Science of the Total Environment 837 (2022) 155887

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

A method for improving the estimation of extreme air temperature by satellite

Rui Yao [a], Lunche Wang [a,b,*], Xin Huang [c,d], Qian Cao [a,b], Yuanyuan Peng [e]

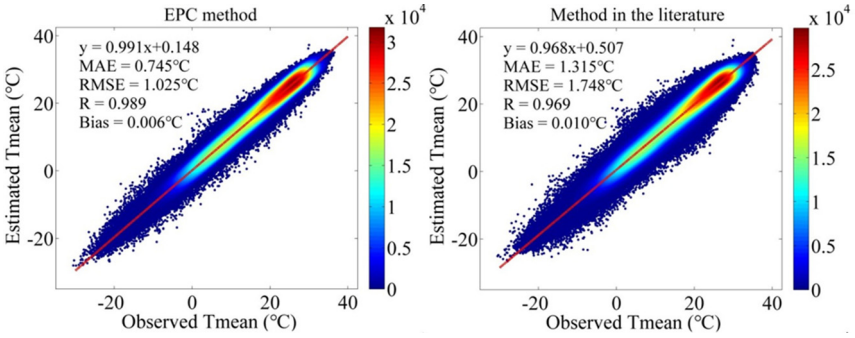
[a] Key Laboratory of Regional Ecology and Environmental Change, China University of Geosciences, Wuhan 430074, China,
[b] State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan 430074, China,
[c] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China,
[d] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China,
[e] School of Global Education and Development, University of Chinese Academy of Social Sciences, Beijing 102488, China

HIGHLIGHTS

- The EPC method was proposed to improve the estimation of extreme Ta.
- The MAEs of the estimated daily Ta using the EPC method ranged from 0.745–1.015 °C.
- The EPC method produced satisfactory accuracy of the estimated extreme Ta.

GRAPHICAL ABSTRACT

Density scatter plots between observed and estimated Ta for the EPC method and the method in the literature.



ARTICLE INFO

Editor: Pavlos Kassomenos

Keywords:
Extreme air temperature
Land surface temperature
Remote sensing
Machine learning
Heat wave
East China

ABSTRACT

Air temperature (Ta) data obtained from meteorological stations were spatially discontinuous. Some satellite data have complete spatial coverage and strong relationships with Ta (e.g., elevation and land surface temperature). Therefore, Ta can be mapped using in situ Ta and satellite data. However, this method may have a large bias when estimating the extreme Ta. In this study, the error prediction and correction (EPC) method, incorporating Cubist machine learning algorithm, was proposed to improve the estimation of extreme Ta. The accuracy of the EPC method was compared with that of the widely used method in previous studies in east China from 2003 to 2012. The mean absolute errors (MAEs) of the estimated daily Ta using the EPC method ranged from 0.75–1.01 °C, which were 0.57–0.96 °C lower than that of the method in the literature. The biases of the estimated Ta obtained using the two methods were close to zero. However, the biases can be as high as 7.10 °C when Ta is extremely low and as low as −3.09 °C when Ta is extremely high. Compared with the method in the literature, the EPC method can reduce the MAE by 1.41 °C, root mean square error by 1.49 °C, and bias by 1.61 °C of the estimated extreme Ta. Additionally, the EPC method produced satisfactory accuracy (MAEs <0.9 °C) of the estimated heat and cold wave magnitudes. Finally, a 1 km resolution daily Ta map in east China from 2003 to 2012 was developed, which will be useful data in multiple research fields.

1. Introduction

Air temperature (Ta) is a key variable in many research fields such as climate change (Huang et al., 2017; Sun et al., 2016), environmental science (Anniballe et al., 2014; Du et al., 2021), and agriculture and forestry

* Corresponding author at: China University of Geosciences, Wuhan 430074, China;
E-mail address: wang@cug.edu.cn (L. Wang).

http://dx.doi.org/10.1016/j.scitotenv.2022.155887
Received 16 February 2022; Received in revised form 14 April 2022; Accepted 8 May 2022
Available online 12 May 2022

(Liu et al., 2020; Zhang et al., 2017). It is traditionally measured using meteorological stations at 2 m above ground. However, meteorological stations are sparsely and unevenly distributed, limiting their application. For example, monitoring Ta variations in mountainous and remote regions is difficult because of the sparse distribution of meteorological stations in these areas (Meyer et al., 2016; Rao et al., 2019; Vancutsem et al., 2010). Additionally, because most meteorological stations are located in urban and their surrounding areas, accurately revealing the regional warming trend (Ren and Zhou, 2014; Sun et al., 2016; Wang et al., 2015) and urban heat island effect (Yao et al., 2021a; Zhou et al., 2019) is difficult.

Some satellite data are spatially continuous and have strong relationships with Ta (e.g., elevation and land surface temperature (Ts)). Therefore, Ta can be mapped using in situ Ta data and satellite variables from local to global scales (Benali et al., 2012; Hrisko et al., 2020; Kloog et al., 2014; Li et al., 2018; Rao et al., 2019; Shen et al., 2020; Vancutsem et al., 2010; Yoo et al., 2018; Zhang et al., 2016; Zhu et al., 2017). For example, Yoo et al. (2018) used a random forest model to estimate the daily maximum and minimum Ta (Tmax and Tmin) in Los Angeles and Seoul from 2006 to 2016. The root mean square errors (RMSEs) of the estimated Tmax and Tmin ranged from 1.1 to 1.7 °C. Hooker et al. (2018) developed the global monthly mean Ta (Tmean) from 2003 to 2016 by geographically weighted regressions and climate space weighted regressions. They found that the RMSEs ranged from 1.14 to 1.55 °C. Estimated Ta data have been successfully used to analyze climate change (Li and Zha, 2019b), the urban heat island effect (Yao et al., 2021a), crop growth (Zhang et al., 2013), and disease transmission (Weiss et al., 2014).

In many previous studies, we found an interesting phenomenon: the slopes of the fitting lines were generally lower than 1 when the observed and estimated Ta were used as the x-axis and y-axis, respectively (Fig. 1) (Chen et al., 2016; Didari et al., 2016; Didari and Zand-Parsa, 2018; Janatian et al., 2017; Jang et al., 2014; Li and Zha, 2019a; Li and Zha, 2019b; Lu et al., 2018; Peón et al., 2014; Yang et al., 2017; Yoo et al., 2018; Zhang et al., 2021; Zou et al., 2021). The slope of the fitting line lower than 1 suggests that: (1) the extremely low Ta is generally overestimated and that (2) the extremely high Ta is normally underestimated (Fig. 1). It suggested that the estimated Ta is conservative. For example, Janatian et al. (2017) used multiple linear regression to estimate daily and weekly Ta in Iran from 2000 to 2004. The slopes of the fitting lines ranged from 0.728 to 0.930. Li and Zha (2019b) used a random forest model to estimate monthly mean Ta in China from 2001 to 2015. The slope of the fitting line was 0.97. Lu et al. (2018) used a hierarchical Bayesian model to estimate monthly mean Ta in northwestern China from 2003 to 2011. The slopes of the fitting lines were all lower than 1. These three studies used different methods to estimate Ta. The temporal resolutions of the estimated Ta in these three studies comprised day, week and month. Additionally, this phenomenon also exists in other fields, for example, estimating solar radiation (Wang et al., 2016; Qin et al., 2018; Feng et al., 2020) and $PM_{2.5}$ (Wei et al., 2020; Zhan et al., 2017; Xue et al., 2019). For example, Wei et al. (2020) found that the slopes of the fitting lines between observed and estimated $PM_{2.5}$ were generally lower than 0.9, and $PM_{2.5}$ was normally underestimated under highly polluted conditions. They attributed this phenomenon to poor quality of aerosol optical depth products and small number of samples for high-pollution cases. Therefore, this phenomenon is widespread. The conservative estimation of Ta leads to uncertainty when the estimated Ta is used for research, especially in fields related to extreme Ta (e.g., spatiotemporal variations in extreme Ta, and the effects of extreme Ta on human health and crop yield). However, to our knowledge, few studies have specifically studied this phenomenon and tried to improve this problem.

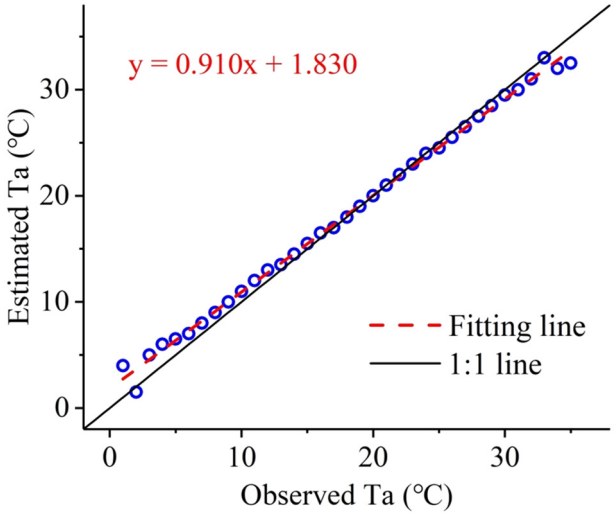


Fig. 1. Diagram of the observed and estimated air temperature (Ta). Note that this is just a diagram, the data in this figure is artificially created.

Therefore, this study aims to systematically discuss this phenomenon and develop a method called error prediction and correction (EPC) to improve the estimation of extreme Ta by satellite. Section 2 presents the study area and data. Section 3 presents the methods for processing the data, and estimating and validating Ta. Section 4 presents the main results, and Section 5 discusses the EPC method. Finally, Section 6 summarizes this study.

2. Study area and data

East China is a relatively developed region with a high population density (Fig. 2). Within this region are three major urban agglomerations (Beijing-Tianjin-Hebei, Yangtze River Delta, and Pearl River Delta urban agglomerations) in east China, making it a hot area for urbanization and climate research (Cao et al., 2018; Hu et al., 2019; Luo and Lau, 2018; Sun et al., 2014; Yang et al., 2011; Zhao et al., 2014). Therefore, east China was selected as the study area. East China contains 11 provinces (Hebei, Henan, Shandong, Hubei, Jiangsu, Anhui, Zhejiang, Jiangxi, Hunan, Fujian and Guangdong) and three municipalities (Beijing, Shanghai, and Tianjin). The total area of east China is 1.77 million km^2, and the total population is 882.1 million. The land cover type in east China is dominated by croplands and forests (Fig. 2a). Finally, east China has a humid and semi-humid climate. Ta and precipitation gradually decrease from south to north in this area (Liu et al., 2018).

