Abstract: Evapotranspiration (ET) is an important element in the water and energy cycle. Potential evapotranspiration (PET) is an important measurement of ET. Its accuracy has significant influence on agricultural water management, irrigation planning, and hydrological modelling. However, whether current PET models are applicable under climate change or not, is still a question. In this study, five frequently used PET models were chosen, including one combination model (the FAO Penman-Monteith model, FAO-PM), two temperature-based models (the Blaney-Criddle and the Hargreaves models) and two radiation-based models (the Makkink and the Priestley-Taylor models), to estimate their appropriateness in the historical and future periods under climate change impact on the Yarlung Zangbo river basin, China. Bias correction methods were not only applied to the temperature output of Global Climate Models (GCMs), but also for radiation, humidity, and wind speed. It was demonstrated that the results from the Blaney-Criddle and Makkink models provided better agreement with the PET obtained by the FAO-PM model in the historical period. In the future period, monthly PET estimated by all five models show positive trends. The changes of PET under RCP8.5 are much higher than under RCP2.6. The radiation-based models show better appropriateness than the temperature-based models in the future, as the root mean square error (RMSE) value of the former models is almost half of the latter models. The radiation-based models are recommended for use to estimate PET under climate change in the Yarlung Zangbo river basin.

Keywords: potential evapotranspiration; climate change; appropriateness; Yarlung Zangbo River Basin

1. Introduction

With climate change and human activities, extreme hydrological events (such as drought and floods) happen more frequently and intensively, leading to a possible reduction of agriculture products and an increase of economic loss [1–4]. By 2050, human food demand will roughly double, due to population growth [5,6]. In response to this, improving the efficiency of use of irrigation water is useful [7,8]. Evapotranspiration (ET) plays a crucial role in both energy and water cycles, and its estimation accuracy has vital influence on irrigation planning, crop yield simulation, drought monitoring and forecasting [9–14].

The most usual expressions of ET are potential evapotranspiration (PET), actual evapotranspiration, and pan evaporation [15–17]. PET is the rate at which evapotranspiration occurs when the surface is well supplied with water [18]. Moreover, PET represents the comprehensive impact of various climate variables on the water use of vegetation [19,20]. Therefore, PET is often used in agricultural water
A number of models have been proposed to estimate PET and these models can be divided into four categories: (1) Temperature-based models, such as the Blaney-Criddle model [24]; (2) radiation-based models, including the Makkink model [25]; (3) mass transfer-based models, such as the Rohwer model [26]; (4) combined energy-mass balanced models, such as the Penman-Monteith (PM) model [27]. The PM model is used widely and has been commonly adopted as a reference, owing to the fact that it is a physical-based model and combines both energy and mass balances [28–30]. Correspondingly, its requirement of climate variables data is much stricter than other models, including relative humidity and wind speed [27]. These climate data may not be available in many regions or even may not be measured at all. Thus, models with less data demand were developed, such as models from categories 2 and 3. These simple models only need temperature data or solar radiation data [31–33]. However, they were proposed under specific climate conditions, which may have impact on their applicability for various conditions. Therefore, understanding the behavior of these models has been a major concern.

Many studies have compared the performance of different PET models under various regions and under different climate conditions for historical periods [19,31–35]. Xu and Singh (2001) [34] compared the accuracy of seven temperature-based equations in the north-western of Canada, and found out that the Blaney-Criddle, Hargreaves, and Thornthwaite models show better performance than others. The applicability of Turc and Priestley-Taylor models in Florida were studied by Douglas et al., (2009) [31] and they found that the Priestley-Taylor model’s performance appears to be superior to other models for estimating PET for a variety of land use types. Liu et al., (2017) [19] investigated the behavior of 16 PET models during a crop growing season at a semiarid site in China, and they found that the combined models (PM) showed the best performance. However, the appropriateness of PET models under climate change impact have rarely been studied, in particular for the Tibetan Plateau. Climate change has a large impact on this area, due to its special geographic location. As Yao et al., (2012) [36] pointed out, the rapid melting of perennial snow and glaciers are affecting the hydrological cycle and threatening the water security of this area. ET plays an important role in the hydrological cycle, thus the appropriateness of PET models in climate change impact analysis should be studied and will be beneficial to agriculture and hydrology research [9,10,14].

The research of climate change impact analysis depends heavily on the meteorological input data used in hydrological model simulation [37–39]. Global Climate Models (GCMs) output provides the basis for related research at various scales, from regional to global [40–42]. However, the raw GCM output is rarely used for impact studies directly, due to errors between GCM simulations and observation data that cannot be ignored [43–45]. Hence, bias correction for raw GCM output is very important for obtaining better hydrological model simulations. The objective of bias correction is to identify possible bias between observation and simulation data, and then adjust the simulation data.

In most studies, bias correction of temperature and precipitation is often paid much attention [46–48]. However, bias also exists in other forcing variables, including solar radiation, relative humidity, and wind speed. Moreover, these biases play a crucial role on hydrological modelling, especially PET estimation [49–52]. The results of Xu et al. (2014) [52] indicated that PET is very sensitive to solar radiation and relative humidity, and these variables are the main contributors to future PET changes. Therefore, bias correction was conducted for all climate variables in this study. The Quantile mapping (QM) [53] method was used to correct bias of maximum and minimum temperature and the monthly correction factor (MCF) [54] method was used for relative humidity, wind speed, and solar radiation.

