS Level 2 data. However, when using a fully connected network for wind speed retrieval, overfitting may occur due to the non-uniform distribution of winds, especially at high wind speeds. To mitigate this issue, Guo et al. introduced an end-to-end convolutional neural network (CNN) model [15], which improved feature extraction, resulting in an RMSE of 1.53 m/s for wind speed retrieval in the range of 0–25 m/s. However, the inherent translation equivariance in CNNs introduces imprecision in feature extraction from Doppler-derived maps. To address this challenge, Zhao et al. proposed a Doppler-derived map model based on the Transformer architecture [15], achieving an improved wind speed retrieval RMSE of 1.43 m/s. These deep learning methods have significantly improved the accuracy of wind speed retrieval compared to traditional methods, avoiding the construction of complex physical models and systematic biases.

Over the past two years, significant progress has been made in the field of meteorological forecasting through deep learning, with several global meteorological forecast models being successively proposed [16–18]. These deep learning models have achieved comparable, if not superior, forecast accuracy compared to numerical weather prediction (NWP) systems, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS). Additionally, they significantly reduce the computational resources required and substantially improve the forecast speed [19]. However, existing deep learning models that are capable of conducting large-scale forecasts still depend on initial fields supplied by NWP systems, due to the limitations of current retrieval methods, which cannot simultaneously retrieve both wind speed and direction, nor can they acquire real-time, large-scale wind field states. As a result, they fail to deliver real-time, accurate large-scale initial wind fields. This constraint hinders the ability of deep learning methods to accomplish an end-to-end forecasting process based solely on observational data.

Driven by the limitations of current wind field retrieval techniques and the demand for large-scale grid wind field retrievals, and inspired by the correlation between clouds and wind [20,21], we propose a novel multi-scale fusion retrieval (MFR) method. MFR is a deep learning approach that uses cloud data from geostationary observation satellites to retrieve wind fields at multiple scales. Direct retrieval of wind fields under clear-sky

or high-level cloud conditions is challenging; however, geostationary satellites provide multi-channel data that encompass various meteorological features [22], making it possible to retrieve wind fields in cloud-free conditions. Furthermore, existing studies [23,24] have demonstrated a correlation between upper-level and near-surface wind fields. Even when the Earth's surface is obscured by high-level clouds, it is still possible to retrieve near-surface wind fields from high-level clouds. To capture the detailed features of wind fields, we developed the cloud-to-wind transformer (C2W-Former) model, which uses a windows self-attention mechanism to extract features from high-resolution mesoscale cloud images. The Multi-encoder Coordinate U-net (M-CoordUnet) model was designed to incorporate the geographical characteristics of large-scale wind fields and avoid inconsistencies among different mesoscale retrieval results by integrating coordinate information and large-scale cloud features. This multi-scale fusion approach reduces the high computational resource requirements of directly modelling high-resolution cloud data at large scales, and preserves observational information by avoiding downsampling. Furthermore, it enables the rapid generation of large-scale grid wind fields, opening up a new direction for deep learning-based initial field generation for forecasting models.

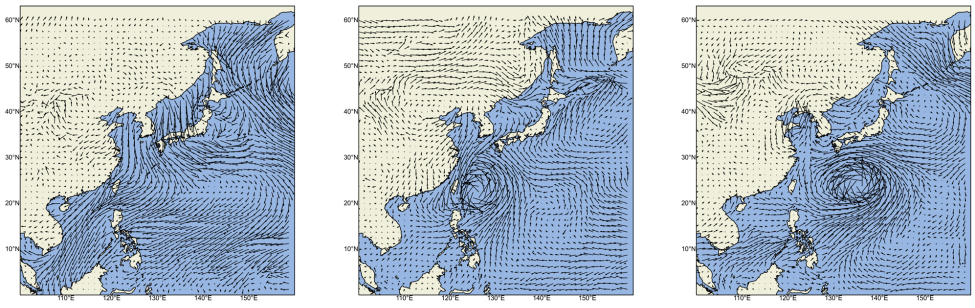
The remainder of this article is organised as follows. Section 2 introduces the study area and the relevant data used. Section 3 outlines the two-stage retrieval method, including the C2W-Former and M-coordUnet deep learning models, as well as the details of sample construction. The evaluation metrics, training details, and experimental results are provided in Section 4. Section 5 presents case studies to analyze the robustness of the proposed model under extreme weather conditions. Finally, the conclusions and future research directions are presented in Section 6.

## 2. Materials

### *2.1. Study Area*

In this study, we focused on the spatial range encompassing the Western North Pacific and the eastern Eurasian continent, covering the geographical area from 0°N to 60°N and from 100°E to 160°E.

In this region, tropical cyclones are frequent occurrences. In 2022 alone, there were 25 named storms recorded in this area. Among these tropical storms, ten developed into typhoons, three of which intensified into super typhoons. Super typhoons Hinnamnor and Nanmadol each caused USD 1 billion in damages (https://en.wikipedia.org/wiki/2022_Pacific_typhoon_season (accessed on 23 December 2024)). These phenomena are clearly illustrated in a vector diagram of ERA5 data for this region (Figure 1).



(a) Normal (b) Typhoon Hinnamnor (c) Typhoon Nanmadol

**Figure 1.** Vector diagrams for (**a**) normal conditions, (**b**) Typhoon Hinnamnor, and (**c**) Typhoon Nanmadol in the study area. Arrow directions and lengths indicate wind direction and speed, respectively.

### 2.2. Data

This study used two primary datasets: Himawari-8 satellite data and ERA5 reanalysis data. Additionally, forecast data from ECMWF IFS, wind field products obtained from two scatterometer satellites and 340 weather stations data were used for comparison with our results.

The Himawari-8 satellite (https://www.eorc.jaxa.jp/ptree/index.html (accessed on 23 December 2024)), operated by the Japan Meteorological Agency, is an advanced meteorological satellite with a primary mission to monitor various domains, such as the atmosphere, climate, oceans, and environment. Equipped with the Advanced Himawari Imager (AHI), the satellite covers multiple spectral channels, including visible light and infrared channels, allowing it to collect visible light and infrared data. The Japan Aerospace Exploration Agency provides Level 1 grid data for AHI with a spatial resolution of 0.02° × 0.02°, covering East Asia, the Western Pacific, Australia, and surrounding waters. During preprocessing, missing values or outliers (<1%) identified by Himawari-8 L1 production were removed to maintain data integrity. In addition, the data may contain some noise, but relevant studies [25–27] indicate that deep learning models can adapt to noise and adjust accordingly. Therefore, we have not performed any special identification or processing of the noise. This study specifically selected channels 7 and 13 from the Level 1 data, as they provide pertinent cloud-related information [28]. Notably, although the original Level 1 data had a high temporal resolution of 10 min, we opted for a 6 h temporal resolution in this study, focusing on data at 0:00, 6:00, 12:00, and 18:00 daily for model construction and analysis, because large-scale wind field variation is not pronounced over short periods, and our analysis revealed that a 6 h interval is suitable for capturing meaningful patterns, striking a balance between data granularity and computational efficiency. Relevant studies [29–31] indicate that radar echo data also have great potential in wind field retrieval, but its wind field coverage is relatively limited and it is difficult to meet the forecast needs of a large area and the entire region. Therefore, we chose the Himawari-8 geostationary satellite to accurately retrieve the wind field through large-scale, high-resolution cloud image data.

The fifth-generation atmospheric reanalysis dataset released by ECMWF, ERA5 [32], provides hourly global weather records from 1940 to the present, covering a wide range of meteorological variables. Through numerous validation studies [33,34], ERA5 has demonstrated high accuracy and reliability under different climate and weather conditions. This study used the 10 m wind components (U and V) provided in ERA5 single-level data (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form (accessed on 23 December 2024)), with a spatial resolution of 0.25° × 0.25°, as the targets for wind field retrieval.

ECMWF IFS is widely recognised as one of the most advanced NWP systems currently available [17]. Powered by high-performance computing, it conducts numerical computations based on specified initial and boundary conditions to forecast the future evolution of wind fields over a designated timeframe. IFS provides wind field forecast data (https://www.ecmwf.int/en/forecasts/datasets/set-i#I-i-a_fc (accessed on 23 December 2024), https://apps.ecmwf.int/shopping-cart/orders/new/subset/162 (accessed on 23 December 2024)) twice daily, covering the initial forecast time and the subsequent 10 days, with global coverage and a spatial resolution of 0.1°. In this study, a bilinear interpolation was applied to downscale the spatial resolution of the IFS data from their original resolution of 0.1° to 0.25°. In this study, IFS forecast outputs for the start of the forecast period were used as initial fields for comparison with the wind field retrieval results obtained through deep learning models.

