

## Article

# Short- to Medium-Term Weather Forecast Skill of the AI-Based Pangu-Weather Model Using Automatic Weather Stations in China

Siyi Xu <sup>1,2</sup>, Yize Zhang <sup>1,2,\*</sup>, Junping Chen <sup>1,2</sup>  and Yunlong Zhang <sup>3,4</sup>

<sup>1</sup> Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China; xusiyi@shao.ac.cn (S.X.); junping@shao.ac.cn (J.C.)

<sup>2</sup> School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> China Railway Design Corporation, Tianjin 300308, China; 13115311@bjtu.edu.cn

<sup>4</sup> Tianjin Key Laboratory of Rail Transit Navigation Positioning and Spatio-Temporal Big Data Technology, Tianjin 300308, China

\* Correspondence: zhyize@shao.ac.cn

**Abstract:** Pangu is an AI-based model designed for rapid and accurate numerical weather forecasting. To evaluate Pangu's short- to medium-term weather forecasting skill over various meteorological parameters, this paper validated its performance in predicting temperature, wind speed, wind direction, and barometric pressure using data from over 2000 weather stations in China. Pangu's performance was compared with ECMWF-HRES and GFS to assess its effectiveness relative to traditional high-precision NWP models under real meteorological conditions. Furthermore, the more recent FuXi and FengWu models were included in the analysis to further validate Pangu's forecasting skill. The study examined Pangu's forecast performance from spatial perspectives, evaluated the dispersion of forecast deviations, and analyzed its performance at different lead times and with various initial fields. The iteration precision of Pangu's four forecast models with lead times of 1 h, 3 h, 6 h, and 24 h was also assessed. Finally, a case study on typhoon track forecasting was conducted to evaluate Pangu's performance in predicting typhoon paths. The results indicate that Pangu surpasses traditional NWP systems in temperature forecasting, while its performance in predicting wind direction, wind speed and pressure is comparable to them. Additionally, the forecast skill of Pangu diminishes as the lead time extends, but it tends to surpass traditional NWP systems with longer lead times. Moreover, FuXi and FengWu demonstrate even higher accuracy compared to Pangu. Pangu's performance is also dependent on initial fields, and the temperature forecasting of Pangu is more sensitive to the initial field compared with other meteorological parameters. Furthermore, the iteration precision of Pangu's 1 h forecast model is significantly lower than that of the other models, but this discrepancy in precision may not be prominently reflected in Pangu's actual forecasting process due to the greedy algorithm employed. In the case study on typhoon forecasting, Pangu, along with FuXi and FengWu, demonstrates comparable performance in predicting Bebinca's track compared to ECMWF and outperforms GFS in its track predictions. This study demonstrated Pangu's applicability in short- to medium-term forecasting of meteorological parameters, showcasing the significant potential of AI-based numerical weather models in enhancing forecast performance.



Academic Editors: Filippo Biondi and Ahmed Shaharyar Khwaja

Received: 2 November 2024

Revised: 2 January 2025

Accepted: 5 January 2025

Published: 8 January 2025

**Citation:** Xu, S.; Zhang, Y.; Chen, J.; Zhang, Y. Short- to Medium-Term Weather Forecast Skill of the AI-Based Pangu-Weather Model Using Automatic Weather Stations in China. *Remote Sens.* **2025**, *17*, 191. <https://doi.org/10.3390/rs17020191>

**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/>).

**Keywords:** numerical weather prediction; deep learning; weather forecast; prediction skill

## 1. Introduction

Accurate weather forecasts and the acquisition of precise meteorological parameters are essential in various fields such as agriculture, transportation, energy generation,

disaster preparedness, and outdoor events. In recent years, weather forecast products have continually undergone advancements in numerical models, initial conditions, resolutions, data assimilation, and computational power [1–4]. Nowadays, traditional numerical weather prediction (NWP) models remain the primary means for obtaining the most accurate weather forecast parameters. Among them, the Integrated Forecast Systems (IFS) of the European Centre for Medium-range Weather Forecast (ECMWF) and the Global Forecast System (GFS) of the National Centers for Environmental Prediction (NCEP) of the United States are considered two of the most renowned leading global NWP systems. They provide forecasts at different resolutions and for multiple time ranges. These traditional numerical weather models (designated as physical models herein) are realized using the laws of physics by optimally merging earth system observations and short-range forecasts through data assimilation. However, improving the data volume and resolution of these models requires a proportionate increase in computational cost, which has persistently hindered their development.

Over the past couple of years, the rapid advancements in the field of artificial intelligence (AI) have led to the emergence of numerical weather models based on deep learning, marking a significant revolution in numerical weather forecasting. Owing to their lower computational cost and relatively high reliability, employing AI-based numerical weather models as an alternative to traditional NWP models for weather parameter prediction under similar initial conditions is a dependable choice as well. The AI-based NWP models have broadly undergone three distinct developmental stages. In the first stage, artificial neural networks (NNs) demonstrated their potential for applications in meteorological forecasting, effectively capturing nonlinearity in systems. Numerous examples exist of NNs being employed to predict specific meteorological variables and phenomena [5–8]. Additionally, ongoing efforts are being made to apply NNs to the post-processing of dynamical model output to enhance operational weather forecasts [9,10]. In the second stage, the neural network frameworks used for model training became increasingly sophisticated, with a gradual increase in training parameters. For instance, Weyn et al. [11] introduced a significantly improved data-driven global weather forecasting framework utilizing a deep convolutional neural network (CNN) to predict several basic atmospheric variables. This model demonstrated superior performance compared to a coarse-resolution dynamical NWP model for short- to medium-range forecasting. Rasp et al. [12] developed a deep residual CNN (Resnet) to predict geopotential, temperature, and precipitation at a  $5.625^\circ$  resolution up to 5 days ahead. When compared to physical models, the Resnet achieved comparable scores at a similar resolution. However, the accuracy of these models still noticeably lagged behind that of traditional NWP models. The third stage, which is the most rapidly evolving period for AI-based NWP models, features more systematized and refined model construction, as well as continuous improvements in resolution and training data volume. Keisler [13] first implemented a graph neural network (GNN)-based model and reported that it performs well, with forecasting capabilities comparable to operational physical models when evaluated at  $1^\circ$  scales and using reanalysis data for initial conditions. Pathak et al. [14] developed FourCastNet based on the Adaptive Fourier Neural Operator model, enabling high-resolution forecasts at  $0.25^\circ$ . They claimed that FourCastNet matches the forecasting accuracy of the ECMWF-IFS at short lead times (the time difference between input and output) for large-scale forecasting while surpassing IFS for small-scale forecasting. Bi et al. [15] introduced the Pangu-Weather (referred to hereafter as Pangu) model, which is based on the design of the 3DEST architecture and the hierarchical temporal aggregation strategy for medium-range forecasting. According to their analysis, Pangu generates superior deterministic forecast results on reanalysis data compared to the ECMWF-IFS and excels at forecasting extreme weather events and ensemble weather forecasts. Lam

et al. [16] presented GraphCast, which is based on a GNN. Their experimental results indicated that GraphCast outperforms the most accurate operational deterministic systems across most verification targets. Similarly, other models, such as SwinRDM [17], FuXi [18], FengWu [19], and ClimaX [20], have also shown promising results. These models are built upon the availability of high-quality, multi-decadal Earth system reanalysis training data, such as the ECMWF ERA5 reanalysis data [21].

Pangu, as a representative AI-based model, is open source for non-commercial use [15,22]. Several studies have been conducted to validate its forecasting skill. Bonavita [23] argued that Pangu's forecasts do not have the fidelity and physical consistency of physics-based models and its advantage in accuracy on traditional deterministic metrics of forecast skill at longer lead times can be partly attributed to these peculiarities (which are called the "double penalty" effect). However, this advantage comes at the cost of reduced interpretability and trustworthiness of the model. Cheng et al. [24] evaluate the compatibility of Pangu with traditional NWP models, and their test results indicate that Pangu can generate results that are comparable or superior to the original NWP products when using the operational analyses from various operational NWP systems as inputs. Furthermore, the quality of Pangu's forecasts improves as the quality of the initial conditions increases. Hakim et al. [25] tested Pangu on a set of four canonical experiments aimed at probing its dynamical response to local perturbations. The results suggest that Pangu encodes realistic physics for the experiments considered, thus motivating future basic research using this tool. Xu et al. [26] verified the performance of Pangu in predicting the unprecedented rainfall in North China during late July and early August. Building on this, they introduced an AI-driven model, Pangu-WRF, which integrates Pangu AI forecasts into the regional Weather Research and Forecasting (WRF) model. This approach demonstrates improved accuracy in predicting rainfall patterns for extreme weather events.

The forecasting and acquisition of meteorological parameters for short- to medium-term periods are crucial in the field of weather forecasting. These parameters are essential not only for daily weather assessments but also for providing foundational data for various related disciplines. By analyzing meteorological factors such as temperature, relative humidity, wind, and pressure, we can deepen our understanding of meteorological phenomena, atmospheric variations, and sky conditions. Additionally, numerous studies can be conducted, such as estimating global solar radiation in a specific region [27,28] and calculating precipitable water vapor in the atmosphere [29,30]. However, to date, few studies have systematically analyzed the feasibility, data quality, and performance of short- to medium-term meteorological parameter forecasts by AI-based models like Pangu, using specific weather station observation data. This paper will focus on the surface meteorological parameters forecasted by Pangu. The performance of Pangu's forecasts will be analyzed based on the observed data from over 2000 weather stations in China. High-precision forecast products from ECMWF and GFS will be used for comparative analysis. Through this examination, we aim to gain further insight into the capabilities of the Pangu model in forecasting meteorological parameters in the short- to medium-term.

Following the introduction of the current Section 1, the paper continues with the data resources and methodology shown in Section 2. The precision analysis of the Pangu model is detailed in Section 3, specifically focusing on its performance in relation to different meteorological parameters, variations based on lead times, and the impact of initial fields on forecast precision. The discussion is given in Section 4.

## 2. Data and Methods

### 2.1. Study Area and Proposed Model

In order to test the forecast performance of Pangu, this paper utilizes the observation data from over 2000 stations published by the National Meteorological Center of mainland China (<http://www.nmc.cn/>) (accessed on 7 June 2024) as true values of the performance test, as shown in Figure 1.