In order to explore the appropriateness of PET models under climate change impact, we examined the applicability of four frequently used PET models in historical and future periods in the Yarlung Zangbo River Basin, China. The FAO-PM model was used as a benchmark. The four PET models included two temperature-based models (Blaney-Criddle and Hargreaves) and two radiation-based models (Makkink and Priestley-Taylor). The first objective of this study was to locally calibrate the parameters of the four PET models. The second objective was to evaluate the accuracy of the four models in the historical period. Correcting the bias of GCM output, including maximum and minimum
temperature, relative humidity, wind speed, and solar radiation, was the third objective. The last objective was to assess the appropriateness of PET models and PET changes under climate change in the Yarlung Zangbo River Basin, and to investigate any uncertainty related to 18 global climate models under RCP2.6 and RCP8.5.

2. Study Area and Data

2.1. Study Area

The Yarlung Zangbo River, located on the north of the Himalaya Mountains, is the highest river over the world, with the elevation varying from 149 to 7100 m. The specific geographic position is between 29°9′ N–31°9′ N and 81°9′ N–97°1′ N (Figure 1). The basin area of the Yarlung Zangbo River Basin is about 246,000 km² and the length of this river is 2057 km (the fifth longest river in China). The dominant climate of this basin is a temperate semi-arid monsoon climate and the koppen-geiger climate classification of this territory is ET (E: polar, T: polar tundra) [55]. The moisture vapor of this basin comes mainly from the Indian Ocean [56]. The annual average precipitation ranges from 378 to 736 mm. The interannual variations of precipitation is quite tiny, but the distribution during the year is uneven. Precipitation throughout the year is mainly concentrated in July, August, and September, reaching 50–80% of annual precipitation. Moreover, the spatial distribution of precipitation differs greatly, with annual precipitation in the Gudi station from the upstream less than 300 mm and Baxika station from the downstream more than 4000 mm.

The annual mean air temperature is 5.3 to 7.6 °C, and it is likely to increase in the 21st century [57]. The studies of Wei and Fang (2013) [58] indicated that the temperature of the Yarlung Zangbo River Basin is rising. The reduction in the area of the glacier is the most significant manifestation of the increase in temperature [58]. Because of limited human activities [36], the Yarlung Zangbo River Basin is a desired study area to analyze the impacts of climate change on hydrological processes, including precipitation, runoff, and evapotranspiration.

The observation data used in this study were collected from 12 meteorological stations operated by the Tibet Meteorological Administration. The time period of data is 1971–2000, also used as the baseline period. A widely used meteorological baseline period is a 30-year “normal” period, as defined by the World Meteorological Organization (WMO) [59]. The main climatic variables include maximum and minimum air temperature, precipitation, wind speed, relative humidity, and sunshine hour at daily time step. Figure 1 shows the location of the 12 meteorological stations, and their elevation ranges from 2992 to 4900 m.
The simulated data by 18 GCMs on monthly time step were downloaded from the website http://cmip-pcmdi.llnl.gov/cmip5/. The detailed information of 18 GCMs are shown in Table 1. The emission scenarios used in this study are RCP2.6 (low emission) and RCP8.5 (high emission). The future time period was chosen as 2041–2070.

Table 1. General information of the chosen Global Climate Models (GCMs).

<table>
<thead>
<tr>
<th>No.</th>
<th>GCMs</th>
<th>Developer</th>
<th>Resolution (lat. × lon.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BCC-CSM1-1</td>
<td>Beijing Climate Center, China Meteorological Administration, China</td>
<td>~2.8125° × 2.8125°</td>
</tr>
<tr>
<td>2</td>
<td>BNU-ESM</td>
<td>Beijing Normal University, China</td>
<td>~2.8125° × 2.8125°</td>
</tr>
<tr>
<td>3</td>
<td>CanESM2</td>
<td>Canadian Centre for Climate Modelling and Analysis, Canada</td>
<td>~2.8125° × 2.8125°</td>
</tr>
<tr>
<td>4</td>
<td>CNRM-CM5</td>
<td>CNRM/Centre Europeen de Recherche et Formation Avancées en Calcul Scientifique, France</td>
<td>~1.406° × 1.406°</td>
</tr>
<tr>
<td>5</td>
<td>CSIRO-Mk3-6-0</td>
<td>CSIRO in collaboration with Queensland Climate Change Centre of Excellence, Australia</td>
<td>~1.875° × 1.875°</td>
</tr>
<tr>
<td>6</td>
<td>GFDL-CM3</td>
<td>Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, USA</td>
<td>~2° × 2.5°</td>
</tr>
<tr>
<td>7</td>
<td>GFDL-ESM2G</td>
<td>Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, USA</td>
<td>~2° × 2.5°</td>
</tr>
<tr>
<td>8</td>
<td>GISS-E2-H</td>
<td>GISS, National Aeronautics and Space Administration, USA</td>
<td>~2° × 2.5°</td>
</tr>
<tr>
<td>9</td>
<td>GISS-E2-R</td>
<td>GISS, National Aeronautics and Space Administration, USA</td>
<td>~2° × 2.5°</td>
</tr>
<tr>
<td>10</td>
<td>HadGEM2-AO</td>
<td>Met Office Hadley Centre, UK</td>
<td>~1.241° × 1.875°</td>
</tr>
<tr>
<td>11</td>
<td>HadGEM2-ES</td>
<td>Met Office Hadley Centre, UK</td>
<td>~1.241° × 1.875°</td>
</tr>
<tr>
<td>12</td>
<td>IPSL-CM5A-LR</td>
<td>Institute Pierre-Simon Laplace, France</td>
<td>~1.875° × 3.75°</td>
</tr>
<tr>
<td>13</td>
<td>MIROC5</td>
<td>National Institute for Environmental Studies, Japan</td>
<td>~1.406° × 1.406°</td>
</tr>
<tr>
<td>14</td>
<td>MIROC-ESM</td>
<td>National Institute for Environmental Studies, Japan</td>
<td>~1.406° × 1.406°</td>
</tr>
<tr>
<td>15</td>
<td>MIROC-ESM-CHEM</td>
<td>National Institute for Environmental Studies, Japan</td>
<td>~1.406° × 1.406°</td>
</tr>
<tr>
<td>16</td>
<td>MPI-ESM-LR</td>
<td>Max Planck Institute for Meteorology, Germany</td>
<td>~1.875° × 1.875°</td>
</tr>
<tr>
<td>17</td>
<td>MRI-CGCM3</td>
<td>Meteorological Research Institute, Japan</td>
<td>~1.125° × 1.125°</td>
</tr>
<tr>
<td>18</td>
<td>NorESM1_M</td>
<td>Norwegian Climate Centre, Norwegian</td>
<td>~1.89° × 2.5°</td>
</tr>
</tbody>
</table>

