Science of the Total Environment 706 (2020) 136037

Contents lists available at ScienceDirect

Science of the Total Environment

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

Developing a temporally accurate air temperature dataset for Mainland China

Rui Yao [a], Lunche Wang [a,*], Xin Huang [b,c], Long Li [d], Jia Sun [a], Xiaojun Wu [a], Weixia Jiang [a]

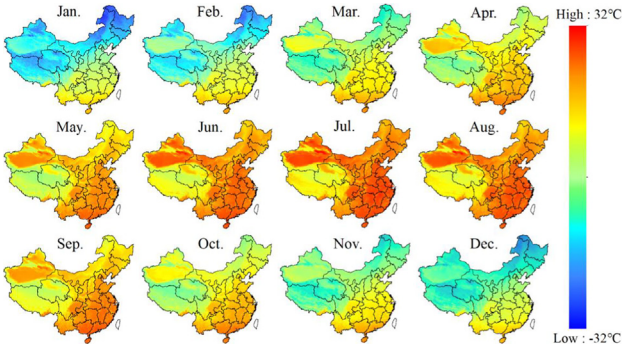
[a] Hubei Key Laboratory of Critical Zone Evolution, School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China
[b] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[c] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[d] Key Laboratory of Virtual Geographic Environment of Ministry of Education, Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, College of Geographic Science, Nanjing Normal University, Nanjing 210023, China

HIGHLIGHTS

- The temporal accuracy of estimated monthly Ta in China was analyzed.
- Inputting data for the same month into the model can generate more accurate Ta.
- Using temporal variables can significantly increase the accuracy of estimated Ta.
- An accurate, national coverage, 1 km resolution Ta dataset was developed.

GRAPHICAL ABSTRACT

Spatial and seasonal variations in monthly mean Ta averaged from 2001 to 2018 for Mainland China.



ARTICLE INFO

Article history:
Received 13 August 2019
Received in revised form 26 November 2019
Accepted 8 December 2019
Available online 09 December 2019

Editor: Pavlos Kassomenos

Keywords:
Air temperature
Land surface temperature
Remote sensing
Machine learning
MODIS
China

ABSTRACT

Spatially continuous satellite data have been widely used to estimate monthly air temperature (Ta). However, it is still not clear whether the estimated monthly Ta is temporally consistent with observed Ta or not. In this study, the accuracies of interannual variations and temporal trends in estimated monthly Ta were systematically analyzed for Mainland China during 2001–2018. The differences in accuracy among five ways to input data into the model were investigated. The Cubist algorithm and ten variables were used to estimate monthly Ta. It was found that inputting data for the same month into the model can generate more accurate results than inputting all data into the model. Using temporal variables (i.e., month and year) can significantly increase the accuracy of estimated Ta. These results can be explained by different relationships between Ta and auxiliary variables that appear at different times. Thus, using temporal variables can help distinguish between different relationships and improve accuracy levels of the estimated Ta. When applying the best method (inputting data for the same month into the model and using the year as a temporal variable), the coefficient of determination (R^2) of estimated monthly mean Ta, interannual variations in monthly mean Ta and temporal trends in monthly mean Ta were recorded as 0.997, 0.731 and 0.848, respectively. The root mean squared errors (RMSEs) of estimated monthly mean Ta, interannual variations in monthly mean Ta and temporal trends in monthly mean Ta were recorded as 0.629 °C, 0.593 °C and 0.201 °C/decade, respectively. An accurate, national coverage, 1 km spatial resolution and long time series

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

https://doi.org/10.1016/j.scitotenv.2019.136037

(2001–2018) monthly mean, maximum and minimum Ta dataset was finally developed. The dataset can be of great use to many fields such as climatology, hydrology and ecology.

1. Introduction

Near surface air temperature (Ta) is a key variable in multiple fields including climatology, hydrology, ecology and epidemiology (Alkama and Cescatti, 2016; Medina-Ramon et al., 2006; Warren et al., 2018; Liu et al., 2018a). Traditionally, Ta is measured by meteorological stations at a height of 2 m. It has high accuracy and temporal resolution but only provides point spatial coverage, which is usually inadequate for urban heat island (UHI), climate change and epidemiological studies (Kloog et al., 2014; Zhu et al., 2019; Zhou et al., 2019; Xu et al., 2018; Liu et al., 2018b). Fortunately, the wall-to-wall coverage of satellite observations can provide spatially continuous data. Land surface temperature (Ts) derived from satellite sensors shows strong relationships with Ta. Thus, spatially continuous Ta data can be developed using satellite Ts data. Moderate Resolution Imaging Spectroradiometer (MODIS) Ts data have been widely used to estimate Ta owing to their high temporal resolution and global coverage (Benali et al., 2012; Kloog et al., 2014; Li and Zha, 2019a; Lu et al., 2018; Vancutsem et al., 2010; Zhu et al., 2017; Zhu et al., 2019).