Level 2B wind field data from two scatterometer satellites, HY-2B (https://osdds.nsoas.org.cn/oceanOne?type=1&id=1673143940646674433 (accessed on 23 December 2024))

and CFOSAT (https://osdds.nsoas.org.cn/oceanOne?type=1&id=1673144819122675714 (accessed on 23 December 2024)), were used to evaluate the accuracy and coverage of the wind field retrievals through comparison with existing satellite wind field products. Due to the non-grid format of wind field products from these two satellites, a spatiotemporally weighted fusion algorithm [35,36] was used to project them onto ERA5 grid points. This approach integrated temporal and spatial interpolation techniques, substituting temporal data for missing spatial data and vice versa. This strategy maximised the utilisation of satellite observation data, as follows:

$$u_n = \frac{\sum_{k=1}^{N} w_k u_k}{\sum_{k=1}^{N} w_k} \tag{1}$$

$$w_k = \frac{2 - \left[\frac{(x_k - x_0)^2 + (y_k - y_0)^2}{R^2} + \frac{(t_k - t_0)^2}{T^2}\right]}{2 + \left[\frac{(x_k - x_0)^2 + (y_k - y_0)^2}{R^2} + \frac{(t_k - t_0)^2}{T^2}\right]} \tag{2}$$

where $u_n$ is the interpolated estimate obtained as a linear combination of $N$ observed values within a spatiotemporal influence radius; subscript $k$ denotes the observation data point; $u_k$ is the observed value at point $k$; $w_k$ is the weight of observation point $k$, determined by the standardised distances in time and space from the data points to the grid interpolation point; $(x_k, y_k, t_k)$ are the spatiotemporal coordinates of observation point $k$, with subscript 0 representing the point to be interpolated; $R$ is the distance influence radius; and $T$ is the time influence radius. In this study, $R$ = 0.5° and $T$ = 6 h.

In addition, we used observational data from weather stations furnished by the National Centers for Environmental Prediction (www.ncei.noaa.gov/products/land-based-station/integrated-surface-database (accessed on 23 December 2024)) (NCEP) and identified 340 stations within the study domain with valid data for 2021, allowing us to assess the error characteristics between the MFR results and in situ observations.

We used Himawari-8 satellite and ERA5 data spanning 2017–2022, with a temporal resolution of 6 h. The dataset was split into training (2017–2020), validation (2022), and test (2021) subsets using a partitioning method similar to that used for Pangu-Weather [17]. Data from IFS, HY-2B, and CFOSAT were sourced from 2021.

During the data preprocessing phase, several critical steps were undertaken to ensure the integrity and consistency of the datasets. Anomalous data points (missing values and outliers) indicated by Himawari-8 L1 production, accounting for less than 1% of the sample, were identified and removed. To address differences in feature scales across datasets, normalization was applied to standardise all variables, which is essential for effective model training. Furthermore, interpolation techniques were utilised to resolve spatial and temporal resolution discrepancies between datasets, ensuring consistent data alignment for model inputs.

## 3. Methods

### *3.1. MFR Method*

To mitigate the issue of periodicity in wind direction, we retrieved the U and V components of the wind field. The observational data used in this study were derived from the Himawari-8 satellite, with the objective of retrieving ERA5 wind field data. Because both datasets were in grid format, at any given moment, they could be considered as $H \times W$ grids, without considering the feature dimensions. The spatial resolutions of the Himawari-8 satellite and ERA5 data were 0.02° and 0.25°, respectively. Therefore, at any given time within the entire study area (0–60°N, 100–160°E), they could be considered to be grid data with resolutions of 3000 × 3000 and 240 × 240, respectively, with the right boundary

discarded for ease of processing. Although the Himawari satellite observational data offer higher spatial resolution, providing richer details, they also require more computational resources for processing. Conducting the retrieval model training across the entire study area directly would require significant computational resources due to the large amount of input data. However, downsampling the Himawari-8 satellite observational data to a spatial resolution of 0.25° before conducting the retrieval model training across the entire study area could lead to a substantial loss of mesoscale observational information during the downsampling process. Therefore, the key issue in the wind field retrieval in this study was extracting and using wind field features from these high-resolution observations more effectively, while maintaining reasonable consumption of computing resources.

The region of interest is on the east coast of Asia where the formation of boundary layer (BL) clouds is at a minimum [37]. Under clear-air conditions, although the area above the Earth's surface is nearly cloud-free, the Himawari-8 satellite's long-wave infrared channel (Band 13) can still observe surface temperature (including sea surface temperature (SST) [38] and land surface temperature (LST) [39]). Studies [40,41] have demonstrated a dynamic relationship between SST and near-surface wind fields; therefore, SST/LST features captured by Band 13 can provide implicit information for wind field retrieval.

Under active weather conditions, the Earth's surface is typically obscured by clouds distributed across various layers of the troposphere [42]. However, the evolution of tropospheric cloud systems is closely related to wind fields within the same layer. Additionally, there is a certain connection between tropospheric and near-surface wind fields. Although explicit physical equations describing this relationship are currently lacking, deep learning models, with their strong nonlinear mapping capabilities (universal approximation ability [43–45]), can capture and learn these implicit relations through large-scale data training [46].

This study integrates multi-scale cloud information and employs a model trained on historical data to infer and reconstruct wind fields under various cloud cover conditions. This approach enables the model to utilise surrounding and historical wind field characteristics (for example, areas without clouds may have been covered in historical data) to accurately predict and reconstruct wind fields in regions that are cloud-free or covered only by high-level clouds.

Figure 2 shows the overall architecture of the multi-scale fusion retrieval method proposed in this study. First, cloud data at time t were obtained from the Himawari-8 satellite (bands 7 and 13) with a resolution of 0.02°. In the mesoscale stage, we used a sliding window sampling method (see Section 3.2.1) to decompose the data into mesoscale cloud data, which were then fed into the proposed C2W-Former model (see Section 3.2.3) to generate the corresponding mesoscale wind field UV components. Subsequently, we applied a mesoscale integration method to obtain the preliminary large-scale UV components at a resolution of 0.25° at time t. This approach enabled us to process high-resolution data in relatively smaller regions, reducing the computational burden while preserving complete mesoscale observational information. Additionally, it provided the model with sufficient samples, avoiding the problem of overfitting in transformer models [47]. In the large-scale stage, we first downsampled the original Himawari cloud data from 0.02° to 0.25° using bilinear interpolation, and then combined them with the initial large-scale UV components (see Section 3.3.1). Finally, we used the M-CoordUnet model (see Section 3.3.2) to generate the final large-scale wind field UV components at time t. The fusion approach combined the strengths of both scales, leveraging both detailed mesoscale information and a large-scale global perspective to obtain more accurate wind field retrievals.

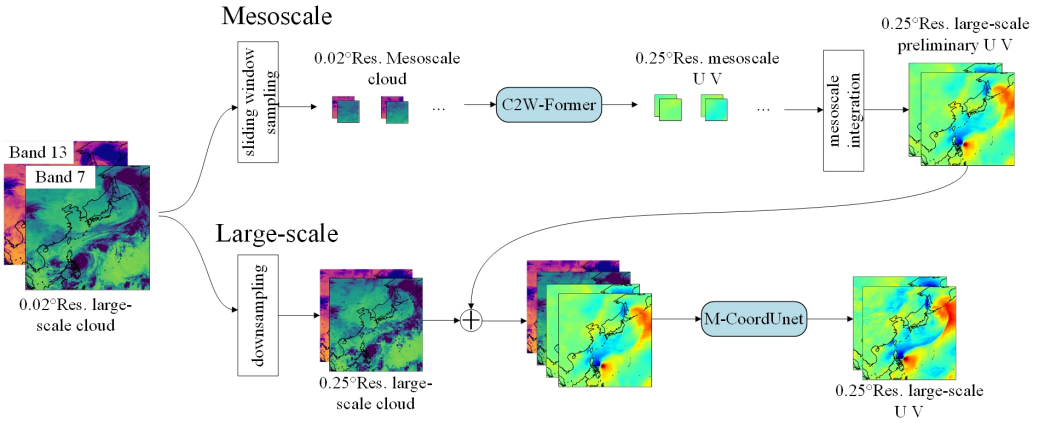



**Figure 2.** Multi-scale fusion retrieval architecture. Res, resolution.

### *3.2. Mesoscale Stage*

In this section, we present the details of the mesoscale stage, encompassing the sliding window sampling method used in the preprocessing phase, the C2W-Former model, and the mesoscale integration method used in the postprocessing phase.

#### 3.2.1. Sliding Window Sampling Method

Spatial sliding window sampling is a commonly used method for partitioning an entire image into smaller sub-images. In this study, we used this method to divide large-scale data into several mesoscale data points, thereby constructing mesoscale retrieval samples. If the original image size is $H \times W$, with a window size (size of the cropped sub-image) of $h \times w$, and horizontal and vertical sampling strides of $S_w$ and $S_h$, respectively, then after sliding window sampling, $N$ sub-images of size $h \times w$ can be obtained, where $N$ is determined as shown in Equation (3). The sampling process is shown in Figure 3. The details of the spatial sliding window sampling applied in this study are shown in Table 1. The actual spatial size corresponding to small images obtained after sliding window sampling for the two types of data was 12° × 12°, which represents the mesoscale sample data.

$$N = \left( \left\lfloor \frac{H - h}{s_h} \right\rfloor + 1 \right) \times \left( \left\lfloor \frac{W - w}{s_w} \right\rfloor + 1 \right) \tag{3}$$

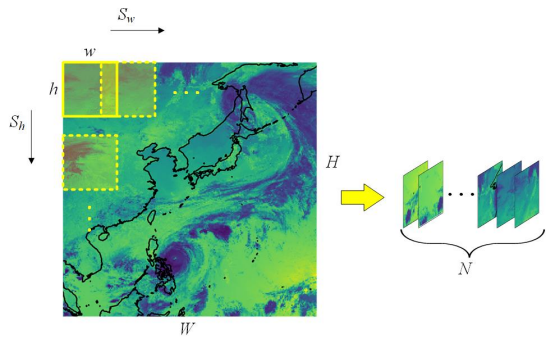



**Figure 3.** Sliding window sampling method.

**Table 1.** Configuration of the sliding window sampling method.

| Data Type | $H \times W$ | $h \times w$ | $S_h \times S_w$ | $Lat \times Lon$ |
|---|---|---|---|---|
| Himawari-8 | 3000 × 3000 | 600 × 600 | 300 × 300 | 12° × 12° |
| ERA5 | 240 × 240 | 48 × 48 | 24 × 24 | 12° × 12° |

#### 3.2.2. Mesoscale Sample Construction

In the mesoscale stage, the construction of mesoscale cloud-to-wind retrieval samples necessitated the use of the original Himawari-8 satellite Level 1 observational data and ERA5 wind field data. First, the study region was clipped from the datasets. Then, Himawari-8 channels 7 and 13 and ERA5 10 m U and V wind components were selected. Spatial sliding window sampling was applied to these datasets. Corresponding mesoscale Himawari satellite and ERA5 data were paired in space and time to create samples, with Himawari satellite data used as the input and ERA5 data used as the learning target.