3. Methodology

3.1. Methodology Framework

The methodology framework of this study is shown in Figure 2. The first step is to locally calibrate parameters from four PET models for each station. The Penman-Monteith model is used here as a benchmark. The calibrated models will be used in the baseline and future periods. In the second step, bias correction is implemented for GCM output. The QM-Gaussian distribution method is used for maximum and minimum temperature, and the monthly correction factor (MCF) method is used for relative humidity, wind speed, and solar radiation. In the third step, the performance of the different PET models is compared in the baseline period, and changes of annual, seasonal, and monthly PET based on the different models are analyzed in the future period. Finally, the appropriateness of the PET models under climate change and uncertainty of global climate models are investigated.
3.2. PET Models

In this study, five widely used PET models were chosen, including one combination model (the FAO Penman-Monteith model), two temperature-based models (the Blaney-Criddle and the Hargreaves models) and two radiation-based models (the Makkink and the Priestley-Taylor models).

3.2.1. FAO Penman-Monteith Model

The Penman family of models is generally thought to be able to estimate PET accurately in most climates [11,29]. The FAO version of the Penman-Monteith model is recommended to calculate PET, owing to its accuracy and reality [27]. The high requirement of meteorological input is the main limitation of this model, thereby its use is often limited in data-sparse areas. The formula for the FAO Penman-Monteith model is:

\[
PET = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{\Delta + 273^2} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34u_2)}
\]

(1)
where \( PET \) is the potential evapotranspiration in \( \text{mm/day} \); \( \Delta \) is the slope of the vapor pressure-temperature curve in \( \text{kPa/}^\circ\text{C} \); \( R_n \) is the net radiation at the crop surface in \( \text{MJm}^{-2}\text{d}^{-1} \); \( G \) is soil heat flux density at the soil surface in \( \text{MJm}^{-2}\text{d}^{-1} \); \( \gamma \) is the psychrometric constant in \( \text{kPa/}^\circ\text{C} \); \( T \) is the mean daily air temperature at 2 m height in \( ^\circ\text{C} \); \( u_2 \) is the wind speed at 2 m height in \( \text{m/s} \); \( e_s \) is the saturation vapor pressure at 2 m height in \( \text{kPa} \) and \( e_a \) is the actual vapor pressure at 2 m height in \( \text{kPa} \).

3.2.2. Blaney-Criddle Model

The Blaney-Criddle (1950) [24] model for estimating PET was originally proposed in the western United States. This model is also widely used in other regions of the world. The usual form of the Blaney-Criddle model is written as:

\[
PET = k \cdot p \cdot (0.46T_a + 8.13)
\]

where \( PET \) is the potential evapotranspiration in \( \text{mm/day} \); \( T_a \) is the mean air temperature in \( ^\circ\text{C} \); \( p \) is the ratio of total daytime hours out of total daytime hours of the year; and \( k \) is the monthly consumptive use coefficient, relating to season, location, and vegetation type. The range of \( k \) is 0.5–1.2 for the growing period and an average value of 0.85 will be used as the initial value. The final value is calibrated locally.

3.2.3. Hargreaves Model

The Hargreaves model is a temperature-based model for estimating PET, and only temperature data and latitude are needed [60,61]. Owing to its low requirements of meteorological input, the Hargreaves model is widely used in many regions. The formula for the Hargreaves model is described as

\[
PET = k \cdot R_A \cdot TD^{1/2}(T_a + 17.8)
\]

where \( PET \) is the potential evapotranspiration in \( \text{mm/day} \); \( R_A \) is the extraterrestrial radiation in \( \text{mm/day} \); \( TD \) is the difference between maximum and minimum temperatures in \( ^\circ\text{C} \); \( T_a \) is the mean air temperature in \( ^\circ\text{C} \); and \( k \) is an empirical coefficient. The initial value of \( k \) is 0.0023, and the final value is calibrated locally.

