

## Article

# Assessment of Daily of Reference Evapotranspiration Using CLDAS Product in Different Climate Regions of China

Li-Feng Wu <sup>1,2,†</sup>, Long Qian <sup>3,†</sup>, Guo-Min Huang <sup>1</sup>, Xiao-Gang Liu <sup>3,\*</sup>, Yi-Cheng Wang <sup>2</sup>, Hua Bai <sup>1</sup> and Shao-Fei Wu <sup>1</sup>

<sup>1</sup> School of Hydraulic and Ecological Engineering, Nanchang Institute of Technology, Nanchang 330099, China; china.sw@163.com (L.-F.W.); g.huang@nit.edu.cn (G.-M.H.); baihua1985@126.com (H.B.); sfw17@nit.edu.cn (S.-F.W.)

<sup>2</sup> State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China; wangych@iwhr.com

<sup>3</sup> Faculty of Agriculture Engineering, Kunming University of Science and Technology, Kunming 650500, China; qianlong@stu.kust.edu.cn

\* Correspondence: liuxiaogangjy@126.com

† These authors contributed equally to this work.

**Abstract:** Reference Crop evapotranspiration ( $ET_0$ ) datasets based on reanalysis products can make up for the time discontinuity and the spatial insufficiency of surface meteorological platform data, which is of great significance for water resources planning and irrigation system formulation. However, a rigorous evaluation must be conducted to verify if reanalysis products have application values. This study first evaluated the ability of the second-generation China Meteorological Administration Land Data Assimilation System (CLDAS) dataset for officially estimating  $ET_0$  (the local meteorological station data is used as the reference dataset). The results suggest that the temperature data of CLDAS have high accuracy in all regions except the Qinghai Tibet Plateau (QTP) region. In contrast, the global solar radiation data accuracy is fair, and the relative humidity and wind speed data quality are poor. The overall accuracy of  $ET_0$  is acceptable other than QTP, but there are also less than 15% (103) of stations with significant errors. In terms of seasons, the error is largest in summer and smallest in winter. Additionally, there are inter-annual differences in the  $ET_0$  of this data set. Overall, the CLDAS dataset is expected to have good applicability in the Inner Mongolia Grassland area for estimating  $ET_0$ , Northeast Taiwan, the Semi Northern Temperate zone, the Humid and Semi Humid warm Temperate zone, and the subtropical region. However, there are certain risks in other regions. In addition, of all seasons, summer and spring have the slightest bias, followed by autumn and winter. From 2017 to 2020, bias in 2019 and 2020 are the smallest, and the areas with large deviation are south of climate zone 3, the coastal area of climate zone 6, and the boundary area of climate zone 7.

**Keywords:** raw reanalysis data; grid data; reference evapotranspiration; meteorological variables; Penman-Monteith equation



**Citation:** Wu, L.-F.; Qian, L.; Huang, G.-M.; Liu, X.-G.; Wang, Y.-C.; Bai, H.; Wu, S.-F. Assessment of Daily of Reference Evapotranspiration Using CLDAS Product in Different Climate Regions of China. *Water* **2022**, *14*, 1744. <https://doi.org/10.3390/w14111744>

Academic Editor: Alban Kuriqi

Received: 9 April 2022

Accepted: 26 May 2022

Published: 29 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

Reference Crop evapotranspiration ( $ET_0$ ) is a critical factor for calculating crop evapotranspiration, the accurate estimation of which plays a vital role in irrigation engineering design and planning, water resources management, and climate change research [1–3]. Due to its large population and rapid economic development, China is facing a severe water shortage problem. The country's per capita water resource is only one-fourth of the world average level [4]. Therefore, an accurate estimation of  $ET_0$  in this region would provide a scientific basis for rationally allocating water resources and minimizing the imbalance between water supply and demand [5]. Currently, the standard estimating method of  $ET_0$  is the Penman-Monteith equation (FAO56 PM) recommended by the Food and Agriculture Organization of the United Nations (FAO) [6,7]. This method combines energy balance and

the aerodynamic theory, which is strongly applicable under different climatic conditions. However, the main drawback of this method is that it requires a high quality of meteorological data, including air temperature, relative humidity or dew temperature, solar radiation, and wind speed [8,9]. In many regions of the world, there are not enough weather stations to monitor the meteorological factors. Additionally, high-quality, long-term observational data are lacking, especially in developing countries, which hinders the application of the PM method for  $ET_0$  estimation on large spatial scales [10–12].

In recent years, reanalysis products have become one of the main grid data sources for water resource management research [13]. Reanalysis data are generated by running a numerical weather-predicting model that assimilates the observed atmospheric and surface data to reconstruct the past surface, ocean, and atmospheric state variables. Unlike geostatistical grid data derived from spatial interpolation, the spatial structure of weather variables (such as temperature and wind speed) synthesizes physical laws embedded in numerical models [8].

Nowadays, many reanalysis data sets have developed rapidly and are used in various fields. Baatz et al. (2020) [14] analyzed state-of-the-art methods, recent developments, and prospects of reanalysis for three subcomponents of the Earth system (atmosphere, ocean, and land), they points out the method's increasing computational capabilities, the growing availability of long-term satellite data with global coverage, and the advancements in model-data fusion methods such as variational and sequential data assimilation. In addition, the above paper discusses the increasing awareness of the drastic changes in the Earth system related to anthropogenic and climatic factors and the way they drive reanalysis development. Recently, networks of distributed in-situ sensors such as buoys and biogeochemical Argo floats [15], eddy covariance stations [16], surface water runoff observations [17], and meteorological station data [18] were used in the reanalysis of physical and biogeochemical Earth system processes. Munoz-Sabater et al. (2021) [19] presented the new global ERA5-Land reanalysis. The quality of ERA5-Land fields was evaluated by direct comparison to many in situ observations collected for the period 2001–2018, and for comparison to additional model or satellite-based global reference datasets. Overall, the water cycle was improved in ERA5-Land compared to ERA5 according to the different variables evaluated, whereas the energy cycle variables showed similar performances. Both ERA5 and ERA5-Land perform substantially better than ERA-Interim.

Reanalysis data have also been applied and compared to estimate evapotranspiration in different regions of the world. Boulard et al. (2016) [20] calculated  $ET_0$  using the ERA-Interim reanalysis data and verified its accuracy in a water balance study in north-eastern France. Srivastava et al. (2016) [21] found that ERA-Interim  $ET_0$  was superior to NCEP/NCAR  $ET_0$  in the UK. Pelosi et al. (2020) [22] also compared two reanalysis datasets for  $ET_0$  estimation in southern Italy. Woldesenbet et al. (2021) [23] evaluated the  $ET_0$  in the Omo-Gibetta watershed and achieved good prediction results. Song et al. (2015) [24] judged the spatiotemporal characteristics of  $ET_0$  in the Shaanxi Province based on NCEP reanalysis data and made future predictions. Liu et al. (2019) [25] estimated the future  $ET_0$  in the Poyang Lake basin based on the CMIP5 model. The results showed that the stepwise regression downscaling model established by the NCEP reanalysis data and the basin  $ET_0$  had better simulation results.  $ET_0$  was assessed in the Iberian Peninsula by Martins et al. (2016) [26]. The focus here is to use the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) hybrid reanalysis product and gridded dataset to calculate  $ET_0$  with good simulation results. Raziie (2021) [27] used the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis, combined with a gridded dataset, to calculate monthly  $ET_0$  for 43 meteorological stations distributed across Iran. The results show that the  $ET_0$  calculated by the mixed reanalysis had a better effect than the  $ET_0$  calculated by the observations at most research stations. Milad and Mehdi (2022) [28] used reanalysis products to estimate  $ET_0$  in areas with sparse data and showed that ERA5 provided more accurate estimates of daily and monthly  $ET_0$ . Some scholars have also

compared satellite grid data with meteorological station values. Wang et al. (2019) [29] comprehensively evaluated and compared this newly released precipitation product (Integrated Multi-satellite Retrievals V05B) and its predecessor TRMM 3B42V7 based upon the ground-based observations under complex topographic and climatic conditions over the Hexi Region in the northwest arid region of China. Their results indicated that compared to ground-based observations, both IMERG and 3B42V7 showed good performance with slight overestimation. Prakash et al. (2016) [30] investigated the capabilities of the Tropical Rainfall Measuring Mission (TRMM), Multi-satellite Precipitation Analysis (TMPA), and the recently released Integrated Multi-satellite Retrievals for GPM (IMERG) in detecting and estimating heavy rainfall across India. The results indicated that the multi-satellite product systematically overestimates its inter-annual variations. With continuous advances in numerical weather models, computing, information, and communication technology (ict) tools, and data assimilation techniques, along with continuous improvements in the quality of atmospheric and ground data obtained from satellites, the spatial and temporal resolution and reliability of reanalysis data have been gradually improved year after year.

The China Meteorological Administration Land Data Assimilation System (CLDAS) is the only real-time service system in land surface data assimilation systems in China. It uses a combined technology of integration and assimilation to fuse data from various sources, such as ground observation, satellite observation, and numerical model products [31]. The output of this system includes high spatial and temporal resolution land surface driving products such as temperature, air pressure, specific humidity, wind speed, precipitation, solar shortwave radiation, and soil moisture. These could be applied in agricultural drought monitoring, mountain flood geological disaster meteorological services, climate system model assessments, and spatial fine grid real data services. Although many studies have evaluated the quality of the CLDAS data, there are limited reports on the estimation of  $ET_0$  by this dataset. In this paper, we used the meteorological reanalysis data of 689 surface meteorological stations in China from 2017 to 2020 and found four grid data points around each meteorological station through calculation and processing. We then calculated the value of the target station using the inverse distance weight method, compared it with the measured data of local meteorological stations and evaluated the accuracy of CLDAS data through statistical indicators. Therefore, this study aims to evaluate the accuracy of  $ET_0$  simulation with CLDAS products for the first time by comparing meteorological data from 689 ground weather stations and to exploring a product that could provide accurate  $ET_0$  for areas lacking meteorological data observation.

## 2. Materials and Methods

### 2.1. Introduction to CLDAS

CLDAS is an isolatitude and longitude mesh fusion analysis product covering the Asian region (0–65° N, 60–160° E) with a resolution of  $0.0625^\circ \times 0.0625^\circ$  and 1 h, and a spatial geometric distance of 9 km between the corresponding grid points, including six variables: 2 m air temperature, 2 m specific humidity, 10 m wind speed, surface pressure, precipitation, and shortwave radiation. This dataset is developed by using the Space and Time Mesoscale Analysis System (STMAS), optimal interpolation (OI), probability density function matching (CDF), physical inversion, terrain correction, and other techniques based on ground and satellite observations from a variety of sources. The dataset's quality and spatio-temporal resolution in China is better and higher than in the international market. The scientific goal of CLDAS is to use data fusion and assimilation technology, on the ground observation, satellite observations, numerical model products, and other sources of data fusion to obtain high space-time resolution and high-quality temperature, pressure, humidity, wind speed, precipitation, and radiation elements such as lattice data to drive the land surface model, obtain soil temperature and humidity, etc. The research focuses on processing and acquiring land surface driving data, realizing the operation and integration of multiple land surface models, and improving the underlying surface data, vegetation parameters, and atmospheric driving data.

### 2.1.1. Data Sources for CLDAS

(1) Ground observation data: hourly temperature, air pressure, humidity, wind speed, precipitation, and other data observed by more than 2400 national automatic weather stations and nearly 40,000 regional automatic weather stations after quality control.

(2) ECMWF (European Center for Mediumrange Weather Forecasts) numerical analysis/forecast products: global 3 h, 0.125° resolution 2 m temperature, 2 m humidity, 10 mU/V wind speed, surface pressure, and other data products released by EC (European Center).

(3) GFS numerical analysis/prediction products: NCEP released global ozone, atmospheric precipitation, surface pressure, and other data products with 3 h and 0.5°.

(4) Satellite precipitation products: FY2 precipitation estimation products (nominal disk chart) of the National Satellite Meteorological Center; East Asia Multi-Satellite Integrated Precipitation Data Product (EMSIP) with a resolution of 1 h and 0.0625° for the Asian region operated by the National Meteorological Information Center.

(5) Fusion precipitation product: the fusion product of FY2/CMORPH precipitation and automatic ground station precipitation with 1 h and 0.1° resolution in China operated by the National Meteorological Information Center.

(6) FY2 satellite entire disk nominal map: multi-channel geostationary satellite observation data with 1 h and 5 km resolution (subsattellite point) of the Service of National Satellite Meteorological Center (nominal disk map).

(7) DEM data: a global 30m spatial resolution topographic data product jointly measured by NASA (National Aeronautics and Space Administration) and NIMA (National Bureau of Surveying and Mapping of the Ministry of Defense) was used to re-sample DEM topographic data with a spatial resolution of 0.0625° in the Asian region using the area weight method.

### 2.1.2. CLDAS Data Processing Methods

The 2 m temperature, 2 m specific humidity, 10 m wind speed, and surface pressure products take ECMWF numerical analysis/forecast products as the background field. Topographic adjustment and multi-grid variational technology (STMAS) are used to integrate the observation data of automatic ground stations in China. The background field outside China is formed by topographic adjustment, variable diagnosis, and interpolation to the analysis grid.

The DISORT radiative transfer model used ozone, atmospheric precipitation, and surface pressure in GFS numerical analysis products as the dynamic input parameters for the radiative transfer model. Additionally, FY2E/G satellite VIS channel complete disk nominal map data inversion was used to form the short-wave radiation product.