Daily Tmean, Tmax and Tmin data from 2003 to 2012 were obtained from the China Meteorological Information Center. Daily Tmean, Tmax and Tmin were defined as daily mean, maximum and minimum Ta, respectively. These data were homogenized by the method used in Xu et al. (2013), and widely used in previous studies (Niu et al., 2019; Sun et al., 2016; Zhu et al., 2019). East China had 1104 meteorological stations with valid Ta data from 2003 to 2012 (Fig. 2b). Daytime and nighttime Ts information was derived from Moderate Resolution Imaging Spectroradiometer (MODIS) MOD11A1 and MYD11A1 data (version 6.1, daily product, 1 km resolution, from 2003 to 2012). The accuracy of the Ts product has been widely validated, and the errors are generally lower than 1 °C (Wan, 2008; Wan, 2014). MODIS Ts products have many gaps, especially for daily Ts products. In this study, these gaps were filled using the enhanced hybrid (EH) method in Yao et al. (2021b). The EH method fully uses three types of information to fill the gaps, and has higher accuracy (mean absolute errors (MAEs) < 1 °C for small gaps) than other gapfilling methods. In this study, Global multi-resolution terrain elevation data 2010 (GMTED2010) (Danielson and Gesch, 2011) was used to derive elevation information. It has 1 km spatial resolution and 1 m vertical resolution, and the accuracy (RMSE) of GMTED2010 is generally lower than 10 m (Athmania and Achour, 2014; Khalid et al., 2016). 1 km resolution MODIS MOD13A2 enhanced vegetation index, 500 m resolution topographic index (Marthews et al., 2015) and 30 m resolution China land cover dataset (CLCD) (Yang and Huang, 2021) were used to derive vegetation greenness, topography and land cover information, respectively. Clear-sky solar radiation, slope and aspect data with 1 km resolution were

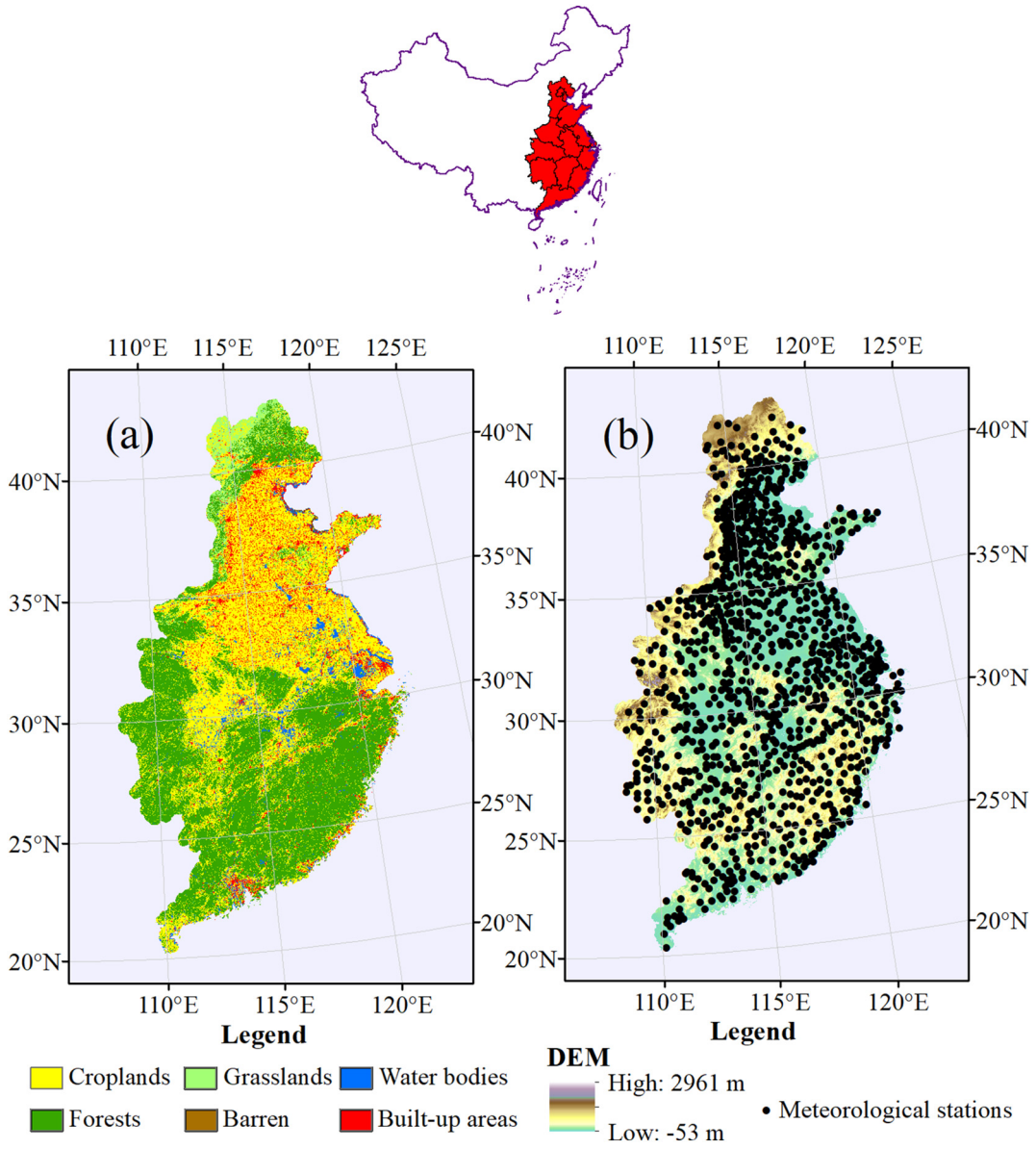


Fig. 2. Study area. (a) Land cover map; (b) Elevation map and spatial distribution of meteorological stations. The background map is China land cover dataset (CLCD) in 2012.

mapped using Area Solar Radiation, Slope and Aspect tools in ArcGIS software, respectively. Latitude and longitude data with 1 km resolution were extracted from MODIS Ts data.

3. Methods

3.1. Variable and model selection

Initially, 18 variables were used as alternatives to estimate Ta (Table 1). Other data such as wind speed and humidity were not used to estimate Ta, because it was difficult to obtain these data with a resolution of 1 km. The daily proportion of missing values in daytime (nighttime) Ts data was calculated using MOD11A1 and MYD11A1 daytime (nighttime) Ts data at the pixel level. The role of this variable is to distinguish between clear and cloudy skies. The 1 km resolution proportions of urban areas and water bodies were mapped using 30 m resolution land cover data.

A simple forward variable selection method was used according to previous researches (Meyer et al., 2016; Xu et al., 2018; Yao et al., 2021a). This method can retain the variables that have a positive effect on the Ta

Table 1
Selection of predictor variables to estimate Ta.

Alternative variable	MOD11A1 daytime and nighttime Ts, MYD11A1 daytime and nighttime Ts, proportions of missing value during the daytime and nighttime in Ts data, MOD13A2 enhanced vegetation index, elevation, slope, aspect, topographic index, clear-sky solar radiation, latitude, longitude, Julian day, year, proportions of urban area and water body
Variables for Tmean estimation	MOD11A1 nighttime Ts, elevation, latitude, longitude, clear-sky solar radiation, Julian day and year
Variables for Tmax estimation	Proportion of missing value during the daytime in Ts data, elevation, latitude, longitude, clear-sky solar radiation, Julian day and year
Variables for Tmin estimation	MOD11A1 nighttime Ts, elevation, latitude, longitude, clear-sky solar radiation, Julian day and year
Not used variables	MOD11A1 daytime Ts, MYD11A1 daytime and nighttime Ts, proportion of missing value during the nighttime in Ts data, MOD13A2 enhanced vegetation index, slope, aspect, topographic index, proportions of urban area and water bodies

estimation, and remove the variables that have no positive effect. After variable selection, seven variables were used to estimate Tmean, Tmax and Tmin (Table 1). Other variables were not used to estimate Ta in this study. It is worth noting that no Ts variables were used estimate Tmax. This is mainly due to the relatively weak correlations between Tmax and Ts. There is a large difference between Tmax and daytime Ts, because daytime Ts is very sensitive to solar radiation and surface characteristics (Yoo et al., 2018; Yao et al., 2021a). The correlation between Tmax and nighttime Ts is relatively weak, because the monitoring time of Tmax and nighttime Ts is quite different. Therefore, other variables may play more important roles in estimating Tmax, and Ts variables were not retained in the final variable set. Cubist machine learning algorithm (Quinlan, 1992) was used to estimate Ta, since previous researches showed that it has higher accuracy than other machine learning algorithms (including random forest, support vector regression, etc.), multiple linear regression and partial least squares regression (Noi et al., 2017; Xu et al., 2018; Zhang et al., 2016).

3.2. Ta estimation

The Ta estimation method used in the previous literature is described as follows. First, the values of some variables (e.g., Ts, proportion of missing values during the daytime in Ts, and clear-sky solar radiation) corresponding to the location of meteorological stations are extracted. Other variables accompanied by Ta data can be used directly (e.g., elevation, latitude, longitude, Julian day, and year). Second, Ta (dependent variable) and the accompanying independent variables were input into the model (machine learning, linear regression, or other models), and the relationship between Ta and independent variables was fit. Subsequently, gridded Ta data were developed using the fitted relationship and gridded independent variable

