water

MDPI

Article

Multi-Scale Evaluation of ERA5 Air Temperature and Precipitation Data over the Poyang Lake Basin of China

Xie Yan [1,2], Meng Zhang [3], Fangxu Yin [3], Jiewen You [4,5], Ying Chen [4,5] and Lu Gao [4,5,*]

[1] College of Resources and Environment, Fujian Agriculture and Forestry University, Fuzhou 350002, China; yanxie1982@126.com
[2] Fujian Provincial Key Laboratory of Soil Environmental Health and Regulation, Fuzhou 350002, China
[3] School of Geographical Sciences, Fujian Normal University, Fuzhou 350117, China; 1255756735@foxmail.com (M.Z.); yfx1998@outlook.com (F.Y.)
[4] Institute of Geography, Fujian Normal University, Fuzhou 350117, China; jiewen.you@hotmail.com (J.Y.); chenying_nju@163.com (Y.C.)
[5] Fujian Provincial Engineering Research Center for Monitoring and Assessing Terrestrial Disasters, Fujian Normal University, Fuzhou 350117, China
* Correspondence: l.gao@foxmail.com; Tel.: +86-591-8346-5214

Abstract: Reanalysis datasets, such as ERA5, are essential for climate research, offering comprehensive spatiotemporal coverage. However, their accuracy needs thorough evaluation for effective regional application, particularly in areas with complex topography like the Poyang Lake Basin (PLB), China's largest freshwater lake. This study evaluated ERA5's accuracy in simulating near-surface air temperature and precipitation in the PLB, using data from 24 meteorological stations. Key metrics, such as the correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE) were applied across daily, monthly, seasonal, and annual scales. The results show that ERA5 performs well for daily mean temperature, with daily R values above 0.98 and RMSEs ranging from 0.95 °C to 3.11 °C. Its highest accuracy was in February and March, with R values exceeding 0.95, and seasonal trends were best captured in spring and autumn (R > 0.99). However, ERA5's performance for precipitation was less accurate, with daily R values between 0.578 and 0.687 and RMSEs between 8.58 mm and 11.10 mm. ERA5 consistently overestimated precipitation, particularly during 1980–2003. These findings highlight ERA5's strengths in temperature modeling and its limitations in precipitation, providing insights for identifying climate events and improving climate simulation in the PLB.

Keywords: Poyang Lake Basin; ERA5 reanalysis; multi-scale evaluation; air temperature; precipitation

check for updates

Citation: Yan, X.; Zhang, M.; Yin, F.; You, J.; Chen, Y.; Gao, L. Multi-Scale Evaluation of ERA5 Air Temperature and Precipitation Data over the Poyang Lake Basin of China. *Water* **2024**, *16*, 3123. https://doi.org/10.3390/w16213123

Academic Editor: Christos S. Akratos

Received: 2 October 2024
Revised: 29 October 2024
Accepted: 30 October 2024
Published: 1 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Poyang Lake Basin (PLB) (113.74–118.47° E, 24.57–30.01° N), located in central Southeast China, spans approximately 162,200 km^2. Poyang Lake itself covers about 3960 km^2, making it the largest freshwater lake in China [1,2]. As a vital flood storage area for the Yangtze River, the basin is densely populated, economically significant, rich in natural resources, and a key biodiversity conservation zone [3–5]. The subtropical monsoon climate leads to significant seasonal and annual variability in temperature and precipitation in the PLB. In recent decades, the PLB has become increasingly vulnerable to climate change, facing more frequent climate extremes, such as droughts and floods, which pose serious challenges to water resources and ecological stability [6–9]. Due to its sensitivity to hydrological cycles, there are growing concerns about temperature and precipitation patterns under a warming climate in the basin [10–12].

Air temperature is a fundamental indicator of climate change and a key metric in understanding regional warming over the PLB in China. Temperature has a strong influence on local and global water and energy cycles, directly shaping the hydrological balance and other crucial environmental processes [13,14]. Temperature changes are significantly

influenced by both global and regional carbon dioxide levels, as evidenced by extensive research showing a strong correlation between CO_2 concentrations and regional temperature variations. As temperatures rise, the ability of the atmosphere to hold water vapor increases. According to the Clausius–Clapeyron (C–C) scaling, the atmospheric water-holding capacity grows by approximately 7% for each degree Celsius of warming [15–17]. This can lead to more intense and variable precipitation, impacting rainfall patterns. Consequently, accurately measuring, estimating, and evaluating temperature and precipitation is critical for understanding climate change impacts and developing effective adaptation strategies to ensure the sustainability of ecosystems and human livelihoods in the PLB.

In situ observations from meteorological stations serve as the primary and most widely utilized source of data for climate change research, offering crucial insights into air temperature, precipitation patterns, and other atmospheric variables. However, in the PLB, the existing meteorological network is sparse and inadequately distributed [18]. This results in several key challenges, including short time series of available data, as well as significant gaps in both temporal and spatial coverage [19]. Such limitations reduce the capacity to track long-term climate trends accurately and to fully understand the regional variability of climate change.

In contrast, reanalysis products have emerged as valuable alternatives to traditional observations for large-scale climate assessments in recent decades. These datasets combine observations with model simulations, providing consistent and spatially complete representations of atmospheric and surface conditions [20]. Reanalysis products offer several advantages, including multivariate gridded data that are both spatially and temporally comprehensive. However, reanalysis products have disadvantages, including observational uncertainties due to changes in systems, limitations in model accuracy, and challenges in merging diverse data sources. ERA5, the most advanced reanalysis product, shows significant improvements over earlier versions [21,22]. However, previous studies indicate that ERA5 can overestimate or underestimate atmospheric variables, particularly precipitation in specific local areas [19,23,24]. Therefore, it is crucial to assess the reliability and applicability of ERA5 to the PLB to ensure accurate climate studies in the region.

While ERA5 reanalysis data have been widely evaluated for various regions across the globe [25], studies focusing on the PLB remain very limited. To our knowledge, only one study has directly evaluated ERA5 for the PLB, focusing solely on precipitation [26]. There is a lack of comprehensive, multi-scale evaluations that assess both air temperature and precipitation in the PLB. Extending the evaluations by incorporating observation stations and conducting assessments across multiple temporal scales would offer a better understanding of the PLB's climate dynamics.

The objective of this study is to evaluate the performance of ERA5 data for air temperature and precipitation in the PLB using observations from 24 meteorological stations. The evaluation is conducted across multiple temporal scales—daily, monthly, seasonal, and annual. This assessment will help determine whether ERA5 data are reliable for local climate studies in this critical region. The structure of the paper is as follows: Section 2 provides an overview of the research area, meteorological observations, ERA5 reanalysis datasets, and evaluation methods used. Section 3 presents the results, while the discussions and conclusions are outlined in Section 4.

2. Materials and Methods

2.1. Meteorological Observations in the PLB

Figure 1 shows the spatial location of the PLB and the 24 meteorological stations within it. The PLB, located in Jiangxi Province, China, has the country's largest freshwater lake and serves as a critical component of the Yangtze River's hydrological system. The PLB features complex topography, with mountainous regions surrounding a flat central plain where the lake is situated. The PLB's climate is characterized by a humid subtropical monsoon climate with hot summers and mild winters, and the region experiences significant seasonal and interannual variability in temperature and precipitation.