The above information comes from the China Meteorological Data Sharing Network (<https://data.cma.cn/>, accessed on 1 May 2021). The data used in this study included temperature, global solar radiation, relative humidity, and wind speed, of which the height of wind speed was 10 m and the height of other meteorological variables was 2 m. The data spanned from 2017 to 2020.

### 2.2. Reference Evapotranspiration

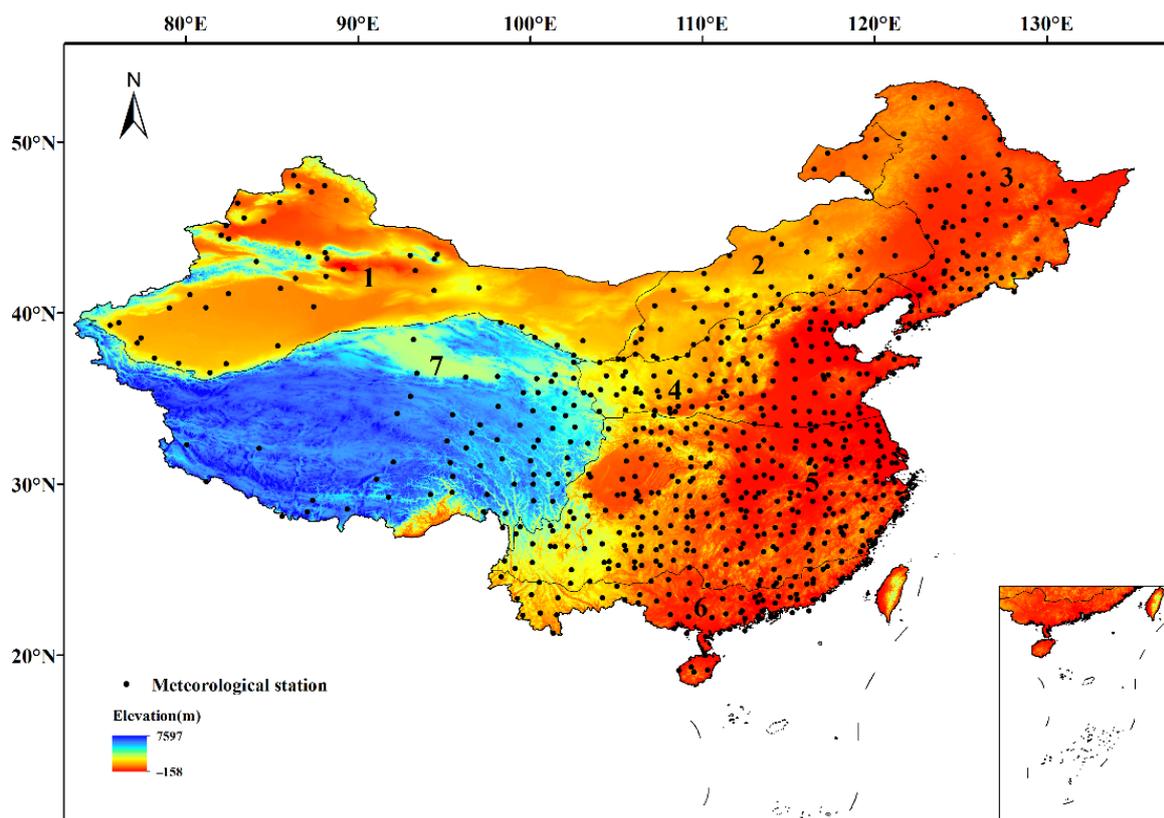
According to the FAO56 PM equation [32], reference evapotranspiration ( $ET_0$ ; mm d<sup>-1</sup>) can be calculated as:

$$ET_0 = \frac{0.408(R_n - G) + \gamma \frac{900}{T_a + 273} U(e_s - e_a)}{\Delta + \gamma(1 + 0.34U)} \quad (1)$$

where  $R_n$  is the net radiation at the crop surface, usually calculated by  $R_s$  (Global solar radiation);  $G$  is the soil heat flux density;  $T_a$  is the mean daily air temperature at 2 m height;  $U$  is the wind speed at 2 m height;  $e_s$  and  $e_a$  are the saturation and actual vapor pressure, respectively;  $\Delta$  is the slope of vapor pressure curve, and  $\gamma$  is the air psychrometric constant. In daily time-step in this study  $G$  can be neglected [33,34].

### 2.3. Data Sources

To examine the performance of this dataset, meteorological data from 689 ground meteorological observation stations of the China Meteorological Administration (CMA) were collected, which included maximum and minimum temperatures at 2 m, global surface radiation or sunshine durations, relative humidity at 2 m, and wind speed at 10 m. If necessary, sunshine durations were converted into global radiation using a formula from a previous study [35]. The stations were divided into seven climate zones [36,37]. The specific distribution is shown in Figure 1, and the names are shown in Table 1.



**Figure 1.** Climate zones of China and geographical distribution of 689 meteorological stations. (See Table 1 for the names of climatic zones 1–7).

**Table 1.** Names of the seven climate zones.

Zone	Area Name
1	Northwest desert zone
2	Inner Mongolia grassland zone
3	Northeast humid and semi humid temperate zone
4	Humid and semi humid warm temperate zone
5	Humid subtropical zone
6	Humid tropical zone
7	Qinghai Tibet Plateau zone

To obtain the daily reanalysis variables for Equation (1) (identified by subscript CLD), the following steps were taken: (a) daily  $T_{maxCLD}$  and  $T_{minCLD}$  were selected as the maximum and minimum of the 24 daily available 1-h values of the  $T_{max}$  and  $T_{min}$  sequences, respectively; (b) daily  $RH_{CLD}$  was obtained by calculating the 24-h average value of 24 RH values per day; (c) calculating the 24-h cumulative value of the 12-h  $R_s$  as the daily  $R_{sCLD}$  value; (d) wind speed at 10 m ( $U_{10CLD}$ ) was calculated as the 24-h

average of 24 1-h values, which were then converted to a height of 2 m ( $U_{CLD}$ ) using Formula (2) as follows, respectively:

$$U = U_z \frac{4.87}{\ln(67.8z - 5.42)} \quad (2)$$

where  $z$  is the height of the wind speed observation instrument (in this paper,  $z$  is equal to 10) for each meteorological station. Grid data from four grid points around it were selected and interpolated to the station by the inverse distance weight (IDW) method. The formula is as follows:

$$V = \frac{\sum_{i=1}^n \frac{v_i}{D_i^2}}{\sum_{i=1}^n \frac{1}{D_i^2}} \quad (3)$$

where  $V$  is the inverse value,  $v_i$  is the value of the control point, and  $D_i$  is the weight coefficient.

#### 2.4. Statistics Indicators

Three common statistical indicators, including the coefficient of determination ( $R^2$ ), root mean square error (RMSE), mean absolute error (MAE), and percent bias (PBias) were chosen to evaluate the accuracy of the CLDAS meteorological variables and  $ET_0$  in this study. The corresponding formulas are:

$$MAE = \frac{1}{n} \sum_{i=1}^n |M_i - P_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (M_i - P_i)^2} \quad (5)$$

$$R^2 = \frac{\left[ \sum_{i=1}^n (M_i - \bar{M})(P_i - \bar{P}) \right]^2}{\sum_{i=1}^n (M_i - \bar{M})^2 \sum_{i=1}^n (P_i - \bar{P})^2} \quad (6)$$

$$PB = \frac{\sum_{i=1}^n (P_i - M_i)}{\sum_{i=1}^n M_i} \quad (7)$$

where  $M_i$  is  $ET_0$  calculated by meteorological station data,  $P_i$  is  $ET_0$  calculated by the CLDAS gridded data,  $\bar{M}$  is average  $ET_0$  calculated by meteorological station data,  $\bar{P}$  is average  $ET_0$  calculated by the CLDAS gridded data, and  $n$  is the number corresponding to  $ET_0$  data. Higher  $R^2$  values (closer to 1) or lower RMSE and MAE values indicate a better estimation performance of the CLDAS dataset. The closer PB is to 0, the better the estimation performance of the CLDAS dataset.

### 3. Results

#### 3.1. Meteorological Factors

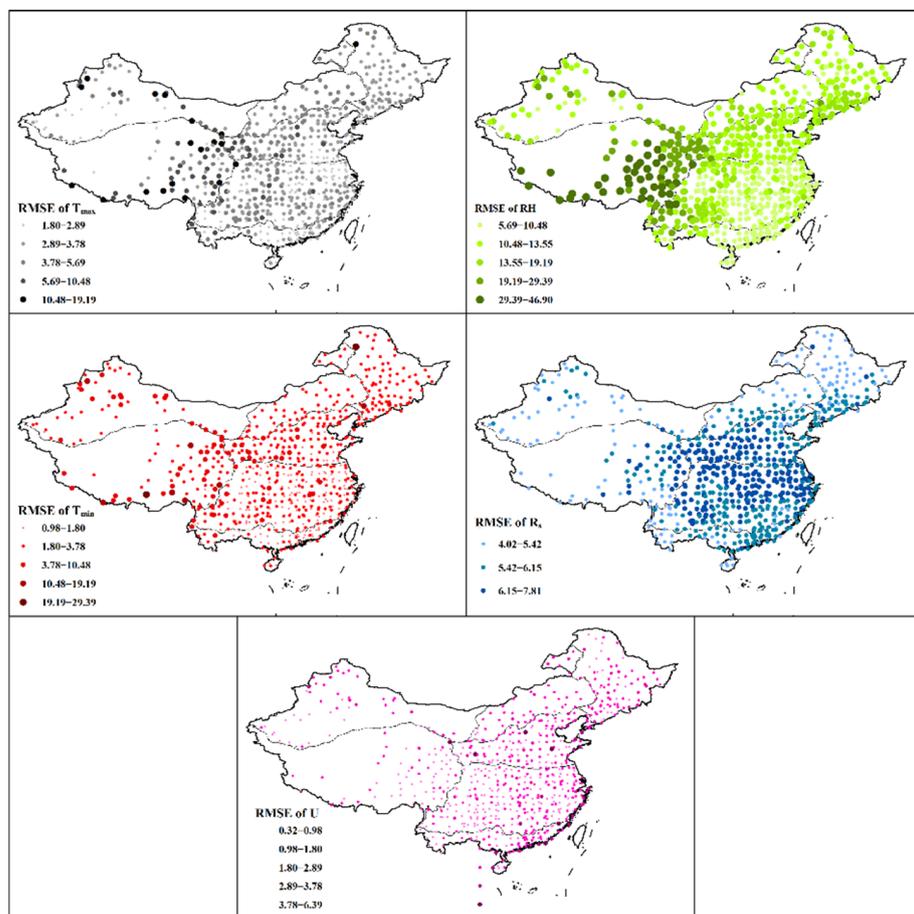
##### 3.1.1. Air Temperature

Table 2 shows the statistical indicators of maximum and minimum temperatures in the CLDAS data for the seven climate zones in China. Results indicated that the accuracy for the maximum and minimum temperatures differed in different climatic regions. For the maximum temperature, CLDAS data showed a high correlation with data from ground stations in the four northern climate zones (i.e., climate zones 1–4), with  $R^2$  larger than 0.9. Climate zone 5 in the humid climate region also yielded a good correlation. In climate zones 6 and 7, the correlations between the two datasets were slightly worse when compared with other climate zones. However, climate zone 6 showed the smallest values in terms of statistical errors, with RMSE and MAE valued at 2.9 and 2.3 °C, respectively. This may be since the range of temperature changes in this region is not as large as that in other regions, and the area of this climate zone is significantly smaller than that in other