3.2.4. Makkink Model

Makkink (1957) [25] proposed a model to calculate grassland potential evapotranspiration under the cold climate of Holland. According to the experimental data of Holland, Hansen (1984) [62] modified the Makkink model. Recently, the Makkink model is widely used in western Europe, the USA, and other regions [63–65]. The formula for the Makkink model is as follows:

\[
PET = k \cdot \frac{\Delta}{\Delta + \gamma} \frac{R_s}{\lambda}
\]

where \( PET \) is the potential evapotranspiration in \( \text{mm/day} \); \( \Delta \) is the slope of the vapor pressure-temperature curve in \( \text{kPa/}^\circ\text{C} \); \( \gamma \) is the psychrometric constant in \( \text{kPa/}^\circ\text{C} \); \( R_s \) is the solar radiation in \( \text{cal/(cm}^2\cdot\text{d)} \); \( \lambda \) is the latent heat in \( \text{cal/g} \); and \( k \) is the empirical coefficient. The initial value of \( k \) is 0.7 and the final value can be calibrated locally.

3.2.5. Priestley-Taylor Model

Priestley and Taylor (1972) [66] proposed a simplified version of the Penman model for estimating PET from an extensive wet surface. In this model, the aerodynamic part was removed and the energy part was multiplied by a coefficient, \( \alpha \). The Priestley-Taylor model is as follows

\[
PET = \alpha \cdot \frac{\Delta}{\Delta + \gamma} \frac{R_n}{\lambda}
\]
where PET is the potential evapotranspiration in mm/day; $\Delta$ is the slope of the vapor pressure-temperature curve in kPa/°C; $\gamma$ is the psychrometric constant in kPa/°C; $R_n$ is the net radiation at the crop surface in cal/(cm$^2\cdot$d); $\lambda$ is the latent heat in cal/g; and $a$ is the empirical coefficient, depending on the wetting condition. The initial value of $a$ is 1.26 and the final value can be calibrated locally.

3.3. Projection of Future PET and Performance Evaluation

Future potential evapotranspiration is obtained by the delta change method [42,67,68], i.e., computing the differences between baseline and future GCM-derived PET, and adding these changes to PET based on observation data in the baseline period. This method is able to directly avoid the possible effects of bias in climate variables from GCMs on PET simulations. The formulas are as follows:

$$PET_{future} = PET_{changes} + PET_{observation}$$  \hspace{1cm} (6)

$$PET_{changes} = PET_{GCM-future} - PET_{GCM-baseline}$$  \hspace{1cm} (7)

where $PET_{GCM-future}$ is the potential evapotranspiration calculated based on GCM simulations in the future period and $PET_{GCM-baseline}$ is the potential evapotranspiration calculated based on GCM simulations in the baseline period.

The performance of the PET models was evaluated using the percentage of absolute bias (PBIAS) and root mean square error (RMSE), which are respectively defined as follows:

$$PBIAS(\%) = \left\{ \frac{100 \times \sum_{i=1}^{n} (E_i - R_i)}{\sum_{i=1}^{n} (R_i)} \right\}$$  \hspace{1cm} (8)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (E_i - R_i)^2}$$  \hspace{1cm} (9)

where $E_i$ and $R_i$ refer to the reference (calculated by the FAO-PM model) and estimated values (calculated by other models) of PET at the $i$th time step, respectively; $n$ refers to the number of time steps; and $\bar{R}$ refers to the mean value of reference PET values. The values of perfect fit for performance evaluation are both 0 for PBIAS and RMSE.

3.4. Bias Correction Method

The output from GCMs or RCMs are generally biased, making correction procedures necessary before they are applied to regional climate impact analysis [44,45]. Bias correction methods identify possible bias between observed and simulated data and then adjust the simulated data. In this study, the quantile mapping (QM) method is used to do bias correction for maximum and minimum temperatures and the monthly correction factor (MCF) method is used for relative humidity, wind speed, and solar radiation.

The QM method is a non-parametric bias correction method proposed by Themessl et al., (2011) [53]. This method was successfully conducted in the bias correction for GCMs or RCMs-simulated precipitation in many studies [44,45]. Wilcke et al. (2013) [69] adopted the QM method to correct the bias of six climate variables from GCMs and results indicated that annual and monthly biases can largely be reduced for all variables. For precipitation, Gamma or double Gamma distribution is often used in the QM method. However, for temperature, a Gaussian distribution is usually assumed to match temperature better [38,67]. Hence, QM-Gaussian distribution is implemented for the bias correction of maximum and minimum temperature in this study.

In most cases, only simulated precipitation and temperature data are bias corrected. However, biases also occur in other variables, such as relative humidity, wind speed, and solar radiation [54]. These variables play vital roles on PET estimation [51,52]. The MCF method proposed by Haddeland...
et al. (2012) [54] is used here to correct the bias in other variables. The MCF method is performed at a monthly time step as follows:

\[
V_{bc} = V_{sim} \cdot \frac{V_{obs}(m)}{V_{sim}(m)}
\]

where \(V_{bc}\) is the variable value after bias correction; \(V_{sim}\) is the raw GCM output variable value; \(V_{obs}(m)\) is the long-term mean monthly observed value of variable; and \(V_{sim}(m)\) is the long-term mean monthly simulated value of variable. The long-term mean differences \((\frac{V_{sim}(m)}{V_{obs}(m)})\) in the period 1971–2000 were used for baseline and future periods. The performance evaluation for the above bias correction methods is root mean square error (RMSE) [70]. The equation is listed as Equation (9).