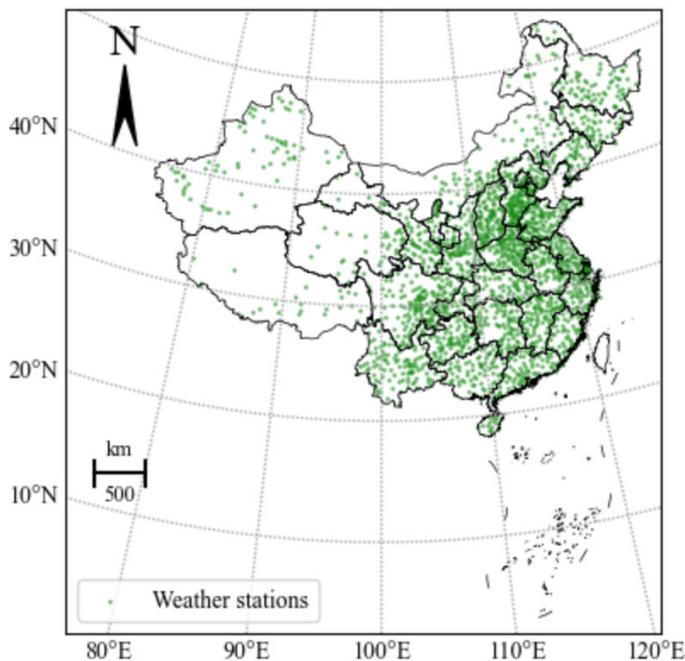


Figure 1. Distribution of regional weather stations in China.

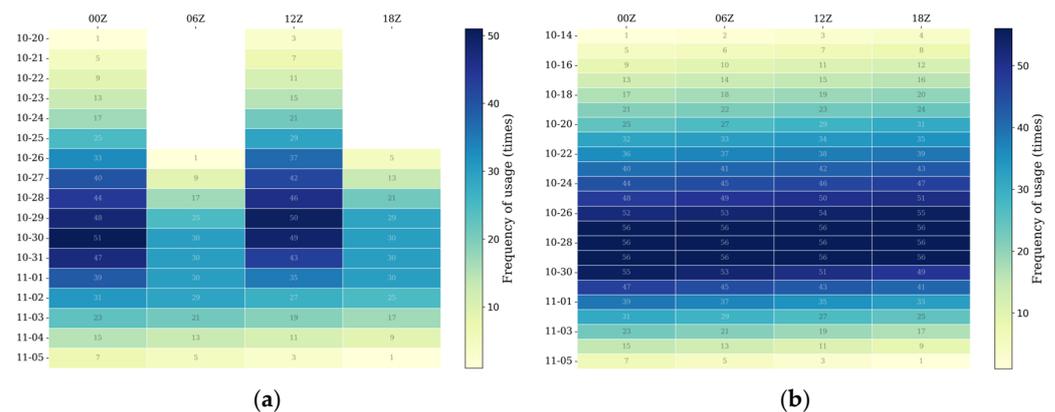
ECMWF-IFS [31] is widely regarded as the most accurate global weather forecast model. A wide range of ECMWF's high-resolution forecast (HRES) data across the globe has been made accessible since 25 January 2022, with a 6-h time interval and a spatial resolution of  $0.4^\circ$ , making these "open data" freely available to anybody interested in them (<https://confluence.ecmwf.int>) (accessed on 7 June 2024). The HRES data used to obtain atmospheric fields can be categorized into two types based on the forecast initiation time, lead time, and specific resolution: the operational high-resolution forecast ('oper'), available at 00Z and 12Z, which has a temporal resolution of 3 h for lead times ranging from 0 to 144 h and 6 h for lead times between 144 and 240 h; and the short cut-off high-resolution forecast ('sca'), available at 06Z and 18Z, which maintains a temporal resolution of 3 h for lead times between 0 and 90 h.

GFS [32], like ECMWF-IFS, also known as one of the world's most accurate NWP models, can provide global forecasts on every 6 h base time (available at 00Z, 06Z, 12Z, and 18Z) and at a spatial resolution of  $0.25^\circ$  (<https://nomads.ncep.noaa.gov/>) (accessed on 7 June 2024). The forecast lead time ranges from 0 to 384 h, with a temporal resolution of 3 h for lead times between 0 and 240 h, and 6 h for lead times between 240 and 384 h.

Pangu [15], part of Huawei Cloud's series of pre-trained AI models, is a robust AI-driven weather forecasting system that delivers strong deterministic results at hourly increments. Trained on 39 years of ERA5 data (1979 to 2017), it employs a three-dimensional Earth-specific transformer (3DEST) architecture to incorporate Earth-specific priors into deep networks. Pangu can generate 69 factors, including 5 upper-air variables at 13 pressure levels and 4 surface variables, with a spatial resolution of  $0.25^\circ$ . When assessing forecast skill, Pangu is considered capable of producing superior deterministic forecasts compared to the world's leading NWP system, ECMWF's operational IFS, while also being much faster.

FuXi [18] and FengWu [19] are two other advanced AI-based weather models developed after Pangu, designed to enhance weather predictions by leveraging machine learning techniques. Both models provide global forecasts on a 6 h time interval, with a spatial resolution of 0.25 degrees. As AI-based weather forecast models continue to evolve rapidly, it is also a point worth discussing whether the accuracy of the later models, FuXi and FengWu, will surpass that of Pangu.

Based on the observed values reported by weather stations, Pangu will be compared with ECMWF and GFS in terms of short- to medium-range predictions to testify whether this AI-based model can outperform traditional NWP models. This study only focuses on the surface variables of the three models, namely, 2 m temperature, u-component of 10 m wind speed, v-component of 10 m wind speed and mean sea level pressure. When it comes to the selection of the time period for the study, the observed value from the weather station is chosen for the period from 30 October 2023 to 5 November 2023, with a temporal resolution of 3 h. The initial fields' timepoints for the ECMWF, GFS, and Pangu models are chosen as shown in Figure 2. It is worth noting that the maximum lead time available for the ECMWF is 10 days, which is shorter than that of the other models. Therefore, to ensure fair comparison in subsequent statistics, the lead time for the forecasts from both the GFS and Pangu models will also be limited to within 10 days. Moreover, the AI-based FuXi and FengWu models will also be used to verify the forecast skill of Pangu across different lead times and during typhoon events.



**Figure 2.** The selected time (where rows represent dates, and columns represent daily timepoints) and frequency of usage of the initial field for ECMWF (a), GFS and Pangu (b).

## 2.2. Data Preprocessing

As mentioned above, this study focuses on the precision of four types of surface variables provided by three weather forecast models. Among all of the surface variables, the 2 m air temperature forecast by Pangu and the traditional NMW models can be directly compared with the observed values in weather stations. The values of the u and v components of the 10 m wind speed can be analyzed to calculate the final wind speed and wind direction for future comparison using the following transformation formula [33]:

$$\begin{cases} \text{Wind speed} = \sqrt{u^2 + v^2} \\ \text{Wind direction} = \text{mod}(180 + \arctan2(u, v), 360) \end{cases} \quad (1)$$

while the mean sea level pressure forecast by the models must be generalized to the station pressure so that they can be compared on the basis of the observed values at weather stations. The barometric formula [34] expresses the dependence of atmospheric pressure on altitude, which can be used to retrieve the pressure at station level:

$$\begin{cases} P_s = P_0 * 10^{-m} \\ m = \frac{h}{18400(1+\alpha t_m)} \end{cases} \quad (2)$$

where  $P_0$  is the pressure at mean sea level,  $P_s$  is pressure at the station level,  $h$  is the height of the barometric pressure sensor,  $\alpha = \frac{1}{273}$  is the coefficient of expansion of air, and  $t_m$  is the mean virtual temperature of the fictitious air column between the sea level and station level. Specifically,  $t_m$  is defined by the average of the sea surface temperature and the temperature at the corresponding station as shown in Equation (3). Here, the station temperature is determined by the average of the current temperature ( $t$ ) and the temperature 12 h ago ( $t_{12}$ ). Since the sea surface temperature cannot be obtained from either Pangu or traditional NWP forecast models—both of which focus on the 2 m surface temperature that more accurately reflects weather changes—it should be retrieved from the station temperature by taking the vertical temperature gradient as 0.5 °C/100 m:

$$t_m = \frac{t + t_{12}}{2} + \frac{h}{400} \quad (3)$$

Since Pangu and GFS generate data-driven forecasts of key atmospheric variables at a resolution of 0.25°, the variables generated by ECMWF will be interpolated to 0.25° using the linear method. In order to correlate the forecast data at grid points with meteorological data from specific latitudes and longitudes, grid point values will be bilinearly interpolated to each weather station. It should be noted that the interpolation method has little impact on the precision metrics established for this study.

### 2.3. Verification Metrics

The forecasting performance of the models is assessed based on the mean absolute error (MAE) and root mean squared error (RMSE); their formulas are as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |R_{fi} - R_{oi}| \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (R_{fi} - R_{oi})^2} \quad (5)$$

where  $N$  is the total number of scored weather stations in the study region, and  $R_{fi}$  and  $R_{oi}$  are the forecasted and observed values at each station, respectively.

In addition, another metric of forecast performance applied here is the anomaly correlation coefficient (ACC), which is a measure of forecast skill in the context of forecasts relative to climatology. The formulation of ACC is defined as follows [35]:

$$\text{ACC} = \frac{\sum_{i=1}^N (R_{fi} - \bar{R}_f) \times (R_{oi} - \bar{R}_o)}{\sqrt{\sum_{i=1}^N (R_{fi} - \bar{R}_f)^2 \times \sum_{i=1}^N (R_{oi} - \bar{R}_o)^2}} \quad (6)$$

where  $\bar{R}_f$  and  $\bar{R}_o$  represent the climatological means of the forecasted and observed values, respectively. In this study, we refer to these as the averages of all the forecasted and observed spatial values of each variable within the selected time period.

The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates. In this study, the empirical bootstrap method is applied to estimate the uncertainty of the verification metrics. Below is the main procedure of the empirical bootstrap method [36–38].

Given the original dataset with  $n$  data points  $[X_1, \dots, X_n]$ , let  $M_n = f(X_1, \dots, X_n)$  be the estimate of the population statistic of interest. The goal is to construct a  $1 - \alpha$  confidence interval for the population statistic as  $(M_n - w_{1-\frac{\alpha}{2}}, M_n - w_\alpha)$ .