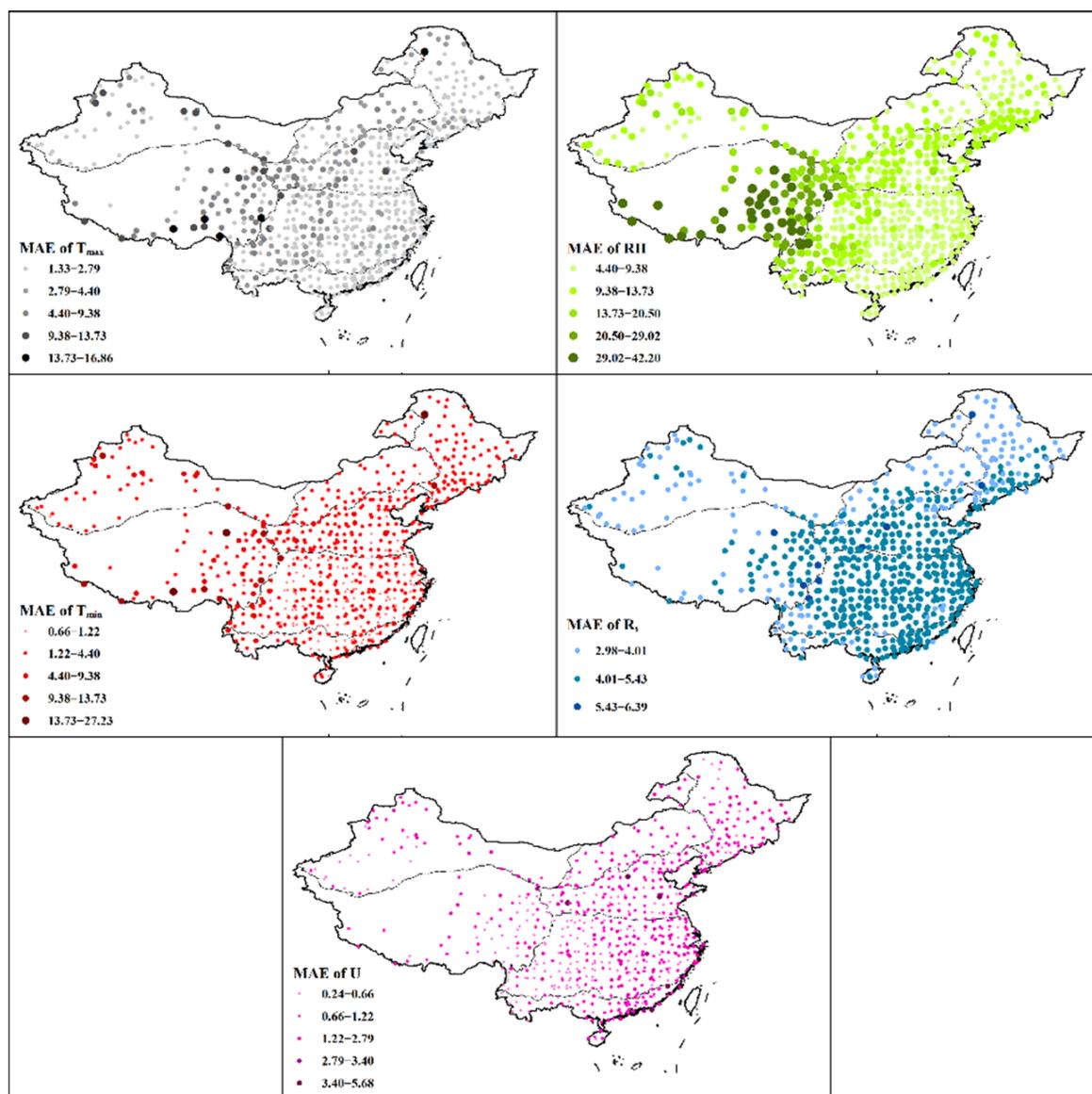
climate zones, so the temperature change in this region is not as dramatic as that in other climate zones. The RMSE and MAE of the high-altitude climate zone (i.e., climate zone 7) were 6.55 °C and 5.83 °C, respectively. Figures 2–4 show the spatial error distribution of the maximum temperature in CLDAS. Overall, the errors at most stations were within a small range. However, in climate zone 7 and the north-central area of climate zone 1, there was a big error in the regions, with RMSE and MAE of many stations more significant than 10 °C, while  $R^2$  was lower than 0.5. Such huge variations in model errors in these stations might be resulted from the regional climate model parameter variations and were unlikely caused by the overall overvalued or undervalued problem of models that may cause significant variation for  $ET_0$  estimation.

**Table 2.** Statistical indicators of maximum and minimum temperatures in different climate zones of China.

Climate Zone	$T_{max}$			$T_{min}$		
	RMSE °C	MAE °C	$R^2$	RMSE °C	MAE °C	$R^2$
	%	%		%	%	
1	4.99	4.32	0.95	4.03	3.45	0.96
2	3.82	2.94	0.93	2.50	1.85	0.97
3	3.75	2.88	0.94	3.02	2.27	0.97
4	3.83	3.05	0.92	2.65	2.11	0.97
5	3.51	2.75	0.88	2.24	1.79	0.96
6	2.90	2.30	0.82	1.87	1.50	0.94
7	6.55	5.83	0.81	5.10	4.53	0.92

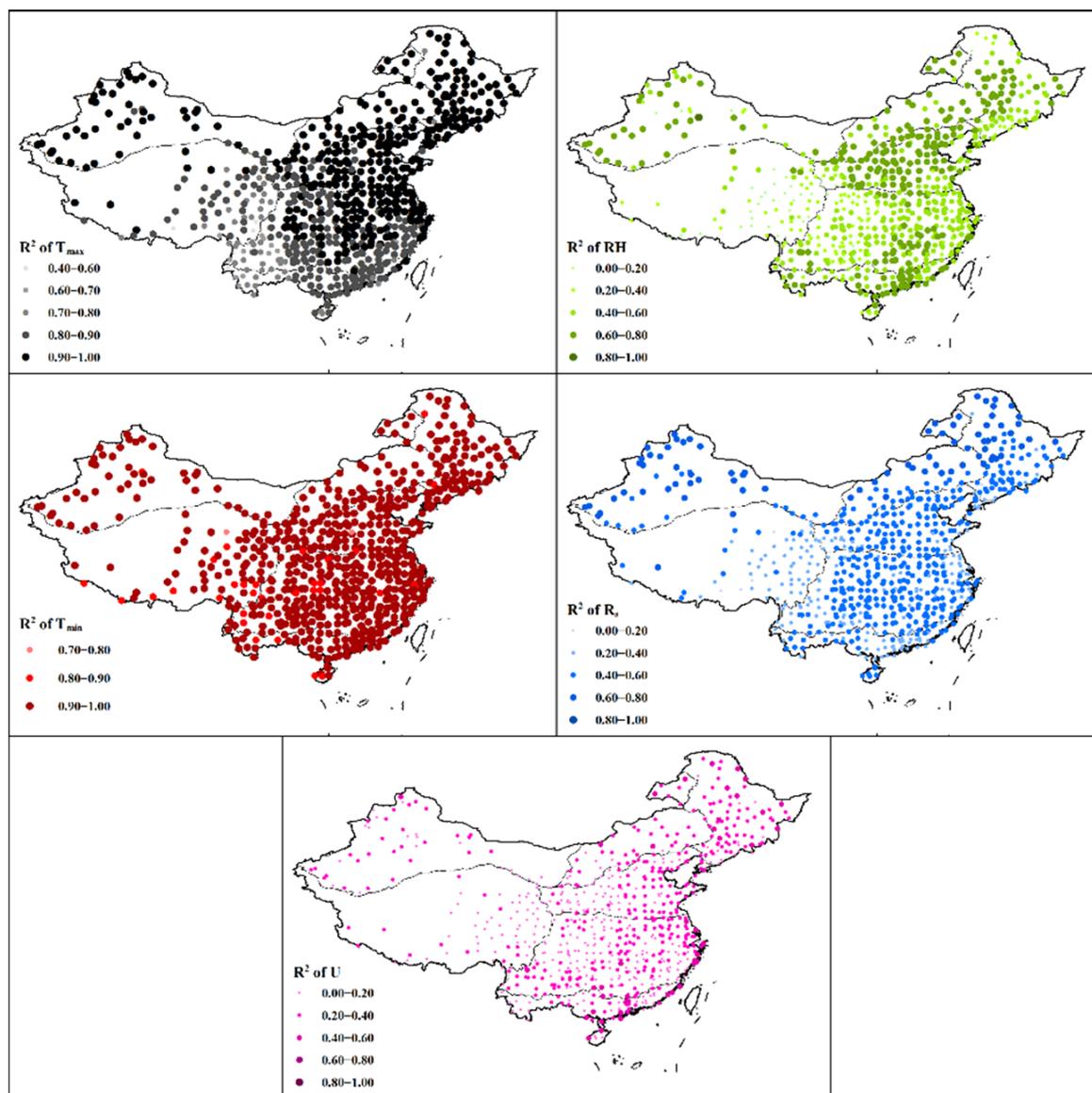


**Figure 2.** RMSE values of the five meteorological factors of in all stations.



**Figure 3.** MAE values of the five meteorological factors of in all stations.

The minimum temperature behavior in the CLDAS data set was similar to the maximum temperature. However, compared with the maximum temperature, correlations between the minimum temperature of CLDAS and the station's temperature were higher in all seven climate zones, with all  $R^2$  larger than 0.9. In climate zone 6, the minimum temperature error was the lowest among all climate zones, with RMSE and MAE valued at 1.87 °C and 1.5 °C, respectively. On the contrary, climate zone 7 had the highest error, with RMSE and MAE reaching 5.1 and 4.53 °C, respectively. In addition, the lowest temperature error in climate zone 1 was also relatively high, with RMSE and MAE valued at 4.03 °C and 3.45 °C, respectively, which might affect the accuracy of  $ET_0$  estimation. As can be seen from Figures 2–4, the performance of  $T_{\min}$  was similar to that of  $T_{\max}$ . Therefore, the accuracy of most stations was within an acceptable range. However, some stations showed significant errors, which were mainly located in the middle of climate zone 1 and climate zone 7. These stations with high error in the minimum temperature had a high coincidence with corresponding high  $T_{\max}$  error stations, indicating severe problems in the temperature simulation of the stations.



**Figure 4.**  $R^2$  values of the five meteorological factors of in all stations.

### 3.1.2. Solar Radiation

Table 3 shows the statistical indicators of solar radiation ( $R_s$ ) in the CLDAS data for the seven climate zones. Across climate zones, RMSE ranged from 5.18 to 6.21  $\text{MJ m}^{-2} \text{d}^{-1}$ , and MAE ranged from 3.83 to 4.54  $\text{MJ m}^{-2} \text{d}^{-1}$ . The differences in  $R_s$  errors among different climate regions were not as apparent as those in air temperature. However, the  $R^2$  of climate zone 7 was significantly lower than that of other climatic regions. These results were similar to the results reported by Liu et al. (2009) [38]. However, their values were generally higher than that of the radiation model based on temperature, where the median RMSE was 3.3  $\text{MJ m}^{-2} \text{d}^{-1}$  in humid regions of China (Fan et al., 2019) [39]. The above phenomenon indicated that the radiation data in the CLDAS data set did not perform well.

Figures 2–4 show the spatial distribution of  $R_s$  errors. Overall, the error of  $R_s$  in climate zones 1–3 was better than in other climate zones. The RMSE of most stations was more significant than 6  $\text{MJ m}^{-2} \text{d}^{-1}$ . This might be due to the severe air pollution in the above areas [40], which would pose particular challenges to accurate simulation.

**Table 3.** Statistical indicators of solar radiation, relative humidity, and wind speed in different climate zones of China.

Climate Zone	$R_s$			RH			U		
	RMSE $\text{MJ m}^{-2} \text{d}^{-1}$	MAE $\text{MJ m}^{-2} \text{d}^{-1}$	$R^2$	RMSE %	MAE %	$R^2$	RMSE $\text{m s}^{-1}$	MAE $\text{m s}^{-1}$	$R^2$
1	5.18	3.83	0.55	14.93	12.11	0.59	1.18	0.88	0.22
2	5.41	3.93	0.65	12.60	9.79	0.55	1.32	1.01	0.29
3	5.33	3.93	0.55	12.81	9.88	0.50	1.28	1.00	0.30
4	6.21	4.51	0.60	13.62	10.73	0.44	1.26	1.01	0.20
5	6.13	4.54	0.48	11.90	9.72	0.42	1.05	0.85	0.21
6	5.54	4.19	0.59	9.19	7.43	0.39	1.38	1.13	0.22
7	5.80	4.43	0.28	31.29	27.62	0.33	1.03	0.81	0.14

### 3.1.3. Relative Humidity

Statistical indicators of CLDAS RH are shown in Table 3. Climate zone 7 had the most significant error among all regions, with RMSE and MAE reaching 31.29% and 27.62%, respectively, close to a random distribution. The consistency between CLDAS and site data in other climatic regions was also not high, with  $R^2$  ranging from 0.39 to 0.59. However, the values of RMSE and MAE indicated that they were still within acceptable limits. Compared climate zone 1 with climate zone 6, the consistency of climate zone 1 was higher, but the RMSE and MAE of climate zone 6 are lower. This result could be attributed to the fact that climate zone 6 is within the humid region with high annual average relative humidity, while climate zone 1 is in the arid area where the relative humidity changes more sharply. Figures 2–4 showed that the overall error of this data set was larger in climate zone 7 than in other climate zones. In addition, compared with the northeast part of climate region 5, the error of RH in the western part of the same climate zone (i.e., areas bordering climate zone 7) was significantly larger. Although there was a relatively large error for RH in some regions, the estimation of  $ET_0$  would unlikely be affected, as previous studies have found that RH would have a low contribution to  $ET_0$  in most regions of China [41].

### 3.1.4. Wind Speed

Statistical indicators of CLDAS U are shown in Table 3. From the perspective of  $R^2$ , the consistency between CLDAS near-surface wind speed and station data was poor in all climate zones, while from the perspective of RMSE and MAE, their accuracies were acceptable. In addition, the mean difference between climate zones was within 30%. However, according to the spatial distribution of the error (Figures 2–4), the RMSE of some stations was more than  $3 \text{ m s}^{-1}$ , of which the errors of most stations in climate zone 7 were significant. Compared with the inland stations, the  $R^2$  of the coastal stations was higher. However, RMSE and MAE were also higher, indicating a problem of overestimation or underestimation. The worldwide modeling for wind speed is challenging and often inaccurate. Similar results were obtained for ERA5 [22], NCEP/NCAR [27], and GLDAS [42]. This is mainly due to the complex terrain changes on the ground, and the wind speed is greatly affected by the roughness of the underlying surface. In addition, it is not easy to simulate the movement direction of winds accurately.