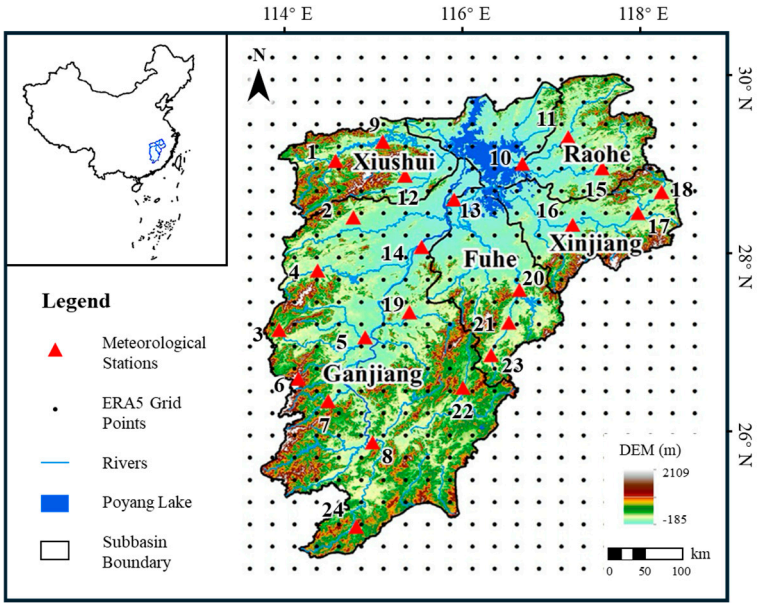


Figure 1. Distribution of meteorological stations and ERA5 grid points over the PLB.

Observational data from 24 meteorological stations in the PLB were obtained from the Daily Surface Meteorological Dataset of China (V3.0) via the China Meteorological Data Network (http://data.cma.cn/ (accessed on 1 October 2024)). The dataset underwent rigorous quality control to ensure completeness and continuity, covering daily total precipitation (mm) and mean temperature (°C) from 1980 to 2020. For seasonal analysis, the year was divided into spring (March to May), summer (June to August), autumn (September to November), and winter (December to February), following the standard seasonal classifications in China. The observational data from these meteorological stations serve as the benchmark for comparing the accuracy of the ERA5 reanalysis dataset across daily, monthly, seasonal, and annual scales.

2.2. ERA5 Reanalysis

ERA5 is the latest generation of atmospheric reanalysis products from the European Centre for Medium-Range Weather Forecasts (ECMWF). It builds upon its predecessor, ERA-Interim, and integrates the four-dimensional variational data assimilation technique within the Integrated Forecasting System (IFS Cycle 41r2) released by ECMWF in 2016. With a horizontal resolution of 0.25° × 0.25°, the ERA5 reanalysis dataset offers high spatial and temporal resolution, frequent updates, and an extensive range of climate parameters. This study utilized hourly ERA5 meteorological grid data from 1980 to 2020 (https://cds.climate.copernicus.eu/datasets/derived-era5-single-levels-daily-statistics (accessed on 1 October 2024)), focusing on two key climate variables: near-surface air temperature and precipitation. Both datasets were interpolated to match the nearest grid points corresponding to the locations of the 24 meteorological stations in the PLB. This was achieved using a nearest-neighbor interpolation approach, which identifies and assigns the values from the closest grid point to each station location without adjusting for elevation differences. This method ensures direct spatial alignment with station points, allowing for consistent comparison with ground-based observations.

2.3. Evaluation Metrics

The evaluation of ERA5 performance was conducted by comparing the ERA5 reanalysis data with observations from the ground stations. To evaluate the accuracy of the ERA5 reanalysis data, we employed three main validation metrics:

(1) Correlation Coefficient (R): This measures the strength of the linear relationship between the ERA5 data and the observations, with values closer to 1 indicating a stronger correlation.
(2) Root Mean Square Error (RMSE): This quantifies the magnitude of errors between ERA5 and the observed data, with lower values indicating better agreement. The RMSE emphasizes larger errors due to its squaring of differences, making it sensitive to extreme discrepancies.
(3) Mean Absolute Error (MAE): This represents the average absolute difference between ERA5 and the observed values, providing a straightforward measure of error magnitude. While both the RMSE and MAE reflect the accuracy of the ERA5 data, RMSE is more responsive to larger errors, whereas MAE offers a general average error. Thus, similar trends in both metrics are expected, though with different sensitivities to extreme values.

Table 1 provides detailed information about the calculation and interpretation of these metrics. The evaluation was conducted across multiple temporal scales (daily, monthly, seasonal, and annual) to assess ERA5's performance in capturing both short-term and long-term climate variability in the PLB.

Table 1. Statistical metrics for performance evaluation in the study.

Metric	Formula	Optimal Value	Unit	
R	$\frac{\sum_{i=1}^{n}(R_i-\overline{R})(O_i-\overline{O})}{\sqrt{\sum_{i=1}^{n}(R_i-\overline{R})^2}\sqrt{\sum_{i=1}^{n}(O_i-\overline{O})^2}}$	1	Dimensionless	(1)
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(R_i-O_i)^2}$	0	°C for temperature; mm for precipitation	(2)
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert R_i-O_i\rvert$	0	°C for temperature; mm for precipitation	(3)

Note: n is the number of stations; R_i is the ERA5 reanalysis data of statistical sample i, O_i is the observed data at station i; $\overline{R}$ is the mean value of the ERA5 reanalysis data; $\overline{O}$ is the mean value of the observed data.

Moreover, a linear regression model was used to analyze the relationship between the ground-based station observations and the ERA5 reanalysis data. For air temperature or precipitation, a univariate linear regression was applied, where the ERA5 predictions (independent variable) were compared to the observations (dependent variable) at different time scales. The linear regression equation is expressed as follows:

$$y = a + bx \tag{4}$$

where y represents the ground station observation, and x is the ERA5 reanalysis value. The regression constant a represents the expected ERA5 value when the observation is zero, while the regression coefficient b indicates the rate of change of the ERA5 values relative to the observations. A b value close to 1 suggests strong agreement, while significant deviations indicate systematic over- or underestimation by ERA5.

The coefficient of determination (R^2) was calculated to assess the strength of the relationship between the ERA5 data and the station observations. The R^2 value ranges from 0 to 1, whereby a value closer to 1 indicates that a larger proportion of the variance in the ground observations is explained by the ERA5 data:

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(R_i-\overline{R}\right)\left(O_i-\overline{O}\right)\right]^2}{\sqrt{\sum_{i=1}^{n}\left(R_i-\overline{R}\right)^2}\sqrt{\sum_{i=1}^{n}\left(O_i-\overline{O}\right)^2}} \tag{5}$$

where n is the number of stations; R_i is the ERA5 reanalysis data of statistical sample i; O_i is the observed data at station i; $\overline{R}$ is the mean value of the ERA5 reanalysis data; and $\overline{O}$ is the mean value of the observed data.

3. Results

3.1. Validation of ERA5 Temperature

3.1.1. Daily Scale

The performance of the ERA5 reanalysis data for simulating daily mean temperature across the PLB was evaluated using data from 24 observation stations. Table 2 summarizes the key validation metrics (R, RMSE, and MAE). The average correlation coefficient R across all stations is 0.985, with 19 stations exceeding 0.98, indicating a strong agreement between ERA5 and the observed data. The RMSE values range from 0.95 °C to 3.11 °C, averaging 1.59 °C, while the MAE values range from 0.62 °C to 2.76 °C, with an average of 0.90 °C. Higher RMSE and MAE values are concentrated at specific stations, notably stations 6, 8, 9, and 11, which may be due to topographical variability, microclimatic effects, and seasonal temperature fluctuations that introduce localized climate variations not fully captured by ERA5's spatial resolution. These factors, especially in areas with complex terrain, contribute to discrepancies between the observed data and the ERA5 estimates.

Table 2. Summary of validation metrics for daily mean temperature across 24 meteorological stations from 1980 to 2020.