During sample construction, any samples with invalid values were discarded. Finally, Z-score standardisation [48] was applied to the inputs and outputs of all samples using the mean and standard deviation of each feature of the Himawari satellite and ERA5 data. This standardisation process accelerated model training and enhanced model performance [49]. The details of the datasets are presented in Table 2. Notably, to facilitate direct comparison with multi-scale fusion and other products, we evaluated mesoscale predictions at the large scale. Specifically, we used the mesoscale-integrated, post-processed preliminary large-scale UV components as our mesoscale test samples.

**Table 2.** Design of the dataset.

| Dataset | Time Range (Year) | Mesoscale Samples | Large-Scale Samples |
|---|---|---|---|
| Training | 2017–2020 | 468,068 | 5778 |
| Validation | 2022 | 111,855 | 1456 |
| Testing | 2021 | 1381 | 1381 |

#### 3.2.3. C2W-Former Model

During the mesoscale retrieval phase, it is crucial to extract wind field characteristics from high-resolution mesoscale observational cloud imagery. To address this challenge, we introduced the C2W-Former model, using the transformer architecture. Compared to Convolutional Neural Networks (CNNs), which have limitations in capturing global context [50], Transformers excel at modeling long-range dependencies due to their flexible receptive fields [51]. Additionally, unlike Recurrent Neural Networks (RNNs), which face issues such as vanishing gradients when processing sequential data [52], Transformers support parallel processing, thereby enhancing computational efficiency. Transformers have demonstrated remarkable feature extraction capability in wind field retrieval and forecasting tasks [17,53,54], particularly due to their ability to capture long-range spatial dependencies through self-attention mechanisms.The C2W-Former model adopted the windows multi-head self-attention (W-MSA) mechanism from the Swin Transformer [55], which not only reduces computational complexity but also efficiently captures interaction dynamics among local features. This hierarchical design allows the model to combine local and global information, ensuring precise retrieval of mesoscale wind field characteristics while maintaining computational efficiency.

Figure 4 shows the architecture of the C2W-Former model. The input was a mesoscale cloud image with dimensions of 2 × 600 × 600, where 600 × 600 represents the image length and width (i.e., the number of observation points in the longitude and latitude directions), and 2 denotes the two channels (7 and 13) of the AHI Level 1 data. First, the input underwent patch partition to divide it into non-overlapping patches, which was similar to the Vision Transformer (ViT) approach. In our implementation, we used a patch size of 3 × 3 and transformed the original input into 200 × 200 patches, each of which was embedded into a latent space vector of length $C = 96$ in this study. Subsequently, we applied stage 1 Swin Transformer blocks (Figure 5), which did not alter the input size. To obtain the desired output size, we cropped four patches from the edges before stage 2, reducing

the input size from 200 × 200 to 192 × 192. Then, we applied a patch merging process, stacking four adjacent patches in the channel dimension to form a new patch, increasing the actual spatial size of each patch and making the model capable of learning at different hierarchical spatial levels. The numbers of Swin Transformer blocks in the three stages were 2, 2, and 6, respectively, and the output sizes of each stage were 200 × 200, 96 × 96, and 48 × 48, respectively, with a consistent feature dimension of 96 throughout. Next, we obtained intermediate feature maps, which were decoded using two convolutional layers. In contrast to the commonly used fully connected decoding method in the transformer, convolutional layers can compress feature dimensions, while also considering neighbouring patch features. Finally, we obtained the corresponding mesoscale UV as a 2 × 48 × 48 output through a 1 × 1 convolutional layer.

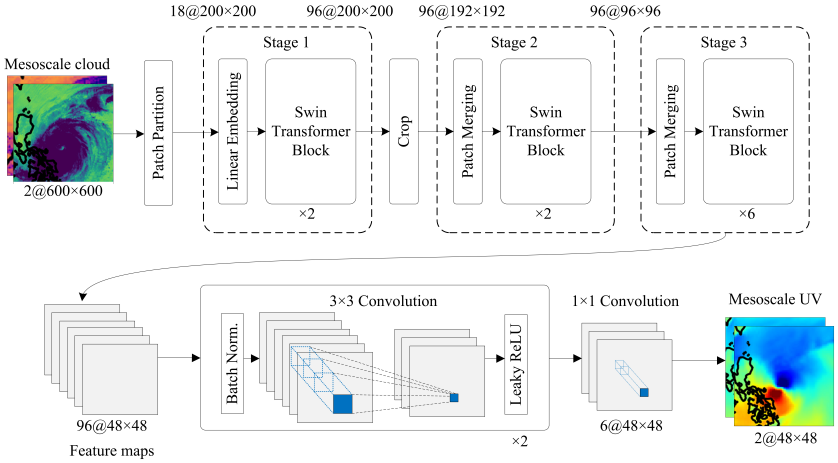
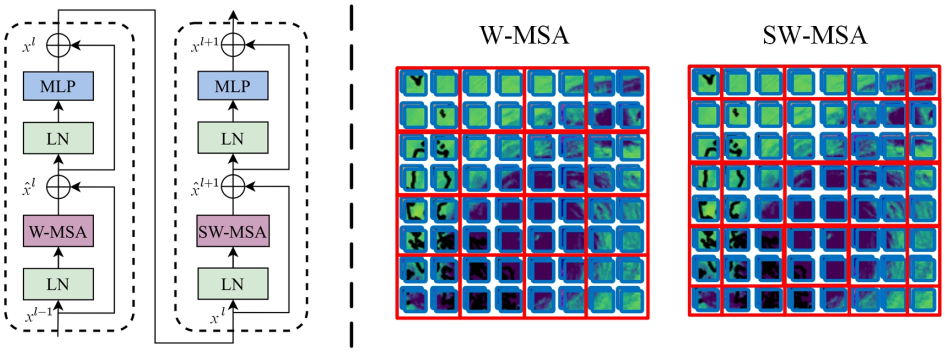



**Figure 4.** Structure of the C2W-Former model.




(a) Swin Tranformer Block (b) W-MSA and SW-MSA

**Figure 5.** (**a**) The Swin Transformer block architecture; (**b**) W-MSA and SW-MSA, which are multi-head self-attention modules with regular and shifted windowing configurations, respectively. Blue and red boxes represent patches and windows, respectively.

Figure 5 shows the architecture of the Swin Transformer block, which was constructed by replacing the standard MSA module in a transformer block with a module based on shifted windows, while the other layers were kept unchanged. As shown in Figure 5a, the Swin Transformer block consisted of a shifted window-based MSA module, followed by a two-layer multilayer perceptron (MLP) with Gaussian error linear unit nonlinearity between the layers. Additionally, a LayerNorm layer was applied before each MSA module and each MLP, and a residual connection was applied after each module. The Swin Transformer blocks were computed as follows:

$$\begin{aligned}
\hat{\mathbf{x}}^{l} &= \text{W-MSA}\left(\text{LN}\left(\mathbf{x}^{l-1}\right)\right) + \mathbf{x}^{l-1}, \\
\mathbf{x}^{l} &= \text{MLP}\left(\text{LN}\left(\hat{\mathbf{x}}^{l}\right)\right) + \hat{\mathbf{x}}^{l}, \\
\hat{\mathbf{x}}^{l+1} &= \text{SW-MSA}\left(\text{LN}\left(\mathbf{x}^{l}\right)\right) + \mathbf{x}^{l}, \\
\mathbf{x}^{l+1} &= \text{MLP}\left(\text{LN}\left(\hat{\mathbf{x}}^{l+1}\right)\right) + \hat{\mathbf{x}}^{l+1},
\end{aligned} \tag{4}$$

where $\hat{\mathbf{x}}^{l}$ and $\mathbf{x}^{l}$ are the output features of the (S)W-MSA module and the MLP module for block $l$, respectively, and W-MSA and SW-MSA are the W-MSA using regular and shifted window partitioning configurations, respectively. Figure 5b shows the W-MSA and SW-MSA, where SW-MSA is a variant of W-MSA with all windows shifted in size by one half-window to the right and downward.

The W-MSA significantly reduced the computational complexity. For an image with $M \times M$ patches within a window, the computational complexities of the W-MSA and traditional MSA were compared using Equations (5) and (6), respectively, as follows:

$$\Omega(MSA) = 4HWC^2 + 2(HW)^2C \tag{5}$$

$$\Omega(W\text{-}MSA) = 4HWC^2 + 2M^2HWC \tag{6}$$

where the former grows quadratically with the number of patches $HW$, and the latter remains linear when $M$ is fixed (default value = 8). Global self-attention computation is generally prohibitively expensive for large $HW$, whereas window-based self-attention is scalable.

#### 3.2.4. Mesoscale Integration Method

Mesoscale integration is the inverse process of sliding window sampling. Because the sliding window step size used in this study was smaller than the window size, the sampled mesoscale data that were obtained overlapped. When it is necessary to restore inverted mesoscale wind field data to large-scale data, a specific method is required to fuse the overlapping parts. The weighted fusion algorithm was used in this study, which ensured the continuity of the fused images. The following section describes the application of the weighted fusion algorithm for mesoscale integration.