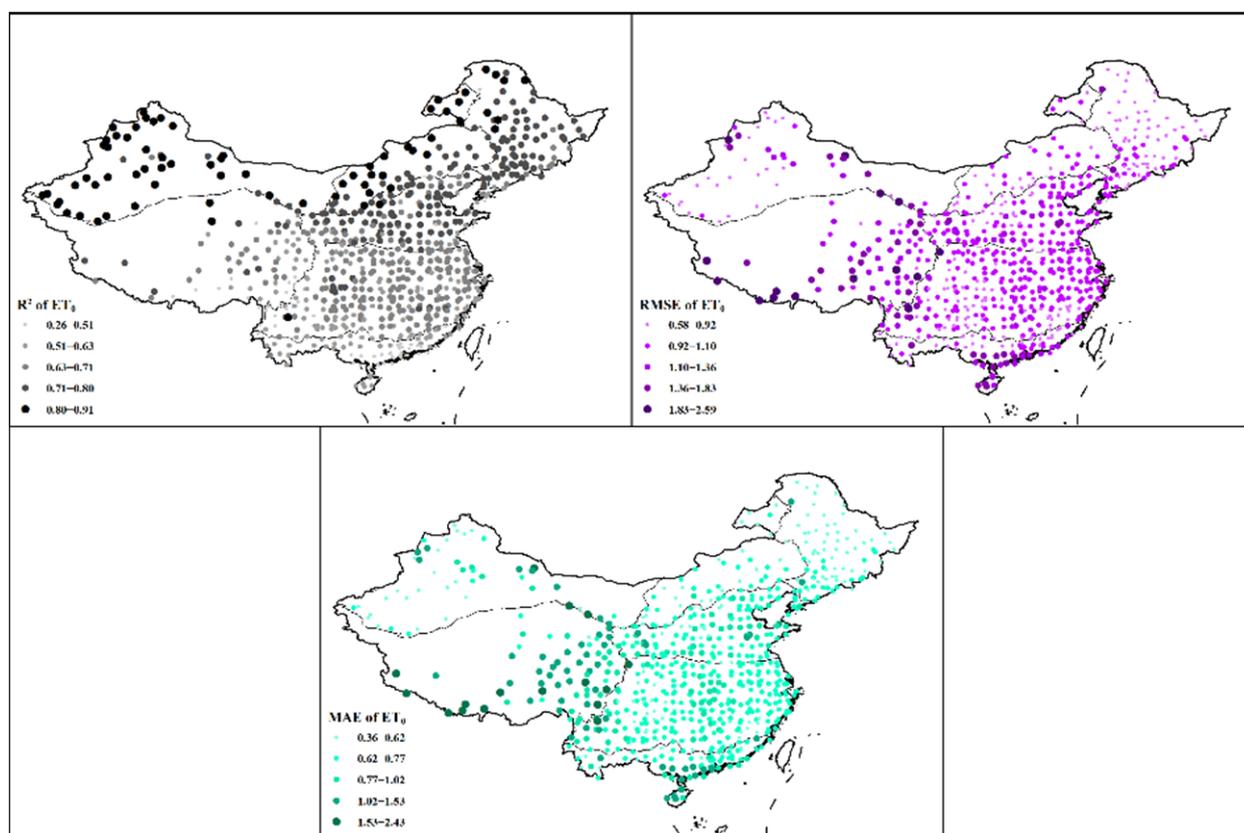
## 3.2. Reference Evapotranspiration

The statistical indicators of calculated  $ET_0$  based on the CLDAS dataset are shown in Table 4. Among all climate zones, climate zone 1 had the best consistency ( $R^2 = 0.84$ ) between CLDAS data and station data, while climate zone 3 showed the lowest errors (RMSE =  $0.87 \text{ mm d}^{-1}$  and MAE =  $0.58 \text{ mm d}^{-1}$ ). For climate zone 7, the values of RMSE ( $1.37 \text{ mm d}^{-1}$ ) and MAE ( $1.19 \text{ mm d}^{-1}$ ) were higher than the corresponding values in any of the other climate zones. Figure 5 shows the spatial distribution of statistical indicators. Across climate zones,  $R^2$  overall showed a decreasing trend from the north to the south,

and the southernmost region (i.e., climate zone 6) had the lowest value of  $R^2$ . However, the spatial distributing patterns of RMSE and MAE were different from  $R^2$ . The stations with significant errors are mainly distributed west of climate zone 1, the coastal areas, and the boundary areas between climate zone 7 and other climate zones. This is mainly due to the more complex climate change between climate zones. In addition, the high wind speed error in the coastal areas often leads to a significant  $ET_0$  error.

**Table 4.** Statistical indicators of reference evapotranspiration in different climate zones of China.

Zone	RMSE	MAE	$R^2$
1	1.10	0.74	0.84
2	0.94	0.63	0.80
3	0.87	0.58	0.74
4	1.03	0.75	0.71
5	0.99	0.72	0.64
6	1.08	0.83	0.52
7	1.37	1.19	0.62



**Figure 5.** Spatial distribution of  $ET_0$  statistical indicators.

To explore the differences in the CLDAS data in different climate regions, one station from each climate zone was randomly selected to fit the correlation between the calculated  $ET_0$  based on CLDAS and the FAO56-PM  $ET_0$  (Figure 6). Although there were a few outliers, the scatter points in climate zone 1 were more concentrated to the 1:1 line than those in other climate zones. Scatter points in climate region 2 were slightly more dispersed than in climate zone 1 and showed some obvious overestimations when  $ET_0$  was more significant than 6. In climate zone 3, the accuracy was excellent when the value of  $ET_0$  was low (<2 mm) but showed a decline as the following scatter points started to discretize. However, no overestimation or underestimation existed. In climate zone 4, the error was relatively large when  $ET_0$  ranged from 3 mm to 6 mm. When the  $ET_0$  of climate zone 5 was less than 2, the

problem of underestimation appeared, and then the points were scattered in the 1:1 line for two measurements, but the distance from the 1:1 line was far. In climate zone 6, the error was significant when  $ET_0$  was greater than 3 mm, and some scatter points were obviously overestimated or underestimated. Although the points were not as discrete as those in climate zones 4 and 5, the  $ET_0$  of climate zone 7 showed a significant underestimation.

Figure 7 shows the  $ET_0$  box diagram of a station randomly selected from each climate zone. From the median value, there are differences in the performance of different climate regions. Among them, the  $ET_0$  prediction bias of climate region 2 and climate region 3 is slight. The bias of climate region 6 and 7 is large. In addition, from the extreme value, the bias of  $ET_0$  estimated in climate zone 4 and climate zone 6 is small, and other regions have overestimated or underestimated in varying degrees. From the quartile line, there are significant differences in  $ET_0$  estimation in climate regions 5, 6, and 7. The predicted  $ET_0$  performance of climate zones 1, 2, 3, and 4 is relatively good.

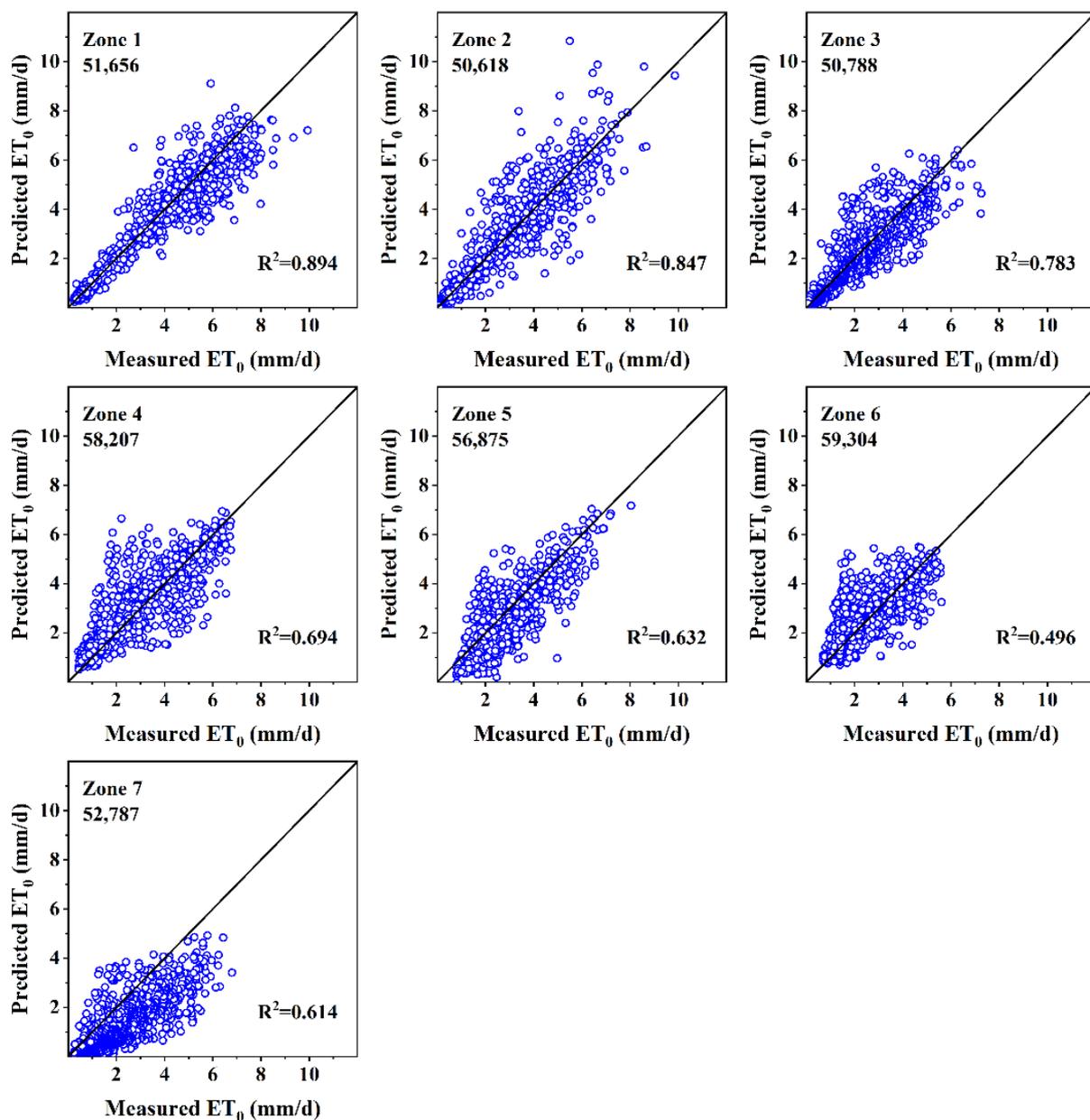
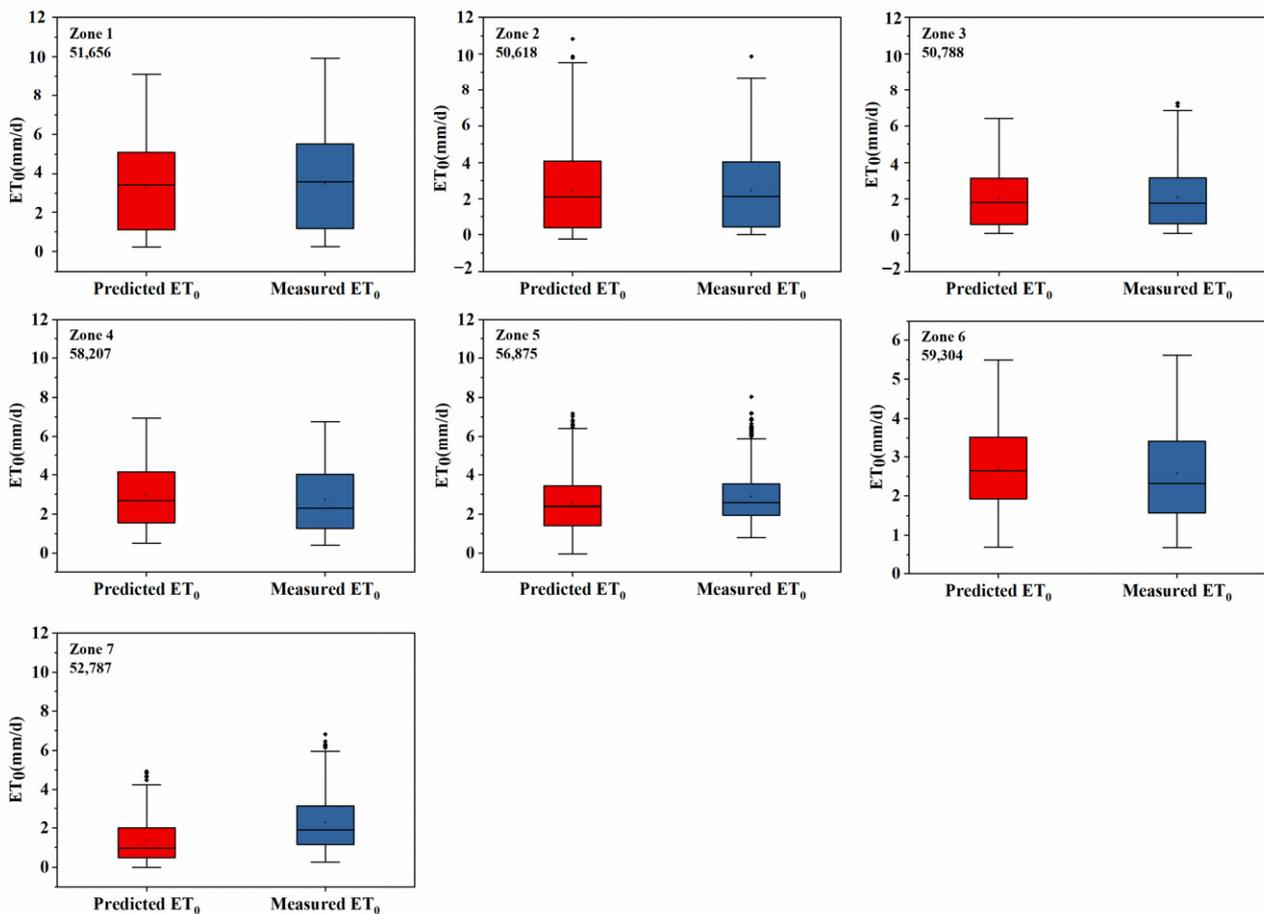


Figure 6. Scatter plots of CLDAS and FAO56 PM values of  $ET_0$  in different climates.