Next, sampling with a replacement from the original dataset generates a new set of data points  $[X_1^{*(1)}, \dots, X_n^{*(1)}]$ . This sampling process is repeated multiple times, and after 10,000 rounds, 10,000 sets of data points are obtained:

$$\begin{aligned} & [X_1^{*(1)}, \dots, X_n^{*(1)}], \\ & [X_1^{*(2)}, \dots, X_n^{*(2)}], \\ & \dots, \\ & [X_1^{*(10,000)}, \dots, X_n^{*(10,000)}]. \end{aligned} \quad (7)$$

Each set of the data points is called a bootstrap sample. From the 10,000 bootstrap samples, 10,000 bootstrap statistics can be calculated:

$$\begin{aligned} M_n^{*(1)} &= f(X_1^{*(1)}, \dots, X_n^{*(1)}), \\ M_n^{*(2)} &= f(X_1^{*(2)}, \dots, X_n^{*(2)}), \\ & \dots, \\ M_n^{*(10,000)} &= f(X_1^{*(10,000)}, \dots, X_n^{*(10,000)}). \end{aligned} \quad (8)$$

The differences between the bootstrap sample statistics and the original sample statistic are then calculated:

$$\begin{aligned} & M_n^{*(1)} - M_n, \\ & M_n^{*(2)} - M_n, \\ & \dots, \\ & M_n^{*(10,000)} - M_n. \end{aligned} \quad (9)$$

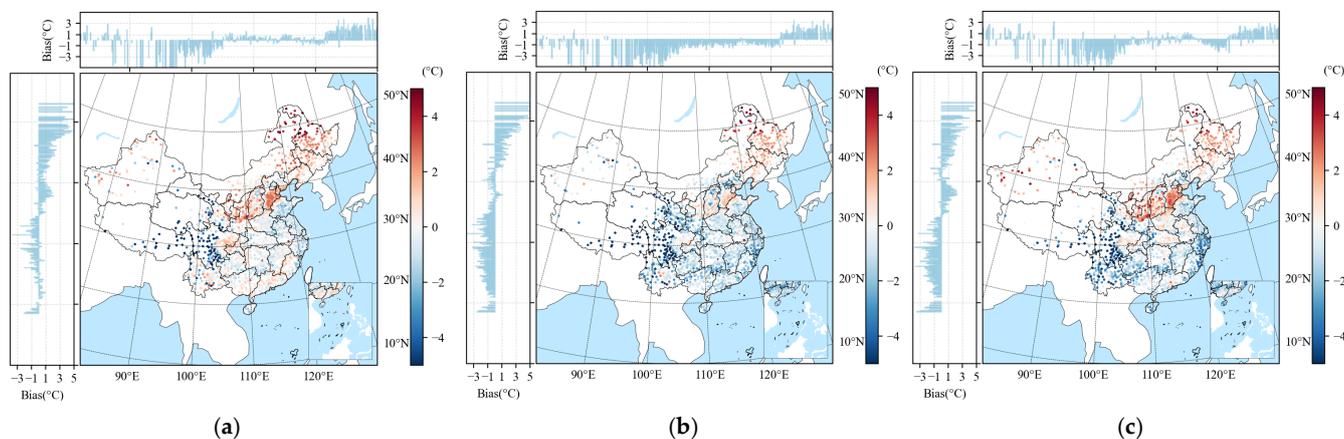
The 10,000 differences are then sorted, and the percentiles  $w_\alpha^*$  and  $w_{1-\frac{\alpha}{2}}^*$  are selected from the lower and upper tails. According to the bootstrap principle, the variation of the bootstrap statistics  $M_n^*$  around the original sample statistic  $M_n$  is a good approximation of how  $M_n$  varies around the true population statistic  $M$ . Therefore, the interval  $(M_n - w_{1-\frac{\alpha}{2}}^*, M_n - w_\alpha^*)$  is used as the confidence interval for the population statistic, replacing  $(M_n - w_{1-\frac{\alpha}{2}}, M_n - w_\alpha)$ .

### 3. Results

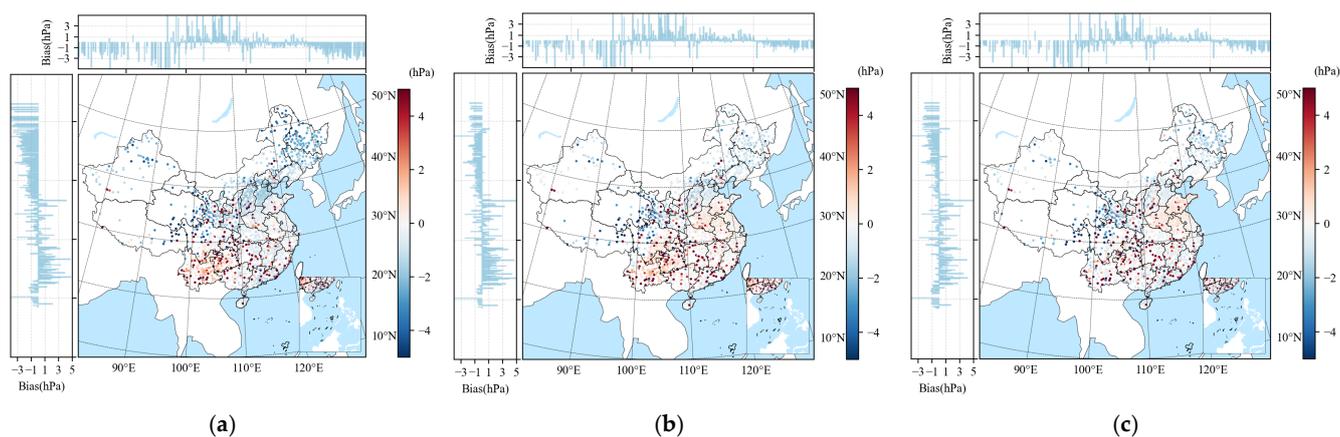
#### 3.1. Spatial Assessment of the Meteorological Parameters

To validate the forecasting characteristics of the Pangu model compared to traditional NWP models for the China region, this paper first analyzes the meteorological parameters (temperature, wind speed, wind direction, and barometric pressure) of Pangu, ECMWF, and GFS forecasts based on the spatial scale of China. Figures 3 and 4 illustrate the spatial distribution of temperature and barometric pressure forecast deviations by Pangu, ECMWF, and GFS during the seven-day period from 30 October to 5 November 2023, based on lead times ranging from 3 h to 10 days. Generally speaking, the temperature deviations of the three forecast models are mostly in the range of  $-5 \sim 5$  °C, and the overall distribution of Pangu temperature deviations is very close to that of ECMWF and GFS. Notably, all three models exhibit a tendency to overestimate temperature deviations in Northeast China and underestimate them in the southwest. All three forecast models present poor temperature forecasts in the western region of China, which may be related to the scarcity and uneven spatial distribution of basic meteorological data in this region. Meanwhile, among the three models, GFS and Pangu exhibit more positive biases in North China (between latitudes  $30^\circ$  and  $40^\circ$  N), indicating a tendency to overestimate the temperature in this region. The

pressure deviations of the three forecast models mostly fall within the range of  $-5\sim 5$  hPa. The overall distribution of Pangu's pressure deviations closely resembles that of traditional NWP forecast models, except in Northeast China, where Pangu exhibits a higher frequency of negative deviation values compared to the other two models. It can be concluded that both AI-based models and traditional NWP models exhibit regional imbalances in their estimated deviations for either temperature or pressure.



**Figure 3.** Temperature forecast mean bias map by Pangu (a), ECMWF (b), and GFS (c) for all days from 30 October to 5 November 2023.



**Figure 4.** Pressure forecast mean bias map by Pangu (a), ECMWF (b), and GFS (c) for all days from 30 October to 5 November 2023.

The speed and direction of the wind can be determined by looking at the 16-sector wind rose, as shown in Figure 5, based on the average forecasting data from 30 October to 5 November 2023. The color indicates wind speed, while the distance from the center indicates the probability in percent that the wind comes from one of 16 directions. As can be seen from the figure, both Pangu and traditional NWP forecasting models exhibit similar characteristics. Although all the three forecasting models align with the actual values at weather stations in terms of the overall wind direction trend, they still exhibit some discrepancies in predicting both the specific wind direction and the ratios of different wind speeds compared to the observed values. This is reasonable, as certain weather conditions are highly unpredictable. Some uncertainties cannot be captured by physical mechanisms in NWP models, nor can they be adequately compensated for by the available information in the training datasets of AI-based models [39].

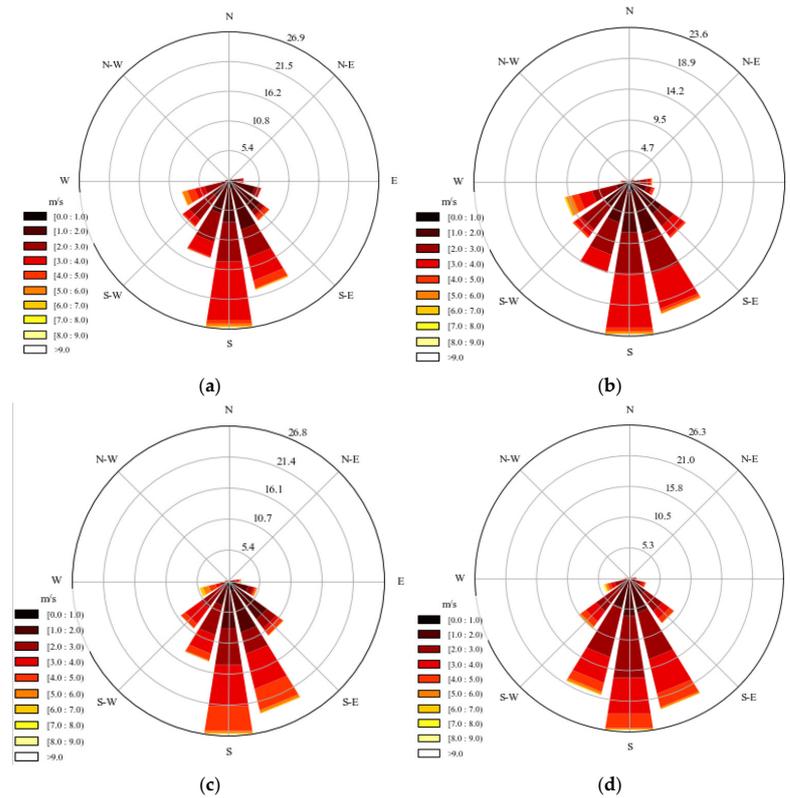


Figure 5. Wind rose of Pangu (a), ECMWF (b), GFS (c) and the observed values (d).