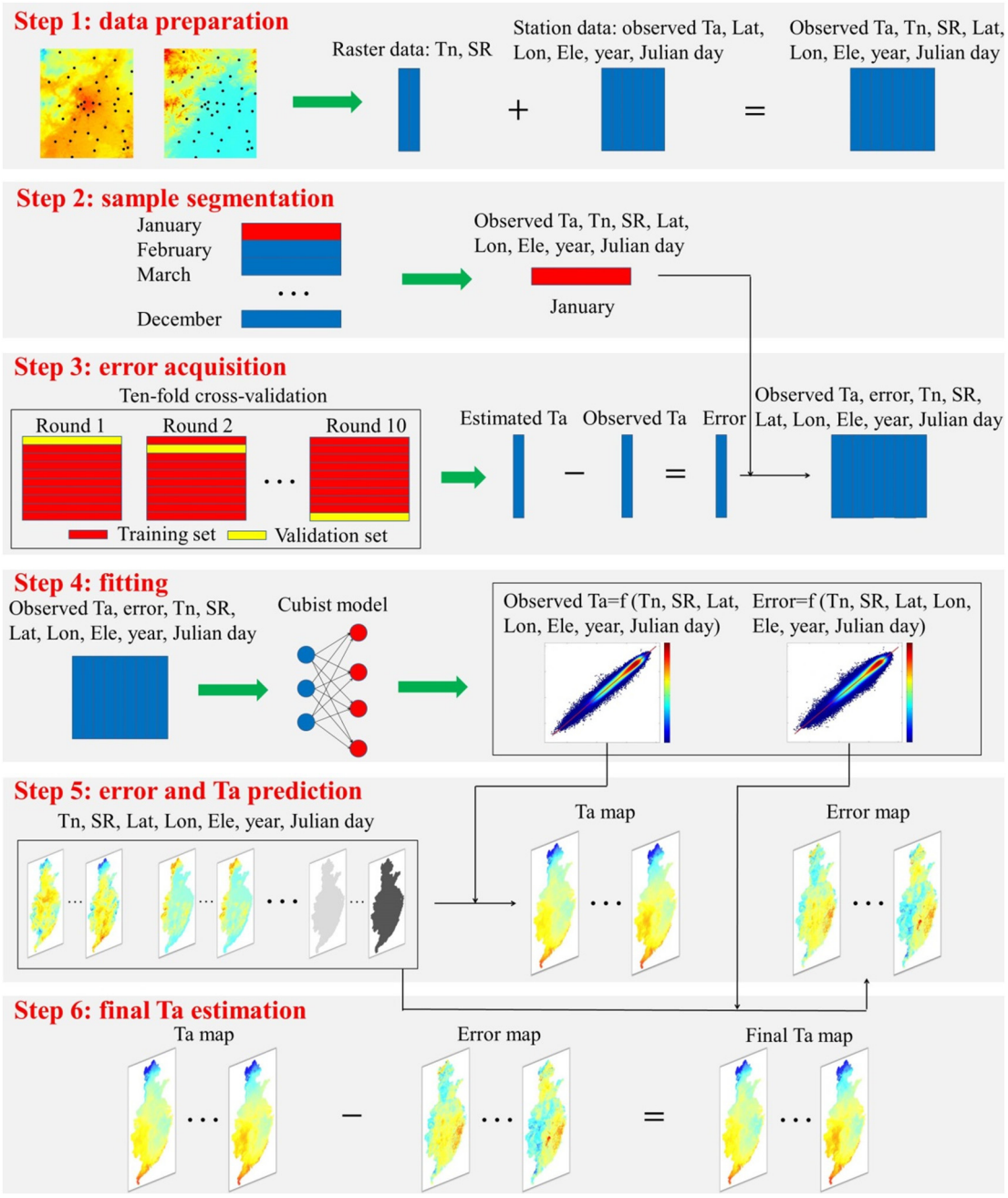


Fig. 3. Diagram of the error prediction and correction (EPC) method. Tn: nighttime Ts. SR: clear-sky solar radiation. Lat: latitude. Lon: longitude. Ele: elevation.

data (Gao et al., 2021; Hereher and El Kenawy, 2020; Li and Zha, 2019a; Rosenfeld et al., 2017).

This study proposes the EPC method to estimate Ta and compares it with the aforementioned method used in the literature. The diagram of the EPC method is shown in Fig. 3. The EPC method comprises six steps:

(1) Data preparation. This step is the same as in the first step of the method in the literature. Ta and the accompanying independent variable data were prepared.

(2) Sample segmentation. All samples (approximately 4 million) were divided into 12 subsets by month. The relationship between Ta and independent variables was fitted, and Ta was predicted by month. This strategy is used because the relationship between Ta and independent variables differed substantially by month (Yao et al., 2020).

(3) Error acquisition. The leave-station-out 10-fold cross-validation method was performed as follows: All meteorological stations were divided into ten subsets. Next, samples from nine subsets of meteorological stations were input into the Cubist model, and samples from the remaining one subset of meteorological stations were used for validation. Subsequently, this procedure was repeated, and all ten subsets were used to validate in turn. Therefore, all samples will have the observed and estimated Ta. Finally, the error of each sample can be calculated as the estimated Ta minus the observed Ta:

$$\text{Error1} = T_{a\_e1} - T_{a\_o} \quad (1)$$

where T_{a-e1} and T_{a-o} are the estimated and observed values of Ta, respectively. Error1 represents the error of the sample.

(4) Fitting. The error and the independent variables were input into the Cubist model, and the relationship between the error and the

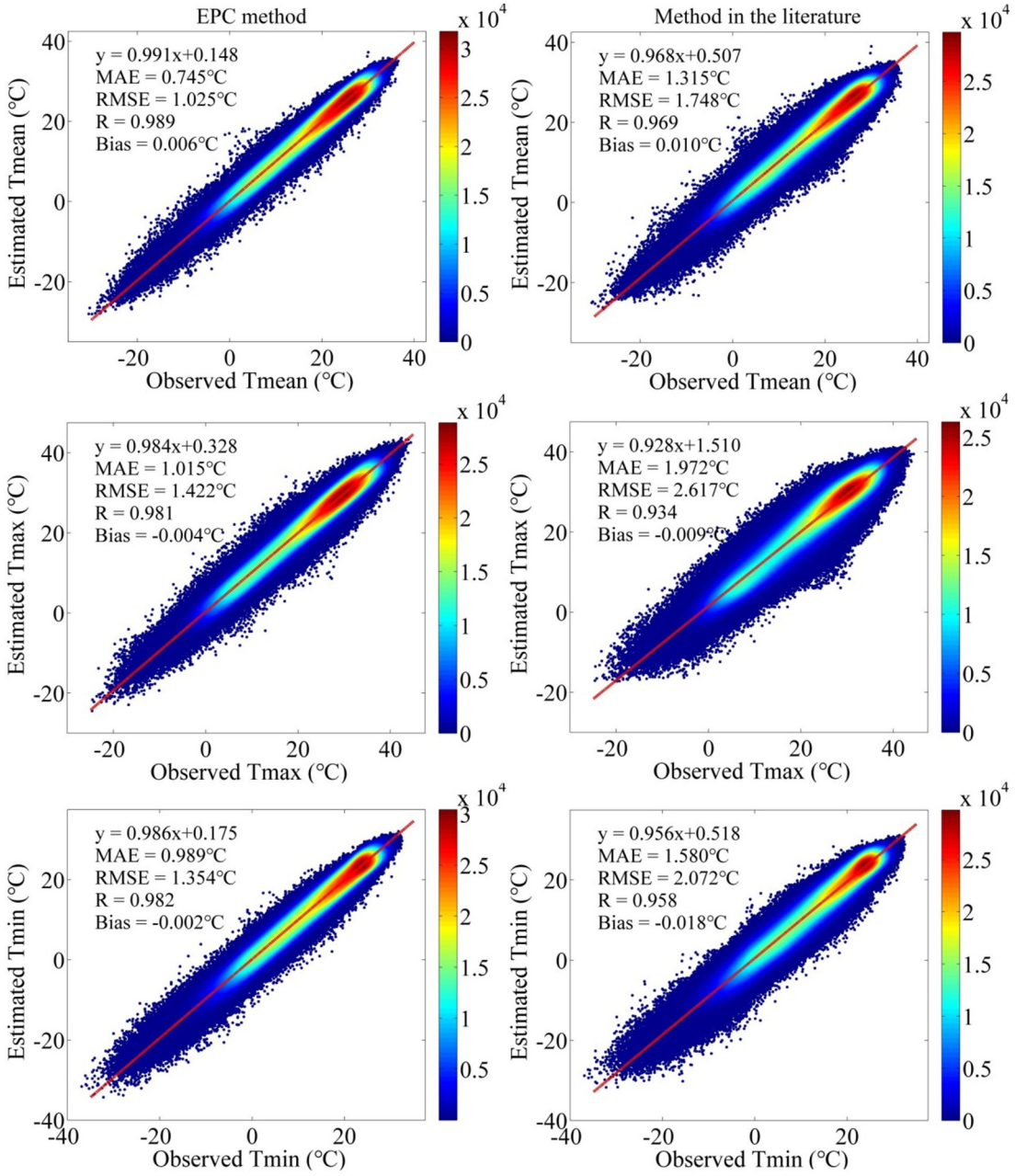


Fig. 4. Density scatter plots between observed and estimated Ta for the EPC method and the method in the literature. Tmean: daily mean Ta. Tmax: daily maximum Ta. Tmin: daily minimum Ta.

independent variables was fitted. Similarly, the relationship between Ta and the independent variables was also fitted.

$$T_{a-o} = f_1(Tn, SR, Lat, Lon, Ele, year, Julian\ day) \quad (2)$$

$$\text{Error1} = f_2(Tn, SR, Lat, Lon, Ele, year, Julian\ day) \quad (3)$$

where Tn represents nighttime Ts (when estimating Tmean and Tmin) or the proportion of missing values during the daytime in Ts data (when estimating Tmax). *SR*, *Lat*, *Lon*, and *Ele* represent clear-sky solar radiation, latitude, longitude and elevation, respectively.

(5) Error and Ta prediction. The 1 km resolution error map was developed using the fitted relationship between the error and the independent variables, and the 1 km resolution independent variables. Similarly, the 1 km resolution original Ta map was developed using the fitted relationship between Ta and the independent variables, and the 1 km resolution independent variables.
(6) Final Ta estimation. The final Ta map can be calculated as the original Ta map minus the error map:

$$T_{a-e3} = T_{a-e2} - Error2 \quad (4)$$

where T_{a-e3} and T_{a-e2} are the final and original Ta maps, respectively. Error2 is an error map.

3.3. Ta validation

Accuracies of the EPC method and the method in the literature were validated using a ten-fold cross-validation method. The validation procedures of the EPC method are described as follows: First, as in steps (1) and (2) of the EPC method, all samples were prepared and divided into 12 subsets by month. Second, all samples in one month were divided into ten subsets. Third, Steps (3)–(6) of the EPC method were performed using nine subsets, and the remaining subset was used for validation. This procedure was repeated, and all ten subsets were used for validation. The MAE, RMSE, coefficient of determination (R^2), and bias were calculated using all the samples to describe the accuracy.