### 3.3. Seasonal Performance of Reference Evapotranspiration from CLDAS

Since the demand for water resources varies significantly between seasons, it is necessary to assess the performance of the CLDAS dataset in different seasons. Figure 8 shows the RMSE performance of CLDAS  $ET_0$  in the four seasons. In spring (March–May), stations with RMSE smaller than  $1.5 \text{ mm d}^{-1}$  accounted for more than 85% of all stations across China. The RMSE was lower in the south of climate zone 1 and the middle and north of climate zone 3, ranging from  $0.5$  to  $1 \text{ mm d}^{-1}$ . For most stations of climate zones 2, 4, 5, and 6, RMSE values ranged from  $1$ – $1.5 \text{ mm d}^{-1}$ . Stations with errors greater than  $1.5 \text{ mm d}^{-1}$  are mainly located in climate zones 1 and 7.



**Figure 7.** The box diagram of CLDAS and FAO56 PM values of  $ET_0$  in different climates.

In summer (June–August), the RMSE of CLDAS  $ET_0$  was generally higher than that of spring. More than 80% of the stations had RMSE ranging between  $1 \text{ mm d}^{-1}$  and  $1.5 \text{ mm d}^{-1}$ . Stations with RMSE smaller than  $1 \text{ mm d}^{-1}$  were mainly concentrated in the southern part of China and near the boundary area between climate zones 5 and 6. Stations with RMSE greater than  $1.5 \text{ mm d}^{-1}$  were distributed in all climatic regions, of which climate zone 7 had the largest RMSE, followed by climate zone 1. Especially for the southwest area of climate zone 7, stations in this area were sparse, and the error was relatively large, with the value of RMSE larger than  $2 \text{ mm d}^{-1}$ .

In autumn (September–November), RMSE was less than  $1 \text{ mm d}^{-1}$  in 80% of all stations, and stations with a significant error were still mainly concentrated in climate zone 7. It is worth mentioning that there were also many stations with RMSE greater than  $1 \text{ mm d}^{-1}$  in the coastal areas of climatic zone 6. This is mainly due to the relatively high temperature of this area in autumn, resulting in a relatively large RMSE.

In winter (December–February), RMSE in northern regions (i.e., climate zones 1–3) was lower than  $0.5 \text{ mm d}^{-1}$  due to the minimal  $ET_0$  value. RMSE of most stations in climate

zones 4 and 5 was less than  $1 \text{ mm d}^{-1}$ . However, the values of RMSE in the southern part of climate zone 7, the coastal part of climate zone 6, and the western part of climatic zone 5 were more outstanding than  $1.5 \text{ mm d}^{-1}$ .

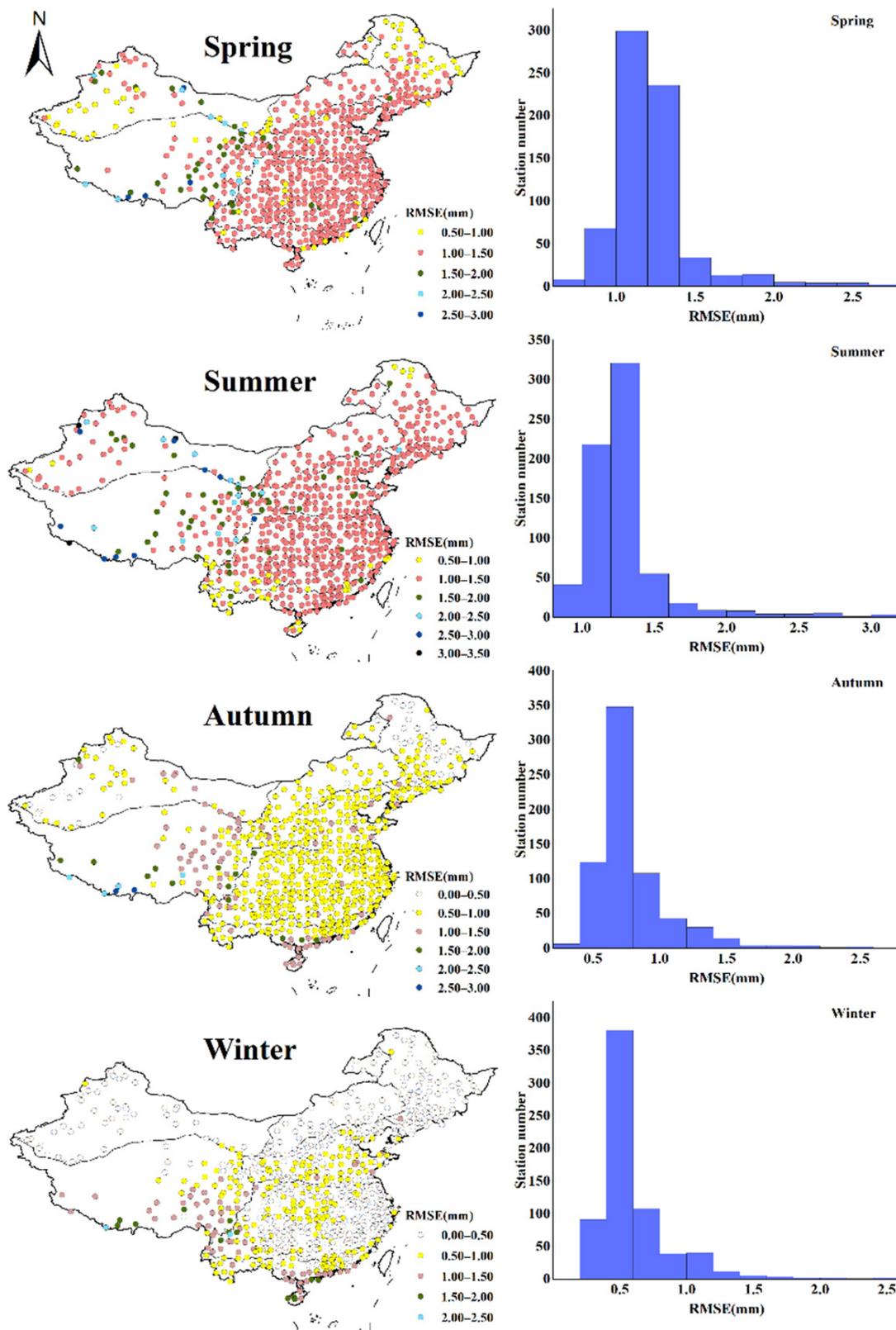


Figure 8. Seasonal RMSE of  $ET_0$  calculated from CLDAS dataset.

Among all seasons, summer had the most significant RMSE error, followed in order by spring, autumn, and winter. The CLDAS dataset performed well in climate zones 2, 3, 4, and 5, but performed poorly in all seasons in climate zone 7. In addition, the coastal areas of climate zone 6 also did not perform well in autumn and winter.

Because the demand for water resources varies significantly in different seasons, it is also necessary to evaluate the specific overestimation or underestimation of the CLDAS dataset in different seasons. This provides a more detailed reference for practical production and life applications. Figure 9 shows the PBias distribution of  $ET_0$  calculated by CLDAS in the four seasons. In spring, the sites with PBias between 0.2 and 0.2 accounted for about 70% of all sites in the country, and the overall forecast stability was good. The values of  $ET_{0CLD}$  in the southern regions of climate zone 1, climate zone 2, the southern part of climate zone 3, most of climate zone 4, and the central and northern parts of climate zone 5 are within 10% of the local station data. In climate zone 7 (Underestimated), numerical biases are generally greater than 30%. The prediction of  $ET_{0CLD}$  in the junction area of climate zone 7 and other climate zones is not very stable, and most of them are underestimated. In addition, the  $ET_{0CLD}$  in coastal areas will have a relatively large bias.

In summer, the bias of  $ET_0$  calculated by CLDAS is generally smaller than that in spring, but some sites have large fluctuations (the bias is greater than 60%), and the PBias of more than 60% of the sites is between  $-10\%$  and  $10\%$ .  $ET_{0CLD}$  is in the climate zone 3. The southeastern coastal areas (overestimated), the southern part of climate zone 5 (overestimated), and the western coastal areas of climate zone 6 (overestimated) have large biases from the local weather station data, with a gap of about 10% to 30%. Numerical bias with zone 7 (underestimation) is generally greater than 30%. The prediction of  $ET_{0CLD}$  for meteorological stations in the junction of climate zone 7 and other climate zones is not very stable, and most of them are underestimated.

In autumn, the bias of  $ET_0$  calculated by CLDAS is generally larger than that in spring and autumn, and only about 50% of the sites have PBias between  $10\%$  and  $10\%$ . The western region (underestimated), the central and western regions of climate zone 5 (underestimated), the southern coastal region of climate zone 6 (underestimated), and the climate zone 7 (underestimated) have large biases from the data of local meteorological stations, with a gap of more than 30%. In addition, the prediction accuracy of the  $ET_{0CLD}$  of the meteorological stations at the junction of climate zone 3 and other climate zones decreased significantly. Most showed an overall underestimation.

In winter, the bias of  $ET_0$  calculated by CLDAS is generally the largest, among which the bias of  $ET_{0CLD}$  from the local station data in the southern region of climate zone 1, the central region of climate zone 2, the central region of climate zone 3, and the central and eastern regions of climate zone 5 is  $10\%$ . Within  $\%$ ;  $ET_{0CLD}$  in the northern region of climate zone 1 (overestimated), the southern coastal region of climate zone 3 (overestimated), most of climate zone 4 (underestimated), the central and western regions of climate zone 5 (underestimated), the southern part of climate zone 6 Coastal areas (underestimated), and climate zone 7 have large biases from local weather station data, with a gap of more than 30%. In addition, the prediction accuracy for the  $ET_{0CLD}$  of the meteorological stations in the transition areas of different climatic zones will drop significantly, and both overestimation and underestimation exist.

Figure 10 shows a boxplot of the calculated PBias for the CLDAS dataset. From the median value of PBias, there are bias in the performance of different seasons. Among these, the estimated bias in spring and summer is smaller, the performance in autumn is second, and the performance in winter is the largest. From the quartile line (aside from winter), the estimated differences in the other three seasons were small. From the perspective of extreme values, the estimated maximum and minimum values in winter are not good. The performance in summer is the best, and the estimated bias is the smallest. In conclusion, of all seasons, summer and spring have the slightest bias, followed by autumn and winter. The CLDAS dataset performs well in climate zones 2, 3, 4, and 5 but not in all seasons in climate zone 7. In addition, the coastal areas of climate zone 3 and climate zone 6 also

performed poorly in autumn and winter, and the performance at the interface of different climate zones was also relatively poor.