To assess the dispersion of forecast deviations across different models, this study introduces the standard deviation  $\sigma$ . The formula is the same as Equation (5), but with the forecast values  $R_{fi}$  replaced by the deviations between the forecasted and observed values, and the observed values  $R_{oi}$  replaced by the mean deviations. The three-sigma rule of thumb is a statistical measure indicating that the data are within three standard deviations from a mean. In the remainder of this subsection, the probabilities of meteorological parameter deviations falling within  $1\sigma$ ,  $2\sigma$ , or  $3\sigma$  from their mean will be statistically analyzed.

As can be seen from Table 1, the deviation distributions for all cases and all meteorological parameters are nearly identical across all models. Notably, for wind direction, the  $3\sigma$  value exceeds the maximum possible deviation ( $180^\circ$ ), resulting in a 100% probability of deviations falling within  $3\sigma$ . In terms of overall dispersion of deviations, pressure deviations are more concentrated compared to other parameters, with over 80% falling within  $1\sigma$ .

Table 1. Probabilities (%) of meteorological parameter deviations within  $1\sigma$ ,  $2\sigma$ , or  $3\sigma$  from the mean.

	Pangu	ECMWF	GFS
$\mu - 1\sigma \leq \text{Temperature bias} \leq \mu + 1\sigma$	76.54	77.44	74.21
$\mu - 2\sigma \leq \text{Temperature bias} \leq \mu + 2\sigma$	93.99	95.05	94.28
$\mu - 3\sigma \leq \text{Temperature bias} \leq \mu + 3\sigma$	98.45	98.79	98.93
$\mu - 1\sigma \leq \text{Pressure bias} \leq \mu + 1\sigma$	83.41	85.38	82.29
$\mu - 2\sigma \leq \text{Pressure bias} \leq \mu + 2\sigma$	94.23	94.49	94.35
$\mu - 3\sigma \leq \text{Pressure bias} \leq \mu + 3\sigma$	97.74	97.64	97.91
$\mu - 1\sigma \leq \text{Wind speed bias} \leq \mu + 1\sigma$	77.85	77.66	77.95
$\mu - 2\sigma \leq \text{Wind speed bias} \leq \mu + 2\sigma$	94.82	94.87	94.05
$\mu - 3\sigma \leq \text{Wind speed bias} \leq \mu + 3\sigma$	98.56	98.58	98.31
$\mu - 1\sigma \leq \text{Wind direction bias} \leq \mu + 1\sigma$	73.00	74.85	72.42
$\mu - 2\sigma \leq \text{Wind direction bias} \leq \mu + 2\sigma$	91.75	91.06	91.88
$\mu - 3\sigma \leq \text{Wind direction bias} \leq \mu + 3\sigma$	100	100	100

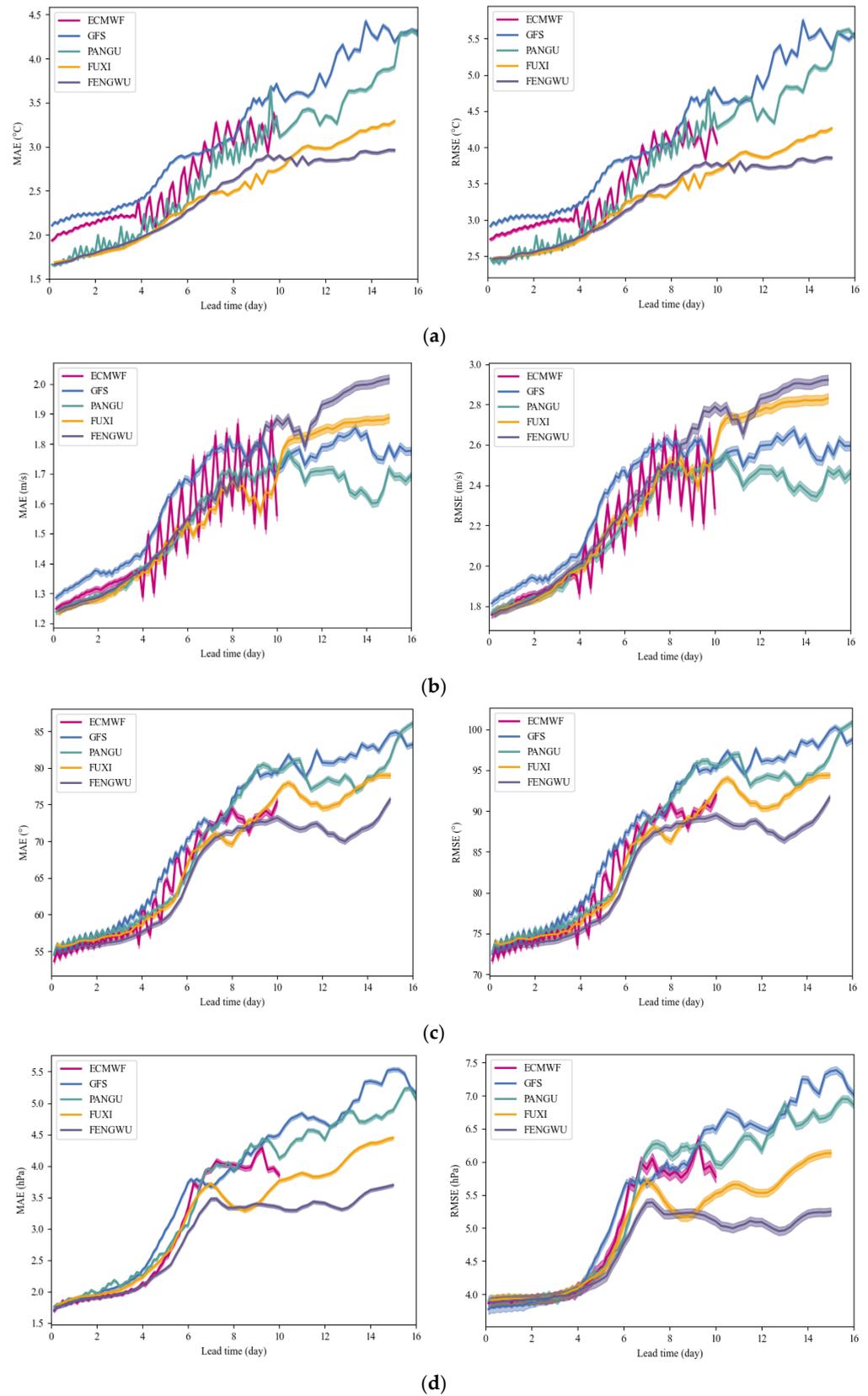
### 3.2. Forecast Skill over Different Lead Times

In short- to medium-term weather forecasting, the selection of lead time significantly influences the reliability of the forecast results. Research indicates that Pangu, based on the ERA5 initial field, has surpassed traditional NWP models in many aspects [15]. To investigate whether this advantage is reflected in the actual forecasting of meteorological parameters, this section will analyze the forecasting performance of Pangu compared to other models at different lead times. We also hope to identify the lead time durations for which Pangu is most suitable for use.

Figure 6 presents a visualization of the 3 h to 16-day forecast MAE and RMSE (lower is better) for Pangu, ECMWF, GFS, FuXi, and FengWu. The 90% confidence intervals, calculated using the empirical bootstrap method, are represented in the figure by the semi-transparent shaded area. Upon reviewing the visualization, it becomes evident that the MAE and RMSE increase with the lead time, which is reasonable as the forecasting skills of forecasting models decrease with lead time. It is noteworthy that the performance of ECMWF forecasts fluctuates significantly after a lead time of 90 h. This is due to the reduced data availability and unevenness caused by the unavailability of “scda” data for ECMWF beyond that point. When compared with traditional NWP models, based on the data presented in Figure 6a, it is clear that Pangu consistently maintains an advantage in temperature forecasting performance across all lead times. Figure 6b demonstrates that Pangu significantly outperforms GFS and achieves performance comparable to ECMWF in wind speed forecasting. Figure 6c,d reveal that the three models exhibit nearly identical forecast skills for wind and pressure prediction. When compared with other AI-based models, it can be concluded that FuXi and FengWu offer further improvements in the forecasting accuracy of surface parameters over Pangu. In most cases, their accuracy exceeds that of GFS.

Moreover, it is easy to see that Pangu, together with FuXi and FengWu, tends to outperform the traditional NWP models with a greater lead time. This may be attributed to the fact that the AI-based models’ forecasts lack the fidelity and physical consistency of physics-based models, a characteristic mentioned in previous research by Bonavita [23].

Table 2 presents the mean MAE and RMSE of Pangu, ECMWF, GFS, FuXi, and FengWu at lead times of 3 h–10 days with 90% confidence intervals. To ensure a fair comparison, the statistical range for lead time is limited to the maximum 10 days available from ECMWF. It indicates that Pangu’s mean RMSE for temperature forecasting is 3.24 °C, which is 0.03 °C lower than ECMWF and 0.41 °C lower than GFS. Additionally, Pangu’s mean RMSE for wind speed, wind direction, and pressure forecasting are 2.13 m/s, 82.51°, and 4.88 hPa, respectively. These values are 0.13 m/s, 1.01°, and 0.02 hPa lower than GFS but 0.06 m/s, 2.82°, and 0.30 hPa higher than ECMWF. In conclusion, Pangu demonstrates superior performance across all meteorological parameters in short- to medium-term weather forecasts compared to GFS, while its performance for meteorological parameters other than temperature falls slightly behind ECMWF. In other words, despite its significantly lower computational and time costs compared to traditional NWP models, Pangu can offer meteorological forecasts comparable to, or even surpassing, those of traditional NWP models. To achieve even higher accuracy in surface parameter forecasts, the FuXi and FengWu models, developed after Pangu, are also excellent choices, as demonstrated in the table.



**Figure 6.** The 3 h to 16-day MAE (left) and RMSE (right) of temperature (a), wind speed (b), wind direction (c), and pressure (d) forecasted by Pangu, ECMWF, GFS, FuXi, and FengWu, along with their 90% confidence intervals.