Station	R	RMSE (°C)	MAE (°C)
1	0.985	1.56	0.90
2	0.972	2.16	0.93
3	0.977	1.93	0.88
4	0.993	1.01	0.70
5	0.994	1.08	0.77
6	0.983	3.11	2.76
7	0.991	1.19	0.83
8	0.955	2.54	0.70
9	0.980	2.16	1.43
10	0.995	1.18	0.94
11	0.967	2.30	0.64
12	0.995	1.10	0.82
13	0.994	1.13	0.87
14	0.988	1.44	0.68
15	0.981	1.85	0.95
16	0.992	1.17	0.88
17	0.988	1.41	0.72
18	0.985	1.53	0.83
19	0.990	1.43	0.87
20	0.993	1.19	0.89
21	0.985	1.51	0.62
22	0.978	1.74	0.63
23	0.994	0.95	0.63
24	0.981	1.53	0.79
Average	0.985	1.59	0.90

Figure 2 shows the spatial distribution of these validation metrics, highlighting regional variations in ERA5 performance. Stations in the central basin characterized by flat terrain show higher correlations and lower error values. In contrast, peripheral areas, particularly in the southern, western, and eastern regions, exhibit lower correlations and higher RMSE values. These spatial patterns suggest that topography and landscape features, such as elevation and proximity to water bodies, can influence the accuracy of the ERA5 temperature simulations.

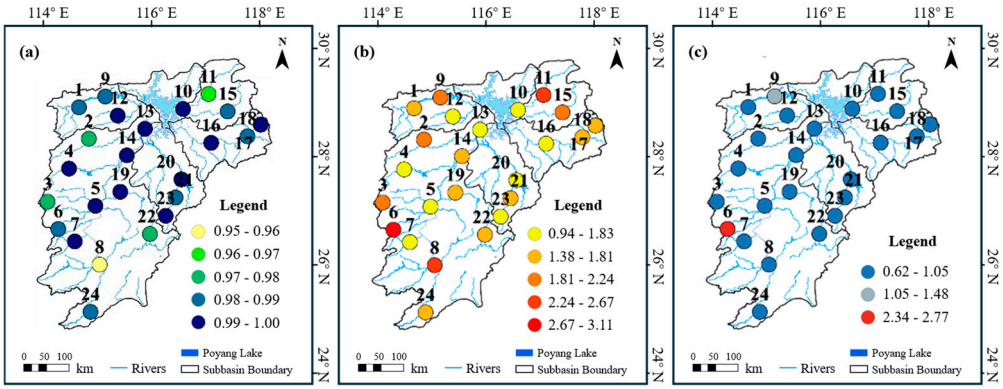


Figure 2. Spatial distribution of (**a**) R, (**b**) RMSE (°C), and (**c**) MAE (°C) between the observation and ERA5 reanalysis data for daily mean temperature.

3.1.2. Monthly Scale

Figure 3 shows the monthly variations, where ERA5 data closely track the month-by-month temperature fluctuations observed at the ground stations, demonstrating the model's ability to reproduce local monthly temperature patterns.

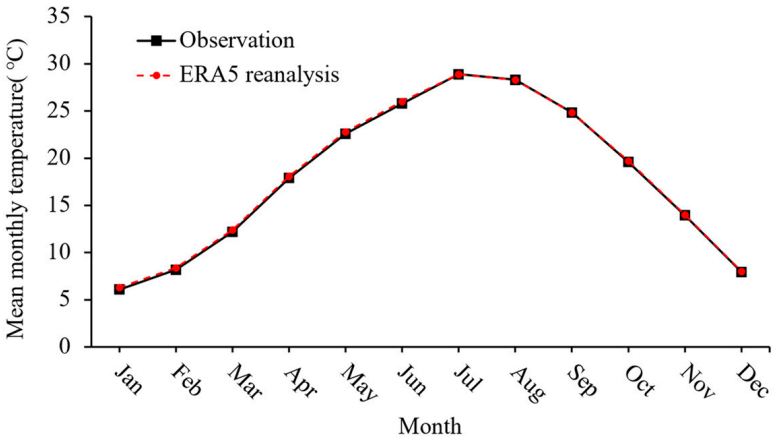


Figure 3. Monthly variations of mean air temperature between the observation and ERA5 reanalysis data.

For each station, as shown in Figure 4a, the correlation coefficients (R) exceed 0.95 at most stations, with the basin-wide monthly average never falling below 0.91. February and March show the strongest alignment, with R values surpassing 0.98 across all stations. The RMSE (Figure 4b) ranges from 0.23 °C to 3.56 °C, with elevated values recorded at station 6, particularly from June to August. The basin-wide average RMSE remains relatively stable between 0.64 °C and 0.79 °C throughout the year. The MAE (Figure 4c) follows a similar pattern, ranging from 0.17 °C to 3.48 °C, with higher errors during warmer months at station 6. Stations 4, 11, 21, and 22 consistently show MAE values below 0.5 °C, indicating high accuracy in these areas. The elevated RMSE and MAE values at station 6 are likely due to a combination of topographical influences, microclimatic effects, and seasonal temperature variability. Seasonal temperature fluctuations are particularly pronounced in the warmer months, further amplifying discrepancies between the observed temperatures and the ERA5 estimates at Station 6.

Overall, ERA5 demonstrates reliable performance on a monthly scale, though certain stations experience larger deviations, especially during warmer months. The alignment between the RMSE and the MAE suggests that these deviations may be systematic, potentially influenced by environmental factors like local climate or terrain. The increased errors in warmer months suggest that elevated temperatures may affect ERA5's accuracy in specific locations.

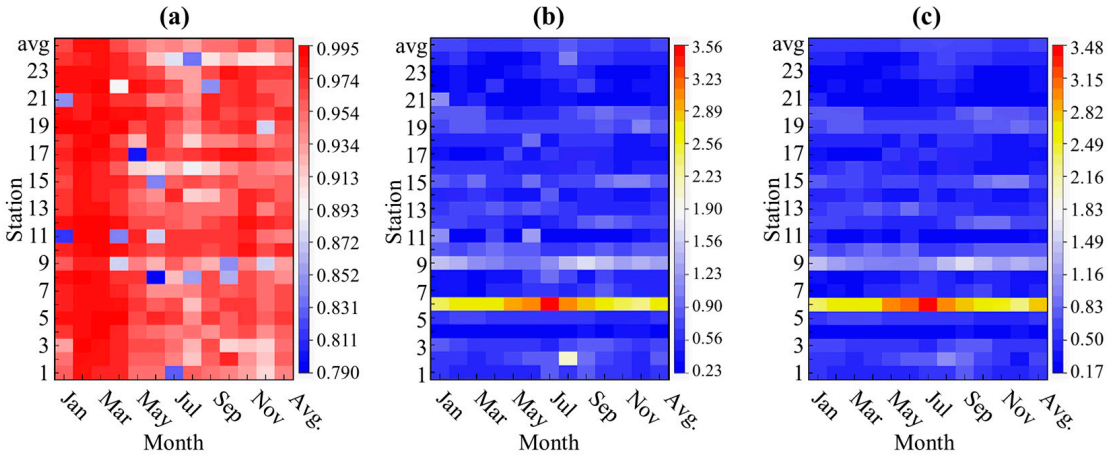


Figure 4. Heatmaps of (**a**) R, (**b**) RMSE (°C), and (**c**) MAE (°C) between the observation and ERA5 reanalysis data for daily mean temperature.

3.1.3. Seasonal Scale

Figure 5 shows the linear regression results between the ERA5 seasonal average temperatures and the observed ground measurements in the PLB. The fit is highest in spring and autumn (R^2 = 0.999), followed by summer (0.989) and winter (0.988), indicating a strong alignment between the ERA5 reanalysis data and the observed seasonal averages.