Suppose there are two mesoscale images $A$ and $B$, to be fused, both of size $m \times n$. These two images must be merged from left to right, with an overlap size of $m \times d$. Image $A$ consists of two parts: $A_x$ and $A_y$, representing the overlapping and non-overlapping parts, respectively. Similarly, image $B$ can be divided into $B_x$ and $B_y$. Then, an index vector $I$ of length $d$ is defined as $I = [1, 2, \ldots, d]$. A weight vector $\vec{w} = [w_1, w_2, \ldots, w_d]$ is defined, where $w_i, i \in (1, 2, \ldots, d)$ is calculated as follows:

$$x_i = \left(\frac{l_i}{d-1} - 0.5\right) \tag{7}$$

$$w_i = \frac{1}{1 + e^{-kx_i}} \tag{8}$$

where $k$ is a hyperparameter controlling the degree of variation in weights, which was set to 0.5 in this study. The weight vector $\vec{w}$ is replicated $m$ times to obtain the weight matrix $W$. Let $C$ represent the merged image, with the overlapping part denoted as $C_x$. Then, $C_x$ can be obtained as follows:

$$C_x = (1 - W) \circ A_x + W \circ B_x \tag{9}$$

where ∘ denotes the Hadamard product. Then, horizontally concatenating $A_y$, $C_x$, and $B_y$ yields the weighted fusion image $C$. Repeating this process enables mesoscale data to be integrated and concatenated into large-scale data, achieving a resolution transformation from $2 \times 48 \times 48$ per window to a continuous large-scale resolution of $2 \times 240 \times 240$, which represents the preliminary large-scale wind field.

### *3.3. Large-Scale Stage*

In this section, we consider the large-scale stage, primarily covering the construction of large-scale samples and the M-CoordUnet model.

#### 3.3.1. Large-Scale Sample Construction

After completing the mesoscale stage, the preliminary wind field at a resolution of 0.25° was obtained through the mesoscale integration process described above, with a resulting output dimension of $2 \times 240 \times 240$, which provided real-time, full-coverage wind fields but still lacked large-scale global information. Although we applied the mesoscale integration method to fuse different mesoscale results, this approach could not fully resolve the issue of inconsistent results across different mesoscale retrievals, leading to blurred or discontinuous boundaries in certain large-scale events (e.g., convection or storms), as shown in Figure 6. There have been studies that addressed similar ambiguity issues using blind source separation (BSS) methods [56], which provides valuable insights for tackling such challenges. However, we simply use the correction of M-CoordUnet on a large scale to mitigate this ambiguity.

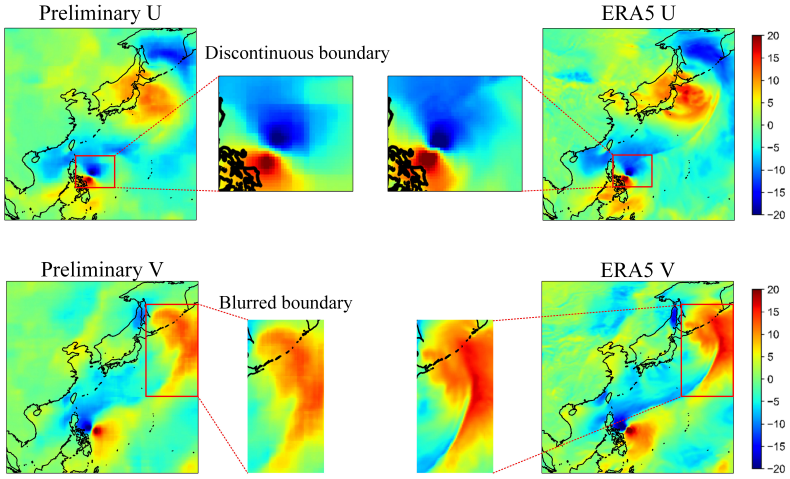



**Figure 6.** Discontinuous and blurred boundaries in the preliminary UV.

To address this limitation, we designed the large-scale stage. In this stage, we incorporated large-scale cloud data to obtain global information at a larger scale. However, the original Himawari satellite cloud observation data had high resolution, which would require excessive computational resources if they were used directly as a model input. Moreover, our aim was to extract large-scale information, rather than detailed features, from the data. Therefore, we applied a bilinear interpolation to downsample the data to a resolution of 0.25°, resulting in a downsampled cloud image of size $2 \times 240 \times 240$, consistent with the output dimensions of the mesoscale stage. Consequently, we applied both downsampled cloud data and preliminary UV as inputs to the large-scale model, using the former to obtain large-scale contour information and the latter to obtain detailed mesoscale information. We continued to use ERA5 UV data as the target, with the same invalid data handling and normalisation strategies as in the mesoscale stage.

### 3.3.2. M-CoordUnet Model

When designing the model for the large-scale stage, we primarily considered the following key issues. First, the input observational cloud data and preliminary UV data are multi-modal, with the cloud data having a dimension of $2 \times 240 \times 240$ and the preliminary UV data having a dimension of $2 \times 240 \times 240$. These were concatenated along the channel dimension, resulting in an input data dimension of $4 \times 240 \times 240$. Using a single encoder for feature extraction could lead to incomplete feature extraction due to the limited capacity of the encoder. Second, large-scale wind fields possess strong regional characteristics [57], but our input data lacked explicit regional information. Third, the number of large-scale samples was limited, and we needed to maintain a lightweight model to avoid overfitting. To address these issues, we designed the M-CoordUnet model, which adopted a multi-encoder design (Figure 7) that used separate encoders to extract features from different variables, thereby avoiding the potential limitations of a single encoder. Furthermore, inspired by the CoordConv-Unet model [58], we incorporated normalised latitude and longitude coordinates into each encoder block, as two new channels to the input. These explicit spatial features facilitated more effective learning and capturing of spatial transformations. Once the features from each variable were separately extracted, they were aggregated by the M-CoordUnet model using the centre block, and then sent to the decoder for decoding. Each decoder block received the output from the previous block and the corresponding layer's encoder block, enabling it to consider both higher-level general features and layer-specific detailed features, thereby facilitating a more accurate reconstruction of the target [59]. Finally, the output was obtained through a $1 \times 1$ convolutional layer, yielding large-scale UV data with a dimension of $2 \times 240 \times 240$.

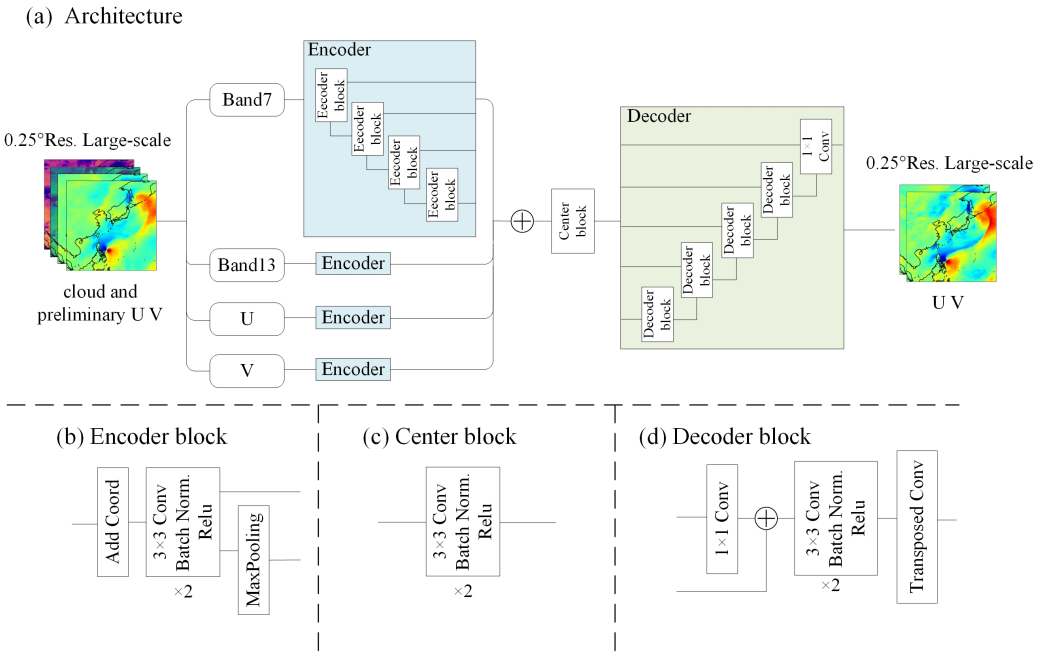



**Figure 7.** Architecture of the Multi-encoder Coordinate U-net (M-CoordUnet) model. (**a**) Overall architecture. (**b**–**d**) Structural details of the encoder, centre, and decoder blocks, respectively.

## 4. Results

To evaluate the performance and versatility of the proposed MFR method compared to the OSR method, scatterometer satellite wind field products, and IFS initial fields, the numerical results for different wind field characteristics (U and V wind field components, wind speed, and wind direction) were compared.

### *4.1. Model Evaluation*

To demonstrate the effectiveness of the MFR method proposed in this study, we also developed a one-stage retrieval (OSR) method for comparison. In the OSR method, we selected similar cloud data from channels 7 and 13 of the Himawari satellite observations. Subsequently, we directly downsampled the spatial resolution of the observation data from the original 0.02° to 0.25° using a bilinear interpolation method, matching the resolution of ERA5 wind field data. The downsampled observation data were then used as the input for the model, with the U and V components of ERA5 wind field data used as targets for model training. We experimented with the OSR method using both the Swin-Unet model [60] and the CoordConv-Unet model [58].