3.4. Analyzing the accuracy of the estimated extreme ta

Three parts of the experiments were conducted to analyze the accuracy of the estimated extreme Ta values. First, the biases of estimated Ta were calculated for each 2.5 °C Ta range. Second, three types of extremely high Ta were defined: the observed Ta was higher than 99, 99.9, and 99.99 percentile of the observed Ta at the same meteorological station. Similarly, three types of extremely low Ta were defined: the observed Ta was lower than the 1, 0.1, and 0.01 percentile of observed Ta at the same meteorological station. The MAE, RMSE, R^2, and bias of the estimated Ta were calculated when the observed Ta fulfilled the requirement for extreme Ta. Third, the performance of using the estimated Ta to study the heat and cold waves was analyzed. In this study, a heat wave event was defined as at least three consecutive days when Tmax was higher than 90 percentile of observed Tmax in the same meteorological station from May to September from 2003 to 2012 (Croitoru et al., 2016; Gaitan et al., 2019; Luo and Lau, 2017). Similarly, a cold wave event was defined as at least three consecutive days when Tmin was lower than 10 percentile of observed Tmin in the same meteorological station from November to March from 2003 to 2012 (Gaitan et al., 2019). Heat wave magnitude (HWM) and cold

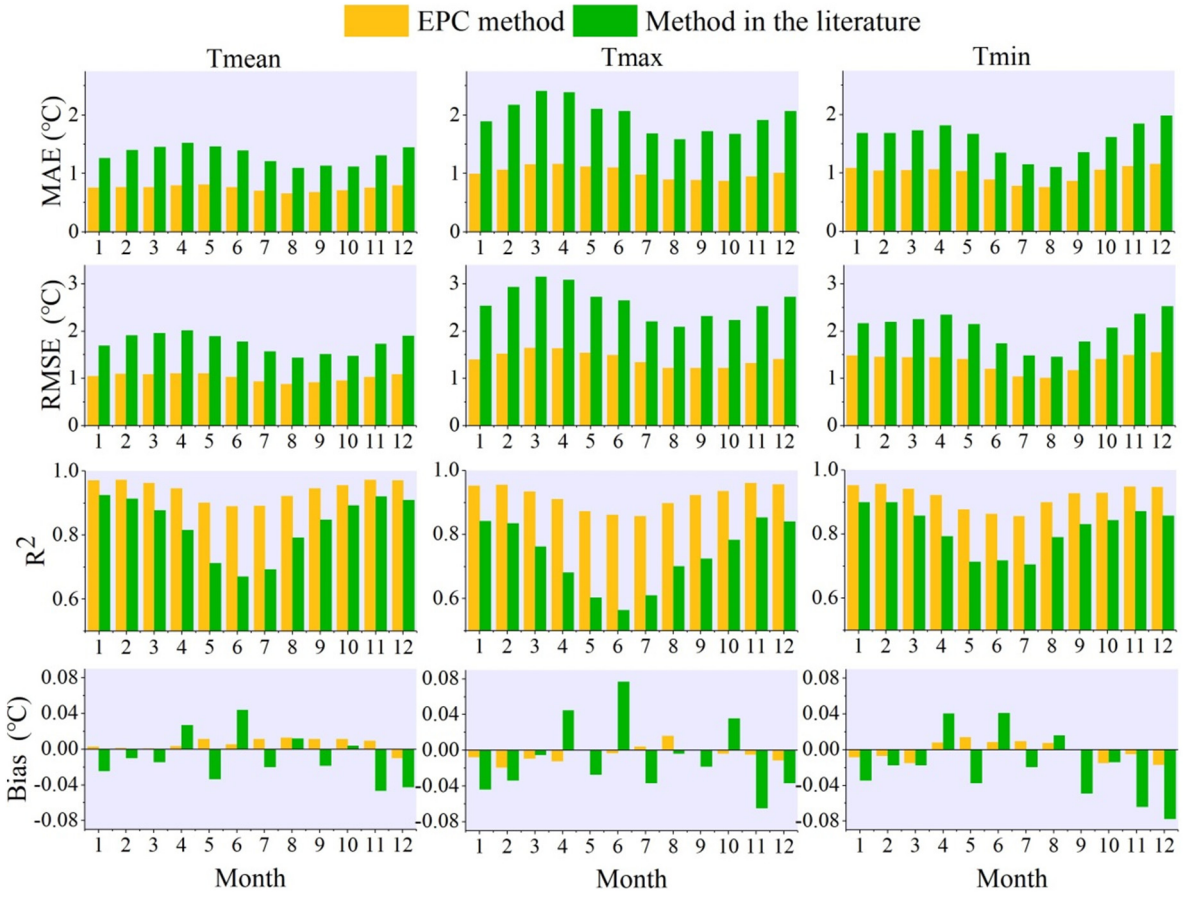


Fig. 5. Mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R^2), and bias of the estimated Ta for the EPC method and the method in the literature by month.

wave magnitude (CWM) were used to quantify the heat and cold waves, respectively. HWM (CWM) was defined as the mean Tmax (Tmin) of all heat (cold) wave events in one year (Croitoru et al., 2016). These indexes were selected because of their suitability for comparing various methods, whereas other indexes such as the number of heat wave events in one year are not. The HWM and CWM were calculated using both observed and estimated Ta values. Next, the MAE, RMSE, R^2, and bias of the estimated HWM and CWM were calculated to assess the accuracy.

4. Results

4.1. Accuracy of Ta estimation

The EPC method significantly outperformed the method in the literature (Figs. 4–6). The MAEs of the estimated Tmean, Tmax, and Tmin of the EPC method were 0.75, 1.01, and 0.99 °C, respectively (Fig. 4). These were 0.57–0.96 °C (37.4–48.6%) lower than those of the method in the literature. In addition, the EPC method produced higher R^2 and lower MAE, RMSE, and bias than the method in the literature in nearly all months and meteorological stations (Figs. 5 and 6). The reasons for the high accuracy of the EPC method are discussed in detail in Section 5.1.

The accuracy of Tmean was the highest, followed by Tmin and Tmax. This result is similar to those in the literature (Lu et al., 2018; Vancutsem et al., 2010; Venter et al., 2020; Yao et al., 2020; Yao et al., 2021a). The reasons are as follows: (1) Tmean is the daily average Ta, and Tmax and Tmin are instantaneous Ta; and (2) the correlation between nighttime Ts and Tmin is stronger than between daytime Ts and Tmax (Lin et al., 2016). Seasonally, the MAEs and RMSEs were generally lower in warm months (e.g., July and August) than those in the cold months, and the R^2 values were also lower in warm months than those in the cold months (Fig. 5). This is because the Ta range in warm months is smaller than that in cold months (Li and Zha, 2019a; Yao et al., 2021a). Spatially, the MAEs were normally lower in plains and higher in mountainous areas (Fig. 6). The

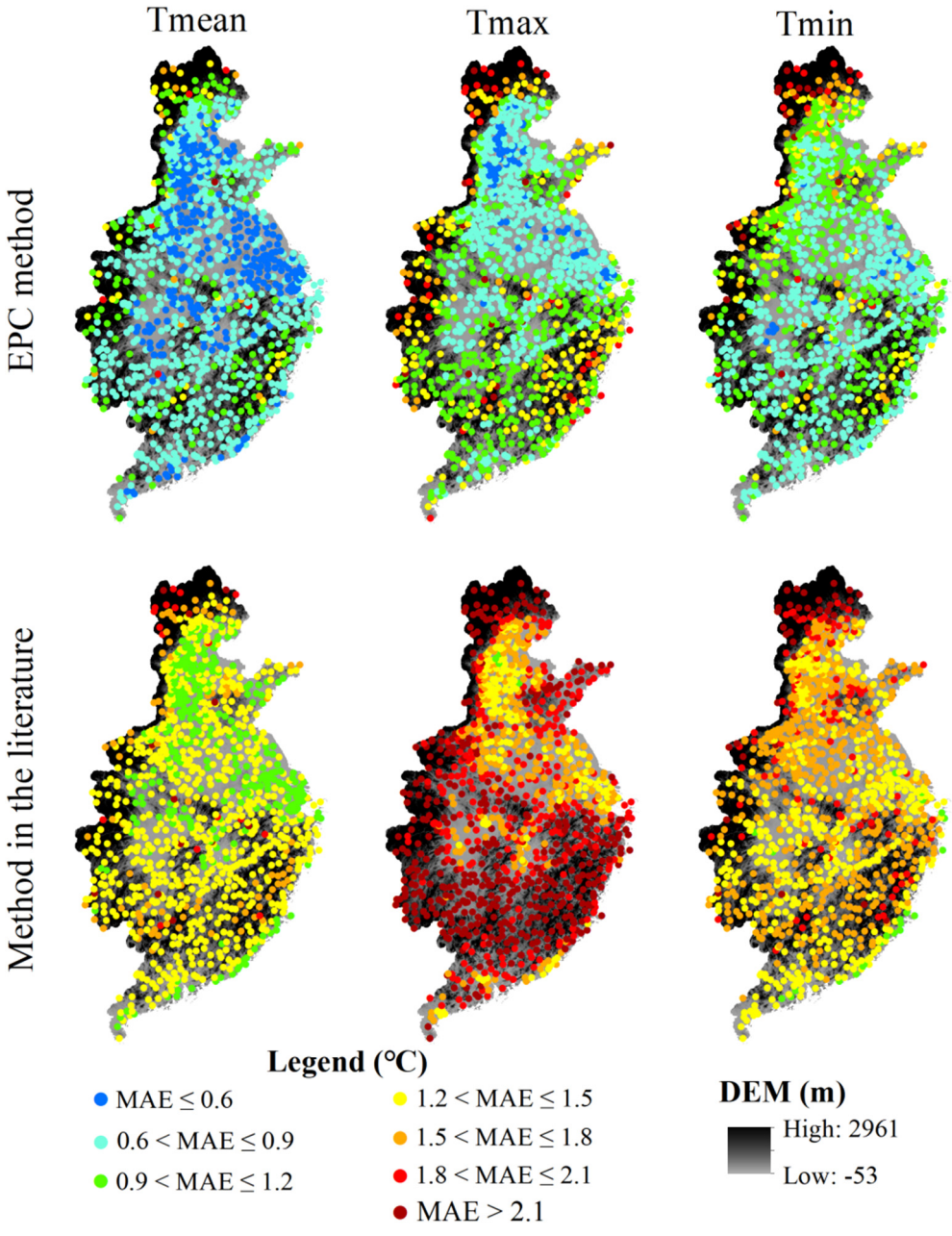


Fig. 6. Spatial distribution of MAEs of the estimated Ta for the EPC method and the method in the literature.

high MAEs in mountain areas can be attributed to the complex terrain and low density of meteorological stations in these areas. Furthermore, the MAEs of a meteorological station in Shandong province (>2.2 °C in all cases, sole red dot in Fig. 6a) were significantly higher than those of other meteorological stations (Fig. 6). Further analysis revealed that the terrain of this area was complex. The elevation of this meteorological station (1533 m) was significantly higher than that of other neighboring meteorological stations (generally lower than 200 m).