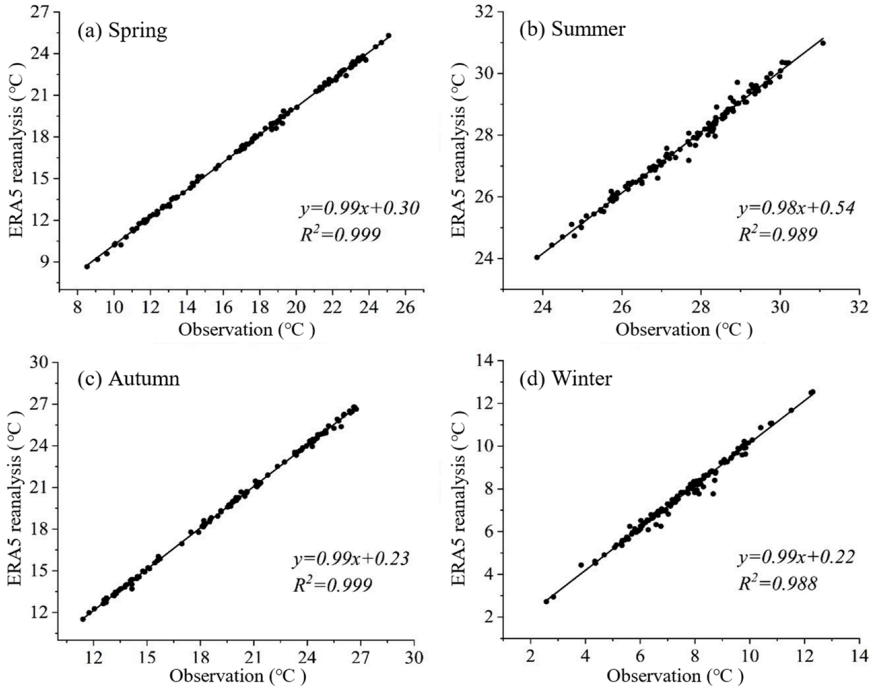


Figure 5. Linear regression of mean seasonal temperature between observation and ERA5 reanalysis data.

Table 3 summarizes the seasonal-scale evaluation metrics for the temperature accuracy of the ERA5 reanalysis data for each station. The analysis of *R* reveals that at approximately 91.7% of the stations, the seasonal *R* value exceeds 0.95, with regional averages in spring and autumn reaching 0.997. In contrast, in summer, the average correlation drops to 0.952, with 62.5% of stations recording *R* values below those of other seasons. The RMSE values fluctuate between 0.28 °C and 3.23 °C across all four seasons. The average RMSE across all stations remains below 1 °C for each season, with station 6 consistently showing higher RMSE levels compared to other stations in all seasons. The MAE has a similar range

(0.21 °C to 3.19 °C), and its seasonal and station averages follow patterns akin to those of RMSE.

Table 3. Comparison of evaluation metrics for mean seasonal temperature between observation and ERA5 reanalysis data from 1980 to 2020.

Station	R				RMSE (°C)				MAE (°C)			
	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter
1	0.998	0.962	0.996	0.972	0.50	0.65	0.69	0.57	0.43	0.53	0.59	0.47
2	0.998	0.773	0.997	0.971	0.61	1.35	0.78	0.57	0.55	0.89	0.70	0.46
3	0.998	0.977	0.993	0.968	0.61	0.62	0.81	0.72	0.55	0.54	0.69	0.62
4	0.998	0.980	0.997	0.982	0.28	0.31	0.38	0.35	0.21	0.24	0.29	0.28
5	0.999	0.985	0.998	0.984	0.66	0.62	0.54	0.63	0.61	0.56	0.48	0.57
6	0.998	0.902	0.997	0.981	2.74	3.23	2.61	2.43	2.72	3.19	2.59	2.40
7	0.998	0.971	0.997	0.983	0.47	0.58	0.66	0.41	0.39	0.50	0.58	0.30
8	0.998	0.853	0.994	0.985	0.34	0.80	0.58	0.42	0.27	0.48	0.42	0.30
9	0.995	0.976	0.994	0.966	1.08	1.24	1.51	1.41	1.02	1.18	1.45	1.32
10	0.999	0.990	0.998	0.985	0.87	0.72	0.78	0.81	0.83	0.77	0.74	0.74
11	0.996	0.925	0.998	0.934	0.48	0.79	0.32	0.70	0.27	0.30	0.26	0.31
12	0.999	0.986	0.998	0.990	0.58	0.54	0.89	0.73	0.53	0.47	0.85	0.68
13	0.998	0.983	0.998	0.982	0.80	0.74	0.53	0.64	0.75	0.66	0.47	0.55
14	0.997	0.947	0.997	0.986	0.49	0.61	0.46	0.49	0.42	0.40	0.37	0.43
15	0.995	0.966	0.997	0.977	0.78	0.77	0.91	0.88	0.65	0.71	0.82	0.80
16	0.996	0.966	0.995	0.979	0.60	0.51	0.52	0.57	0.53	0.42	0.39	0.49
17	0.995	0.988	0.998	0.988	0.51	0.52	0.45	0.36	0.34	0.45	0.39	0.30
18	0.997	0.956	0.995	0.977	0.49	0.69	0.57	0.49	0.43	0.46	0.44	0.42
19	0.998	0.983	0.998	0.973	0.76	0.74	0.78	0.88	0.72	0.68	0.74	0.79
20	0.996	0.986	0.998	0.981	0.63	0.56	0.88	0.84	0.53	0.49	0.82	0.80
21	0.996	0.979	0.998	0.939	0.39	0.36	0.28	0.72	0.24	0.29	0.23	0.28
22	0.996	0.982	0.995	0.984	0.43	0.40	0.49	0.34	0.24	0.33	0.31	0.26
23	0.998	0.977	0.998	0.987	0.33	0.56	0.36	0.31	0.27	0.49	0.27	0.23
24	0.997	0.854	0.992	0.970	0.37	0.75	0.60	0.54	0.30	0.49	0.48	0.43
Avg	0.997	0.952	0.997	0.976	0.66	0.78	0.72	0.70	0.58	0.65	0.64	0.59

3.1.4. Annual Scale

The validation of ERA5's annual temperature accuracy from 1980 to 2020 is shown in Figure 6. As shown in Figure 6a, the linear regression between the ground station data and the ERA5 annual mean temperatures yields an R^2 of 0.988, indicating a strong linear relationship. Figure 6b compares the annual temperature trends of the ERA5 estimates with observations over the 40-year period. The high correlation is particularly evident from the mid-1990s onward, when both datasets closely align. In certain years, such as 2000 and 2010, ERA5 slightly overestimates temperatures compared to the observations. Overall, ERA5 reanalysis reliably captures the interannual variability in the PLB.

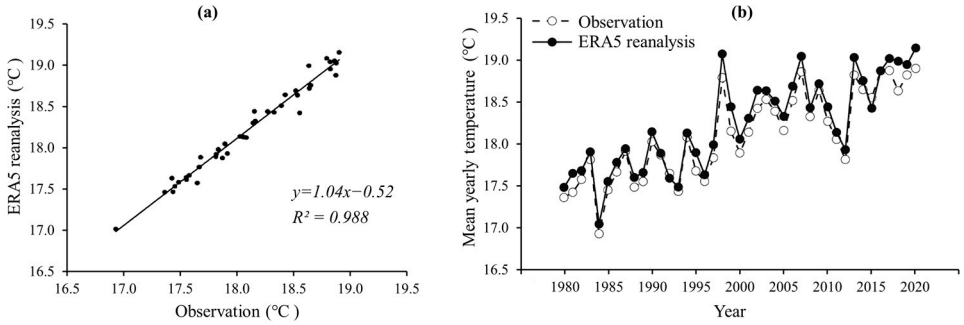


Figure 6. Validation of mean annual temperature between observation and ERA5 reanalysis data: (**a**) linear regression and (**b**) annual time series.