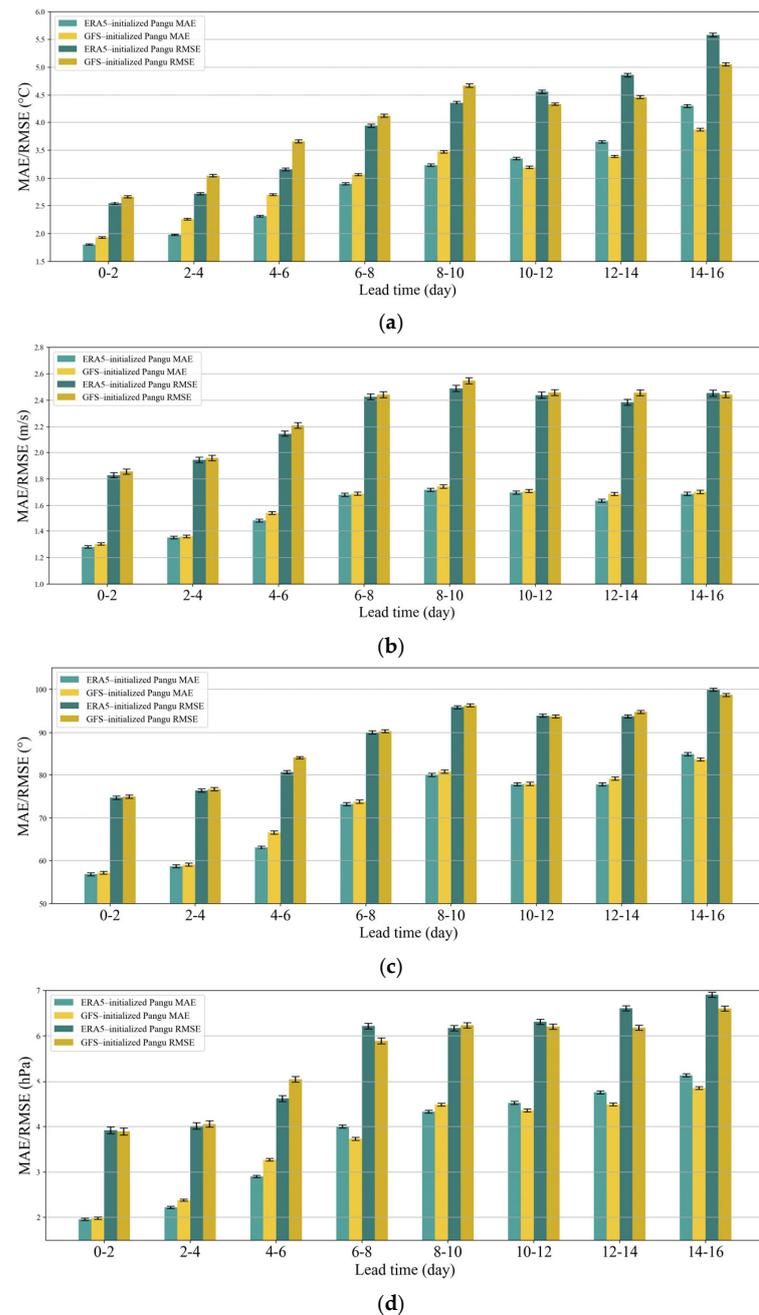
**Table 2.** Average MAE and RMSE of Pangu, ECMWF, GFS, FuXi, and FengWu at lead times of 3 h to 16 days with 90% confidence intervals (CIs). CI lower represents the lower bound of the confidence interval, while CI upper represents the upper bound.

Parameter	Model	MAE	MAE CI Lower	MAE CI Upper	RMSE	RMSE CI Lower	RMSE CI Upper
Temperature (°C)	Pangu	2.36	2.34	2.37	3.24	3.22	3.27
	ECMWF	2.44	2.42	2.45	3.27	3.24	3.30
	GFS	2.74	2.72	2.76	3.65	3.62	3.67
	FuXi	2.15	2.17	2.18	2.98	3.00	3.02
	FengWu	2.19	2.21	2.22	3.01	3.03	3.06
Wind speed (m/s)	Pangu	1.48	1.47	1.49	2.13	2.11	2.15
	ECMWF	1.45	1.44	1.46	2.07	2.05	2.10
	GFS	1.57	1.55	1.58	2.26	2.24	2.28
	FuXi	1.44	1.45	1.47	2.12	2.14	2.17
	FengWu	1.48	1.50	1.51	2.17	2.19	2.22
Wind direction (°)	Pangu	65.26	64.90	65.63	82.51	82.14	82.89
	ECMWF	62.17	61.73	62.61	79.69	79.23	80.15
	GFS	66.32	65.95	66.68	83.52	83.15	83.89
	FuXi	63.44	63.80	64.18	80.72	81.10	81.49
	FengWu	62.48	62.85	63.22	79.85	80.23	80.62
Pressure (hPa)	Pangu	2.97	2.94	2.99	4.88	4.82	4.95
	ECMWF	2.62	2.58	2.65	4.58	4.50	4.66
	GFS	3.01	2.98	3.04	4.90	4.84	4.97
	FuXi	2.71	2.73	2.76	4.57	4.64	4.71
	FengWu	2.57	2.60	2.63	4.45	4.52	4.59

### 3.3. Forecast Skill over Different Initial Fields

Trained on 39 years of global ERA5 reanalysis data at a spatial resolution of 0.25°, Pangu can obtain reliable deterministic forecast results on reanalysis data in all tested variables. However, it is currently unknown whether using the high-precision ERA5 reanalysis data as the initial field for Pangu will naturally provide Pangu with an advantage. Given this context, this section will compare the forecast performance of the Pangu model using ERA5 reanalysis data and GFS analysis data as the initial field, respectively.

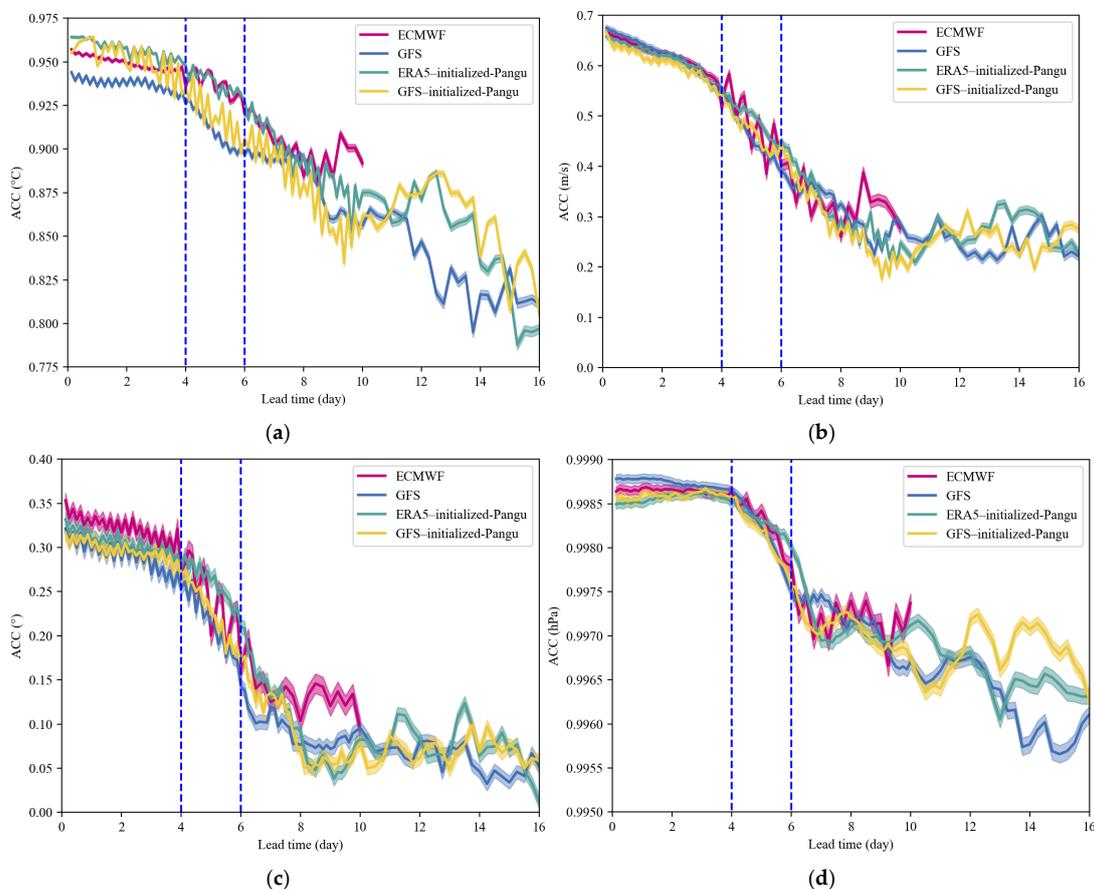
Figure 7 presents the MAE and RMSE of Pangu using both ERA5 and GFS as the initial field. Generally, the MAE and RMSE for most meteorological parameters exhibit a linear increase with lead time within nearly a week, after which the precision loss for all parameters begins to slow. Notably, the choice of initial field has a greater impact on temperature compared to other meteorological parameters. For lead times within a week, the GFS-initialized Pangu demonstrates noticeably poorer forecasting ability for temperature compared to the ERA5-initialized Pangu. However, as lead time increases, the advantage of the ERA5-initialized Pangu diminishes and may even fall behind the GFS-initialized Pangu. This does not imply that the ERA5-initialized Pangu performs worse with longer lead times, as the Pangu model typically exhibits higher uncertainty in medium-range forecasts. When it comes to other meteorological parameters, the impact of the initial field on Pangu's performance is less significant. It is worth mentioning that for short-term forecasts, particularly within the 4–6-day range, the ERA5-initialized Pangu consistently exhibits a slight advantage, a point that will be elaborated upon in the subsequent ACC score analysis.



**Figure 7.** MAE and RMSE of temperature (a), wind speed (b), wind direction (c), and pressure (d) forecasted by Pangu using ERA5 and GFS as the initial field, with the 90% confidence intervals displayed as black vertical error bars.

ACC scores (higher is better) for each model (ERA5-initialized Pangu, GFS-initialized Pangu, ECMWF, GFS) and their 90% confidence intervals are shown in Figure 8 for different lead times to illustrate aspects of the short- to medium-term weather forecast skill of Pangu with different initial fields. As the lead time increases, the ACC gradually decreases for all tested variables. Across all the subplots, it is evident that while the ACC of traditional NWP models (GFS) continues to fluctuate downward beyond a week's lead time, the ACC of the AI-based model (Pangu) exhibits irregular fluctuations and even rises. This visually reinforces the notion that the lack of fidelity and physical consistency inherent in AI-based models, often considered a drawback, might actually become an advantage in medium-range forecasting. Overall, except for the 4–6-day lead time range, where the ERA5-initialized Pangu consistently outperforms the GFS-initialized Pangu for all

meteorological parameters, the ACC scores between the two initializations show little difference in other lead time intervals. Additionally, for temperature forecasting, the GFS-initialized Pangu exhibits greater volatility in its ACC scores, suggesting that using the ERA5-initialized Pangu may be a more stable choice.