Notably, the overall biases of the estimated Ta were all close to zero (> −0.08 °C and <0.08 °C in all cases) for the EPC method and the method in the literature (Figs. 4 and 5). However, the slopes of the fitting lines between the observed and estimated Ta were all lower than 1, consistent with the literature (Chen et al., 2016; Didari et al., 2016; Didari and Zand-Parsa, 2018; Janatian et al., 2017; Jang et al., 2014; Li and Zha, 2019a; Li and Zha, 2019b; Lu et al., 2018; Peón et al., 2014; Yang et al., 2017; Yoo et al., 2018; Zou et al., 2021). Therefore, we suspected that the Ta estimation would have a large bias and error when Ta was extremely high or low.

4.2. Accuracy of the estimated extreme Ta

Although the overall biases of the estimated Ta were close to zero (Figs. 4 and 5), the biases of the estimated Ta differed greatly by Ta range (Fig. 7). The biases of the estimated Ta were positive (negative) when the observed Ta was extremely high (low). When the observed Tmax was lower than −17.5 °C, the bias of the estimated Tmax using the method in the literature can exceed 7 °C. Additionally, the MAEs and RMSEs of the estimated extreme Ta (Fig. 8) were higher than those of the overall MAEs and RMSEs (Fig. 4). These results suggest that the estimated extreme Ta has a large bias and error. Furthermore, the more extreme the Ta, the lower the accuracy (Figs. 7 and 8). Indeed, it is reasonable that the accuracy of the estimated extreme Ta is low. The extreme Ta is generally significantly higher or lower than the spatiotemporal neighboring Ta and thus difficult to predict. The large bias of estimated Ta leads to uncertainty when the estimated Ta is used for research, especially in fields related to extreme Ta (e.g., spatiotemporal variations in extreme Ta and its effects on human health and crop yield). Finally, the absolute values of biases of the estimated extremely low Ta (>4 °C in many cases) were significantly higher than the absolute values of biases of the extremely high Ta (generally <3 °C). The primary reason was that the number of samples was large when Ta was high (Fig. 7), and vice versa. Therefore, the relationship between Ta and the independent variables can be more accurately fitted when Ta is high.

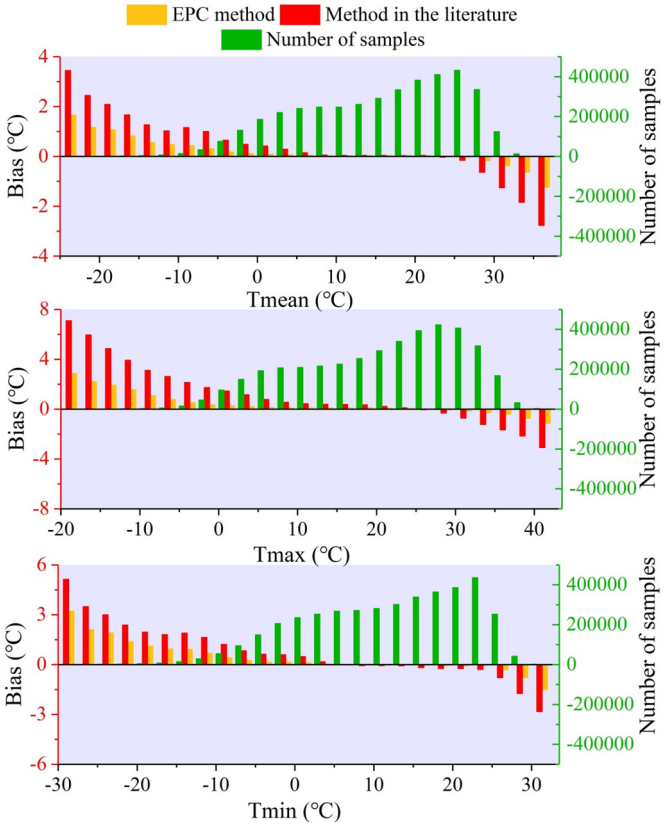


Fig. 7. Biases of the estimated Ta for the EPC method and the method in the literature and numbers of samples in different Ta ranges.

The EPC method can significantly reduce the bias and increase the accuracy of the estimated extreme Ta (Table 2, and Figs. 7 and 8). Biases of the EPC method were approximately 50% lower than those of the method in the literature (Fig. 7). For example, when the observed Tmean was lower than −22.5 °C, the biases of the estimated Tmean using the EPC method and the method in the literature were 1.65 and 3.45 °C, respectively. When the observed Tmean >99 percentile of the observed Tmean, the biases of the estimated Tmean for the EPC method and the method in the literature were 0.75 and 1.70 °C, respectively. Compared with the method in the literature, the EPC method can reduce the MAE by 0.67–3.66 °C (average: 1.41 °C, 24.1–60.1%), RMSE by 0.75–3.46 °C (average: 1.49 °C, 22.4–54.2%) and bias by 0.74–4.21 °C (average: 1.61 °C, 37.3–72.9%) of the estimated extreme Ta (Fig. 8 and Table 2). These results further suggest that EPC is a robust method for estimating Ta.

4.3. Accuracy of the estimated HWM and CWM

The EPC method produced a satisfactory accuracy for the estimated HWM and CWM (Table 3). The MAEs of the estimated HWM and CWM of the EPC method were 0.62 and 0.89 °C, respectively. Comparatively, the MAEs of the method in the literature were significantly higher than those of the EPC method (HWM: 1.75 °C; CWM: 2.03 °C). Additionally, it should be noted that the MAEs and RMSEs of the estimated HWM and CWM using the EPC method were lower than the overall MAEs and RMSEs of the estimated Ta. This finding primarily occurs because HWM and CWM are the multi-day average Ta. The positive and negative errors on different days can be offset. However, the MAEs and RMSEs of the estimated HWM and CWM using the method in the literature were close to the overall MAEs and RMSEs of the estimated Ta. This is because the bias of the method in the literature is large (Figs. 7 and 8). The positive and negative errors on different days are difficult to offset.

Fig. 9 shows the temporal variations in the observed and estimated Ta averaged for 1104 meteorological stations in the summer (June, July, and August) of 2003. The estimated Ta had a positive bias on relatively cold days and a negative bias on relatively hot days. On relatively hot days, Ta was significantly higher than Ta on neighboring days. Therefore, Ta on relatively hot days was difficult to predict and had a negative bias. Furthermore, the accuracy of the EPC method was significantly higher than that of the method in the literature. The MAEs of the averaged estimated Ta using the EPC and the method in the literature in this period were 0.06 and 0.45 °C, respectively. This phenomenon occurred because the bias of the EPC method was significantly lower than that of the method in the literature (Figs. 7 and 8). Therefore, the positive and negative errors of the EPC method are offset. Finally, the 0.06 °C of MAE was promising, indicating that the MAE of the estimated spatial average Ta using the EPC method was close to zero. Therefore, the estimated Ta using the EPC method can be used to accurately reveal the regional warming trend and urban heat island effect.

4.4. Mapping Ta using the EPC method and the method in the literature

The final Ta map was developed using the EPC method and the method in the literature. The spatial and seasonal variations in the estimated Ta using the EPC method were similar to those of the method in the literature

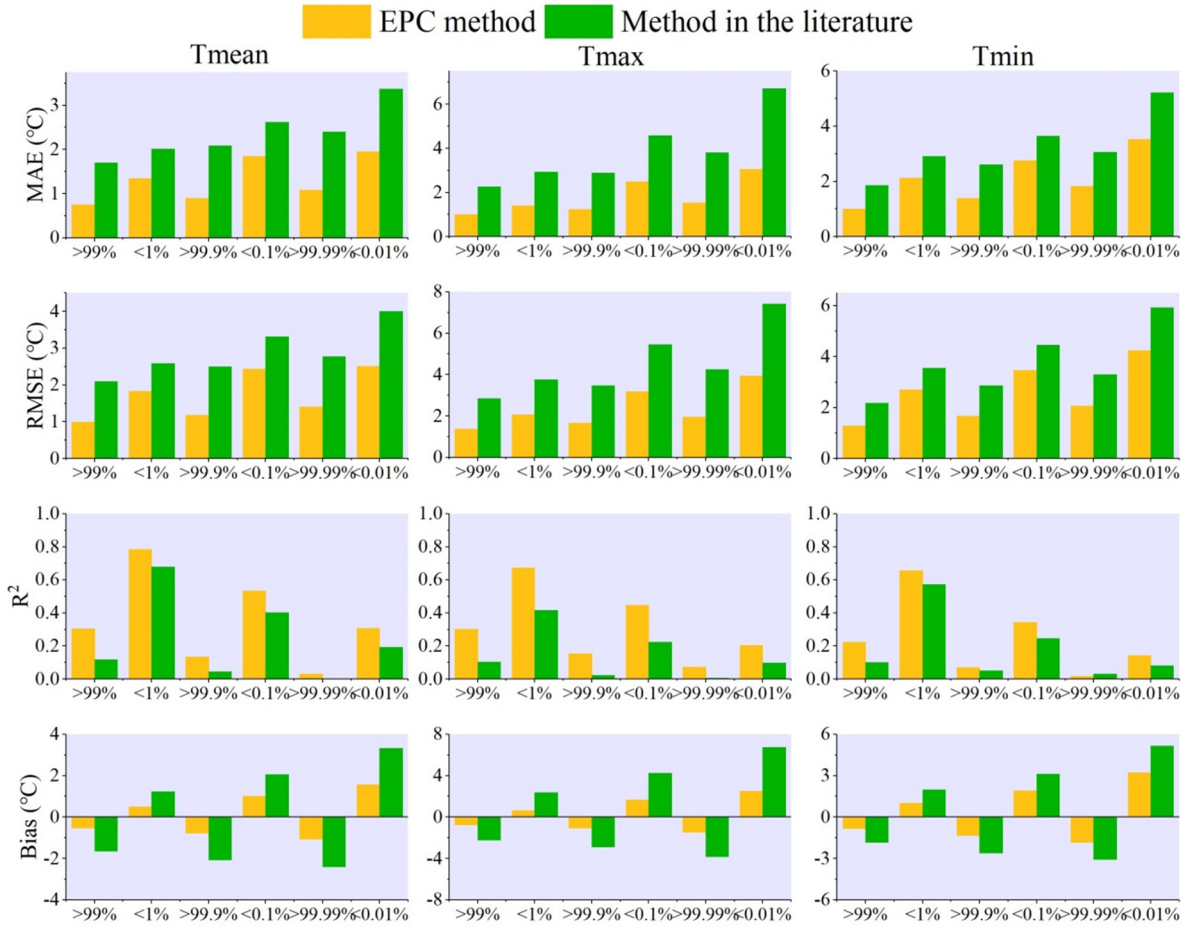


Fig. 8. MAE, RMSE, R^2, and bias of the estimated extreme Ta for the EPC method and the method in the literature.