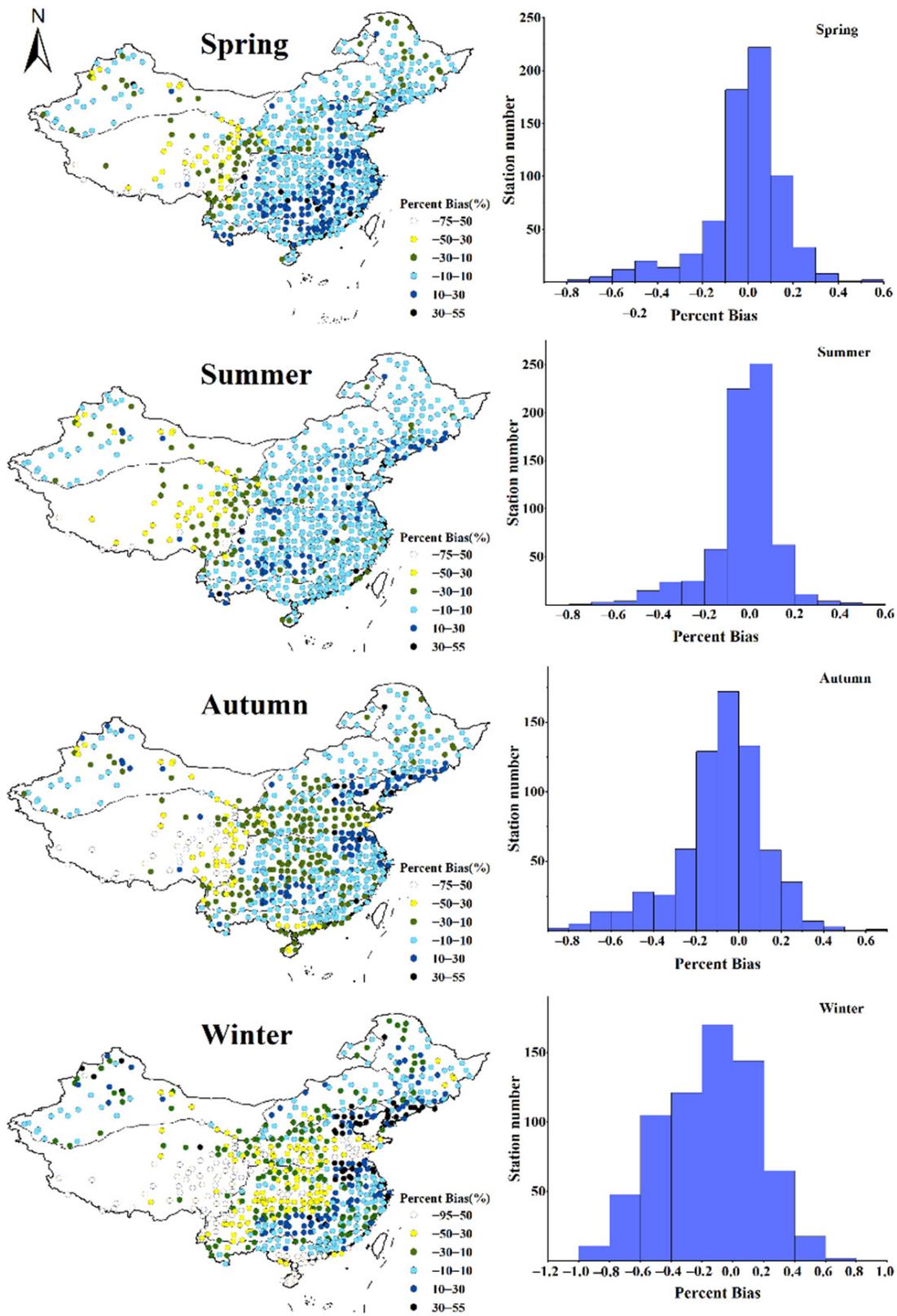
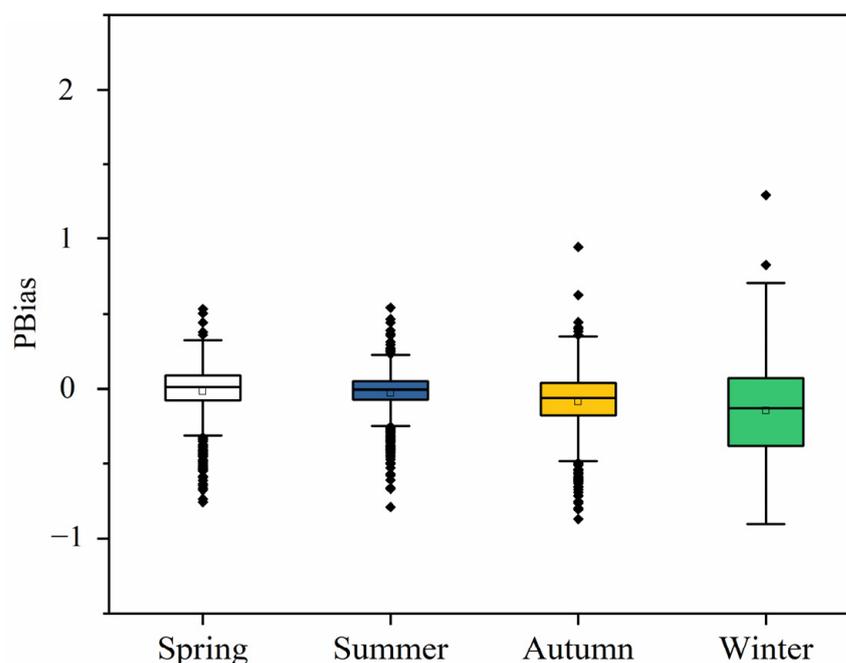


Figure 9. Seasonal PBias of  $ET_0$  calculated from CLDAS dataset.



**Figure 10.** The box diagram of Seasonal PBias of  $ET_0$  calculated from CLDAS dataset.

### 3.4. Annual Performance of Reference Evapotranspiration from CLDAS

China is a country with frequent droughts and floods. Water demand also varies widely between years. Therefore, it is necessary to evaluate the difference in CLDAS  $ET_0$  error in different years. The RMSE of CLDAS  $ET_0$  in 2017–2020 is shown in Figure 11. Overall, RMSE in 2019 was lower than that in other years, with more than 85% of all stations having a value less than  $1 \text{ mm d}^{-1}$ . In 2020, stations with RMSE less than  $1 \text{ mm d}^{-1}$  accounted for 60% of the total stations. Regarding the spatial distribution of errors, climate zones 3, 4, 5, and 6 overall performed better than other climate zones. Significant errors were in the southern part of climate zone 7 and the eastern part of climate zone 1, with RMSE generally more significant than  $1.5 \text{ mm d}^{-1}$ . This may be related to the special geographical location of these stations, such as at the boundary of different climate zones. The above results further confirmed that the data set had good performance in some regions. At the same time, there were also significant uncertainties in other regions, which could bring certain risks to the application.

Figure 12 shows the PB distribution of  $ET_0$  calculated by CLDAS in 2019–2020. In 2017, the sites with PBias between 0.1 and 0.1 accounted for about 60% of all sites in the country, and the overall forecast stability was good. The values of  $ET_{0\text{CLD}}$  in the southern region of climate zone 1, climate zone 2, the central and northern regions of climate zone 3, the central region of climate zone 4, the central and northern regions of climate zone 5, and the central region of climate zone 6 are compared with local weather station data. The bias is within 10%; the bias of  $ET_{0\text{CLD}}$  from the local weather station data is larger in the northern area of climate zone 1, the southern coastal area of climate zone 3 (overestimated), and the southern coastal area of climate zone 6 (underestimated), with a gap of 10% to 30%. However, the numerical bias of climate zone 7 (underestimated) is generally greater than 30%. The prediction of  $ET_{0\text{CLD}}$  in the junction area between climate zone 7 and other climate zones is not very stable, and most of them are underestimated.

In 2018, the bias of  $ET_0$  calculated by CLDAS was similar to that in 2017, with about 60% of sites having a PBias between 10% and 10%, and  $ET_{0\text{CLD}}$  in the central region of climate zone 1 and a few in the southern part of climate zone 3. The coastal areas (overestimated), parts of the southern part of climate zone 5 (overestimated), and the southern coastal areas of climate zone 6 (overestimated) have large biases from local weather station data, with a gap of about 10% to 30%. The (underestimated) numerical

bias is generally greater than 30%. Similarly, the prediction of  $ET_{0CLD}$  for meteorological stations in the junction of climate zone 7 with other climate zones is not very stable, and most of them are underestimated.

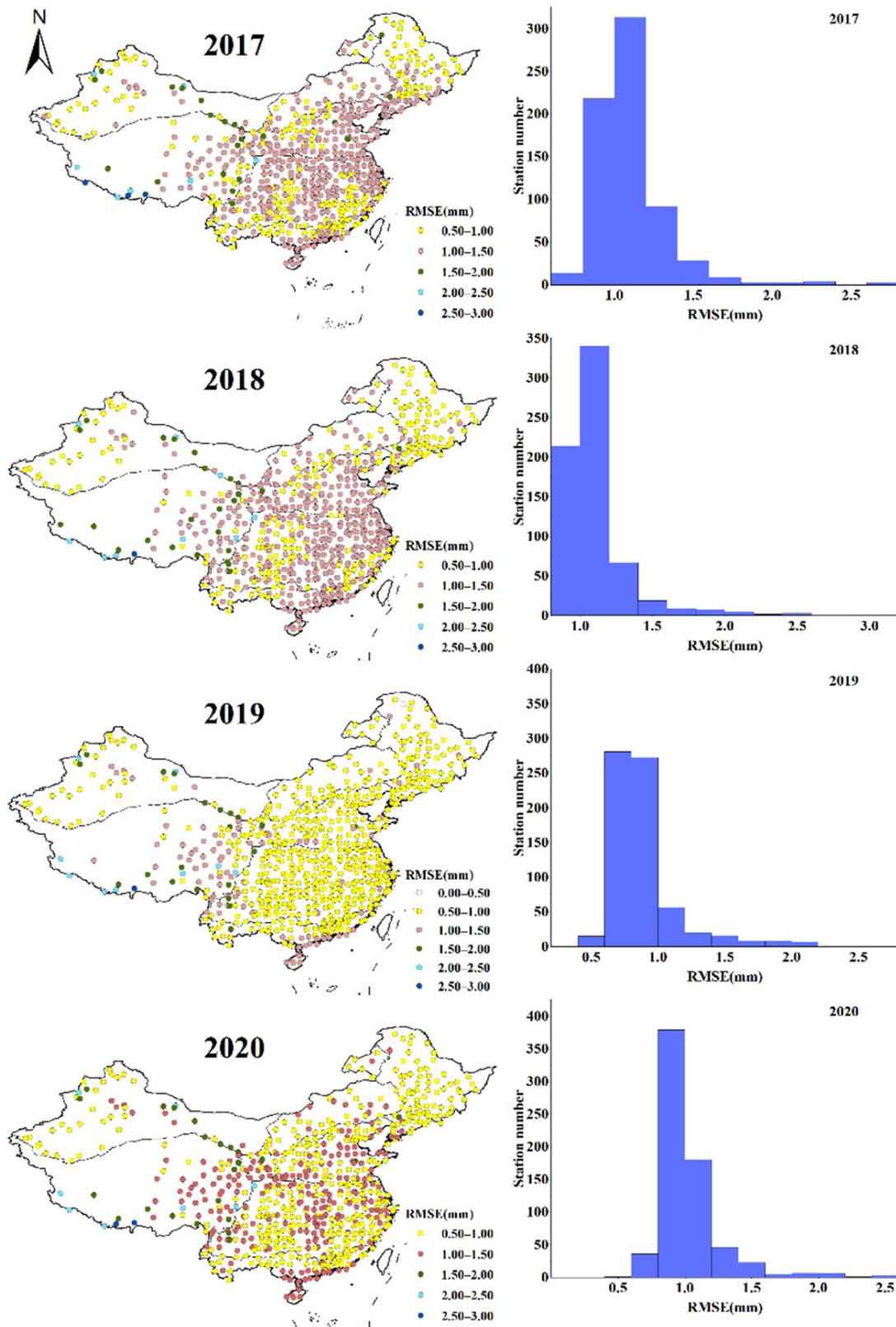


Figure 11. Annual RMSE of  $ET_0$  calculated from CLDAS dataset.

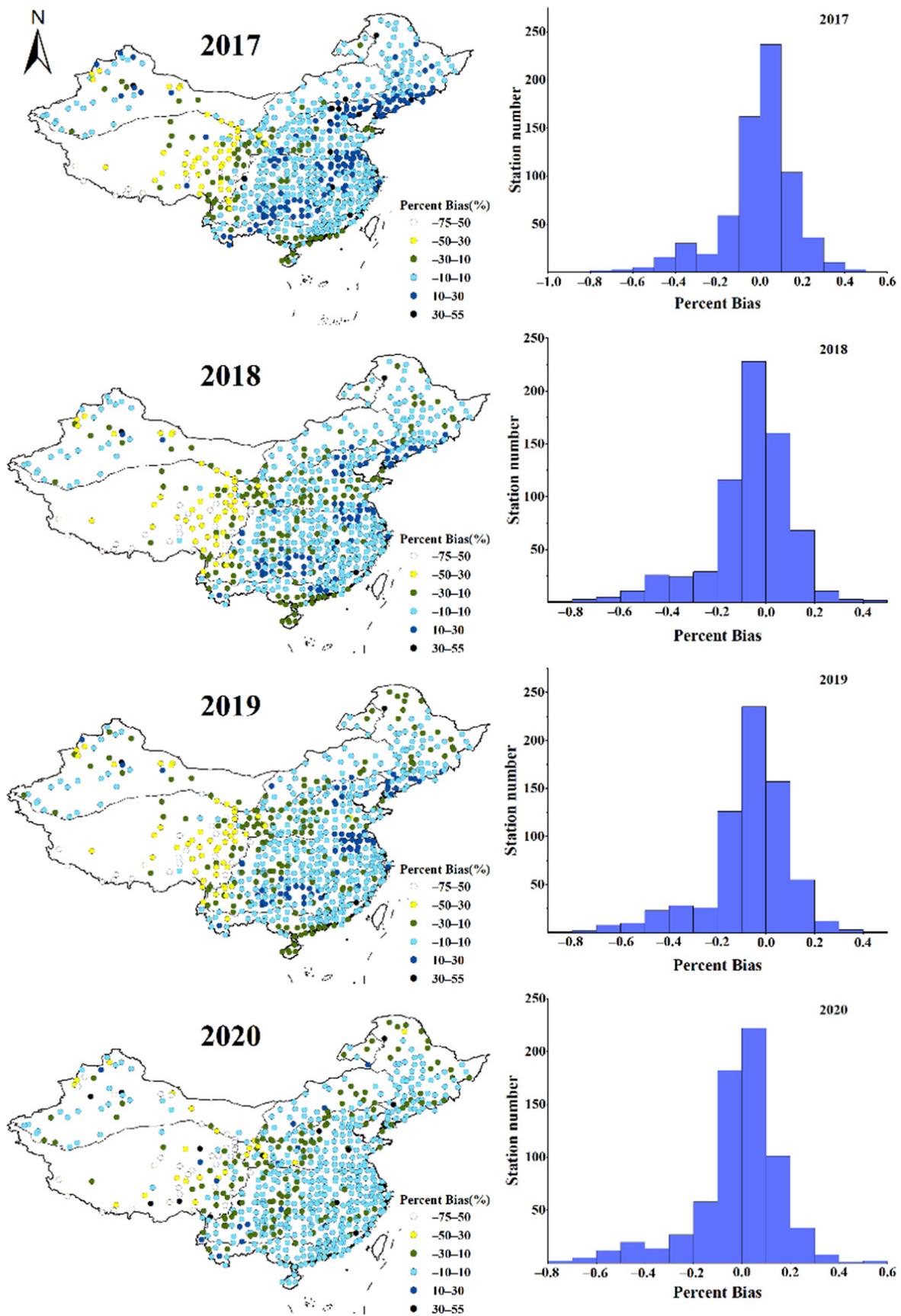
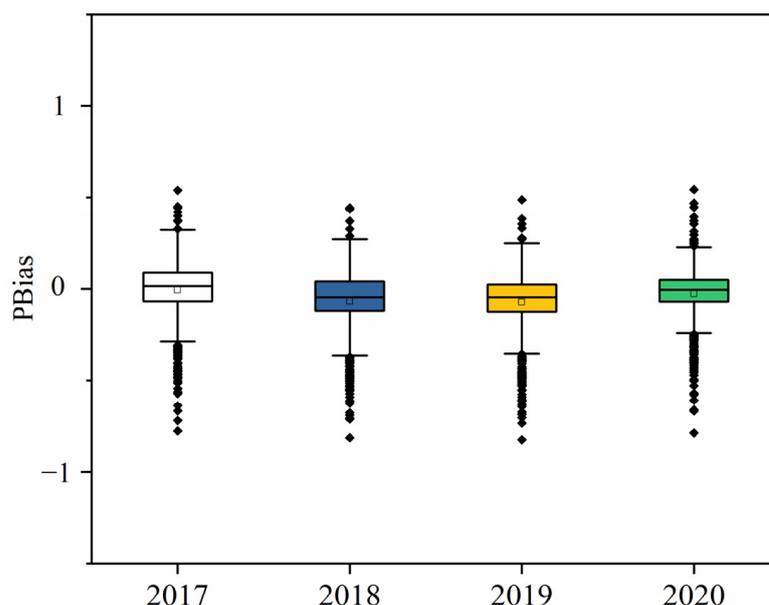