To better demonstrate the wind field retrieval performance in this study, evaluations were conducted separately for the U and V wind field components, as well as for wind speed and wind direction. The mean absolute error (*MAE*) and *RMSE*, which were used to assess scalar data such as wind speed and U and V wind field components, were calculated as follows:

$$MAE = \frac{1}{L}\Sigma_{k=1}^{K}\Sigma_{i=1}^{I}\Sigma_{h=1}^{H}\Sigma_{w=1}^{W}|y_{k,i,h,w} - \widehat{y}_{k,i,h,w}| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{L}\Sigma_{k=1}^{K}\Sigma_{i=1}^{I}\Sigma_{h=1}^{H}\Sigma_{w=1}^{W}(y_{k,i,h,w} - \hat{y}_{k,i,h,w})^2} \tag{11}$$

where $y$ is the true value, which in this experiment corresponds to the ERA5 wind field, and $\widehat{y}$ is the model prediction, representing the retrieved wind field in this experiment. $K$ is the total number of samples, $I$ is the number of features, and $H$ and $W$ are the height and width of the wind field image, respectively. In this experiment, $I$ was 2 (U and V components), and $H$ and $W$ were 240, resulting in a total of $L = K \times I \times H \times W$ elements.

However, for wind direction, which exhibits periodicity, using *MAE* or *RMSE* may lead to issues at discontinuous points. For example, an error between 10° and 350° should be treated as 20° rather than 340°. Therefore, for wind direction evaluation, we used their deviations ($MAE_d$ and $RMSE_d$, respectively), which were calculated as follows:

$$E = \begin{cases} \hat{y} - y, -180^\circ \le \hat{y} - y \le 180^\circ \\ \hat{y} - y + 360^\circ, \hat{y} - y \le -180^\circ \\ \hat{y} - y - 360^\circ, \hat{y} - y \ge 180^\circ \end{cases} \tag{12}$$

$$MAE_d = \frac{1}{L}\Sigma_{k=1}^{K}\Sigma_{i=1}^{I}\Sigma_{h=1}^{H}\Sigma_{w=1}^{W}|E_{k,i,h,w}| \tag{13}$$

$$RMSE_d = \sqrt{\frac{1}{L}\Sigma_{k=1}^{K}\Sigma_{i=1}^{I}\Sigma_{h=1}^{H}\Sigma_{w=1}^{W}(E_{k,i,h,w})^2} \tag{14}$$

### *4.2. Training Details*

The C2W-Former and M-CoordUnet models were operated using PyTorch [61]. Both models were trained on the training set on a single NVIDIA Tesla A40 48G graphics processing unit (GPU), with C2W-Former training lasting for approximately 7 days and M-CoordUnet training lasting for approximately 3 h. The loss functions used by both models were MAE + MSE, with the AdamW optimiser [62]. The initial learning rate was set to $5 \times 10^{-4}$, and the cosine learning rate adjustment strategy [63] was adopted, where the learning rate follows a cosine function, oscillating between a maximum of $5 \times 10^{-4}$ and a minimum of $5 \times 10^{-6}$ within each cycle. The batch size was set to 16.

The C2W-Former model was trained for a total of 30 epochs. Based on the validation set results, the model from the 22nd epoch was selected as the optimal model. The M-

CoordUnet model was trained for a total of 70 epochs, with the model from the 28th epoch chosen as the optimal model based on the validation set performance. The difference in the number of training epochs between the two models primarily arose from the difference in sample size. Because training samples used during the mesoscale retrieval phase were obtained by sampling from large-scale whole images, each yielding 81 mesoscale sub-images (nine in each of the horizontal and vertical directions), the difference in sample sizes between the two stages was approximately 81-fold.

### *4.3. Experimental Results*

We compared the wind field features retrieved by different methods from the test dataset and existing wind field products with the ERA5 wind field. For convenience, we denoted the large-scale preliminary UV results obtained from the mesoscale stage of the MFR method as MFR (S1), and the final results obtained from the large-scale stage of the MFR method as MFR (S2).

Table 3 and Figure 8 present the retrieval results of the MFR and OSR methods, where the U and V wind field components and wind speed were calculated using MAE/RMSE (m/s), and the wind direction was calculated using $MAE_d$ (°) and $RSME_d$ (°). In MFR (S2), minimal errors and maximum correlations were observed in the UV, wind speed, and wind direction (bold, Table 3), with MFR (S1) exhibiting the second-lowest errors (underscore, Table 3). The wind field retrievals significantly improved across various features from MFR (S1) to MFR (S2), confirming the effectiveness of multi-scale fusion. Furthermore, the MFR method outperformed the OSR method in all aspects, demonstrating its efficiency and superiority in wind field retrieval tasks. There were two main advantages of the MFR. First, dividing the entire large area into mesoscale regions and constructing retrieval models both considerably reduced the computational resources required for direct modelling at large-scales, and maintained the high resolution of the observational data, facilitating the generation of finer wind field retrieval information. Second, after completing mesoscale retrieval, wind field features were further extracted from large-scale cloud data, enhancing the geographical spatial perception of the model by incorporating coordinate information, and combining them with mesoscale wind field features to achieve multi-scale feature integration. The main limitation of the OSR method is that it extracted cloud features only from large scales, thereby losing significant mesoscale observational details.

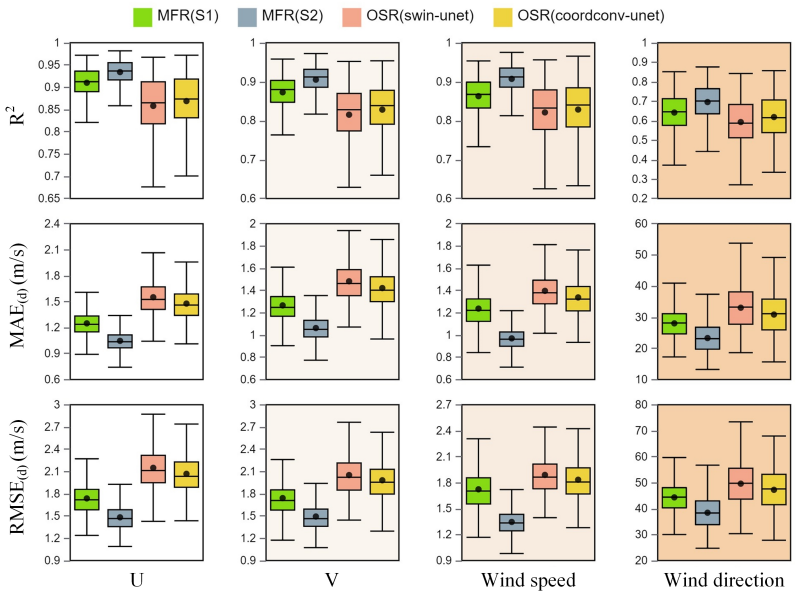



**Figure 8.** Analysis of MFR and OSR in comparison to ERA5 for land and sea regions.

**Table 3.** Error analysis of the MFR and one-stage retrieval (OSR) wind features in comparison to the ERA5 wind field for land and sea regions. Bold and underscore indicate optimal and suboptimal performance, respectively.

| | U | | V | | Wind Speed | | Wind Direction | |
|---|---|---|---|---|---|---|---|---|
| **Model** | **MAE** | **RMSE** | **MAE** | **RMSE** | **MAE** | **RMSE** | **MAE** | **RMSE** |
| MFR (S2) | **1.05** | **1.48** | **1.06** | **1.49** | **0.97** | **1.35** | **23.31** | **38.41** |
| MFR (S1) | <u>1.25</u> | <u>1.74</u> | <u>1.27</u> | <u>1.74</u> | <u>1.23</u> | <u>1.72</u> | <u>28.03</u> | <u>44.32</u> |
| OSR (CoordConv-Unet) | 1.48 | 2.07 | 1.42 | 1.98 | 1.33 | 1.83 | 30.86 | 47.23 |
| OSR (Swin-Unet) | 1.55 | 2.15 | 1.48 | 2.05 | 1.40 | 1.89 | 33.05 | 49.60 |

Next, we compared wind field retrievals from the MFR with interpolated wind field data from the HY-2B and CFOSAT satellite products. Because both scatterometer satellites are dedicated to oceanic observations, providing data solely on ocean surface winds, this comparison focused exclusively on marine areas within the study region. Table 4 presents errors with respect to the ERA5 wind fields. When evaluated based on the U and V components, the MFR wind field outperformed those of HY-2B and CFOSAT. Meanwhile, when considering wind speed and direction, the MFR results exhibited similarity to HY-2B, with the MFR performing better for wind speed, and HY-2B performing slightly better for wind direction. Both MFR and HY-2B significantly outperformed CFOSAT. These results show that the MFR wind field accuracy rivalled or even surpassed some existing operational satellite wind field products. In terms of regional coverage, MFR achieved complete coverage of both land and sea in the target area, whereas existing scatterometer wind field products only partially covered marine regions.