3.2. Validation of ERA5 Precipitation

3.2.1. Daily Scale

Table 4 summarizes the evaluation of the ERA5 reanalysis data for daily total precipitation across the basin using three validation metrics: R, RMSE, and MAE. The R values for the 24 meteorological stations range from 0.578 to 0.687, with an average of 0.637, which is significantly lower than the R values for daily temperature, indicating that ERA5 is less accurate in simulating precipitation. The RMSE values range between 8.58 mm and 11.10 mm, while the MAE values fall between 3.41 mm and 4.66 mm, with relatively small variations across the stations.

The spatial distribution of these metrics is shown in Figure 7. Stations 6 and 7 have the lowest R values, both below 0.6, which can be attributed to complex topographical features and localized microclimatic effects in these southern regions. These factors introduce temperature variations that are not fully captured by ERA5 data, resulting in reduced correlation. In contrast, stations in the northern part of the basin, particularly stations 11, 15, and 17, exhibit the highest R values. These stations are situated in more consistent, open terrain, allowing ERA5 data to more accurately represent the observed temperature patterns. The RMSE and MAE values follow a similar spatial trend, with elevated errors observed at stations 6 and 11, while station 9 shows the lowest error values across the basin.

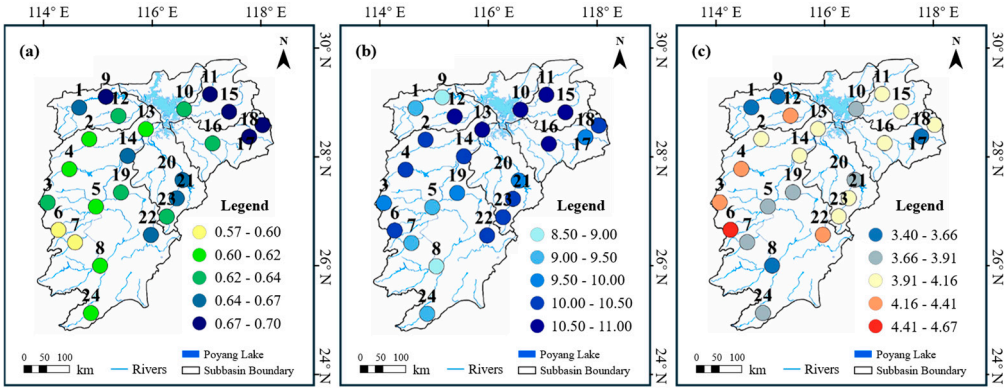


Figure 7. Spatial distribution of (**a**) R, (**b**) RMSE (°C), and (**c**) MAE (°C) between the observation and ERA5 reanalysis data for daily total precipitation.

Table 4. Summary of validation metrics for daily total precipitation across 24 meteorological stations from 1980 to 2020.

Station	R	RMSE (mm)	MAE (mm)
1	0.645	9.45	3.48
2	0.621	10.37	4.07
3	0.627	9.66	4.23
4	0.601	10.29	4.31
5	0.611	9.43	3.67
6	0.578	10.1	4.66
7	0.579	9.14	3.79
8	0.601	8.89	3.63
9	0.676	8.58	3.41
10	0.634	10.92	3.9
11	0.687	11.1	4.16
12	0.633	10.73	4.22
13	0.617	10.66	3.92
14	0.661	10.05	3.95
15	0.687	10.93	4.08
16	0.637	11.08	4.06
17	0.687	10.13	4.08

Table 4. *Cont.*

Station	R	RMSE (mm)	MAE (mm)
18	0.679	9.72	3.61
19	0.627	9.9	3.8
20	0.65	9.83	3.88
21	0.658	10.12	4.02
22	0.654	10.27	4.22
23	0.637	10.42	4.1
24	0.611	9.28	3.72
Average	0.637	10.04	3.96

3.2.2. Monthly Scale

Figure 8 shows the comparison between ERA5 and the observed monthly precipitation trends. We find that ERA5 is able to capture the overall intra-annual variation of monthly precipitation in the basin. However, it tends to overestimate precipitation in certain months. The smallest discrepancies occur in January, February, November, and December. In contrast, from March to October, ERA5 consistently overestimates precipitation, with the most significant differences observed from April to June.

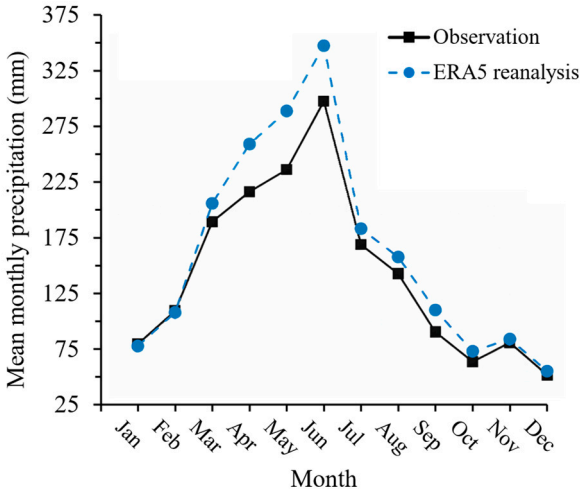


Figure 8. Monthly variations of precipitation between observation and ERA5 reanalysis data.

For each station, as shown in Figure 9a, R values across the stations range from 0.407 to 0.964, with an average of 0.777, indicating considerable spatial variation. Stations 17 and 19–23 show the highest mean R values, all above 0.80. January, February, and October to December have higher R values, with most stations exceeding 0.80 and some surpassing 0.90. In contrast, May, June, and September have lower R values, with most stations below 0.70. The RMSE and MAE values vary significantly, ranging from 12.0 mm to 183.0 mm (Figure 9b) and 9.0 mm to 151.0 mm (Figure 9c), respectively. The highest errors are observed at station 10 in June and station 8 in December. Station 4 has the largest average errors, with an RMSE of 72.79 mm and an MAE of 57.91 mm, followed by stations 11 and 12. Monthly comparisons reveal that January, February, and October to December have the lowest errors, while June shows the highest errors overall. The elevated RMSE and MAE values in June across stations 10 to 17 are likely due to the onset of the monsoon season, during which heavy and variable rainfall increases discrepancies between the ERA5 precipitation estimates and the observed data.

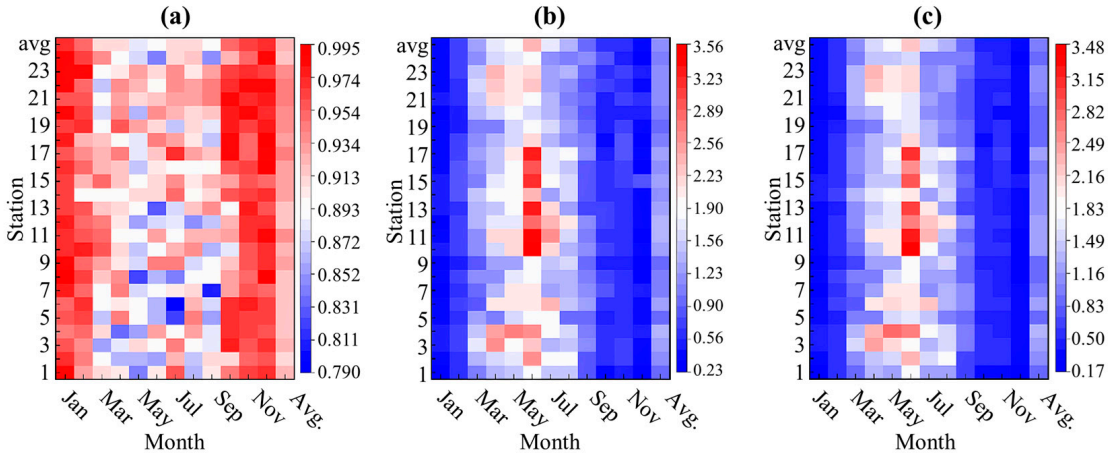


Figure 9. Heatmaps of (**a**) R, (**b**) RMSE (°C), and (**c**) MAE (°C) between the observation and ERA5 reanalysis data for daily total precipitation.