**Figure 8.** ACC with 90% confidence intervals of temperature (a), wind speed (b), wind direction (c), and pressure (d) forecasted by ERA5-initialized Pangu, GFS-initialized Pangu, ECMWF, and GFS, respectively.

Table 3 displays the MAE and RMSE of the GFS-initialized Pangu. It can be seen that the mean RMSE of temperature, wind speed, wind direction, and pressure forecasts using GFS as the initial field are 3.53 °C, 2.16 m/s, 83.46°, and 4.94 hPa, respectively. These values are respectively higher by 0.29 °C, 0.03 m/s, 0.95°, and 0.06 hPa compared to the ERA5-initialized Pangu. This further indicates that the GFS-initialized Pangu is slightly inferior in precision compared to the ERA5-initialized Pangu.

**Table 3.** MAE and RMSE of GFS-initialized-Pangu at lead times of 3 h to 16 days with 90% confidence intervals.

Parameter	MAE	MAE CI Lower	MAE CI Upper	RMSE	RMSE CI Lower	RMSE CI Upper
Temperature (°C)	2.60	2.59	2.62	3.53	3.50	3.55
Wind speed (m/s)	1.50	1.49	1.51	2.16	2.14	2.18
Wind direction (°)	66.37	66.00	66.73	83.46	83.09	83.84
Pressure (hPa)	3.06	3.03	3.09	4.94	4.87	5.00

### 3.4. Iteration Precision of Pangu's Forecast Models

Pangu offers four types of forecast models with lead times of 1 h, 3 h, 6 h, and 24 h. To reduce the cumulative forecast errors, Pangu uses hierarchical temporal aggregation, a greedy algorithm that always calls for the deep network with the largest affordable lead time, which greatly reduces the number of iterations. For example, when the lead time is 56 hours, it would execute the 24 h forecast model two times, the 6 h forecast model one time, and the 1 h forecast model two times. However, this algorithm can lead to discontinuities in forecasts if the iteration precision of models with different lead times varies significantly. Therefore, the iteration precision of Pangu's four forecast models will be examined.

Figure 9 displays the MAE and RMSE of Pangu's forecasts with 90% confidence intervals on 30 October at 12:00 UTC, utilizing 1 h, 3 h, 6 h, and 24 h forecast models. It clearly indicates that for all meteorological parameters, the iteration precision of the 1 h forecast model is significantly lower than that of the other models. For temperature, wind speed, and pressure (Figure 9a,b,d), the disadvantages of the 1 h model in terms of precision and stability are particularly pronounced, while the precision differences among the 3 h, 6 h, and 24 h models are minimal. In the case of wind direction (Figure 9c), the precision of all four models shows greater variability, which may be attributed to the inherent uncertainty of the wind direction itself.

Table 4 displays the average hourly cumulative MAE and RMSE of Pangu's four forecast models on 30 October at 12:00 UTC. It can be seen from the table that the hourly cumulative RMSE of the 1 h forecast model is nearly double that of the other models for temperature and wind speed, and the precision gap is even more pronounced for pressure. Although the precision difference for wind direction is relatively smaller, there remains a gap between the 1 h forecast model and the other models. Among all the forecast models, the 24 h forecast model should be the most accurate; however, this advantage is not clearly evident in the results. Despite the significant discrepancy in hourly cumulative precision between the 1 h forecast model and the others, this gap may not manifest noticeably in Pangu's actual forecasting process due to the greedy algorithm employed. Since the hourly cumulative errors of the 3 h, 6 h, and 24 h models are similar and the 1 h forecast model will be used at most twice in a single forecast, we can approximate that the way Pangu utilizes these four models does not lead to discontinuities in forecasts.

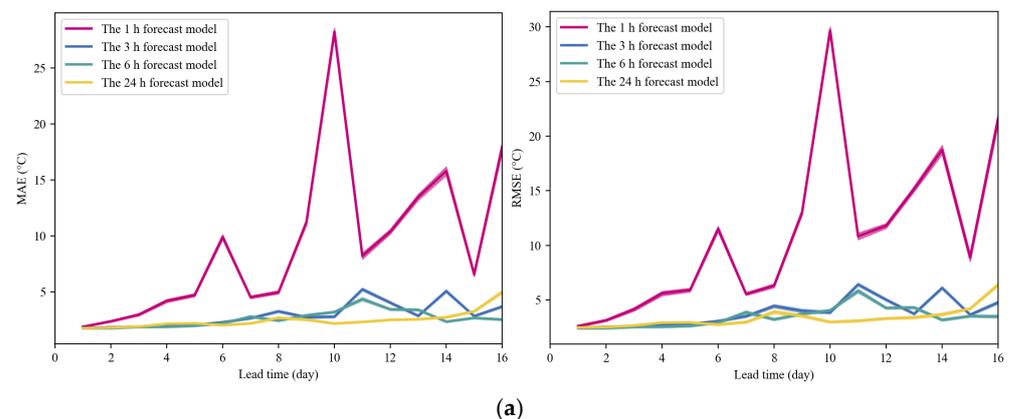
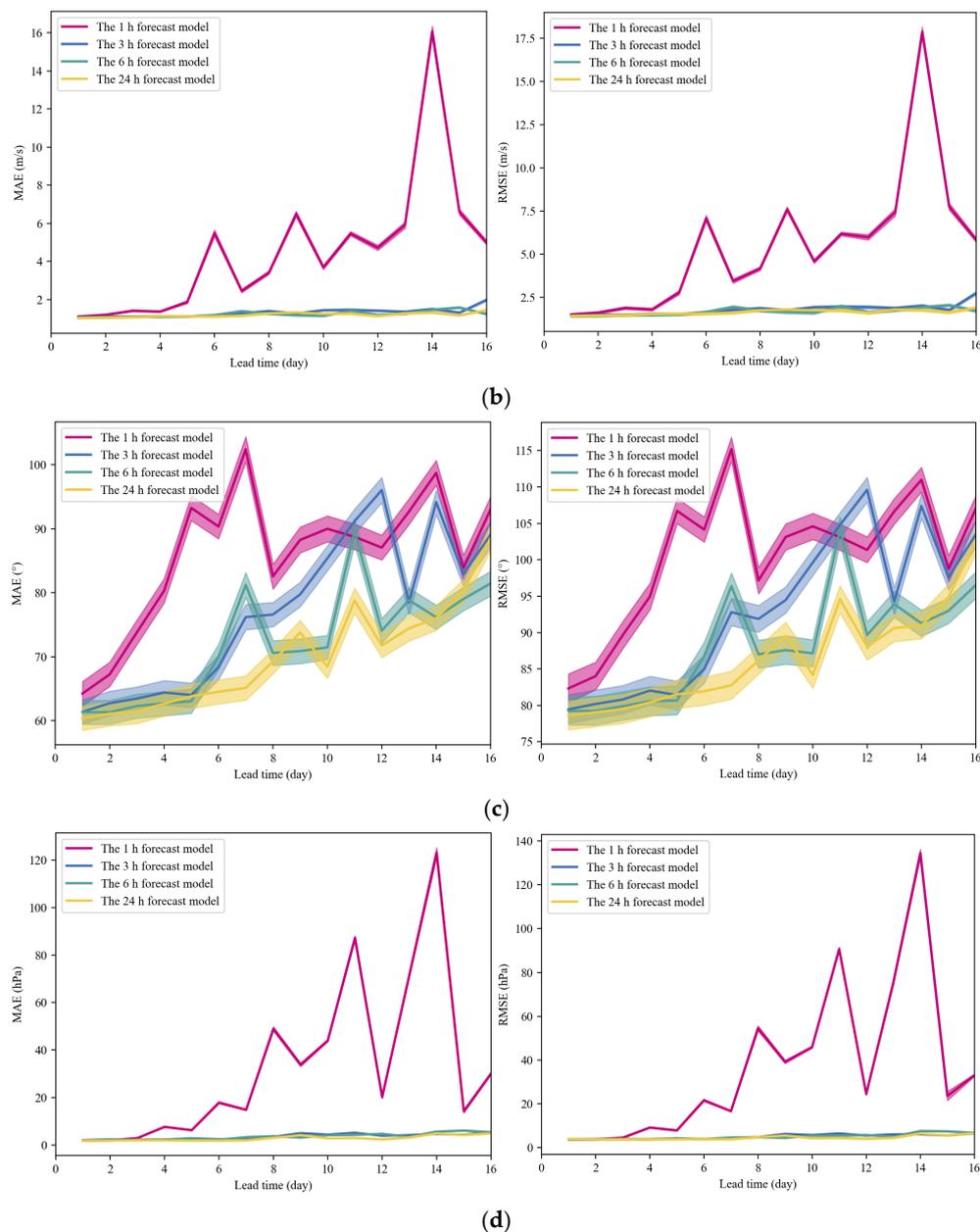


Figure 9. Cont.



**Figure 9.** MAE and RMSE of Pangu’s forecasts for temperature (a), wind speed (b), wind direction (c), and pressure (d), with 90% confidence intervals on 30 October at 12:00 UTC, using 1 h, 3 h, 6 h, and 24 h forecast models.

**Table 4.** Average hourly cumulative MAE and RMSE for Pangu forecast models (1 h, 3 h, 6 h, 24 h) on 30 October at 12:00 UTC.

Parameter	The 1 h Forecast Model		The 3 h Forecast Model		The 6 h Forecast Model		The 24 h Forecast Model	
	MAE	RMS	MAE	RMS	MAE	RMS	MAE	RMS
Temperature (°C)	0.048	0.058	0.020	0.027	0.019	0.025	0.019	0.025
Wind speed (m/s)	0.023	0.029	0.010	0.014	0.010	0.014	0.010	0.014
Wind direction (°)	0.679	0.819	0.605	0.753	0.584	0.735	0.570	0.723
Pressure (hPa)	0.139	0.162	0.025	0.039	0.024	0.038	0.020	0.037

### 3.5. Case Study: Tracking the Path of Typhoon Bebinca

To evaluate Pangu’s forecasting performance in extreme weather scenarios, it is essential to conduct comparative evaluations of forecast results based on representative cases.