**Table 4.** Error analysis of the MFR and satellite products over the sea region. Bold and underscore indicate optimal and suboptimal performance, respectively.

| | U | | V | | Wind Speed | | Wind Direction | |
|---|---|---|---|---|---|---|---|---|
| **Model/Satellite** | **MAE** | **RMSE** | **MAE** | **RMSE** | **MAE** | **RMSE** | **MAE** | **RMSE** |
| MFR (S2) | **1.21** | **1.68** | **1.22** | **1.69** | **1.11** | **1.52** | <u>18.73</u> | <u>31.92</u> |
| HY-2B | <u>1.35</u> | <u>2.01</u> | <u>1.41</u> | <u>2.10</u> | <u>1.20</u> | <u>1.74</u> | **18.32** | **31.19** |
| CFOSAT | 1.54 | 2.24 | 1.61 | 2.29 | 1.36 | 2.00 | 23.04 | 38.03 |

Table 5 presents comparisons of the wind fields retrieved by MFR (S2) and the initial wind fields from IFS [64] conducted daily throughout 2021 at 00:00 and 12:00 UTC, covering both land and sea regions. Figure 9 shows scatterplots of each model (IFS, MFR, and OSR) versus ERA5 at 00:00 on 19 April 2021, during the passage of Super Typhoon Surigae. Combining these results shows that IFS initial wind fields outperformed MFR across all features. This gap was primarily attributed to differences in the volume of observational data utilised. IFS uses data from multiple meteorological observation satellites, buoys, and observation stations, and processes the information on supercomputers, whereas MFR utilises data from a single satellite and a single GPU. This discrepancy highlights the significance of our findings and underscores the importance of exploring the potential of using multifaceted observational data in wind field retrieval.

Table 6 presents the RMSE (m/s) performance of different models across varying wind speed ranges, including 3–10 m/s, 10–20 m/s, and >20 m/s. The results show that MFR(S2) achieves the lowest RMSE values in all wind speed categories (bold, Table 6), demonstrating its superior accuracy in wind field retrieval. In particular, for wind speeds above 20 m/s, there is a significant improvement from MFR(S1) to MFR(S2), highlighting the considerable benefit of multi-scale feature fusion in capturing wind field dynamics in

high-wind conditions. These findings further validate the effectiveness of the MFR method in accurately modeling wind fields across diverse scenarios.

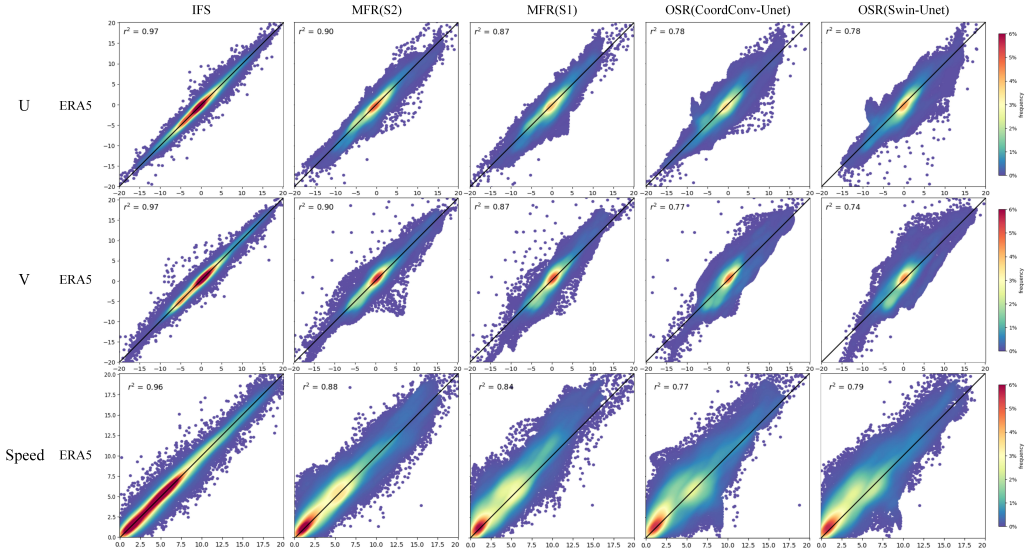



**Figure 9.** Scatter plots of each model versus ERA5 at 00:00 on 19 April 2021 (Super Typhoon Surigae). Closer proximity to the x = y line indicates better agreement between the two models. Warmer colours indicate higher frequency.

**Table 5.** Error analysis of the MFR and IFS wind initial field products over land and sea regions (00:00 and 12:00 daily). Bold indicate optimal performance.

| | U | | V | | Wind Speed | | Wind Direction | |
|---|---|---|---|---|---|---|---|---|
| **Model** | **MAE** | **RMSE** | **MAE** | **RMSE** | **MAE** | **RMSE** | **MAE** | **RMSE** |
| MFR (S2) | 1.04 | 1.47 | 1.05 | 1.48 | 0.96 | 1.34 | 23.81 | 39.19 |
| IFS | **0.62** | **0.89** | **0.64** | **0.91** | **0.58** | **0.83** | **16.03** | **30.41** |

**Table 6.** Error analysis of the MFR and one-stage retrieval (OSR) models at different wind speed intervals. Bold and underscore indicate optimal and suboptimal performance, respectively.

| **Method** | **3–10 m/s** | **10–20 m/s** | **>20 m/s** |
|---|---|---|---|
| MFR (S2) | **1.403** | **2.228** | **4.721** |
| MFR (S1) | <u>1.746</u> | 3.338 | 7.264 |
| OSR (CoordConv-Unet) | 1.912 | 3.322 | <u>6.137</u> |
| OSR (Swin-Unet) | 1.982 | <u>3.317</u> | 7.028 |

Figure 10 presents a comparative analysis of ERA5, IFS, and MFR against observational data from meteorological stations. ERA5 and IFS exhibited an extremely high level of agreement across all metrics. Although the MFR outcomes slightly lagged behind those of ERA5 and IFS, with the exception of a slightly larger deviation in wind speed $R^2$, the other variances were comparatively minor. This result suggests that from the perspective of in situ observations, the MFR results approached those of ERA5 and IFS.

Figure 11 shows the MAE values for the UV components of MFR-retrieved wind fields in the test dataset in different months. Errors in oceanic regions were higher than those in land areas, which was attributed to the prevalence of strong winds over oceans. This discrepancy represents a significant challenge to successful retrieval. Land area errors exhibited a distinct seasonal pattern, with lower errors in autumn and winter and higher errors in spring and summer. In contrast, seasonal patterns were less pronounced in

oceanic regions, primarily due to the dominance of monsoon climates over much of the eastern Eurasian continent, where land wind field characteristics were closely linked to seasonal variation.

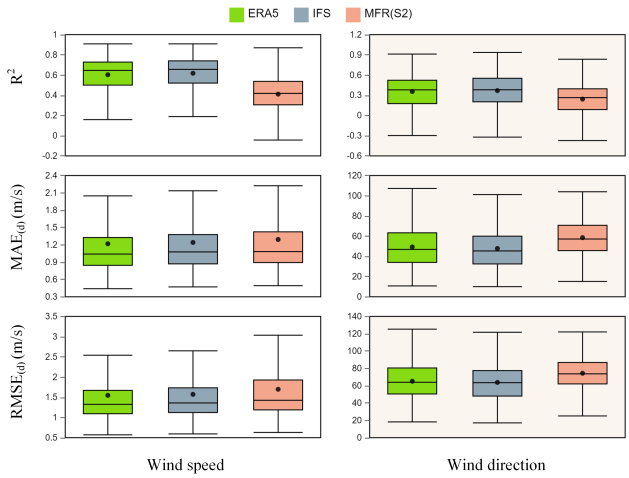



**Figure 10.** Analysis of ERA5, IFS, and MFR results in comparison to weather station data.

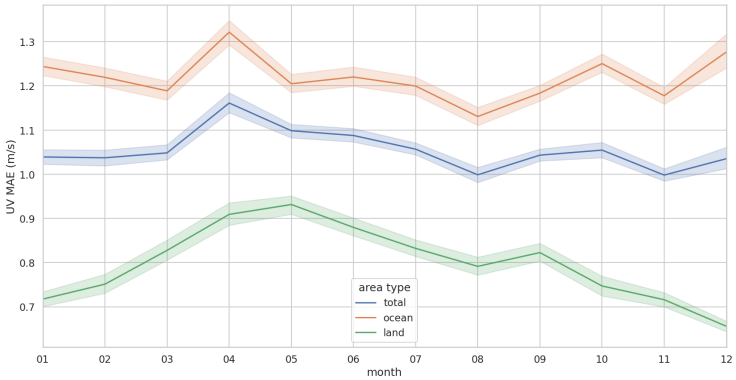



**Figure 11.** UV MAE statistics of MFR wind fields for land (green), sea (orange), and the total study area (blue) across different months and regions in the test data. Solid line indicates average error; shaded area indicates the 95% confidence interval.

## 5. Discussion

To better illustrate the effectiveness of the different retrieval methods and wind field products, we conducted a case study of a wind field retrieval during the landfall of Super Typhoon Surigae in the Philippines on April 19, 2021 at 0:00 UTC (Figure 12). All models successfully retrieved the storm (columns 3–6), with data positions closely matching those of the ERA5 wind field data. The MFR (S2) results were closer to the ERA5 data in terms of intensity and detail compared to MFR (S1), OSR (CoordConv-Unet), and OSR (Swin-Unet). For example, for the U component near the Japanese islands and the V component in the eastern Japanese islands, MFR (S2) produced significantly better contour details and intensity than MFR (S1), OSR (CoordConv-Unet), and OSR (Swin-Unet), all of which produced noticeable underestimations in these two regions. However, in terms of contour details, the IFS initial field outperformed MFR (S2). HY-2B and CFOSAT captured the most intricate features; however, both suffered from insufficient data coverage, notably CFOSAT, which did not observe the storm area in the 6 h period before and after the time frame of the interpolation. This comparison of wind speeds clearly showed that in storm retrieval, the edge details of the IFS initial field were the most accurate and clear, followed by MFR (S2), with OSR (Swin-Unet) showing the worst performance.