3.2.3. Seasonal Scale

Figure 10 shows the linear regression results between the observed precipitation data and the ERA5 reanalysis precipitation data at the seasonal scale in the basin. The R^2 values for spring, summer, autumn, and winter are 0.74, 0.88, 0.86, and 0.89, respectively. This confirms that ERA5 performs best in simulating winter precipitation and least accurately in spring, consistent with previous findings.

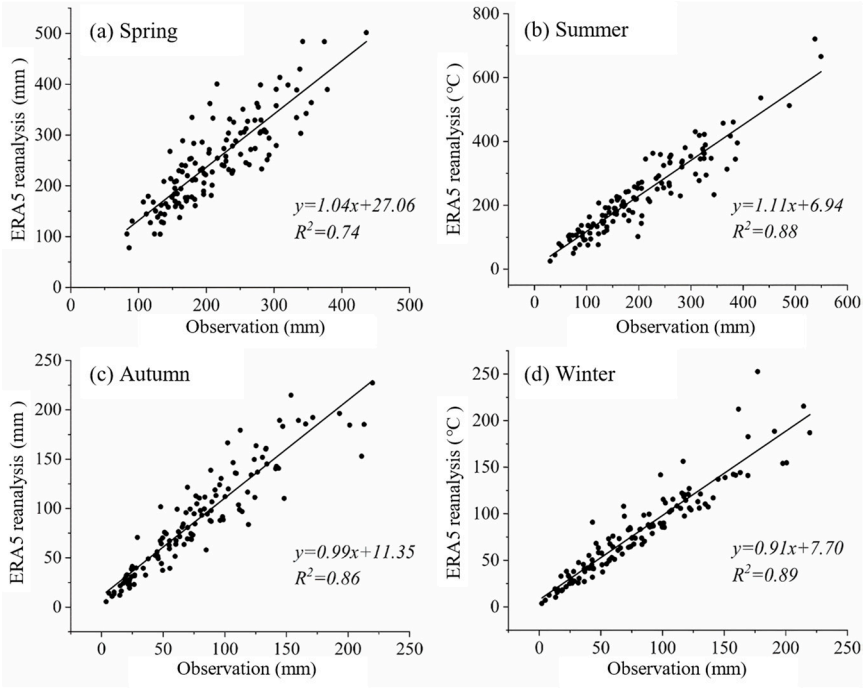


Figure 10. Linear regression of mean seasonal precipitation between the observation and ERA5 reanalysis data.

Table 5 presents the R, RMSE, and MAE values between the observed and the ERA5 seasonal precipitation data for the basin. The R values range from 0.576 to 0.925, with an average of 0.795. The highest R value is observed at station 8 in winter, and the lowest at Station 6 in summer. Winter shows the highest consistency, with all stations exceeding an R of 0.820 and an average of 0.882, while spring shows the lowest consistency, with all R values below 0.75. The RMSE values range from 20.44 mm to 135.66 mm, and the MAE values range from 14.93 mm to 106.06 mm. The largest errors are at station 11 in

summer, and the smallest are at station 9 in winter. Across the seasons, the errors are largest in summer and smallest in winter, where the RMSE and MAE values are 27.11 mm and 18.49 mm, respectively.

Table 5. Comparison of validation metrics for precipitation between observation and ERA5 reanalysis data at the seasonal scale from 1980 to 2020.

Station	R				RMSE (mm)				MAE (mm)			
	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn	Winter
1	0.682	0.789	0.808	0.875	68.24	83.33	31.21	23.43	51.03	63.05	22.29	16.39
2	0.645	0.748	0.780	0.880	82.09	107.79	37.31	25.38	60.56	83.83	26.46	18.92
3	0.666	0.803	0.825	0.874	109.34	91.47	44.28	28.40	87.66	67.93	31.83	20.36
4	0.614	0.802	0.814	0.828	117.75	107.18	41.63	30.72	94.62	86.16	30.47	20.39
5	0.712	0.722	0.810	0.888	67.28	83.63	36.65	24.60	52.18	59.92	25.74	17.57
6	0.716	0.576	0.825	0.835	91.60	104.45	46.06	28.40	71.63	82.23	33.66	19.31
7	0.707	0.724	0.745	0.857	78.50	84.03	46.69	25.51	61.40	66.03	33.55	17.71
8	0.683	0.659	0.772	0.925	71.54	80.17	41.45	26.25	53.51	61.14	28.47	17.44
9	0.720	0.832	0.753	0.904	59.35	85.20	36.40	20.44	46.15	68.82	26.71	14.93
10	0.734	0.800	0.714	0.872	87.09	129.31	45.39	27.74	64.69	91.81	31.64	18.27
11	0.738	0.792	0.807	0.897	96.86	135.66	40.90	26.57	72.32	106.06	28.27	18.09
12	0.721	0.720	0.763	0.890	83.94	128.17	45.51	27.73	64.47	99.84	33.43	19.13
13	0.726	0.719	0.824	0.881	82.88	124.64	34.90	26.21	62.03	90.49	25.82	17.54
14	0.719	0.836	0.823	0.837	91.03	94.78	37.80	32.12	69.92	70.97	27.70	21.77
15	0.725	0.822	0.810	0.829	80.38	120.45	38.60	38.58	57.61	87.79	27.94	22.53
16	0.717	0.814	0.817	0.872	79.65	109.80	37.68	28.41	59.46	78.80	26.54	18.48
17	0.751	0.869	0.816	0.887	84.66	121.66	45.93	27.03	61.56	94.29	32.66	18.27
18	0.695	0.882	0.792	0.890	73.35	84.11	41.20	27.78	56.76	64.82	26.57	17.88
19	0.719	0.789	0.821	0.918	65.09	81.66	39.82	23.20	51.54	59.84	27.91	17.07
20	0.716	0.863	0.867	0.914	84.26	72.68	33.92	22.99	63.75	55.61	23.48	15.20
21	0.751	0.860	0.873	0.890	98.32	76.27	37.71	29.32	75.11	57.73	26.25	19.56
22	0.743	0.819	0.886	0.916	103.16	84.19	36.11	25.42	81.47	63.31	26.15	19.05
23	0.722	0.822	0.853	0.894	104.26	86.25	39.75	28.34	81.84	64.84	28.76	19.71
24	0.719	0.717	0.774	0.919	78.46	72.15	44.57	26.05	57.42	51.49	30.60	18.09
Avg	0.710	0.782	0.807	0.882	84.96	97.88	40.06	27.11	64.95	74.03	28.45	18.49

3.2.4. Annual Scale

Figure 11a shows the linear regression between the observed and ERA5 annual precipitation data, with an R^2 of approximately 0.61, indicating a moderate linear relationship. However, the fitness is relatively weak compared to ERA5's performance in simulating temperature, but a relatively weak fit compared to ERA5's performance in simulating temperature. Figure 11b shows the annual precipitation trends for both ERA5 and the ground observations. While ERA5 captures the interannual variability well, it tends to overestimate precipitation during certain periods, particularly between 1980 and 2003. After 2003, the trends and annual amounts align more closely, consistent with other studies that report that ERA5 tends to overestimate precipitation in the region [27–29].