4. Results

4.1. Comparison of PET in the Baseline Period

The parameters from the four models were calibrated locally for each station. The Penman-Monteith model was used here as a benchmark. Root mean square error (RMSE) was used as the objective for parameter calibration of the four models shown in Section 3.2. The parameter corresponding to the smallest RMSE was regarded as the optimal parameter. Since the Blaney-Criddle model was originally developed for seasonal use [24], its parameter was then calibrated at seasonal time step. The other three models were calibrated at yearly time step. Table 2 shows the calibrated parameters in the four models for each station. As shown in Table 2, big differences exist between the values of the original parameter and the calibrated parameter.

Table 2. Calibrated parameters of the four potential evapotranspiration (PET) estimation models (Season definition: March–April–May (MAM), June–July–August (JJA), September–October–November (SON), and December–January–February (DJF)).

<table>
<thead>
<tr>
<th>Models</th>
<th>Blaney-Criddle</th>
<th>Hargreaves</th>
<th>Makkink</th>
<th>Priestley-Taylor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original Stations No.</td>
<td>0.85 (0.5–1.2)</td>
<td>0.0023</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>MAM JJA SON DJF</td>
<td>Annual Annual Annual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dangxiong</td>
<td>1 1.16 0.91 0.89 1.20 0.0025 0.62 1.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dingri</td>
<td>2 1.34 0.97 1.01 1.35 0.0026 0.65 1.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaize</td>
<td>3 1.37 1.12 1.18 1.70 0.0026 0.69 1.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jiali</td>
<td>4 1.05 0.85 0.85 1.17 0.0023 0.61 1.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jiangzi</td>
<td>5 1.23 0.93 0.95 1.19 0.0024 0.65 1.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasa</td>
<td>6 1.07 0.89 0.84 0.96 0.0024 0.66 1.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linzi</td>
<td>7 0.81 0.70 0.68 0.74 0.002 0.66 1.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longzi</td>
<td>8 1.12 0.90 0.91 1.02 0.0024 0.66 1.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nielamu</td>
<td>9 0.96 0.78 0.80 1.02 0.0026 0.58 0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulan</td>
<td>10 1.23 0.98 1.07 1.17 0.0027 0.64 0.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rikaze</td>
<td>11 1.16 0.90 0.86 1.09 0.0023 0.64 1.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zedang</td>
<td>12 1.13 0.91 0.89 1.0 0.0024 0.70 1.25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 shows a comparison of performance evaluation indices from annual PET before and after calibration. From this figure, it can be shown that the PBIAS and RMSE decrease largely after calibration, particularly for temperature-based models which display large biases before calibration. For instance, the performance evaluation of the Blaney-Criddle model at Longzi Station (No. 8) improved greatly after calibration, decreasing from 13.53% to 0.79% and from 0.42 to 0.096 mm/day for PBIAS and RMSE, respectively. This fact demonstrates that parameter calibration is indeed necessary. It was found out that the performance from the twelve stations was different, especially at Pulan Station (No. 10) where the four models revealed bad performance with the FAO-PM model. The reason was that wind speed is very important in estimating PET [51] but was not included in the four models, and the average value of wind speed at the Pulan station is much bigger than other stations. Generally, the performance evaluation indices (PBIAS and RMSE) indicate that annual PET estimated by the
Nielamu 9 0.96 0.78 0.80 1.02 0.0026 0.58 0.99 
Pulan 10 1.23 0.98 ... and autumn, but not in winter. The above conclusion is also verified by the other performance evaluation index, RMSE.

The reason was that wind speed is very important in estimating PET [51] but was not included in the
also verified by the other performance evaluation index,
satisfactory performance in spring, summer and autumn, but not in winter. The above conclusion is
in spring, summer, and winter, but not autumn. The Hargreaves and Priestley-Taylor models have
the closest agreement with the FAO-PM model among all models. The Makkink model acts very well
than autumn (SON) and winter (DJF). The seasonal PET estimated by the Blaney-Criddle model shows
Blaney-Criddle model was originally developed for seasonal use. For four seasons, the appropriateness
the Makkink, Hargreaves, and Priestley-Taylor models. The main reason for this may be that the
the best overall appropriateness with the FAO-PM model in a seasonal PET estimation, followed by
the FAO-PM model than Hargreaves and Priestley-Taylor models.

Figure 3. Performance evaluation indices of annual PET before and after calibration using four PET models.

Figure 4 presents PBIAS between seasonal PET estimated by the FAO-PM model and the other four
models in the baseline period. From this figure, it is observed that the Blaney-Criddle model shows
the best overall appropriateness with the FAO-PM model in a seasonal PET estimation, followed by
the Makkink, Hargreaves, and Priestley-Taylor models. The main reason for this may be that the
Blaney-Criddle model was originally developed for seasonal use. For four seasons, the appropriateness
of each model is different. Overall, all models performed better in spring (MMA) and summer (JJA)
than autumn (SON) and winter (DJF). The seasonal PET estimated by the Blaney-Criddle model shows
the closest agreement with the FAO-PM model among all models. The Makkink model acts very well
in spring, summer, and winter, but not autumn. The Hargreaves and Priestley-Taylor models have
satisfactory performance in spring, summer and autumn, but not in winter. The above conclusion is
also verified by the other performance evaluation index, RMSE.