Figure 12.  $ET_0$  annual PBias distribution calculated from CLDAS dataset.

In 2019, about 55% of the sites had PBias between  $-10\%$  and  $10\%$ . The  $ET_{0CLD}$  in the southern part of climate zone 1, most of climate zone 2, the middle part of climate zone 3, small parts of the central region of climate zone 4, the central and eastern regions of climate zone 5, and the central region of climate zone 6 within  $10\%$  of the data from local weather stations.  $ET_{0CLD}$  is in the central region of climate zone 1, the southern coastal (overestimated) and northern regions (underestimated) of climate zone 3, the central and western regions of climate zone 4 (underestimated), and the southern part of climate zone 5 (overestimated). The bias between the southern coastal areas of climate zone 6 (overestimated) and the local weather station data is relatively large, about  $10\%$  to  $30\%$ . The numerical bias of climate zone 7 (underestimated) is generally greater than  $30\%$ .

In 2020, the bias of  $ET_0$  calculated by CLDAS was generally the smallest, and about  $60\%$  of the sites have PBias between  $10\%$  and  $10\%$ .  $ET_{0CLD}$  is in the southern part of climate zone 1 and the southwest of climate zone 2. The central region of climate zone 3, the central and southern regions of climate zone 4, most of climate zone 5, and most of climate zone 6 were within  $10\%$  of data from local weather stations.  $ET_{0CLD}$  is in the central region of climate zone 1, a few southern coastal (overestimated) and northern regions (underestimated) of climate zone 3, the central and western regions of climate zone 4 (underestimated), and a small number of regions in climate zone 5 (underestimated). The bias of local weather station data is large, with a gap of about  $10\%$  to  $30\%$ . The numerical bias of climate zone 7 (underestimated) was generally greater than  $30\%$ . Similarly, the prediction of  $ET_{0CLD}$  for meteorological stations in the junction of climate zone 7 with other climate zones is not very stable, and most of them are underestimated.

Figure 13 shows a boxplot of annual PBias calculated for the CLDAS dataset. From the median value of PBias, there is a bias in the performance of different years. The estimated bias in 2020 and 2017 is smaller, and the performance of the other two years is relatively poor. From the quartile line, the estimated bias in 2020 is the smallest and more compact, and the estimated differences in other years have different degrees of fluctuation. From the perspective of extreme values, there was a clear overestimation in 2017 and a clear underestimation in 2018. Additionally, 2020 had the best performance, and the estimated bias was the smallest. In conclusion, from 2017 to 2020, bias in 2019 and 2020 was the smallest. The CLDAS dataset performs well in climate zones 2, 3, 4, and 5 but not in all seasons in climate zone 7. In addition, the coastal areas of climate zone 3 and climate zone 6 also performed poorly except in 2020. The performance of the boundary areas of different climate zones was also relatively poor.



**Figure 13.** The box diagram of annual PBias of  $ET_0$  calculated from CLDAS dataset.

### 3.5. Main Factors Affecting Reference Crop Evapotranspiration

The evapotranspiration process is affected by many factors [43], and its changes are mainly attributed to the changes in meteorological factors. The country's climate is complex and diverse. From a geographical point of view, the eastern part is mostly a monsoon climate zone with a complex and changeable climate. The air above it is severely polluted, affecting solar radiation and surface wind speed. Therefore, the performance of estimated  $ET_0$  in coastal areas will decline. The northwest region is far from the sea, is a non-monsoon region, and belongs to a temperate continental climate. The ground topography in this region (the junction of climate zones 2, 4, and 5) is complex and changeable, and the wind speed is greatly affected by the roughness of the underlying surface. The direction of wind movement is accurately simulated, so the reduction in the accuracy of wind speed is likely to lead to a decline in the performance of estimating  $ET_0$  in some areas. The Qinghai-Tibet Plateau belongs to the plateau climate zone. Due to its complex and changeable terrain, the climate itself on the Qinghai-Tibet Plateau will fluctuate depending on the region, which greatly affects the estimation of  $ET_0$ . To sum up, the closer are to inland areas (such as climate zones 1, 2, and 3), the higher the accuracy of  $ET_0$  estimation will be. The performance of  $ET_0$  estimation in coastal areas, the Qinghai-Tibet Plateau, and the junction of climatic zones will be negative effects [44–46]. From the seasonal point of view, the summer is affected by the warm and humid air from the ocean, with high temperature, humidity, and rain. The climate is oceanic, so the estimation error in summer is the largest. In winter, affected by the dry and cold airflow from the continent, the climate is cold, dry, and less rainy, and the climate is continental, and the estimation error will be relatively small [47,48]. In addition, specific regions need to be further analyzed according to the actual situation.

## 4. Discussion

The calculation of  $ET_0$  is affected by a variety of climatic factors. Ma et al. (2010) [49] studied the influence of main climatic factors on  $ET_0$  in mountainous plateau areas and found that the change of wind speed had the most significant impact on  $ET_0$  at each site. Luo et al. (2010) [50] conducted a sensitivity analysis on  $ET_0$  and main meteorological factors in the main agricultural areas of Tibet, and the results showed that  $ET_0$  in the whole region had a declining trend over the past 50 years. The meteorological element that had the most significant impact on  $ET_0$  was  $R_s$ . Similar results were obtained in our study, where the accuracy of  $ET_0$  was affected by the error of  $R_s$ . Xie et al. (2017) [51] analyzed the impact of changes in meteorological factors on  $ET_0$  in China's main grain-producing areas from 1961 to 2013, in which  $ET_0$  showed a saw-tooth decline. The changing characteristics of main meteorological factors in the study area and the response of  $ET_0$  are similar to the results of our study, showing regional and seasonal variations. Overall, our study suggests that the errors of meteorological factors in the Qinghai-Tibet Plateau region and the boundary region of the climate zones are more significant than in other regions, with the highest errors observed in summer.

Due to the incomplete understanding of the physical mechanism of weather changes and limited observational data, there is still a specific error in the reanalysis data [52], and the magnitude of this error tends to vary with different climatic factors. Temperature is a meteorological variable with minor errors, usually less than 10% [53,54]. Similar results were found in our study, in which the  $R^2$  of  $T_{\max}$  and  $T_{\min}$  are generally greater than 0.9 in the seven climatic zones. Due to the influence of topography, the errors of wind speed and relative humidity are usually large [26], and similar results were obtained in our study.

It is worth mentioning that the weather stations in our study are affiliated with the China Meteorological Administration. The ground of the weather station is usually covered with short grass under adequate irrigation conditions. However, areas in the grid system do not necessarily have lush vegetation. Therefore, there might be some differences in the environmental factors between the two types of systems, especially for the radiant energy (i.e.,  $R^2 < 0.65$  in the seven climate zones for the  $R_{s,CLD}$  estimation in our study).

This may lead to the problem of overestimating or underestimating the reanalysis data, which indirectly explains why the estimated  $ET_0$  values in some areas fluctuate severely in our study. In addition, the variation of wind speed is greatly affected by the terrain and the type of underlying surface, and it is not easy to obtain the average wind speed in a specific area. Similar results were obtained in our study, where the overall  $U_{CLA}$  accuracy is not satisfactory.

Finally, this study can provide an idea for economically underdeveloped countries and contribute to improving the reanalysis data set and the accuracy of  $ET_0$  estimation. Therefore, when other developing countries establish regional climate models, they can consider their own terrain and climate characteristics and establish a more local model.

## 5. Conclusions

$ET_0$  data set based on reanalysis products can make up for the time discontinuity and spatial insufficiency of surface meteorological platform data, which is significant for water resources planning and irrigation system formulation. However, a rigorous evaluation of reanalysis products must be carried out to see if they have value in application. This study evaluated the ability of the CLDAS dataset officially published by the Chinese meteorological system for  $ET_0$  estimation. Results indicate that the temperature data of CLDAS have high accuracy in all regions except the Qinghai Tibet Plateau (QTP) region. In contrast, the accuracy of the total radiation data is average, and the quality of relative humidity and wind speed data is poor. The overall accuracy of  $ET_0$  is acceptable except for QTP, but there are many stations with large errors. Among seasons, RMSE is the largest in summer and smallest in winter. There are also inter-annual differences in the  $ET_0$  of this data set. Overall, the CLDAS dataset is expected to have good applicability in the Inner Mongolia Grassland area, Northeast Taiwan, the Semi-Northern Temperate zone, the Humid and Semi Humid warm Temperate zone, and the subtropical region. However, there are certain risks in other regions. In addition, of all seasons, summer and spring have the slightest bias, followed by autumn and winter. From 2017 to 2020, bias in 2019 and 2020 are the smallest, and the areas with large deviation are in the south of climate zone 3, the coastal area of climate zone 6, and the boundary area of climate zone 7.

**Author Contributions:** Conceptualization, L.-F.W. and L.Q.; methodology, L.-F.W., G.-M.H. and L.Q.; data curation, L.Q. and X.-G.L.; writing—original draft preparation, L.-F.W., Y.-C.W., G.-M.H. and L.Q.; writing—review and editing, L.-F.W., H.B., X.-G.L. and L.Q.; project administration, L.-F.W. and S.-F.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Science and Technology Project of Jiangxi Provincial Department of Education (GJJ180925), National Natural Science Foundation of China (51979133 and 51769010) and Natural Science Foundation of Jiangxi province in China (20181BBG78078 and 20212BDH80016).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

Meaning of main acronyms:

CLDAS: China Meteorological Administration Land Data Assimilation System;  $ET_0$ : Reference Crop evapotranspiration; FAO: Food and Agriculture Organization of the United Nations; ECMWF: European Center for Medium Weather Forecasting; NCEP: National Centers for Environmental Prediction;  $R_s$ : Global solar radiation; U: wind speed at 2 m;

RH: relative humidity;  $T_{\max}$ : maximum temperature;  $T_{\min}$ : minimum temperature. (When the subscript CLD exists in these meteorological data, it is the corresponding CLDAS meteorological data); RMSE: Root Mean Square Error; MAE: Mean Absolute Error; PBias: percent bias;  $R^2$ : coefficient of determination.