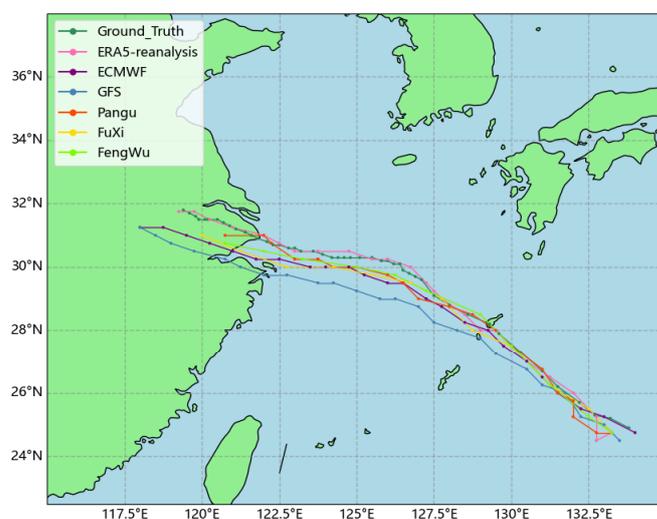
In this study, Typhoon Bebinca (ID: 2413) was selected as the representative case. Bebinca formed over the South China Sea in September 2024 and was characterized by its erratic path and significant intensity variations. The storm brought widespread heavy rainfall and strong winds to Southern China and neighboring regions, highlighting the critical need for accurate forecasting during such extreme weather events.

The identification of the eye of tropical cyclones follows the relevant criteria published by ECMWF [40], with the primary rules summarized as follows.

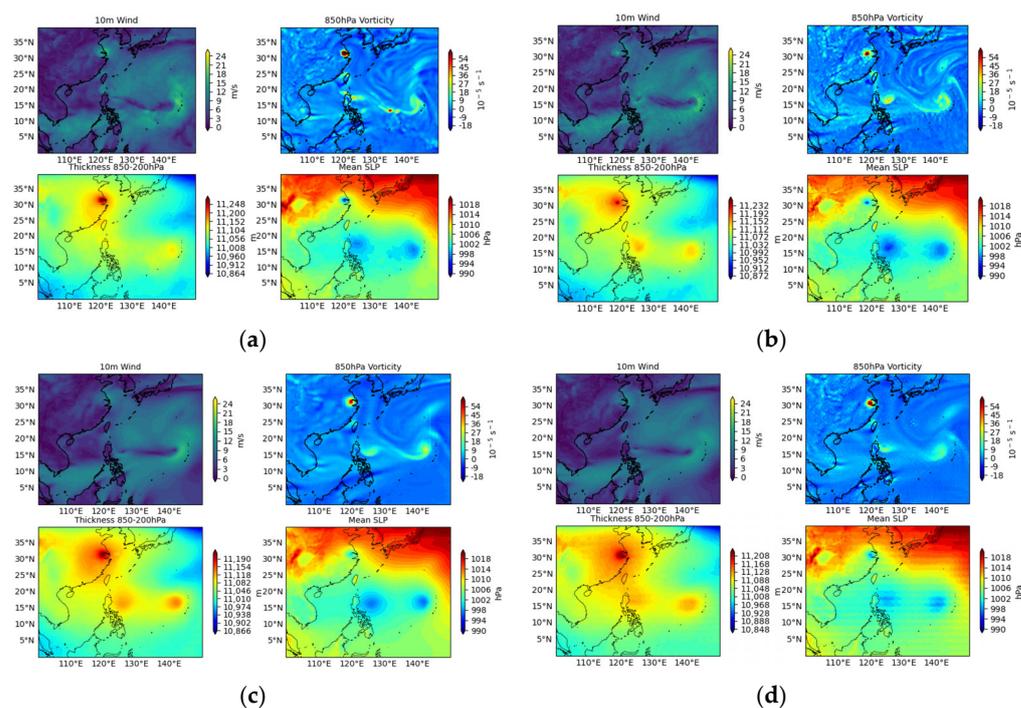
Locate the local minimum of mean sea level pressure (MSLP) that is below 1015 hPa, and ensure the following conditions are met:

1. There is a maximum relative vorticity at 850 hPa exceeding  $5 \times 10^{-5}$  within a radius of 278 km for the Northern Hemisphere, or a minimum smaller than  $-5 \times 10^{-5}$  for the Southern Hemisphere;
2. There is a maximum thickness between 850 hPa and 200 hPa within a radius of 278 km if the cyclone is extratropical;
3. The maximum 10 m wind speed exceeds 8 m/s within a radius of 278 km if the cyclone is over land.

Figure 10 presents the track forecasts for Typhoon Bebinca generated by Pangu, ECMWF, GFS, FuXi, and FengWu. These forecasts were initialized with the ERA5 reanalysis data from 13 September 2024 at 12:00 UTC. The comparison includes ground truth data published by the China Meteorological Administration and ERA5 reanalysis data as reference points. Figure 11 illustrates the intermediate results for tracking Typhoon Bebinca at 06:00 UTC on 16 September, focusing on key variables such as 10 m wind, 850 hPa vorticity, thickness between 850–200 hPa, and mean sea level pressure (MSLP). From the figures, it is evident that both Pangu and traditional NWP models exhibit variations in their predicted tracks, with none achieving perfect alignment with the ground truth. The AI-based models Pangu, FengWu, and FuXi produced typhoon track forecasts that were relatively close to ECMWF's predictions. However, GFS demonstrated significant deviations from the ground truth in its predicted typhoon tracks, particularly underperforming in the case of typhoon Bebinca. After the typhoon made landfall, Pangu, along with FuXi and FengWu, exhibited inaccuracies in predicting meteorological parameters such as the u and v components of 10 m wind speed and mean sea level pressure (MSLP), as reflected in Figure 11. These inaccuracies led to the premature weakening and dissipation of the predicted typhoon track.



**Figure 10.** Track forecast for typhoon Bebinca, based on initial conditions from 13 September 2024 at 12:00 UTC.



**Figure 11.** Intermediate results for tracking Typhoon Bebinca on 16 September 2024 at 06:00 UTC: (a) ERA5 reanalysis data, (b) ECMWF forecast, (c) Pangu forecast, (d) FuXi forecast.

#### 4. Conclusions

This study examined the performance of Pangu in predicting meteorological parameters including temperature, wind speed, wind direction, and barometric pressure based on the observed data from over 2000 weather stations in China. Pangu was primarily compared with ECMWF-HRES and GFS to assess whether it could surpass these high-precision traditional NWP models under real-world meteorological conditions. Additionally, the more recent FuXi and FengWu models, developed after Pangu, were included to verify Pangu's accuracy. Finally, a case study on typhoon track forecasting was conducted to evaluate Pangu's performance in predicting typhoon paths. The primary conclusions are as follows.

The overall distribution of Pangu's meteorological forecasts is very close to that of ECMWF and GFS. According to the precision statistics, the mean RMSE of Pangu in forecasting temperature is 3.24 °C, which is 0.03 °C lower than ECMWF and 0.41 °C lower than GFS. The mean RMSE of Pangu in forecasting wind speed, wind direction, and pressure are 2.13 m/s, 82.51°, and 4.88 hPa, respectively, which are 0.13 m/s, 1.01°, and 0.02 hPa lower than GFS but 0.06 m/s, 2.82°, and 0.30 hPa higher than ECMWF. It can be concluded that the temperature forecast performance of Pangu exceeds that of the traditional NWP models, while its performance in other meteorological parameters is comparable to the traditional NWP models.

The forecast skill of Pangu decays as the lead time increases, but Pangu tends to outperform the traditional NWP models with a greater lead time. Additionally, FuXi and FengWu further improve the forecasting accuracy of surface parameters compared to Pangu. The performance of Pangu also depends on initial fields to some extent, especially for temperature forecasting. For lead times within a week, the GFS-initialized Pangu demonstrates noticeably poorer forecasting ability for temperature compared to the ERA5-initialized Pangu, and the situation reverses for lead times exceeding one week. When considering the iteration precision of Pangu's four forecast models with lead times of 1 h, 3 h, 6 h, and 24 h, the iteration precision of the 1 h forecast model is significantly lower

than that of the other models. In the case study on typhoon forecasting, Pangu, along with FuXi and FengWu, demonstrates comparable performance in predicting Bebinca's track compared to ECMWF and outperforms GFS in its track predictions.

Despite the aforementioned work, this study has several limitations. It solely examined and tested Pangu's forecasting performance over surface variables, without considering upper variables. Regarding the types of forecasted parameters, Pangu does not provide predictions for water vapor or precipitation, even though the abundance and distribution of water vapor play a crucial role in the formation and distribution of rainfall, which are vital for predicting weather changes. Notably, other AI-based weather models, such as FuXi, have begun to provide forecasts for total precipitation. Therefore, evaluating the accuracy of precipitation predictions from AI models is an area worthy of further exploration. Moreover, the research primarily concentrated on comparing Pangu with traditional NWP models in short- to medium-term forecasting, while overlooking Pangu's capabilities in medium- to long-term forecasting, and even in subseasonal-to-seasonal (S2S) forecasting. Additionally, the performance of Pangu in extreme weather forecasting and ensemble forecasting still requires assessment.

Once trained, data-driven models like Pangu are orders of magnitude faster than traditional NWP models in generating forecasts, thus enabling large ensembles of thousands of members to be generated in seconds. For instance, Pangu requires only 1.4 s to complete a 24 h global weather forecast on a single NVIDIA V100 GPU. This implies that Pangu can accomplish 10,000 ensemble global forecasts with a 24 h lead time in just about 4 h, significantly improving the efficiency of weather forecasting. Although the forecast skill of Pangu has outperformed the NWP models for some key variables, the algorithm of Pangu is not combined with actual physical laws, and it still relies on analysis data generated by traditional NWP models for initial conditions. It is believed that the data-driven models have the potential to update their algorithm and iterate themselves in the near future and thus continue to play a more important role in the field of weather forecasting.

**Author Contributions:** Conceptualization, Y.Z. (Yize Zhang); methodology, Y.Z. (Yize Zhang); formal analysis, S.X.; data curation, S.X.; writing—original draft preparation, S.X.; writing—review and editing, Y.Z. (Yize Zhang) and J.C.; visualization, S.X. funding acquisition, Y.Z. (Yize Zhang) and Y.Z. (Yunlong Zhang). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China, (Grant Nos. 12403075 and 42474034), the Science and Technology Research and Development Project of China State Railway Group Co., Ltd., (Grant No. Q2023T004), and the Tianjin Key Laboratory of Rail Transit Navigation Positioning and Spatio-temporal Big Data Technology (Grant No. TKL2024A01).