Figure 4. Seasonal percentage of absolute bias (PBIAS) of the four PET models at the twelve stations in
the baseline period.
Figure 5 shows average monthly PET values in the baseline period from all stations estimated by the FAO-PM and four other models. Calibrated PET models performed very well at most stations, especially the Jiali and Linzhi stations. The four models slightly underestimated average PET for November–April and overestimated for the other months for the majority of stations. At Pulan station, deviations between PET by the FAO-PM and other four models were higher than those at other stations. Taking the sensitivity of PET to climate variables into consideration, the reason for the higher deviation is probably because the four models do not consider wind speed, which has a significant impact on PET [51]. Moreover, the average value of wind speed at the Pulan station is much bigger than other stations, which leads to the higher deviation. In general, the Blaney-Criddle and Makkink models revealed better appropriateness than the Hargreaves and Priestley-Taylor models at a monthly time step. This conclusion is the same for the seasonal performance of the four models.

Figure 5. Comparison of average monthly PET computed by the FAO-PM model and the four models in the baseline period.
As presented in Figure 6, it can be found that spatial heterogeneity exists in the appropriateness of the four models for estimating monthly PET. In January, the Hargreaves and Priestley-Taylor models obtained lower PET values in the entire river basin, which is consistent with their performance in winter (DJF). In April, the Hargreaves model obtained lower PET values in the upstream and midstream of the river basin. In July, the Blaney-Criddle model showed the closest agreement with the FAO-PM model, while the Hargreaves model overestimated the PET values in the upstream and midstream of the river basin. In October, the spatial patterns based on the five PET models were rather similar, except the Makkink model.

![Figure 6](image.png)

**Figure 6.** Spatial pattern of annual PET estimated by the five models.

### 4.2. Bias Correction Results for GCMs Data

Figure 7 shows the RMSE value between observation data with raw and corrected GCM maximum temperature and solar radiation in the baseline period. As presented in Figure 7, the RMSE values of two climate variables decrease significantly after bias correction for all GCMs. The range of RMSE is 2.3–17.7 °C for raw GCM maximum temperature data, while the range is 1.5–8.8 °C after bias correction. The ecdf of observed maximum temperature and corrected GCMs data also show close agreements for all GCM models (see Figure S1). This confirms the capability of the QM-Gaussian method for the bias correction of temperature data. For other climate variables (solar radiation, relative humidity, and wind speed), the QM method was actually adopted to correct bias of them, but the effect was unsatisfactory. Thus, the MCF method proposed by Haddeland et al., (2012) [54] was chosen for these climate variables. After the MCF method used, the values of RMSE from all GCM models were greatly reduced. The value of RMSE was the range of RMSE of 1.98–7.67 MJ/m²/day for raw GCM solar radiation data, while this value was close to zero at most stations after bias correction. The bias correction of wind speed and relative humidity by the MCF method also yielded superior results. It is found that ecdf of corrected GCMs data are much closer to observation data for most GCM models (see Figure S2).

![Figure 7](image.png)

**Figure 7.** Root mean square error (RMSE) between observation data with raw and corrected GCM maximum temperature and solar radiation in the baseline period.
4.3. Changes in Annual, Seasonal and Monthly PET

Figure 8 presents the relative changes of future annual PET estimated by the five models at each station under two RCPs. It is easy to found that the relative changes of annual PET under RCP8.5 are much higher than RCP2.6. The Blaney-Criddle and Hargreaves models both project positive changes in PET at all GCMs. However, the Makkink and Priestley-Taylor models project negative future PET changes at a few GCMs, such as MIROC-ESM (No.14) and MIROC-ESM-CHEM (No.15). As shown in Figure 8, it can be observed that the difference among stations is much smaller than differences among GCMs. Generally, the change signs of future annual PET under two RCPs using the five models are consistent at most GCMs. In conclusion, the relative change of future annual PET under RCP2.6 estimated by the Priestley-Taylor model shows the closest agreement with those by the FAO-PM model, followed by Makkink, Hargreaves, and Blaney-Criddle models. Additionally, the performance of the five models under RCP8.5 is difficult to judge.

Figure 8. Relative change percentages of future annual PET based on five models ((a) FAO-PM, (b) Blaney-Criddle, (c) Hargreaves, (d) Makkink, (e) Priestley-Taylor) for the 12 stations and the 18 GCMs under RCP2.6 (left column) and RCP8.5 (right column).
Figure 9 presents RMSE between seasonal PET change estimated by the FAO-PM model and the other four models in the future period under two RCPs. Similar to seasonal PET, RMSE under RCP8.5 is much higher than under RCP2.6. The appropriateness of each model performed differently with four seasons. The four models performed better in autumn (SON) and winter (DJF) than spring (MMA) and summer (JJA), which was contrary to that of baseline period. As presented in Figures 9 and 10 the sum RMSE values of four seasons from the radiation-based models is lower than those of the temperature-based models. Similar to the annual time step, the appropriateness of radiation-based models at seasonal time step was much better than that of temperature-based models in the future period. The other performance evaluation index PBIAS also confirmed this conclusion.

Figure 9. Average seasonal PET RMSE of 18 GCMs for the 12 stations in the future period under RCP8.5 (Season definition: March–April–May (MAM), June–July–August (JJA), September–October–November (SON), and December–January–February (DJF)).