(Fig. 10), because the Ta was seasonally averaged. Ta was strongly related to elevation, latitude, and season. Areas with high elevation (e.g., Taihang Mountains and Wuyi Mountains) or latitude (e.g., Beijing-Tianjin-Hebei urban agglomerations) generally have low Ta. Ta was higher in summer than in winter (December, January, and February). The Tmax was the highest, followed by Tmean and Tmin. These results were reasonable, which further demonstrate the reliability of the estimated Ta data and the EPC method. The developed Ta map provides useful data to multiple research fields such as meteorology and climatology, environmental science and ecology, and human health and epidemiology.

5. Discussion

5.1. Strength and limitation of the EPC method

In this study, for estimating Ta, the EPC method significantly outperformed the method in the literature. The EPC method has a higher overall accuracy of the estimated Ta than the method in the literature (Fig. 4). Additionally, the EPC method produced better accuracy of the estimated extreme Ta, HWM, and CWM than the method in the literature (Fig. 8, Tables 2 and 3). The reasons for the high accuracy of the EPC are

Table 2
MAE, RMSE, and bias of the estimated extreme Ta reduced by the EPC method relative to the method in the literature. Percentage reduction is shown in parentheses.

		Tmean	Tmax	Tmin
>99.99 percentile of observed Ta	MAE (°C)	1.32 (54.9%)	2.29 (60.1%)	1.22 (40.1%)
	RMSE (°C)	1.36 (49.3%)	2.31 (54.2%)	1.24 (37.5%)
	Bias (°C)	1.34 (55.8%)	2.37 (62.2%)	1.23 (40.3%)
>99.9 percentile of observed Ta	MAE (°C)	1.20 (57.2%)	1.66 (57.3%)	1.22 (46.9%)
	RMSE (°C)	1.32 (52.8%)	1.80 (51.9%)	1.19 (41.5%)
	Bias (°C)	1.28 (62.1%)	1.82 (63.1%)	1.28 (49.3%)
>99 percentile of observed Ta	MAE (°C)	0.96 (56.1%)	1.27 (56.0%)	0.84 (45.2%)
	RMSE (°C)	1.10 (52.5%)	1.48 (52.1%)	0.88 (40.6%)
	Bias (°C)	1.11 (68.1%)	1.45 (65.8%)	0.99 (54.2%)
<1 percentile of observed Ta	MAE (°C)	0.67 (33.2%)	1.50 (51.5%)	0.79 (27.0%)
	RMSE (°C)	0.75 (28.9%)	1.71 (45.4%)	0.86 (24.0%)
	Bias (°C)	0.74 (60.2%)	1.74 (72.9%)	0.98 (49.0%)
<0.1 percentile of observed Ta	MAE (°C)	0.77 (29.4%)	2.10 (45.8%)	0.88 (24.1%)
	RMSE (°C)	0.88 (26.6%)	2.28 (41.8%)	1.00 (22.4%)
	Bias (°C)	1.04 (50.6%)	2.61 (61.3%)	1.21 (38.5%)
<0.01 percentile of observed Ta	MAE (°C)	1.42 (42.0%)	3.66 (54.5%)	1.69 (32.4%)
	RMSE (°C)	1.49 (37.2%)	3.46 (46.7%)	1.69 (28.6%)
	Bias (°C)	1.75 (52.6%)	4.21 (62.7%)	1.92 (37.3%)

Table 3
MAE, RMSE, R^2, and bias of the estimated HWM and CWM using the EPC method and the method in the literature.

		HWM	CWM
EPC method	MAE (°C)	0.62	0.89
	RMSE (°C)	0.84	1.22
	R^2	0.81	0.96
	Bias (°C)	−0.35	0.43
Method in the literature	MAE (°C)	1.75	2.03
	RMSE (°C)	2.06	2.40
	R^2	0.66	0.95
	Bias (°C)	−1.71	1.98

as follows: First, the relationship between Ta and the independent variables was fitted, and Ta was predicted by month (step (2) of EPC). This strategy was used because the relationship between Ta and the independent variables differed greatly by month. Therefore, using only one fitted relationship to predict Ta for all months would result in large uncertainties. Second, the EPC method obtains, predicts, and corrects the error of the estimated Ta (steps (3)–(6) of EPC). These steps can reduce errors and increase the accuracy of the estimated Ta. As can be seen from Fig. 9, the estimated Ta using the method in the literature had a positive bias on relatively cold days and a negative bias on relatively hot days. It suggests that the variation of error has certain regularity and Ta estimated by the Cubist model is conservative. One possible reason for the conservative estimation is that the extreme Ta is generally significantly higher (lower) than the spatiotemporal neighboring Ta and thus difficult to predict. The other reason may be that the model treats extremely high (low) values as outliers. The

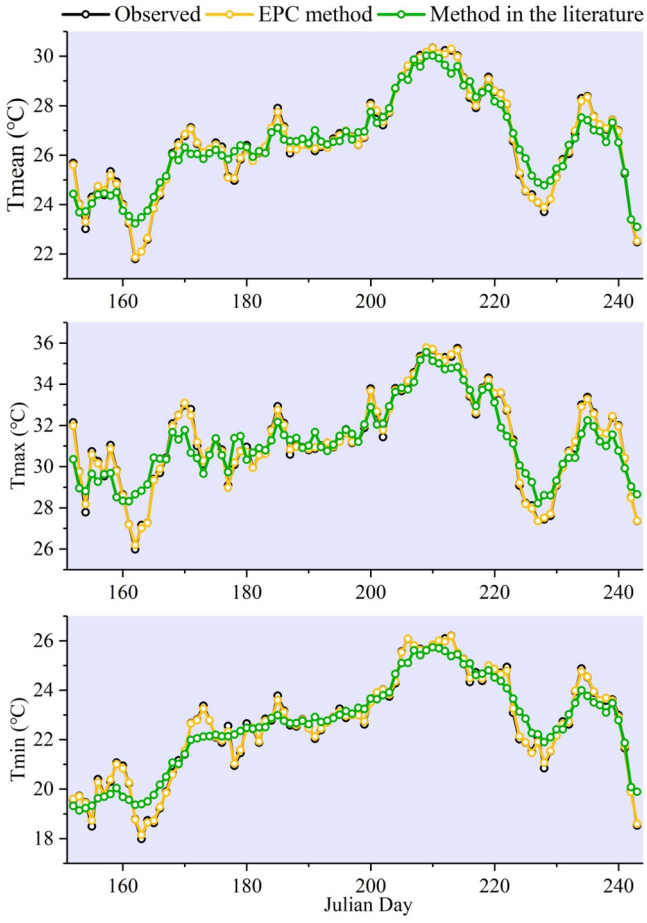


Fig. 9. Temporal variations in observed and estimated Ta averaged for 1104 meteorological stations in the summer of 2003.

EPC method can obtain the positive bias on relatively cold days and negative bias on relatively hot days by performing step (3) and then predict and correct the error by performing steps (5) and (6). After correcting the error, the estimated Ta was closer to the observed Ta (Fig. 9). Therefore, the EPC method had a low error (MAE and RMSE) and bias than the method in the literature.

The main limitation of EPC method is that the computing costs are significantly higher than those of the method in the literature. The computing time of the EPC method for training is approximately ten times that of the method in the literature, primarily because the former uses a tenfold cross-validation method to obtain the error (step (3) of EPC). Fortunately, only once is training necessary to predict the Ta for the whole study period. The computing time of the EPC method for predicting is approximately twice that of the method in the literature. This is because the EPC method needs to predict the Ta and error simultaneously (step (5) of EPC). Therefore, future research should consider the computing costs when using the EPC method.

5.2. Using the EPC method in research related to extreme ta

Climate change has become a serious problem. Global surface Ta has been increasing and will continue to increase (Huang et al., 2017; Zeng et al., 2017). As global warming continues, extreme weather events gradually increase, which harms human beings and the environment (Chen et al., 2020; Luo and Lau, 2017; Robine et al., 2008). For example, heat waves in Europe caused more than 70,000 deaths in the summer of 2003 (Robine et al., 2008). Chen et al. (2020) found that maize and rice yields were significantly affected by extreme Ta in the Yangtze River Basin, China. Accurate, high spatiotemporal resolution Ta data are necessary to accurately reveal spatiotemporal variations in extreme Ta and its negative impacts. The EPC method proposed in this study can produce 1 km resolution daily Ta data with high accuracy. Additionally, this study demonstrated that the EPC method can produce a satisfactory estimation of the extreme Ta. Most importantly, the EPC method performs well in practical applications: the MAEs of the estimated HWM and CWM using the EPC method were 0.62 °C and 0.89 °C, respectively. Therefore, further research could use the EPC method to map Ta data and in studies related to extreme Ta. For example, it can be used to analyze the spatiotemporal variations in heat and cold waves at a 1 km resolution (Chung et al., 2020). Additionally, it can be used to study the urban heat island effect and its associated determinants (Yao et al., 2021a). Furthermore, it can be used to map the spatial distribution of heat health risk (Hu et al., 2017). Using an accurate Ta map at a resolution of 1 km for these studies is more advanced than using meteorological stations, because the meteorological stations are sparsely and unevenly distributed (Wang et al., 2015).

The time period of this study is from 2003 to 2012. The estimated Ta was not validated from a more recent year (e.g. 2020), due to data limitations. Daily Ta data from meteorological stations were available from 1961 to 2014. MODIS version 6.1 MYD11 Ts data were available from July 2002 to December 2012 when this manuscript was submitted. Therefore, the time period of this study is from 2003 to 2012. The time period can be extended when longer time series of satellite and meteorological data are available in the future.