**Data Availability Statement:** All the datasets used in this study are publicly available. The data sources are as follows: (1) The trained models and inference code of Pangu-Weather, available at <https://github.com/198808xc/Pangu-Weather> (accessed on 7 June 2024); (2) the observation data from weather stations, available at <http://www.nmc.cn> (accessed on 7 June 2024); (3) the ECMWF-HRES forecasts, available at <https://confluence.ecmwf.int/display/DAC/ECMWF+open+data:+real-time+forecasts+from+IFS+and+AIFS> (accessed on 7 June 2024); (4) the GFS forecasts and analysis data, available at <https://rda.ucar.edu/datasets/d084006> (accessed on 7 June 2024); (5) the ERA5 reanalysis initial condition data, available at <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form> (accessed on 7 June 2024); and (6) the verification code, available at <https://github.com/Clarmy/pangu-weather-verify/commits?author=Clarmy> (accessed on 7 June 2024).

**Acknowledgments:** We express our sincere gratitude to meteorological technician Wentao Li for providing a portion of the open-source verification code used in this study.

**Conflicts of Interest:** Author Yunlong Zhang was employed by the company China Railway Design Corporation. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Bauer, P.; Thorpe, A.; Brunet, G. The quiet revolution of numerical weather prediction. *Nature* **2015**, *525*, 47–55. [[CrossRef](#)] [[PubMed](#)]
2. Geer, A.J.; Lonitz, K.; Weston, P.; Kazumori, M.; Okamoto, K.; Zhu, Y.; Liu, E.H.; Collard, A.; Bell, W.; Migliorini, S.; et al. All-sky satellite data assimilation at operational weather forecasting centres. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 1191–1217. [[CrossRef](#)]
3. Ben Bouallègue, Z.; Magnusson, L.; Haiden, T.; Richardson, D.S. Monitoring trends in ensemble forecast performance focusing on surface variables and high-impact events. *Q. J. R. Meteorol. Soc.* **2019**, *145*, 1741–1755. [[CrossRef](#)]
4. Ben Bouallègue, Z.; Clare, M.C.; Magnusson, L.; Gascon, E.; Maier-Gerber, M.; Janoušek, M.; Rodwell, M.; Pinault, F.; Dramsch, J.S.; Lang, S.T. The rise of data-driven weather forecasting: A first statistical assessment of machine learning-based weather forecasts in an operational-like context. *Bull. Am. Meteorol. Soc.* **2024**, *105*, E864–E883. [[CrossRef](#)]
5. Dueben, P.D.; Bauer, P. Challenges and design choices for global weather and climate models based on machine learning. *Geosci. Model Dev.* **2018**, *11*, 3999–4009. [[CrossRef](#)]
6. Weyn, J.A.; Durran, D.R.; Caruana, R. Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height From Historical Weather Data. *J. Adv. Model. Earth Syst.* **2019**, *11*, 2680–2693. [[CrossRef](#)]
7. Wang, J.; Wang, X.; Guan, J.; Zhang, L.; Zhang, F.; Chang, T. STPF-Net: Short-Term Precipitation Forecast Based on a Recurrent Neural Network. *Remote Sens.* **2023**, *16*, 52. [[CrossRef](#)]
8. Liang, Z.; Lee, Y.-K.; Grassotti, C.; Lin, L.; Liu, Q. Machine Learning-Based Estimation of Tropical Cyclone Intensity from Advanced Technology Microwave Sounder Using a U-Net Algorithm. *Remote Sens.* **2023**, *16*, 77. [[CrossRef](#)]
9. Rasp, S.; Lerch, S. Neural Networks for Postprocessing Ensemble Weather Forecasts. *Mon. Weather Rev.* **2018**, *146*, 3885–3900. [[CrossRef](#)]
10. Han, L.; Chen, M.; Chen, K.; Chen, H.; Zhang, Y.; Lu, B.; Song, L.; Qin, R. A deep learning method for bias correction of ECMWF 24–240 h forecasts. *Adv. Atmos. Sci.* **2021**, *38*, 1444–1459. [[CrossRef](#)]
11. Weyn, J.A.; Durran, D.R.; Caruana, R. Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere. *J. Adv. Model. Earth Syst.* **2020**, *12*, e2020MS002109. [[CrossRef](#)]
12. Rasp, S.; Thuerey, N. Data-Driven Medium-Range Weather Prediction With a Resnet Pretrained on Climate Simulations: A New Model for WeatherBench. *J. Adv. Model. Earth Syst.* **2021**, *13*, e2020MS002405. [[CrossRef](#)]
13. RDMKeisler, R. Forecasting global weather with graph neural networks. *arXiv* **2022**, arXiv:2202.07575. [[CrossRef](#)]
14. Pathak, J.; Subramanian, S.; Harrington, P.; Raja, S.; Chattopadhyay, A.; Mardani, M.; Kurth, T.; Hall, D.; Li, Z.; Azizzadenesheli, K. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv* **2022**, arXiv:2202.11214. [[CrossRef](#)]
15. 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](#)]
16. 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](#)]
17. Chen, L.; Du, F.; Hu, Y.; Wang, Z.; Wang, F. Swinrdm: Integrate swinrrn with diffusion model towards high-resolution and high-quality weather forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023.
18. Chen, L.; Zhong, X.; Zhang, F.; Cheng, Y.; Xu, Y.; Qi, Y.; Li, H. FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. *NPJ Clim. Atmos. Sci.* **2023**, *6*, 190. [[CrossRef](#)]
19. Chen, K.; Han, T.; Gong, J.; Bai, L.; Ling, F.; Luo, J.-J.; Chen, X.; Ma, L.; Zhang, T.; Su, R. Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. *arXiv* **2023**, arXiv:2304.02948. [[CrossRef](#)]
20. Nguyen, T.; Brandstetter, J.; Kapoor, A.; Gupta, J.K.; Grover, A. Climax: A foundation model for weather and climate. *arXiv* **2023**, arXiv:2301.10343. [[CrossRef](#)]
21. 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](#)]
22. Ebert-Uphoff, I.; Hilburn, K. The outlook for AI weather prediction. *Nature* **2023**, *619*, 473–474. [[CrossRef](#)] [[PubMed](#)]
23. Bonavita, M. On some limitations of data-driven weather forecasting models. *arXiv* **2023**, arXiv:2309.08473. [[CrossRef](#)]
24. Cheng, W.; Yan, Y.; Xia, J.; Liu, Q.; Qu, C.; Wang, Z. The Compatibility between the Pangu Weather Forecasting Model and Meteorological Operational Data. *arXiv* **2023**, arXiv:2308.04460. [[CrossRef](#)]
25. Hakim, G.J.; Masanam, S. Dynamical tests of a deep-learning weather prediction model. *arXiv* **2023**, arXiv:2309.10867. [[CrossRef](#)]

26. Xu, H.; Zhao, Y.; Zhao, D.; Duan, Y.; Xu, X. Improvement of disastrous extreme precipitation forecasting in North China by Pangu-weather AI-driven regional WRF model. *Environ. Res. Lett.* **2024**, *19*, 054051. [[CrossRef](#)]
27. Ismail, A.H.; Dawi, E.; Almokdad, N.; Abdelkader, A.; Salem, O. Estimation and Comparison of the Clearness Index using Mathematical Models-Case study in the United Arab Emirates. *Evergreen* **2023**, *10*, 863–869. [[CrossRef](#)]
28. Yahiaoui, S.; Assas, O. Comparison of solar radiation models using meteorological parameters. *Energy Syst.* **2024**, *15*, 863–897. [[CrossRef](#)]
29. Li, H.; Zhu, G.; Huang, L.; Mo, Z.; Kang, Q. An improved method for developing the Precipitable Water Vapor vertical correction global grid Model. *Atmos. Res.* **2024**, *311*, 107664. [[CrossRef](#)]
30. Yao, Y.; Shan, L.; Zhao, Q. Establishing a method of short-term rainfall forecasting based on GNSS-derived PWV and its application. *Sci. Rep.* **2017**, *7*, 12465. [[CrossRef](#)] [[PubMed](#)]
31. Owens, R.G.; Hewson, T.D. *ECMWF Forecast User Guide*; ECMWF: Reading, UK, 2018. [[CrossRef](#)]
32. National Centers for Environmental Prediction; National Weather Service; NOAA; U.S. Department of Commerce. NCEP Global Forecast System (GFS) Analyses and Forecasts. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. 2007. Available online: <https://rda.ucar.edu/datasets/d084006/> (accessed on 7 June 2024).
33. Zhang, W.; Jiang, Y.; Dong, J.; Song, X.; Pang, R.; Guoan, B.; Yu, H. A deep learning method for real-time bias correction of wind field forecasts in the Western North Pacific. *Atmos. Res.* **2023**, *284*, 106586. [[CrossRef](#)]
34. Lente, G.; Ósz, K. Barometric formulas: Various derivations and comparisons to environmentally relevant observations. *ChemTexts* **2020**, *6*, 13. [[CrossRef](#)]
35. Liu, Y.; Wu, Q.; Zhang, Y.; Jiang, L. Evaluating precipitation prediction skill for the pre- and post-rainy seasons in South China in ECMWF subseasonal forecasts. *Geosci. Lett.* **2024**, *11*, 11. [[CrossRef](#)]
36. Efron, B. Computers and the theory of statistics: Thinking the unthinkable. *SIAM Rev.* **1979**, *21*, 460–480. Available online: <https://www.jstor.org/stable/2030104> (accessed on 1 January 2025). [[CrossRef](#)]
37. Gentle, J.E. *Elements of Computational Statistics*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2002.
38. Chen, Y.-C. *STAT/Q SCI 403: Introduction to Resampling Methods*; University of the Sunshine Coast: Sunshine Coast, QLD, Australia, 2017.
39. Alan Thorpe, F.P. Machine Learning and Physics in Weather Forecasting: A Discussion Between Alan Thorpe and Florian Pappenberger. 17 June 2024. Available online: <https://www.ecmwf.int/en/about/media-centre/focus/2024/machine-learning-and-physics-weather-forecasting-discussion-0> (accessed on 1 January 2025).
40. White, P. Newsletter No. 102—Winter 2004/05, European Centre for Medium Range Weather Forecasts. 2005. Available online: <https://www.ecmwf.int/node/14623> (accessed on 1 January 2025).

**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.