Figure 10 shows the relative changes of the average monthly PET values of 18 GCMs estimated by the five models. Obviously, the relative changes of PET calculated by the five models under RCP8.5 are much higher than RCP2.6, especially the Blaney-Criddle and Hargreaves models. This may be because these two models are temperature-based models and are, therefore, more sensitive to change of temperature in the future. As presented in Figure 10, the change signs of PET under
Atmosphere 2019, 10, x FOR PEER REVIEW 16 of 25

The relative changes of monthly PET under two RCPs estimated by the five models based on all 12 stations. Monthly PET values estimated by the five models under RCP8.5 are higher than those under RCP2.6. In January, the Hargreaves and Priestley-Taylor models obtained lower PET values than the FAO-PM model in the entire river basin under

Figure 10. Average monthly PET relative changes of 18 GCMs at 12 stations under two RCPs.
Figure 11 shows the spatial pattern of the average monthly PET of 18 GCMs under two RCPs estimated by the five models based on all 12 stations. Monthly PET values estimated by the five models under RCP8.5 are higher than those under RCP2.6. In January, the Hargreaves and Priestley-Taylor models obtained lower PET values than the FAO-PM model in the entire river basin under RCP2.6, but only the Hargreaves model showed lower values under RCP8.5. In April, the spatial patterns based on the five PET models under two RCPs were rather similar. In July, the Hargreaves model obtained higher PET values in the upstream and midstream of the river basin. In October, PET values estimated by the Priestley-Taylor model was lower than that of the FAO-PM model under two RCPs, while the Blaney-Criddle model projected higher PET values under RCP8.5.

![Figure 11](image_url)

**Figure 11.** Spatial pattern of average PET values of the 18 GCMs in January, April, July, and October in the future period under RCP2.6 and RCP8.5.

4.4. Uncertainty of Global Climate Models

To investigate the uncertainty associated with the 18 GCMs, boxplots of monthly areal PET relative change values estimated by the five models in the future period under two RCPs are presented in Figure 12. In this figure, the line in the middle of the box refers to the median value. The whiskers indicate the maximum and minimum values. The crosses outside the boxes represent outliers. As shown in Figure 12, it can be observed that the uncertainty under RCP8.5 was slightly higher than those under RCP2.6. For PET estimated by the FAO-PM model, higher uncertainty occurred in summer and winter. For the Blaney-Criddle model, relative changes of PET from June to October showed lower uncertainty than in other months, which was the same for the Hargreaves model. PET estimated by Makkink and Priestley-Taylor models illustrated similar characteristics of uncertainty, i.e., higher uncertainty occurs in summer and winter. This means that the uncertainty of PET from the same category reveals similar characteristics. The uncertainty from the temperature-based models (Blaney-Criddle and Hargreaves) were the smallest, followed by the radiation-based models (Makkink and Priestley-Taylor) and the combined model (FAO-PM).
Figure 12. Uncertainty of future monthly areal PET change (%) under two RCPs.
5. Discussion

To detect the appropriateness of PET models under climate change, this study evaluated the performance of five PET models in baseline and future periods, including one combination model (FAO-PM), two temperature-based models (Blaney-Criddle and Hargreaves), and two radiation-based models (Makkink and Priestley-Taylor). FAO-PM is used here as a benchmark to locally calibrate the parameters of the other four models. Local calibration for reference crop evapotranspiration models is also suggested from another study [19]. The results revealed that the Blaney-Criddle and Makkink models show better appropriateness in the baseline period. However, the appropriateness of PET models alters under climate change, i.e., the radiation-based models (Makkink and Priestley-Taylor) showed closer agreement with FAO-PM than the temperature-based models (Blaney-Criddle and Hargreaves). This confirms the conclusion from a previous study by Liu et al., (2017) [19] that radiation-based models perform better than temperature-based models for reference crop evapotranspiration at a semiarid site in China.

As shown in Figure 10, it can be concluded that future monthly PET estimated by temperature-based models is much higher than that obtained by FAO-PM, while radiation-based models show lower values. To illustrate this phenomenon, we take Zedang Station as an example. Figure 13 presents the future average monthly change of the 18 GCMs for five climate variables at Zedang Station under two RCPs. Figure 14 shows the contribution percentage of climate variables to PET (estimated by FAO-PM) at Zedang Station based on the sensitivity analysis from another study by the authors [51]. As shown in Figure 14, PET is most sensitive to wind speed, followed by relative humidity, maximum temperature, solar radiation, and minimum temperature. Moreover, the contribution percentages vary at different months. Therefore, it can be figured out that the higher PET values calculated by the temperature-based models are mainly owing to the fact that these models only consider temperature and the significant increase of temperature in the future. The total contribution percentage of maximum and minimum temperatures ranges from 16% to 48% and a lower value occurs from May to August, demonstrating that the variations of temperature has limited impact on future PET estimation. Hence, the future PET values calculated by temperature-based models is higher than that obtained by the FAO-PM model, particularly from May to August. As presented in Figure 13, wind speed, relative humidity, and solar radiation have small decreases in the future period. The decrease of solar radiation and wind speed make PET decrease, and the decrease of relative humidity makes PET increase. Considering that the contribution percentage of solar radiation is small and its slight decrease, the change impact of solar radiation can be ignored in the future period. Nevertheless, mean temperature data is needed for the computation of variable Δ (slope of the vapor pressure-temperature curve), which is used in the radiation-based and combination models. Thus, future PET estimated by the radiation-based models revealed a positive change trend, but the values were lower than that obtained by FAO-PM. For the other two variables, the decreased extent of wind speed and its contribution percentage were bigger than relative humidity, therefore, the impact of these two variables make future PET increase slightly. This interprets why PET calculated by FAO-PM is higher than that obtained by the radiation-based models.
In this study, the QM-Gaussian method was used to correct the bias in maximum and minimum temperatures from GCMs and the results are very satisfactory. Although bias correction for temperature and precipitation is often made in other studies, bias correction for other climate variables is rarely conducted. Several previous studies have developed new tools or used existing approaches to bias correct for other climate variables, such as the distribution-based scaling (DBS) approach for relative humidity and wind speed [71], the QM method for all climate variables [44], and the MCF method for solar radiation, relative humidity, and wind speed [54]. In this study, the QM-Gaussian method was also applied for other climate variables, including solar radiation, relative humidity, and wind.
speed. But, the effect was not as good as that from temperature as the MCF method proposed by Haddeland et al., (2012) [54] is very simple and its bias correction effect had already been confirmed by other studies [72–74]. Therefore, the MCF method was adopted for the bias correction of other climate variables. As shown in Figure 7, the hydrological index RMSE was greatly improved after bias correction by the MCF method. Moreover, a Delta change approach was used to project the future PET in this study. Overall, the bias correction and Delta change method achieved the double effect of avoiding the possible impact of bias in climate variables from GCMs on PET estimation, which increased the accuracy of PET estimation in the future period.