6. Conclusions

In this study, a method called EPC incorporating the Cubist algorithm was proposed to improve the estimation of extreme Ta, and compared with method in the literature. The MAEs of the estimated daily Ta using the EPC method ranged from 0.75–1.01 °C, which were 0.57–0.96 °C lower than those of the method in the literature. However, the biases of the estimated Ta using the method in the literature can be as high as 7.096 °C when Ta is extremely low and as low as −3.09 °C when Ta is extremely high. Compared with the method in the literature, the EPC method can reduce the MAE by 1.41 °C, RMSE by 1.49 °C and bias by 1.61 °C of the estimated extreme Ta. Additionally, the EPC method produced satisfactory

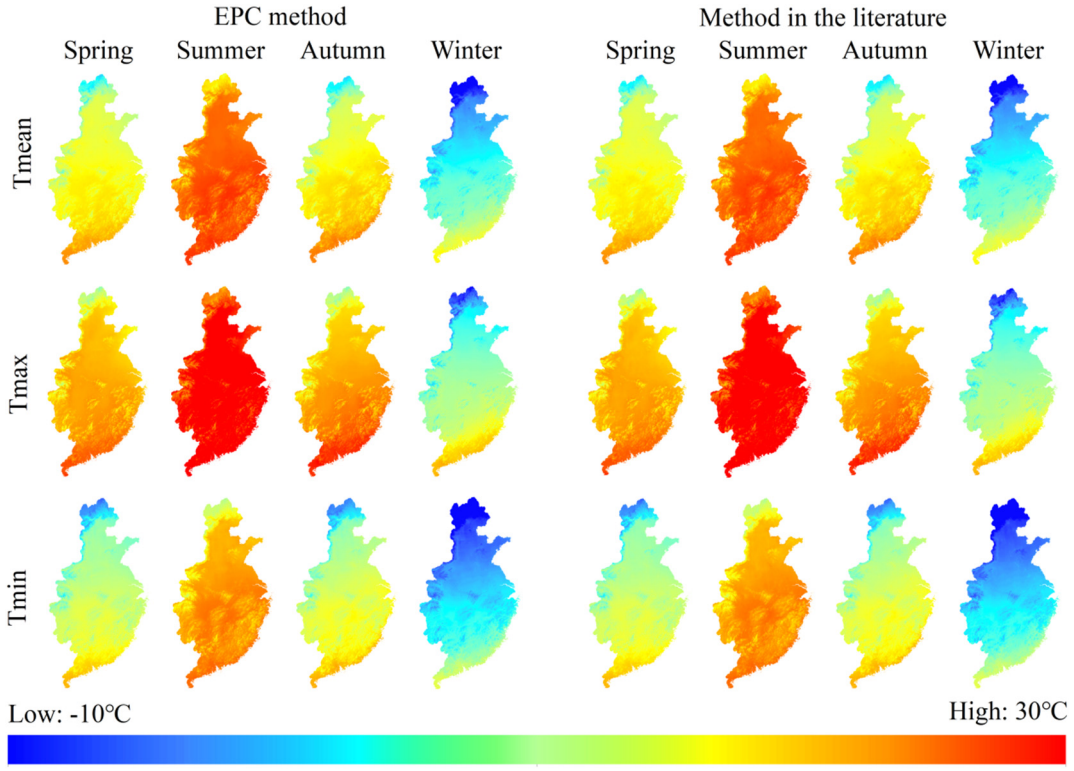


Fig. 10. Spatial and seasonal variations in estimated Ta using the EPC method and the method in the literature. Spring: March, April, May. Summer: June, July, August. Autumn: September, October, November. Winter: December, January, February.

accuracy of the estimated HWM and CWM (MAEs <0.9 °C). The reasons for the high accuracy of the EPC method are as follows: (1) the relationship between Ta and independent variables was fitted, and the Ta was predicted by month; and (2) the EPC method uses an error correction procedure to reduce the error and bias.

This paper for the first time studied the phenomenon that the slopes of the fitting lines between observed and estimated Ta are generally lower than 1. Additionally, the EPC method was proposed to improve this problem. Furthermore, the developed 1 km resolution daily Ta map in east China from 2003 to 2012 using the EPC method will be a useful data source in multiple research fields. Future research could use the estimated Ta data to analyze the spatiotemporal variations in extreme Ta and the effects of extreme Ta on crop yield and human health.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was financially supported by the National Natural Science Foundation of China (41771360, 41975044 and 41801021), the Special Fund for Basic Scientific Research of Central Colleges, China University of Geosciences (Wuhan) (CUGL170401 and CUGCJ1704), and the Fundamental Research Funds for National Universities, China University of Geosciences (Wuhan).

CRediT authorship contribution statement

Rui Yao: Data curation, Writing- Original draft preparation, revision. **Lunche Wang:** Conceptualization, Methodology, Software, Reviewing and Editing, Supervision, Project administration, Funding acquisition, revision. **Xin Huang**: Methodology, Software, Supervision. **Qian Cao:** Methodology, Software, revision. **Yuanyuan Peng:** Software, Methodology.

References

Anniballe, R., Bonafoni, S., Pichierri, M., 2014. Spatial and temporal trends of the surface and air heat island over Milan using MODIS data. Remote Sens. Environ. 150, 163–171.

Athmania, D., Achour, H., 2014. External validation of the ASTER GDEM2, GMTED2010 and CGIAR-CSI- SRTM v4.1 free access Digital Elevation Models (DEMs) in Tunisia and Algeria. Remote Sens. 6, 4600–4620.

Benali, A., Carvalho, A.C., Nunes, J.P., Carvalhais, N., Santos, A., 2012. Estimating air surface temperature in Portugal using MODIS LST data. Remote Sens. Environ. 124, 108–121.

Cao, Q., Yu, D., Georgescu, M., Wu, J., Wang, W., 2018. Impacts of future urban expansion on summer climate and heat-related human health in eastern China. Environ. Int. 112, 134–146.

Chen, Y., Quan, J., Zhan, W., Guo, Z., 2016. Enhanced statistical estimation of air temperature incorporating nighttime light data. Remote Sens. 8, 656.

Chen, X., Wang, L., Niu, Z., Zhang, M., Li, C.a., Li, J., 2020. The effects of projected climate change and extreme climate on maize and rice in the Yangtze River Basin, China. Agr. Forest Meteorol. 282–283 107867.

Chung, J., Lee, Y., Jang, W., Lee, S., Kim, S., 2020. Correlation analysis between air temperature and MODIS land surface temperature and prediction of air temperature using TensorFlow long short-term memory for the period of occurrence of cold and heat waves. Remote Sens. 12, 3231.

Croitoru, A.-E., Piticar, A., Ciupertea, A.-F., Roşca, C.F., 2016. Changes in heat waves indices in Romania over the period 1961–2015. Glob. Planet. Chang. 146, 109–121.

Danielson, J.J., Gesch, D.B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010). U.S. Geological Survey Open-File Report. 2011–1073.

Didari, S., Zand-Parsa, S., 2018. Enhancing estimation accuracy of daily maximum, minimum, and mean air temperature using spatio-temporal ground-based and remote-sensing data in southern Iran. Int. J. Remote Sens. 39, 6316–6339.

Didari, S., Norouzi, H., Zand-Parsa, S., Khanbilvardi, R., 2016. Estimation of daily minimum land surface air temperature using MODIS data in southern Iran. Theor. Appl. Climatol. 130, 1149–1161.

Du, H., Zhan, W., Liu, Z., Li, J., Li, L., Lai, J., et al., 2021. Simultaneous investigation of surface and canopy urban heat islands over global cities. ISPRS J. Photogramm. 181, 67–83.

Feng, Y., Hao, W., Li, H., Cui, N., Gong, D., Gao, L., 2020. Machine learning models to quantify and map daily global solar radiation and photovoltaic power. Renew. Sustain. Energy Rev. 118, 109393.