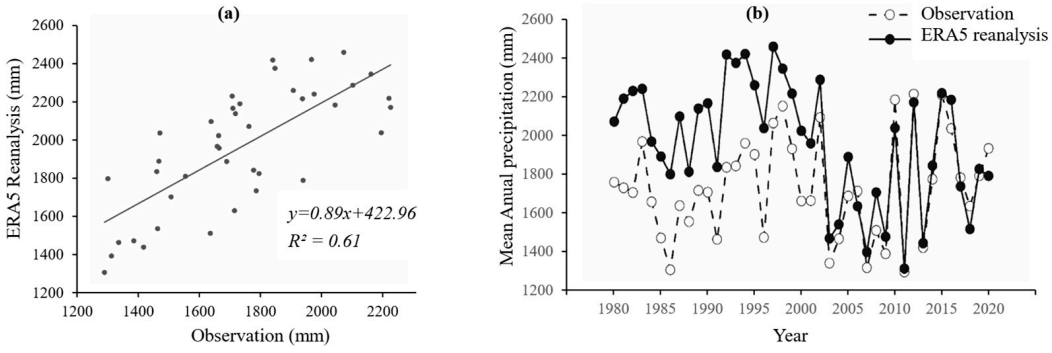


Figure 11. Validation of annual precipitation between the observation and ERA5 reanalysis data: (**a**) linear regression and (**b**) annual time series.

4. Discussions and Conclusions

The ERA5 reanalysis datasets offer improvements over traditional observations for comprehensive spatiotemporal coverage of meteorological variables. However, local climate variability requires thorough evaluation of ERA5 accuracy, particularly for the PLB in China. Previous studies have provided limited evaluation of ERA5's performance in capturing temperature and precipitation patterns in the PLB region. This study addressed this gap by assessing ERA5 data across daily, monthly, seasonal, and annual scales. The multi-scale assessment results offer valuable insights into ERA5's accuracy and limitations, which has important implications for improving climate modeling, resource management, and adaptation strategies in the basin.

Our results show that ERA5 performs well in simulating air temperature in the PLB. On a daily scale, the correlation coefficient R averaged 0.985, with 19 of 24 stations showing R values above 0.95. The monthly results were also strong, especially for February and March. On the seasonal scale, most stations had R values above 0.90, with better performance in spring and autumn. ERA5 effectively captured interannual temperature changes, achieving an R^2 of 0.988 for annual mean temperatures. Spatially, ERA5 showed better accuracy in the flat central basin than in the surrounding mountainous areas. This suggests that topography influences the accuracy of the model, with simpler terrain yielding more precise simulations. Overall, ERA5 proves to be a reliable tool for temperature modeling in the PLB.

In terms of precipitation, ERA5's performance was less accurate but still captured the general trends of precipitation patterns. We found that ERA5 reanalysis data tend to overestimate precipitation in the PLB region. This overestimation is likely due to the limitations of ERA5's spatial resolution, which struggles to capture localized precipitation patterns influenced by complex topography and convective processes. Unlike temperature, which varies more gradually, precipitation is highly localized and affected by orographic lifting and convective activity, especially during the monsoon season, leading to larger discrepancies between ERA5 estimates and the observations. While variations in the absolute precipitation values were observed, especially in the initial years of the study, the ERA5 data consistently capture the overall trends in precipitation. Over the study period, a general pattern of increasing or decreasing precipitation is evident. These trends underscore the reliability of ERA5 data in capturing the overall precipitation trends despite some discrepancies in absolute values.

This study provides a multi-scale evaluation of ERA5 reanalysis data in the PLB, assessing its performance across daily, monthly, seasonal, and annual scales. Such an evaluation is crucial for understanding how well ERA5 captures localized climate patterns and variability. The findings demonstrate that ERA5 performs well in simulating temperature but faces greater challenges with precipitation, particularly in terms of accuracy. While ERA5 proves useful for temperature-related studies in the PLB, caution is needed when using its precipitation data, which may benefit from model improvements or bias correction [30–33]. This study serves as a reference for future use of ERA5 data in characterizing climate events in the PLB and highlights the need for ongoing refinement to improve its ability to capture regional climate dynamics.

Author Contributions: Conceptualization, L.G.; methodology, L.G., X.Y., and M.Z.; software, F.Y., X.Y., and M.Z.; validation, F.Y., X.Y., and M.Z.; formal analysis, F.Y., X.Y., and M.Z.; investigation, F.Y., J.Y., X.Y., and M.Z.; resources, Y.C. and L.G.; data curation, Y.C. and L.G.; writing—original draft preparation, X.Y. and M.Z.; writing—review and editing, F.Y., J.Y., X.Y., and Y.C.; visualization, X.Y., F.Y., and J.Y.; supervision, L.G.; project administration, L.G.; funding acquisition, L.G. 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 No. 42271030), Fujian Provincial Funds for Distinguished Young Scientists (Grant No. 2022J06018), and "Young Eagle Plan" Top Talents of Fujian Province, Natural Science Foundation of Fujian Province (Grant No. 2022J01604).

Data Availability Statement: All data in this study are publicly available. The observational data from the 24 meteorological stations used in this analysis are provided by the National Meteorological Science Data Center via the China Meteorological Data Network (http://data.cma.cn/ (accessed on 1 October 2024)). The gridded temperature and precipitation data from the ERA5 reanalysis dataset can be accessed via https://cds.climate.copernicus.eu/datasets/derived-era5-single-levels-daily-statistics (accessed on 1 October 2024).