As shown in Figure 8, the future PET values calculated by the FAO-PM, Makkink and Priestley-Taylor models from NorESM1_M (No.18) are much higher than other GCMs under the two RCPs. The common variable included in these three models but not included in temperature-based models is solar radiation. The future change of annual solar radiation from each GCM at the 12 stations are presented in Figure 15. In this figure, it can be observed that the variations of solar radiation from NorESM1_M are much bigger than those from the other 17 GCMs, which leads to the higher PET values. In addition, as shown in Figure 8, annual PET changes at the Linzhi and Longzi stations (No. 7 and No. 8) are smaller than other stations under NorESM1_M, which is owed to their smaller change of solar radiation.

![Figure 15. Future change of annual solar radiation of all GCMs at all stations under the two RCPs.](image)

Data from 18 GCMs under the two RCPs were used to calculate PET in the future period and the uncertainty was also estimated in this study, which is beneficial for the analysis of the possible impact of climate change on PET and its contribution analysis. The results are helpful in understanding future PET changes and the selection of optimal PET models in the future period. The study of the appropriateness analysis of PET models under climate change is very important for many fields, including hydrologic modelling, agricultural water management, irrigation planning, crop yield simulation, drought monitoring and forecasting.

6. Conclusions

This study investigated the appropriateness of potential evapotranspiration models for climate change impact analysis, including one combination model (the FAO Penman-Monteith model) as a benchmark model, two temperature-based models (the Blaney-Criddle and the Hargreaves models) and two radiation-based models (the Makkink and the Priestley-Taylor models). The key findings of this study are summarized below.
Local calibration for the parameters of PET models is necessary, as performance evaluation indices (PBIAS and RMSE) are significantly improved after calibration.

In the baseline period, the PET estimations from the Blaney-Criddle and Makkink models are in better appropriateness with those obtained by the FAO-PM model. The Hargreaves and Priestley-Taylor models show a slightly poor performance.

The bias correction methods, i.e., the QM-Gaussian method for maximum and minimum temperatures and the MCF method for the other three climate variables, demonstrate their effectiveness by strongly decreasing RMSE after bias correction.

In the future period, monthly PET estimated by all models show positive trends. The changes of PET under RCP8.5 are much higher than RCP2.6. The radiation-based models (Priestley-Taylor and Makkink) show better appropriateness than the temperature-based models (Hargreaves and Blaney-Criddle) in the future, as the RMSE value of the former models is almost half of the latter models.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4433/10/8/453/s1, Figure S1: Comparison of empirical cumulative distribution functions (CDF) of areal maximum air temperature from gauge observations and raw, bias-corrected (BC) 18 GCM models; Figure S2: Comparison of empirical cumulative distribution functions (CDF) of areal solar radiation from observations and raw, bias-corrected (BC) CMIP5 18 models.

Author Contributions: Conceptualization, S.P.; Data curation, W.X.; Formal analysis, S.P.; Funding acquisition, Y.-P.X.; Investigation, H.G. and Z.B.; Methodology, S.P., W.X.; and Z.B.; Project administration, Y.-P.X.; Resources, Y.-P.X.; Software, H.G.; Validation, S.P.; Visualization, S.P.; Writing—original draft, S.P.; Writing—review & editing, Y.-P.X.

Funding: The work here is financially supported by National Nature Science Foundation of China (91547106) and National Key Research and Development Plan “Inter-governmental Cooperation in International Scientific and Technological Innovation”(2016YFE0122100).

Acknowledgments: We also would like to appreciate the National Climate Center of China Meteorological Administration for providing climate data used in this study.

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

References


21. Constantin, J.; Willaume, M.; Murgue, C.; Lacroix, B.; Therond, O. The soil-crop models STICS and AqYield predict yield and soil water content for irrigated crops equally well with limited data. *Agricultural Meteorol.* 2015, 206, 55–68. [CrossRef]


58. Wei, Y.; Fang, Y. Spatio-temporal characteristics of global warming in the Tibetan Plateau during the last 50 years based on a generalised temperature zone-elevation model. *Plos ONE* 2013, 8, 60044. [CrossRef] [PubMed]
61. Hansen, S. Estimation of potential and actual evapotranspiration. *Nordic Hydrol.* 1984, 15, 205–212. [CrossRef]