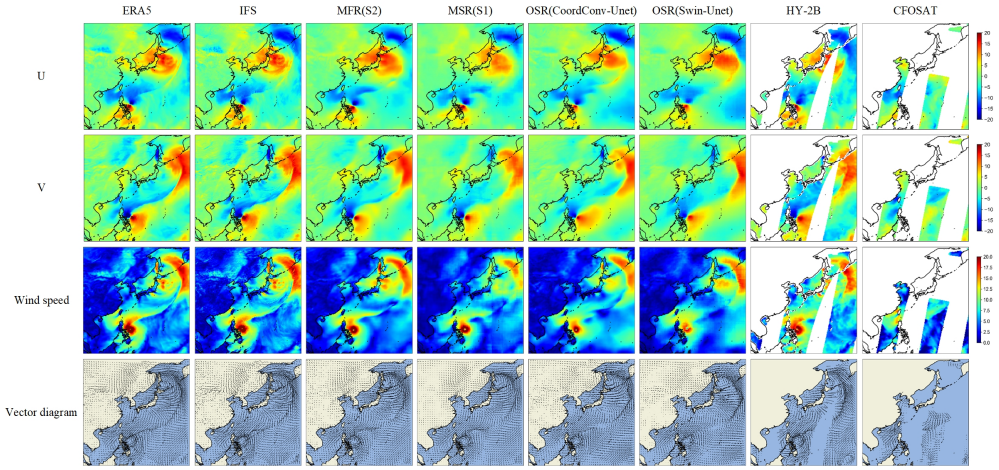



**Figure 12.** Comparison of wind field characteristics among different models and data products during Super Typhoon Surigae (19 April 2021, 00:00 UTC).

To avoid interpretations based on a single case, we examined another wind field retrieval case for Super Typhoon Mindulle on 28 September 2021, 12:00 UTC (Figure 13). The IFS initial field produced the richest details and was closely aligned with ERA5 data for the U component in the eastern Japanese islands and the V component in the northeastern region. The event was closely represented by MFR (S2), with only minor deviations. MFR (S1), OSR (CoordConv-Unet), and OSR (Swin-Unet) produced noticeable over- and underestimations in the U and V components, respectively, in this area. In terms of wind speed, MFR (S2) was closely aligned with both the IFS initial field and ERA5; however, it still exhibited smoothing issues. In contrast, MFR (S1), OSR (CoordConv-Unet), and OSR (Swin-Unet) had smoothing issues and also produced significant underestimations, particularly in land areas and the eastern region of the storm itself, as shown in the vector diagram. Furthermore, HY-2B did not reproduce the storm area before and after the 6 h time frame.

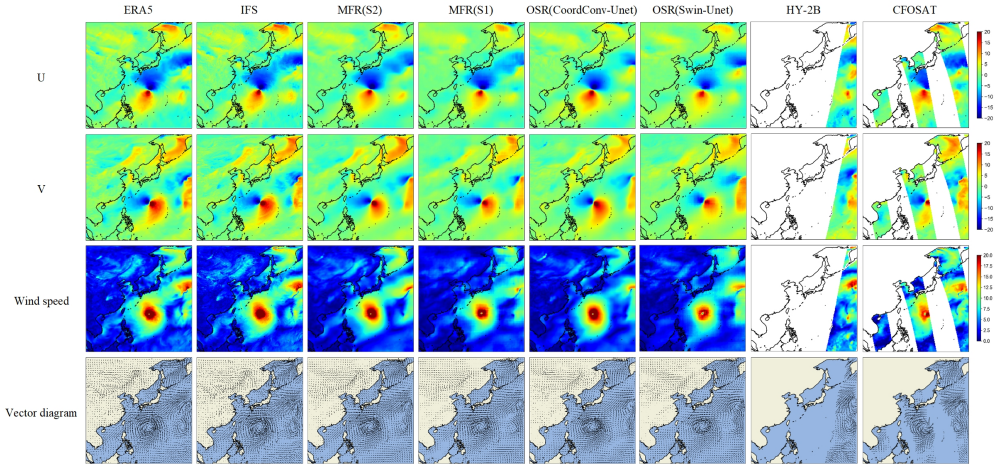



**Figure 13.** Comparison of wind field characteristics among different models and data products during Super Typhoon Mindulle (28 September 2021, 12:00 UTC).

Overall, the retrieval results of the MFR method showed good performance during both super typhoons, indicating its stability and reliability even under extreme weather conditions.

## 6. Conclusions

We developed an MFR wind field retrieval method. By analysing and modelling data at both the mesoscale and large scales, extracting wind field features from Himawari

satellite cloud observations, and integrating data from different scales, we achieved high-quality wind field retrievals with minimal resource requirements. The MAE/MAE$_d$ of wind speed and wind direction in the retrieved wind field were 0.97 m/s and 23.31°, respectively. The accuracy of the retrieved wind field matched or exceeded that of some scatterometer satellite wind field products, while providing complete coverage of land and sea areas. In terms of in situ observations, the MFR results nearly attained the same level of accuracy as ERA5 and IFS. Additionally, we demonstrated the effectiveness of the MFR method in retrieving wind fields under extreme weather conditions, highlighting its robust performance. Wind fields obtained through this method can be used as initial fields for large-scale grid forecasting, facilitating the development of deep learning-based wind field forecasting models using observational data.

In future studies, we will explore the incorporation of multi-source observational data into the retrieval process. Single observational satellites have limitations in their observation methods and biases in data processing. Therefore, integrating additional satellite observational data, as well as data from buoys and observation stations, into the retrieval process will mitigate the limitations of single observation methods and enhance the quality of the final retrieved wind fields. We will also investigate a forecasting model based on the retrieved wind field, with the aim of achieving deep learning large-scale grid wind field forecasting based on observational data.

**Author Contributions:** Conceptualization, W.Z. and X.S.; methodology, W.Z., Y.W. and K.F.; software, W.Z., Y.W., R.P. and B.G.; validation, W.Z., Y.W. and K.F.; formal analysis, W.Z., Y.W. and X.S.; investigation, W.Z. and Y.W.; resources, X.S.; data curation, W.Z.; writing—original draft preparation, W.Z., Y.W. and K.F.; writing—review and editing, W.Z., Y.W., K.F. and X.S.; visualization, Y.W. and K.F.; supervision, W.Z.; project administration, X.S.; funding acquisition, W.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (grant 42276202).

**Data Availability Statement:** The data supporting the findings of this study are available in the corresponding website links on the relevant page. Access to the data can be obtained through the provided link.

**Acknowledgments:** We sincerely thank the editors and reviewers for their constructive feedback and the time and effort they dedicated to the review process.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Lerner, J.; Grundmeyer, M.; Garvert, M. The importance of wind forecasting. *Renew. Energy Focus* **2009**, *10*, 64–66. [CrossRef]
2. Marchigiani, R.; Gordy, S.; Cipolla, J.; Adams, R.C.; Evans, D.C.; Stehly, C.; Galwankar, S.; Russell, S.; Marco, A.P.; Kman, N.; et al. Wind disasters: A comprehensive review of current management strategies. *Int. J. Crit. Illn. Inj. Sci.* **2013**, *3*, 130–142. [PubMed]
3. Ernst, B.; Oakleaf, B.; Ahlstrom, M.L.; Lange, M.; Moehrlen, C.; Lange, B.; Focken, U.; Rohrig, K. Predicting the wind. *IEEE Power Energy Mag.* **2007**, *5*, 78–89. [CrossRef]
4. Herbert, G.J.; Iniyan, S.; Sreevalsan, E.; Rajapandian, S. A review of wind energy technologies. *Renew. Sustain. Energy Rev.* **2007**, *11*, 1117–1145. [CrossRef]
5. Dagestad, K.F.; Horstmann, J.; Mouche, A.; Perrie, W.; Shen, H.; Zhang, B.; Li, X.; Monaldo, F.; Pichel, W.; Lehner, S.; et al. Wind retrieval from synthetic aperture radar—An overview. In *Proceedings of the 4th SAR Oceanography Workshop: Advances in SAR Oceanography*; European Space Agency: Paris, France, 2013.
6. Gao, H.; Yang, D.; Wang, F.; Wang, Q.; Li, X. Retrieval of ocean wind speed using airborne reflected GNSS signals. *IEEE Access* **2019**, *7*, 71986–71998. [CrossRef]
7. Zavorotny, V.U.; Voronovich, A.G. Scattering of GPS signals from the ocean with wind remote sensing application. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 951–964. [CrossRef]
8. Zuffada, C.; Elfouhaily, T.; Lowe, S. Sensitivity analysis of wind vector measurements from ocean reflected GPS signals. *Remote Sens. Environ.* **2003**, *88*, 341–350. [CrossRef]