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

References

1. Liu, H.; Yuan, H.; Wang, S.; Zheng, L.; Liao, M. Spatiotemporal Dynamics of Water Body Changes and Their Influencing Factors in the Seasonal Lakes of the Poyang Lake Region. *Water* **2021**, *13*, 1539. [CrossRef]
2. Deng, C.; Chi, Y.; Huang, Y.; Jiang, C.; Su, L.; Lin, H.; Jiang, L.; Guan, X.; Gao, L. Sensitivity of WRF Multiple Parameterization Schemes to Extreme Precipitation Event over the Poyang Lake Basin of China. *Front. Environ. Sci.* **2023**, *10*, 1102864. [CrossRef]
3. Li, B.; Yang, G.; Wan, R.; Zhang, Y.; Dai, X.; Chen, Y. Spatiotemporal Variability in the Water Quality of Poyang Lake and Its Associated Responses to Hydrological Conditions. *Water* **2016**, *8*, 296. [CrossRef]
4. Li, X.; Hu, Q. Spatiotemporal Changes in Extreme Precipitation and Its Dependence on Topography over the Poyang Lake Basin, China. *Adv. Meteorol.* **2019**, *2019*, 1253932. [CrossRef]
5. Zhao, N. A Method for Merging Multi-Source Daily Satellite Precipitation Datasets and Gauge Observations over Poyang Lake Basin, China. *Remote Sens.* **2023**, *15*, 2407. [CrossRef]
6. Li, J.; Zheng, Z.; Liu, G.; Chen, N.; Lei, S.; Du, C.; Xu, J.; Li, Y.; Zhang, R.; Huang, C. Estimating Effects of Natural and Anthropogenic Activities on Trophic Level of Inland Water: Analysis of Poyang Lake Basin, China, with Landsat-8 Observations. *Remote Sens.* **2023**, *15*, 1618. [CrossRef]
7. Zhang, Q.; Ye, X.; Werner, A.D.; Li, Y.; Yao, J.; Li, X.; Xu, C. An Investigation of Enhanced Recessions in Poyang Lake: Comparison of Yangtze River and Local Catchment Impacts. *J. Hydrol.* **2014**, *517*, 425–434. [CrossRef]
8. Yan, G.; Wu, Z.; Li, D.; Xiao, H. A Comparative Frequency Analysis of Three Standardized Drought Indices in the Poyang Lake Basin, China. *Nat. Hazards* **2018**, *91*, 353–374. [CrossRef]
9. Gu, C.; Mu, X.; Zhao, G.; Gao, P.; Sun, W.; Yu, Q. Changes in Stream Flow and Their Relationships with Climatic Variations and Anthropogenic Activities in the Poyang Lake Basin, China. *Water* **2016**, *8*, 564. [CrossRef]
10. Zhang, M.; Xu, W.; Hu, Z.; Merz, C.; Ma, M.; Wei, J.; Guan, X.; Jiang, L.; Bao, R.; Wei, Y.; et al. Projection of Future Climate Change in the Poyang Lake Basin of China Under the Global Warming of 1.5–3 °C. *Front. Environ. Sci.* **2022**, *10*, 985145. [CrossRef]
11. Ye, X.; Zhang, Q.; Bai, L.; Hu, Q. A Modeling Study of Catchment Discharge to Poyang Lake under Future Climate in China. *Quat. Int.* **2011**, *244*, 221–229. [CrossRef]
12. Lei, X.; Gao, L.; Wei, J.; Ma, M.; Xu, L.; Fan, H.; Li, X.; Gao, J.; Dang, H.; Chen, X. Contributions of Climate Change and Human Activities to Runoff Variations in the Poyang Lake Basin of China. *Phys. Chem. Earth Parts ABC* **2021**, *123*, 103019. [CrossRef]
13. Karl, T.R.; Trenberth, K.E. Modern Global Climate Change. *Science* **2003**, *302*, 1719–1723. [CrossRef] [PubMed]
14. Hoegh-Guldberg, O.; Jacob, D.; Taylor, M.; Guillén Bolaños, T.; Bindi, M.; Brown, S.; Camilloni, I.A.; Diedhiou, A.; Djalante, R.; Ebi, K.; et al. The Human Imperative of Stabilizing Global Climate Change at 1.5 °C. *Science* **2019**, *365*, eaaw6974. [CrossRef] [PubMed]
15. Trenberth, K.E.; Dai, A.; Rasmussen, R.M.; Parsons, D.B. The Changing Character of Precipitation. *Bull. Am. Meteorol. Soc.* **2003**, *84*, 1205–1218. [CrossRef]
16. Held, I.M.; Soden, B.J. Robust Responses of the Hydrological Cycle to Global Warming. *J. Clim.* **2006**, *19*, 5686–5699. [CrossRef]
17. Lepore, C.; Veneziano, D.; Molini, A. Temperature and CAPE Dependence of Rainfall Extremes in the Eastern United States. *Geophys. Res. Lett.* **2015**, *42*, 74–83. [CrossRef]
18. Li, X.; Zhang, Q.; Ye, X. Dry/Wet Conditions Monitoring Based on TRMM Rainfall Data and Its Reliability Validation over Poyang Lake Basin, China. *Water* **2013**, *5*, 1848–1864. [CrossRef]
19. Lei, X.; Xu, W.; Chen, S.; Yu, T.; Hu, Z.; Zhang, M.; Jiang, L.; Bao, R.; Guan, X.; Ma, M.; et al. How Well Does the ERA5 Reanalysis Capture the Extreme Climate Events Over China? Part I: Extreme Precipitation. *Front. Environ. Sci.* **2022**, *10*, 921658. [CrossRef]
20. Zhao, P.; Gao, L.; Wei, J.; Ma, M.; Deng, H.; Gao, J.; Chen, X. Evaluation of ERA-Interim Air Temperature Data over the Qilian Mountains of China. *Adv. Meteorol.* **2020**, *2020*, 7353482. [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. Tarek, M.; Brissette, F.P.; Arsenault, R. Evaluation of the ERA5 Reanalysis as a Potential Reference Dataset for Hydrological Modelling over North America. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 2527–2544. [CrossRef]
23. Amjad, M.; Yilmaz, M.T.; Yucel, I.; Yilmaz, K.K. Performance Evaluation of Satellite-and Model-Based Precipitation Products over Varying Climate and Complex Topography. *J. Hydrol.* **2020**, *584*, 124707. [CrossRef]
24. Hu, X.; Yuan, W. Evaluation of ERA5 Precipitation over the Eastern Periphery of the Tibetan Plateau from the Perspective of Regional Rainfall Events. *Int. J. Climatol.* **2021**, *41*, 2625–2637. [CrossRef]

25. Xu, W.; Lei, X.; Chen, S.; Yu, T.; Hu, Z.; Zhang, M.; Jiang, L.; Bao, R.; Guan, X.; Ma, M.; et al. How Well Does the ERA5 Reanalysis Capture the Extreme Climate Events Over China? Part II: Extreme Temperature. *Front. Environ. Sci.* **2022**, *10*, 921659. [CrossRef]
26. Li, X.; Qin, X.; Yang, J.; Zhang, Y. Evaluation of ERA5, ERA-Interim, JRA55 and MERRA2 Reanalysis Precipitation Datasets over the Poyang Lake Basin in China. *Int. J. Climatol.* **2022**, *42*, 10435–10450. [CrossRef]
27. Vega-Durán, J.; Escalante-Castro, B.; Canales, F.A.; Acuña, G.J.; Kaźmierczak, B. Evaluation of Areal Monthly Average Precipitation Estimates from MERRA2 and ERA5 Reanalysis in a Colombian Caribbean Basin. *Atmosphere* **2021**, *12*, 1430. [CrossRef]
28. Xu, X.; Frey, S.K.; Boluwade, A.; Erler, A.R.; Khader, O.; Lapen, D.R.; Sudicky, E. Evaluation of Variability Among Different Precipitation Products in the Northern Great Plains. *J. Hydrol. Reg. Stud.* **2019**, *24*, 100608. [CrossRef]
29. Faramarzi, M.; Srinivasan, R.; Iravani, M.; Bladon, K.D.; Abbaspour, K.C.; Zehnder, A.J.B.; Goss, G.G. Setting up a Hydrological Model of Alberta: Data Discrimination Analyses Prior to Calibration. *Environ. Model. Softw.* **2015**, *74*, 48–65. [CrossRef]
30. Cucchi, M.; Weedon, G.P.; Amici, A.; Bellouin, N.; Lange, S.; Müller Schmied, H.; Hersbach, H.; Buontempo, C. WFDE5: Bias-Adjusted ERA5 Reanalysis Data for Impact Studies. *Earth Syst. Sci. Data* **2020**, *12*, 2097–2120. [CrossRef]
31. Cao, D.; Jiang, X.; Liu, S.; Chai, F.; Liu, Y.; Lai, C. Utility of Open-Access Long-Term Precipitation Data Products for Correcting Climate Model Projection in South China. *Water* **2023**, *15*, 2906. [CrossRef]
32. Beck, H.E.; Wood, E.F.; Pan, M.; Fisher, C.K.; Miralles, D.G.; van Dijk, A.I.J.M.; McVicar, T.R.; Adler, R.F. MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 473–500. [CrossRef]
33. Casanueva, A.; Kotlarski, S.; Herrera, S.; Fernández, J.; Gutiérrez, J.M.; Boberg, F.; Colette, A.; Christensen, O.B.; Goergen, K.; Jacob, D.; et al. Daily Precipitation Statistics in a EURO-CORDEX RCM Ensemble: Added Value of Raw and Bias-Corrected High-Resolution Simulations. *Clim. Dyn.* **2016**, *47*, 719–737. [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.