Gaitan, E., Monjo, R., Portoles, J., Pino-Otin, M.R., 2019. Projection of temperatures and heat and cold waves for Aragon (Spain) using a two-step statistical downscaling of CMIP5 model outputs. Sci. Total Environ. 650, 2778–2795.
Gao, M., Li, Z., Tan, Z., Li, H., Peng, J., 2021. Use of Google Earth Engine to generate a 20-year 1 km × 1 km monthly air temperature product over Yellow River Basin. IEEE J-STARS. 14, 10079–10090.
Hereher, M.E., El Kenawy, A., 2020. Extrapolation of daily air temperatures of Egypt from MODIS LST data. Geocarto Int. 1–17.
Hooker, J., Duveiller, G., Cescatti, A., 2018. A global dataset of air temperature derived from satellite remote sensing and weather stations. Sci. Data 5, 180246.
Hrisko, J., Ramamurthy, P., Yu, Y., Yu, P., Melecio-Vázquez, D., 2020. Urban air temperature model using GOES-16 LST and a diurnal regressive neural network algorithm. Remote Sens.Environ. 237.
Hu, K., Yang, X., Zhong, J., Fei, F., Qi, J., 2017. Spatially explicit mapping of heat health risk utilizing environmental and socioeconomic data. Environ. Sci. Technol. 51, 1498–1507.
Hu, Y., Hou, M., Jia, G., Zhao, C., Zhen, X., Xu, Y., 2019. Comparison of surface and canopy urban heat islands within megacities of eastern China. ISPRS J. Photogramm. 156, 160–168.
Huang, J., Yu, H., Dai, A., Wei, Y., Kang, L., 2017. Drylands face potential threat under 2 °C global warming target. Nat. Clim. Chang. 7, 417–422.
Janatian, N., Sadeghi, M., Sanaeinejad, S.H., Bakhshian, E., Farid, A., Hasheminia, S.M., et al., 2017. A statistical framework for estimating air temperature using MODIS land surface temperature data. Int. J. Climatol. 37, 1181–1194.
Jang, K., Kang, S., Kimball, J., Hong, S., 2014. Retrievals of all-weather daily air temperature using MODIS and AMSR-E data. Remote Sens. 6, 8387–8404.
Khalid, N.F., Din, A.H.M., Omar, K.M., Khanan, M.F.A., Omar, A.H., Hamid, A.I.A., Pa'suya, M.F., 2016. Open-source Digital Elevation Model (DEMs) evaluation with GPS and LiDAR data. ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 299–306.
Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., 2014. Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the northeastern USA. Remote Sens. Environ. 150, 132–139.
Li, L., Zha, Y., 2019a. Estimating monthly average temperature by remote sensing in China. Adv. Space Res. 63, 2345–2357.
Li, L., Zha, Y., 2019b. Satellite-based regional warming hiatus in China and its implication. Sci. Total Environ. 648, 1394–1402.
Li, X., Zhou, Y., Asrar, G.R., Zhu, Z., 2018. Developing a 1 km resolution daily air temperature dataset for urban and surrounding areas in the conterminous United States. Remote Sens. Environ. 215, 74–84.
Lin, X., Zhang, W., Huang, Y., Sun, W., Han, P., Yu, L., et al., 2016. Empirical estimation of near-surface air temperature in China from MODIS LST data by considering physiographic features. Remote Sens. 8, 629.
Liu, Z., Liu, Y., Wang, S., Yang, X., Wang, L., Baig, M.H.A., et al., 2018. Evaluation of spatial and temporal performances of ERA-interim precipitation and temperature in mainland China. J. Clim. 31 (11), 4347–4365.
Liu, D., Mishra, A.K., Ray, D.K., 2020. Sensitivity of global major crop yields to climate variables: a non-parametric elasticity analysis. Sci. Total Environ. 748, 141431.
Lu, N., Liang, S., Huang, G., Qin, J., Yao, L., Wang, D., et al., 2018. Hierarchical Bayesian space-time estimation of monthly maximum and minimum surface air temperature. Remote Sens. Environ. 211, 48–58.
Luo, M., Lau, N.-C., 2017. Heat waves in southern China: synoptic behavior, long-term change, and urbanization effects. J. Clim. 30, 703–720.
Luo, M., Lau, N.-C., 2018. Increasing heat stress in urban areas of eastern China: acceleration by urbanization. Geophys. Res. Lett. 45, 13060–13069.
Marthews, T.R., Dadson, S.J., Lehner, B., Abele, S., Gedney, N., 2015. High-resolution global topographic index values for use in large-scale hydrological modelling. Hydrol. Earth Syst. Sci. 19, 91–104.
Meyer, H., Katurji, M., Appelhans, T., Müller, M., Nauss, T., Roudier, P., et al., 2016. Mapping daily air temperature for Antarctica based on MODIS LST. Remote Sens. 8, 732.
Niu, Z., Wang, L., Fang, L., Li, J., Yao, R., 2019. Analysis of spatiotemporal variability in temperature extremes in the Yellow and Yangtze River basins during 1961–2014 based on high-density gauge observations. Int. J. Climatol. 40 (1), 1–21.
Noi, P., Degener, J., Kappas, M., 2017. Comparison of multiple linear regression, cubist regression, and Random Forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. Remote Sens. 9, 398.
Peón, J., Recondo, C., Calleja, J.F., 2014. Improvements in the estimation of daily minimum air temperature in peninsular Spain using MODIS land surface temperature. Int. J. Remote Sens. 35, 5148–5166.
Qin, W., Wang, L., Lin, A., Zhang, M., Xia, X., Hu, B., et al., 2018. Comparison of deterministic and data-driven models for solar radiation estimation in China. Renew. Sustain. Energy Rev. 81, 579–594.
Quinlan, R., 1992. Learning with continuous classes. Proceedings of the 5th Australia Joint Conference on Artificial Intelligence, Hobart, Tasmania, pp. 343–348.
Rao, Y., Liang, S., Wang, D., Yu, Y., Song, Z., Zhou, Y., et al., 2019. Estimating daily average surface air temperature using satellite land surface temperature and top-of-atmosphere radiation products over the Tibetan Plateau. Remote Sens. Environ. 234, 111462.
Ren, G., Zhou, Y., 2014. Urbanization effect on trends of extreme temperature indices of national stations over mainland China, 1961–2008. J. Clim. 27, 2340–2360.
Robine, J.M., Cheung, S.L., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J.P., et al., 2008. Death toll exceeded 70,000 in Europe during the summer of 2003. C. R. Biol. 331, 171–178.
Rosenfeld, A., Dorman, M., Schwartz, J., Novack, V., Just, A.C., Kloog, I., 2017. Estimating daily minimum, maximum, and mean near surface air temperature using hybrid satellite models across Israel. Environ. Res. 159, 297–312.
Shen, H., Jiang, Y., Li, T., Cheng, Q., Zeng, C., Zhang, L., 2020. Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data. Remote Sens. Environ. 240, 111692.
Sun, Y., Zhang, X., Zwiers, F.W., Song, L., Wan, H., Hu, T., et al., 2014. Rapid increase in the risk of extreme summer heat in eastern China. Nat. Clim. Chang. 4, 1082–1085.
Sun, Y., Zhang, X., Ren, G., Zwiers, F.W., Hu, T., 2016. Contribution of urbanization to warming in China. Nat. Clim. Chang. 6, 706–709.
Vancutsem, C., Ceccato, P., Dinku, T., Connor, S.J., 2010. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sens. Environ. 114, 449–465.
Venter, Z.S., Brousse, O., Esau, I., Meier, F., 2020. Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data. Remote Sens.Environ. 242.
Wan, Z., 2008. New refinements and validation of the MODIS land-surface temperature/emissivity products. Remote Sens. Environ. 112, 59–74.
Wan, Z., 2014. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. Remote Sens. Environ. 140, 36–45.
Wang, F., Ge, Q., Wang, S., Li, Q., Jones, P.D., 2015. A new estimation of urbanization's contribution to the warming trend in China. J. Clim. 28, 8923–8938.
Wang, L., Kisi, O., Zounemat-Kermani, M., Salazar, G.A., Zhu, Z., Gong, W., 2016. Solar radiation prediction using different techniques: model evaluation and comparison. Renew. Sustain. Energy Rev. 61, 384–397.
Wei, J., Li, Z., Cribb, M., Huang, W., Xue, W., Sun, L., et al., 2020. Improved 1km resolution PM2:5 estimates across China using enhanced space–time extremely randomized trees. Atmos. Chem. Phys. 20, 3273–3289.
Weiss, D.J., Bhatt, S., Mappin, B., Boeckel, T.P.V., Smith, D.L., Hay, S.I., Gething, P.W., 2014. Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000–2012: a high-resolution spatiotemporal prediction. Malaria J. 13, 171.
Xu, W., Li, Q., Wang, X.L., Yang, S., Cao, L., Feng, Y., 2013. Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. J. Geophys. Res.-Atmos. 118, 9708–9720.
Xu, Y., Knudby, A., Shen, Y., Liu, Y., 2018. Mapping monthly air temperature in the Tibetan Plateau from MODIS data based on machine learning methods. IEEE J-STARS. 11, 345–354.
Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., et al., 2019. Spatiotemporal continuous estimates of PM2.5 concentrations in China, 2000–2016: a machine learning method with inputs from satellites, chemical transport model, and ground observations. Environ. Int. 123, 345–357.
Yang, J., Huang, X., 2021. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. Earth Syst. Sci. Data 13, 3907–3925.
Yang, X., Hou, Y., Chen, B., 2011. Observed surface warming induced by urbanization in east China. J. Geophys. Res. 116, D14113.
Yang, Y., Cai, W., Yang, J., 2017. Evaluation of MODIS land surface temperature data to estimate near-surface air temperature in northeast China. Remote Sens. 9.
Yao, R., Wang, L., Huang, X., Li, L., Sun, J., Wu, X., et al., 2020. Developing a temporally accurate air temperature dataset for mainland China. Sci. Total Environ. 706, 136037.
Yao, R., Wang, L., Huang, X., Liu, Y., Niu, Z., Wang, S., et al., 2021a. Long-term trends of surface and canopy layer urban heat island intensity in 272 cities in the mainland of China. Sci. Total Environ. 772, 145607.
Yao, R., Wang, L., Huang, X., Sun, L., Chen, R., Wu, X., et al., 2021b. A robust method for filling the gaps in MODIS and VIIRS land surface temperature data. IEEE T. Geosci. Remote 1–15.
Yoo, C., Im, J., Park, S., Quackenbush, L.J., 2018. Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data. ISPRS J. Photogramm. 137, 149–162.
Zeng, Z., Piao, S., Li, L.Z.X., Zhou, L., Ciais, P., Wang, T., et al., 2017. Climate mitigation from vegetation biophysical feedbacks during the past three decades. Nat. Clim. Chang. 7, 432–436.
Zhan, Y., Luo, Y., Deng, X., Chen, H., Grieneisen, M.L., Shen, X., et al., 2017. Spatiotemporal prediction of continuous daily PM2.5 concentrations across China using a spatially explicit machine learning algorithm. Atmos. Environ. 155, 129–139.
Zhang, L.W., Huang, J.F., Guo, R.F., Li, X.X., Sun, W.B., Wang, X.Z., 2013. Spatio-temporal reconstruction of air temperature maps and their application to estimate rice growing season heat accumulation using multi-temporal MODIS data. J. Zhejiang Univ. Sci. B 14, 144–161.
Zhang, H., Zhang, F., Ye, M., Che, T., Zhang, G., 2016. Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data. J. Geophys. Res.-Atmos. 121, 11425–11441.
Zhang, Y., Song, C., Band, L.E., Sun, G., Li, J., 2017. Reanalysis of global terrestrial vegetation trends from MODIS products: browning or greening? Remote Sens.Environ. 191, 145–155.
Zhang, W., Zhang, B., Zhu, W., Tang, X., Li, F., Liu, X., et al., 2021. Comprehensive assessment of MODIS-derived near-surface air temperature using wide elevation-spanned measurements in China. Sci. Total Environ. 800, 149535.
Zhao, P., Jones, P., Cao, L., Yan, Z., Zha, S., Zhu, Y., et al., 2014. Trend of surface air temperature in eastern China and associated large-scale climate variability over the last 100 years. J. Clim. 27, 4693–4703.
Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., et al., 2019. Satellite remote sensing of surface urban heat islands: progress, challenges, and perspectives. Remote Sens. 11, 48.
Zhu, W., Lű, A., Jia, S., Yan, J., Mahmood, R., 2017. Retrievals of all-weather daytime air temperature from MODIS products. Remote Sens. Environ. 189, 152–163.
Zhu, X., Zhang, Q., Xu, C.Y., Sun, P., Hu, P., 2019. Reconstruction of high spatial resolution surface air temperature data across China: a new geo-intelligent multisource data-based machine learning technique. Sci. Total Environ. 665, 300–313.
Zou, L., Li, G., Xu, S., 2021. A novel method for optimizing air temperature estimation and quantifying canopy layer heat island intensity in eastern and Central China. Adv. Space Res. 68, 3291–3301.