9. Ruf, C.S.; Balasubramaniam, R. Development of the CYGNSS geophysical model function for wind speed. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *12*, 66–77. [CrossRef]
10. Ricciardulli, L.; Wentz, F.J. A scatterometer geophysical model function for climate-quality winds: QuikSCAT Ku-2011. *J. Atmos. Ocean. Technol.* **2015**, *32*, 1829–1846. [CrossRef]
11. Thoss, A. Cloud motion winds, validation and impact on numerical weather forecasts. In Proceedings of the Workshop on Wind Extraction from Operational Meteorological Satellite Data, Washington, DC, USA, 17–19 September 1991; pp. 105–112.
12. Nonaka, K.; Nishimura, S.; Igarashi, Y. Utilization of estimated sea surface wind data based on Himawari-8/9 low-level AMVs for tropical cyclone analysis. *RSMC Tokyo Typhoon Cent. Tech. Rev.* **2019**, *21*.
13. Horváth, Á.; Davies, R. Feasibility and error analysis of cloud motion wind extraction from near-simultaneous multiangle MISR measurements. *J. Atmos. Ocean. Technol.* **2001**, *18*, 591–608. [CrossRef]
14. Li, X.; Yang, D.; Yang, J.; Zheng, G.; Han, G.; Nan, Y.; Li, W. Analysis of coastal wind speed retrieval from CYGNSS mission using artificial neural network. *Remote Sens. Environ.* **2021**, *260*, 112454. [CrossRef]
15. Guo, W.; Du, H.; Guo, C.; Southwell, B.J.; Cheong, J.W.; Dempster, A.G. Information fusion for GNSS-R wind speed retrieval using statistically modified convolutional neural network. *Remote Sens. Environ.* **2022**, *272*, 112934. [CrossRef]
16. Kurth, T.; Subramanian, S.; Harrington, P.; Pathak, J.; Mardani, M.; Hall, D.; Miele, A.; Kashinath, K.; Anandkumar, A. Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. In Proceedings of the Platform for Advanced Scientific Computing Conference, Davos, Switzerland, 26–28 June 2023; pp. 1–11.
17. Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; Tian, Q. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **2023**, *619*, 533–538. [CrossRef] [PubMed]
18. Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W.; et al. Learning skillful medium-range global weather forecasting. *Science* **2023**, *382*, 1416–1421. [CrossRef]
19. Chen, K.; Han, T.; Gong, J.; Bai, L.; Ling, F.; Luo, J.J.; Chen, X.; Ma, L.; Zhang, T.; Su, R.; et al. Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. *arXiv* **2023**, arXiv:2304.02948.
20. Menzel, W.; Smith, W.; Stewart, T. Improved cloud motion wind vector and altitude assignment using VAS. *J. Clim. Appl. Meteorol.* **1983**, *22*, 377–384. [CrossRef]
21. He, L.; Bai, H.; Ouyang, D.; Wang, C.; Wang, C.; Jiang, Y. Satellite Cloud-Derived Wind Inversion Algorithm Using GPU. *Comput. Mater. Contin.* **2019**, *60*, 599–613. [CrossRef]
22. Bessho, K.; Date, K.; Hayashi, M.; Ikeda, A.; Imai, T.; Inoue, H.; Kumagai, Y.; Miyakawa, T.; Murata, H.; Ohno, T.; et al. An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. *J. Meteorol. Soc. Jpn. Ser. II* **2016**, *94*, 151–183. [CrossRef]
23. Wallace, J.M.; Hobbs, P.V. *Atmospheric Science: An Introductory Survey*; Elsevier: Amsterdam, The Netherlands, 2006; Volume 92.
24. Frank, W.M.; Ritchie, E.A. Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes. *Mon. Weather Rev.* **2001**, *129*, 2249–2269. [CrossRef]
25. Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; Vinyals, O. Understanding deep learning requires rethinking generalization. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
26. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
27. Xie, Q.; Luong, M.T.; Hovy, E.; Le, Q.V. Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 14–19 June 2020; pp. 10687–10698.
28. Yu, X.; Lou, X.; Yan, Y.; Yan, Z.; Cheng, W.; Wang, Z.; Zhao, D.; Xia, J. Radar echo reconstruction in oceanic area via deep learning of satellite data. *Remote Sens.* **2023**, *15*, 3065. [CrossRef]
29. Zhu, M.; Liao, Q.; Wu, L.; Zhang, S.; Wang, Z.; Pan, X.; Wu, Q.; Wang, Y.; Su, D. Multiscale Representation of Radar Echo Data Retrieved through Deep Learning from Numerical Model Simulations and Satellite Images. *Remote Sens.* **2023**, *15*, 3466. [CrossRef]
30. Mouche, A.A.; Collard, F.; Chapron, B.; Dagestad, K.F.; Guitton, G.; Johannessen, J.A.; Kerbaol, V.; Hansen, M.W. On the use of Doppler shift for sea surface wind retrieval from SAR. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2901–2909. [CrossRef]
31. Vachon, P.W.; Dobson, F.W. Wind retrieval from RADARSAT SAR images: Selection of a suitable C-band HH polarization wind retrieval model. *Can. J. Remote Sens.* **2000**, *26*, 306–313. [CrossRef]
32. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
33. Ramon, J.; Lledó, L.; Torralba, V.; Soret, A.; Doblas-Reyes, F.J. What global reanalysis best represents near-surface winds? *Q. J. R. Meteorol. Soc.* **2019**, *145*, 3236–3251. [CrossRef]
34. Çalışır, E.; Soran, M.B.; Akpınar, A. Quality of the ERA5 and CFSR winds and their contribution to wave modelling performance in a semi-closed sea. *J. Oper. Oceanogr.* **2023**, *16*, 106–130. [CrossRef]

35. Zeng, L.; Levy, G. Space and time aliasing structure in monthly mean polar-orbiting satellite data. *J. Geophys. Res. Atmos.* **1995**, *100*, 5133–5142. [CrossRef]
36. Zou, J.; Lin, M.; Zou, B.; Guo, M.; Cui, S. Fusion of sea surface wind vector data acquired by multi-source active and passive sensors in China sea. *Int. J. Remote Sens.* **2017**, *38*, 6477–6491. [CrossRef]
37. Stull, R.B. *An Introduction to Boundary Layer Meteorology*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 13.
38. Kurihara, Y.; Murakami, H.; Kachi, M. Sea surface temperature from the new Japanese geostationary meteorological Himawari-8 satellite. *Geophys. Res. Lett.* **2016**, *43*, 1234–1240. [CrossRef]
39. Yamamoto, Y.; Ichii, K.; Ryu, Y.; Kang, M.; Murayama, S. Uncertainty quantification in land surface temperature retrieved from Himawari-8/AHI data by operational algorithms. *ISPRS J. Photogramm. Remote Sens.* **2022**, *191*, 171–187. [CrossRef]
40. Small, R.d.; deSzoeke, S.P.; Xie, S.; O'neill, L.; Seo, H.; Song, Q.; Cornillon, P.; Spall, M.; Minobe, S. Air–sea interaction over ocean fronts and eddies. *Dyn. Atmos. Ocean.* **2008**, *45*, 274–319. [CrossRef]
41. Hashizume, H.; Xie, S.P.; Fujiwara, M.; Shiotani, M.; Watanabe, T.; Tanimoto, Y.; Liu, W.T.; Takeuchi, K. Direct observations of atmospheric boundary layer response to SST variations associated with tropical instability waves over the eastern equatorial Pacific. *J. Clim.* **2002**, *15*, 3379–3393. [CrossRef]
42. Houze, R.A., Jr. Clouds in tropical cyclones. *Mon. Weather Rev.* **2010**, *138*, 293–344. [CrossRef]
43. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
44. Lin, H.; Jegelka, S. Resnet with one-neuron hidden layers is a universal approximator. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018; Volume 31.
45. Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. How powerful are graph neural networks? *arXiv* **2018**, arXiv:1810.00826.
46. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef]
47. Zhang, W.; Vaidya, I. Mixup training leads to reduced overfitting and improved calibration for the transformer architecture. *arXiv* **2021**, arXiv:2102.11402.
48. Shalabi, L.A.A.; Shaaban, Z.; Kasasbeh, B. Data Mining: A Preprocessing Engine. *J. Comput. Sci.* **2006**, *2*, 735–739. [CrossRef]
49. Al-Faiz, M.Z.; Ibrahim, A.A.; Hadi, S.M. The effect of Z-Score standardization (normalization) on binary input due the speed of learning in back-propagation neural network. *Iraqi J. Inf. Commun. Technol.* **2018**, *1*, 42–48. [CrossRef]
50. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part I 13; Springer: Berlin/Heidelberg, Germany, 2014; pp. 818–833.
51. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
52. Hochreiter, S. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
53. Zhao, D.; Heidler, K.; Asgarimehr, M.; Arnold, C.; Xiao, T.; Wickert, J.; Zhu, X.X.; Mou, L. DDM-Former: Transformer networks for GNSS reflectometry global ocean wind speed estimation. *Remote Sens. Environ.* **2023**, *294*, 113629. [CrossRef]
54. Nascimento, E.G.S.; de Melo, T.A.; Moreira, D.M. A transformer-based deep neural network with wavelet transform for forecasting wind speed and wind energy. *Energy* **2023**, *278*, 127678. [CrossRef]
55. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
56. Chang, S.; Deng, Y.; Zhang, Y.; Zhao, Q.; Wang, R.; Zhang, K. An advanced scheme for range ambiguity suppression of spaceborne SAR based on blind source separation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5230112. [CrossRef]
57. Li, T.; Wang, B. REVIEW A review on the western north Pacific monsoon: Synoptic-to-interannual variabilities. *TAO Terr. Atmos. Ocean. Sci.* **2005**, *16*, 285. [CrossRef]
58. El Jurdi, R.; Petitjean, C.; Honeine, P.; Abdallah, F. CoordConv-unet: Investigating CoordConv for organ segmentation. *IRBM* **2021**, *42*, 415–423. [CrossRef]
59. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; proceedings, part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
60. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 205–218.

61. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32.
62. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
63. Loshchilov, I.; Hutter, F. SGDR: Stochastic Gradient Descent with Warm Restarts. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
64. ECMWF. *IFS Documentation CY48R1—Part I: Observations*; Number 1; ECMWF: Reading, UK, 2023. [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.