## References

- Mao, Y.P.; Fang, S.Z. Research of reference evapotranspiration's simulation based on machine learning. *J. Geo-Inf. Sci.* **2020**, *22*, 1692–1701. [[CrossRef](#)]
- Feng, Y.; Peng, Y.; Cui, N.; Gong, D.; Zhang, K. Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Comput. Electron. Agric.* **2017**, *136*, 71–78. [[CrossRef](#)]
- Feng, Y.; Jia, Y.; Cui, N.; Zhao, L.; Li, C.; Gong, D. Calibration of Hargreaves model for reference evapotranspiration estimation in Sichuan basin of southwest China. *Agric. Water Manag.* **2017**, *181*, 1–9. [[CrossRef](#)]
- Wang, Y.; Sheng, L.X.; Li, K.; Sun, H.Y. Analysis of present situation of water resources and countermeasures for sustainable development in China. *J. Water Resour. Water Eng.* **2008**, *3*, 10–14.
- Jia, Y.; Cui, N.; Wei, X.; Gong, D.; Hu, X. Applicability evaluation of different algorithms for reference crop evapotranspiration in Yangtze River Basin based on inverse distance weighted method. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 130–138. [[CrossRef](#)]
- Lei, G.; Zeng, W.; Zhu, J.; Zha, Y.; Fang, Y.; Song, Y.; Chen, M.; Qian, Y.; Wu, J.; Huang, J. Quantification of leaf growth, height increase, and compensatory root water uptake of sunflower in heterogeneous saline soils. *Agron. J.* **2019**, *111*, 1010–1027. [[CrossRef](#)]
- Li, L.A.; Qiu, R.J.; Liu, C.W. Comparison and improvement of estimation models for the reference evapotranspiration using temperature data. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 123–130. [[CrossRef](#)]
- Paredes, P.; Martins, D.S.; Pereira, L.S.; Cadima, J.; Pires, C. Accuracy of daily estimation of grass reference evapotranspiration using ERA-Interim reanalysis products with assessment of alternative bias correction schemes. *Agric. Water Manag.* **2018**, *210*, 340–353. [[CrossRef](#)]
- Li, C. Performance of revised Hargreaves models in hilly area of Central Sichuan Basin. *Water Sav. Irrig.* **2021**, *11*, 88–96. [[CrossRef](#)]
- Dong, J.H.; Liu, X.G.; Wu, L.F.; Huang, G.M.; Yang, Q.L. Cross-station adaptability of  $ET_0$  based on machine learning. *J. Northwest Agric. For. Univ. (Nat. Sci. Ed.)* **2021**, *49*, 144–154. [[CrossRef](#)]
- Zhu, B.; Feng, Y.; Gong, D.Z.; Jiang, S.Z.; Zhao, L.; Cui, N.B. Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. *Comput. Electron. Agric.* **2020**, *173*, 105430. [[CrossRef](#)]
- Huang, C.X.; Zhao, D.M.; Wang, B.F. A comparative study on calculation methods of reference crop evapotranspiration in the semi-arid region of Central Eastern Gansu. *Agric. Res. Arid Areas* **2018**, *36*, 41–47. [[CrossRef](#)]
- Mahto, S.S.; Mishra, V. Does ERA-5 outperform other reanalysis products for hydrologic applications in India? *J. Geophys. Res. Atmos.* **2019**, *124*, 9423–9441. [[CrossRef](#)]
- Baatz, R.; Franssen, H.J.H.; Euskirchen, E.; Sihi, D.; Vereecken, H. Reanalysis in Earth System Science: Toward Terrestrial Ecosystem Reanalysis. *Rev. Geophys.* **2021**, *59*, e2020RG000715. [[CrossRef](#)]
- Fennel, K.; Gehlen, M.; Brasseur, P.; Brown, C.W.; Ciavatta, S.; Cossarini, G. Advancing marine biogeochemical and ecosystem reanalyses and forecasts as tools for monitoring and managing ecosystem health. *Front. Mar. Sci.* **2019**, *6*, 89. [[CrossRef](#)]
- Peylin, P.; Bacour, C.; MacBean, N.; Leonard, S.; Rayner, P.; Kuppel, S. A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle. *Geosci. Model Dev.* **2016**, *9*, 3321–3346. [[CrossRef](#)]
- Hobeichi, S.; Abramowitz, G.; Evans, J.; Beck, H.E. Linear optimal runoff aggregate (LORA): A global gridded synthesis runoff product. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 851–870. [[CrossRef](#)]
- Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horanyi, A.; Muñoz-Sabater, J. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
- Sabater, J.M.; Dutra, E.; Panareda, A.; Albergel, C.; Thepaut, J.N. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [[CrossRef](#)]
- Boulard, D.; Castel, T.; Camberlin, P.; Sergent, A.S.; Bréda, N.; Badeau, V.; Rossi, A.; Pohl, B. Capability of a regional climate model to simulate climate variables requested for water balance computation: A case study over northeastern France. *Clim. Dyn.* **2016**, *46*, 2689–2716. [[CrossRef](#)]
- Srivastava, P.K.; Han, D.; Islam, T.; Petropoulos, G.P.; Gupta, M.; Dai, Q. Seasonal evaluation of evapotranspiration fluxes from MODIS satellite and mesoscale model downscaled global reanalysis datasets. *Theor. Appl. Climatol.* **2016**, *124*, 461–473. [[CrossRef](#)]
- Pelosi, A.; Terribile, F.; D'Urso, G.; Chirico, G.B. Comparison of ERA5-Land and UERRA MESCAN-SURFEX reanalysis data with spatially interpolated weather observations for the regional assessment of reference evapotranspiration. *Water* **2020**, *12*, 1669. [[CrossRef](#)]
- Woldesenbet, T.; Elagib, N. Spatial-temporal evaluation of different reference evapotranspiration methods based on the climate forecast system reanalysis data. *Hydrol. Process.* **2021**, *35*, e14239. [[CrossRef](#)]
- Song, Y.; Su, X.L.; Niu, J.P.; Cui, C.F. Temporal and spatial characteristics and forecasting of reference crop evaporation in Shaanxi. *J. Northwest Agric. For. Univ. (Nat. Sci. Ed.)* **2015**, *43*, 225–234. [[CrossRef](#)]

25. Liu, Z.; Lu, J.Z.; Huang, J.; Chen, X.; Zhang, L.; Sheng, Y. Prediction and trend of future reference crop evapotranspiration in the Poyang Lake Basin based on CMIP5 Models. *J. Lake Sci.* **2019**, *31*, 1685–1697. [[CrossRef](#)]
26. Martins, D.S.; Paredes, P.; Razei, T.; Pires, C.; Cadima, J.; Pereira, L.S. Assessing reference evapotranspiration estimation from reanalysis weather products. An application to the Iberian Peninsula. *Int. J. Climatol.* **2017**, *37*, 2378–2397. [[CrossRef](#)]
27. Razei, T.; Pehkar, A. Performance evaluation of NCEP/NCAR reanalysis blended with observation-based datasets for estimating reference evapotranspiration across Iran. *Theor. Appl. Climatol.* **2021**, *144*, 885–903. [[CrossRef](#)]
28. Milad, N.; Mehdi, H. Reference crop evapotranspiration for data-sparse regions using reanalysis products—ScienceDirect. *Agric. Water Manag.* **2022**, *262*, 107319. [[CrossRef](#)]
29. Wang, X.; Ding, Y.; Zhao, C.; Wang, J. Similarities and improvements of GPM IMERG upon TRMM 3B42 precipitation product under complex topographic and climatic conditions over Hexi region, Northeastern Tibetan Plateau. *Atmos. Res.* **2019**, *218*, 347–363. [[CrossRef](#)]
30. Prakash, S.; Mitra, A.; Pai, D.S.; Aghakouchak, A. From TRMM to GPM: How well can heavy rainfall be detected from space? *Adv. Water Resour.* **2016**, *88*, 1–7. [[CrossRef](#)]
31. Liu, J.; Shi, C.; Sun, S.; Liang, J.; Yang, Z.L. Improving land surface hydrological simulations in China using CLDAS meteorological forcing data. *J. Meteorol. Res.* **2019**, *33*, 1194–1206. [[CrossRef](#)]
32. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*; FAO Irrigation and Drainage Paper; FAO: Rome, Italy, 1998.
33. Mobilia, M.; Longobardi, A. Prediction of Potential and Actual Evapotranspiration Fluxes Using Six Meteorological Data-Based Approaches for a Range of Climate and Land Cover Types. *Int. J. Geo-Inf.* **2021**, *10*, 192. [[CrossRef](#)]
34. Takakura, T.; Kubata, C.; Sase, S.; Hayashi, M.; Ishii, M.; Takayama, K.; Nishina, H.; Kurata, K.; Giacomelli, G.A. Measurement of evapotranspiration rate in a single-span greenhouse using the energy–balance equation. *Biosyst. Eng.* **2009**, *702*, 298–304. [[CrossRef](#)]
35. Fan, J.; Wu, L.; Zhang, F.; Cai, H.; Wang, X.; Lu, X.; Xiang, Y. Evaluating the effect of air pollution on global and diffuse solar radiation prediction using support vector machine modeling based on sunshine duration and air temperature. *Renew. Sustain. Energy Rev.* **2018**, *94*, 732–747. [[CrossRef](#)]
36. Yin, Y.H.; Li, B.Y. A New Scheme for Climate Regionalization in China. *Acta Geogr. Sin.* **2010**, *65*, 3–12. [[CrossRef](#)]
37. Zheng, J.Y.; Bian, J.J.; Ge, Q.S.; Hao, Z.X.; Yin, Y.H. Liao, Y.M. The climate regionalization in China for 1981–2010. *Geogr. Res.* **2013**, *32*, 987–997. [[CrossRef](#)]
38. Liu, X.; Mei, X.; Li, Y.; Wang, Q.; Jensen, J.R.; Zhang, Y.; Porter, J.R. Evaluation of temperature-based global solar radiation models in China. *Agric. For. Meteorol.* **2009**, *149*, 1433–1446. [[CrossRef](#)]
39. Fan, J.; Wu, L.; Zhang, F.; Cai, H.; Zeng, W.; Wang, X. Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China. *Renew. Sustain. Energy Rev.* **2019**, *100*, 186–212. [[CrossRef](#)]
40. Wang, Y.Q.; Zhang, X.Y.; Sun, J.Y.; Zhang, X.C.; Che, H.Z.; Li, Y.J.A.C. Spatial and temporal variations of the concentrations of PM 10, PM 2.5 and PM 1 in China. *Atmos. Chem. Phys.* **2015**, *15*, 13585–13598. [[CrossRef](#)]
41. Wang, Z.; Xie, P.; Lai, C.; Chen, X.; Wu, X.; Zeng, Z.; Li, J. Spatiotemporal variability of reference evapotranspiration and contributing climatic factors in China during 1961–2013. *J. Hydrol.* **2017**, *544*, 97–108. [[CrossRef](#)]
42. Blankenau, P.A.; Kilic, A.; Allen, R. An evaluation of gridded weather data sets for the purpose of estimating reference evapotranspiration in the United States. *Agric. Water Manag.* **2020**, *242*, 106376. [[CrossRef](#)]
43. Zhang, Y.S.; Zhao, X.Q.; Zhao, S.X.; Feng, C.B. Correlation between Evapotranspiration and Climate Factors in Warm Steppe in Source Region of Yangtze, Yellow and Yalu Tsangpo Rivers. *J. Desert Res.* **2010**, *30*, 363–368.
44. Li, S.C.; Liu, F.Y. Climate Complexity and Spatial Variation in China. *Clim. Environ. Res.* **2008**, *13*, 1. [[CrossRef](#)]
45. Dong, Y.; Hu, J.L.; Wang, J.; Chen, X. Study of Temporal and Spatial Variation of the Reference Crop Evapotranspiration in Xinjiang Uygur Autonomous Region During the Period from 1961 to 2013. *Res. Soil Water Conserv.* **2016**, *23*, 304–308 + 313. [[CrossRef](#)]
46. Huo, Z.L.; Shi, H.B.; Chen, Y.X.; Wei, Z.M.; Qu, Z.Y. Spatio-temporal variation and dependence analysis of  $ET_0$  in north arid and cold region. *Trans. Chin. Soc. Agric. Eng.* **2004**, *6*, 60–63. [[CrossRef](#)]
47. Fu, J.; Qin, J.X.; Li, Z.X.; Zhang, Z.B.; Hu, S.S. Changing Reference Evapotranspiration and Effects of Climatic Factors. *Resour. Environ. Yangtze Basin* **2018**, *27*, 7.
48. Hu, Q.; Dong, B.; Pan, X.B.; Jiang, H.F.; Pan, Z.H.; Qiao, Y.; Shao, C.X.; Ding, M.L.; Yin, M.L.; Hu, L.T. Spatiotemporal variation and causes analysis of dry-wet climate over period of 1961–2014 in China. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 124–132. [[CrossRef](#)]
49. Ma, X.; Li, J.; Gu, S.X.; Wang, J.; Liu, T.; Duan, S.; Luo, S. Research on the Impact of Main Climatic Factors on  $ET_0$  in Plateau Areas. *China Rural Water Hydropower* **2010**, *10*, 9–12.
50. Luo, H.; Cui, Y.; Duan, Z. Analysis of the Sensitivity of  $ET_0$  and the Main Meteorological Factors in Major Agricultural Regions in Tibet. In *Modern Water-Saving and Efficient Agriculture and Ecological Irrigation Area Construction*; Yunnan University Press: Kunming, China; pp. 130–136.
51. Xie, R.; Cui, N.; Li, Z.; Zhao, L.; Hu, X.; Gong, D. Spatiotemporal Variation of Main Meteorological Factors and Their Impact on Reference Crop Evapotranspiration in Main Agricultural Production Areas in China. *J. Irrig. Drain.* **2017**, *36*, 81–89. [[CrossRef](#)]

52. Nacar, S.; Kankal, M.; Okkan, U. Evaluation of the suitability of NCEP/NCAR, ERA-Interim and, ERA5 reanalysis data sets for statistical downscaling in the Eastern Black Sea Basin, Turkey. *Meteorol. Atmos. Phys.* **2022**, *134*, 39. [[CrossRef](#)]
53. Yan, S.; Wu, L.; Fan, J.; Zhang, F.; Zou, Y.; Wu, Y. A novel hybrid WOA-XGB model for estimating daily reference evapotranspiration using local and external meteorological data: Applications in arid and humid regions of China. *Agric. Water Manag.* **2021**, *244*, 106594. [[CrossRef](#)]
54. Tarek, M.; Brissette, F.; 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](#)]