The Ta estimation using MODIS Ts data can generally be grouped into two categories. Some studies have estimated daily Ta (Kloog et al., 2012; Kloog et al., 2014; Li et al., 2018; Lin et al., 2012; Yoo et al., 2018; Zhu et al., 2013). For example, Li et al. (2018) used gapfilled MODIS Ts data and a geographically weighted regression method to map daily maximum and minimum Ta in the conterminous United States for 2010. The authors found that the root mean squared errors (RMSEs) of predicted daily maximum and minimum Ta were 2.1 and 1.9 °C, respectively. Other studies have mapped monthly Ta (Hooker et al., 2018; Li and Zha, 2019a; Lin et al., 2016; Lu et al., 2018; Xu et al., 2018; Zhu et al., 2019). For example, Lu et al. (2018) used a hierarchical Bayesian method to estimate monthly Ta in Qinghai province (China) for the period of 2003–2011, and recorded the RMSEs of estimated monthly maximum and minimum Ta as 2.15 and 1.97 °C, respectively. Using the Cubist model and 11 variables, Xu et al. (2018) mapped monthly mean Ta for the Tibetan Plateau during 2001–2015. Mean absolute error (MAE) and RMSE were recorded as 0.73 and 1 °C, respectively. Hooker et al. (2018) used geographically and climate space weighted regressions to map monthly mean Ta at a global scale during 2003–2016, finding RMSEs of 1.14–1.55 °C.

However, two questions still have not been comprehensively addressed by the previous studies. First, are the interannual variations and long-term trends of estimated monthly Ta consistent with those of observed monthly Ta? Accurate interannual variations in estimated Ta are needed when investigating climate variability and its relationships with other variables (e.g., vegetation dynamics, animal growth and other climate variables) (Honsey et al., 2019; Piao et al., 2014; Wang et al., 2019; Zhang et al., 2017; Liu et al., 2017), and accurate long-term trends of estimated monthly Ta are needed to evaluate patterns of climate change (e.g., global warming and changes in UHI) (Li and Zha, 2019b; Li and Zha, 2019c; Xu et al., 2018). However, to our knowledge, only Li and Zha (2019b) have shown that the accuracy of temporal trends of estimated monthly Ta was acceptable (coefficient of determination (R^2): 0.630, MAE: 0.00132 °C/month, RMSE: 0.00178 °C/month) for the period of 2001–2015 in China. Second, previous studies generally input all data (all years and months) to build a model when estimating monthly Ta (Li and Zha, 2019a; Lu et al., 2018; Xu et al., 2018; Zhu et al., 2019). The model can also be built using: (1) data for only one month and (2) data for the same month from multiple years. To our knowledge, few studies have compared the accuracy of different methods to input data. Thus, the following question is raised as well: do different approaches to inputting data vary in accuracy?

The present study aims at answering the two questions raised above, and at developing an accurate monthly Ta dataset. Mainland China (Fig. 1) was selected as our study area, as it is a hot spot of climate change and UHI research (Huang and Wang, 2019; Luo and Lau, 2017, 2018; Niu et al., 2019; Yang et al., 2019; Yao et al., 2017; Zhou et al., 2019). Specifically, the objectives of this study are: (1) to investigate the accuracy of estimated monthly Ta for Mainland China for the period of 2001–2018; (2) to analyze the consistency of interannual variations in estimated and observed monthly Ta; (3) to reveal the consistency of long-term trends of estimated and observed monthly Ta; and (4) to develop a spatially and temporally accurate monthly mean, maximum and minimum Ta dataset for Mainland China for the period of 2001–2018.

2. Data and methods

2.1. Data information and preprocessing

The data used in the present study are presented in Table 1. Daily mean, maximum and minimum Ta data drawn from 699 meteorological stations during 2001–2018 were collected from the China Meteorological Data Service Center (CMDC). 5 stations have more than 7 missing daily values in a month. They were excluded from this study and the remaining 694 meteorological stations were thus used (Fig. 1). Subsequently, daily mean, maximum and minimum Ta were averaged to monthly mean, maximum and minimum Ta. There are 0.03% missing daily values in the remaining 694 meteorological stations. These missing daily values were ignored when daily values averaged to monthly values. For example, if there were 30 valid and 1 missing Ta values in a month, the monthly mean Ta was calculated as the average Ta of the 30 valid Ta values.

Ts and vegetation information for the period of 2001–2018 were derived from MODIS MOD11A2 and MOD13A3 enhanced vegetation index (EVI) data, respectively (Table 1). Data drawn from the Terra satellite rather than those from the Aqua satellite were used, as they cover longer time series (Terra satellite: February 2000 to the present; Aqua satellite: July 2002 to the present) (He et al., 2017; Yao et al., 2019; Yao et al., 2018). These data have been widely validated and have shown to exhibit good performance (Huete et al., 2002; Wan, 2008; Wan, 2014). Eight-day composite Ts was averaged into monthly Ts (missing values were not processed (Lai et al., 2018; Li et al., 2017)). In addition, latitude and longitude information was obtained from MODIS data.

Monthly solar radiation data with a 1 km spatial resolution were mapped using the Area Solar Radiation Tool available in ArcGIS software (Xu et al., 2018; Yoo et al., 2018). Topographic index data were downloaded directly from the Environmental Information Data Centre. Positive and negative topographic index values of this dataset represent valley and ridge landforms, respectively. Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Digital Elevation Model (DEM) data were derived from Geospatial Data Cloud. Topographic index and DEM data were resampled to a 1 km spatial resolution. Monthly solar radiation was assumed to remain the same in different years. Latitude, longitude, topographic index and DEM data were assumed to remain the same over the whole study period (2001–2018).

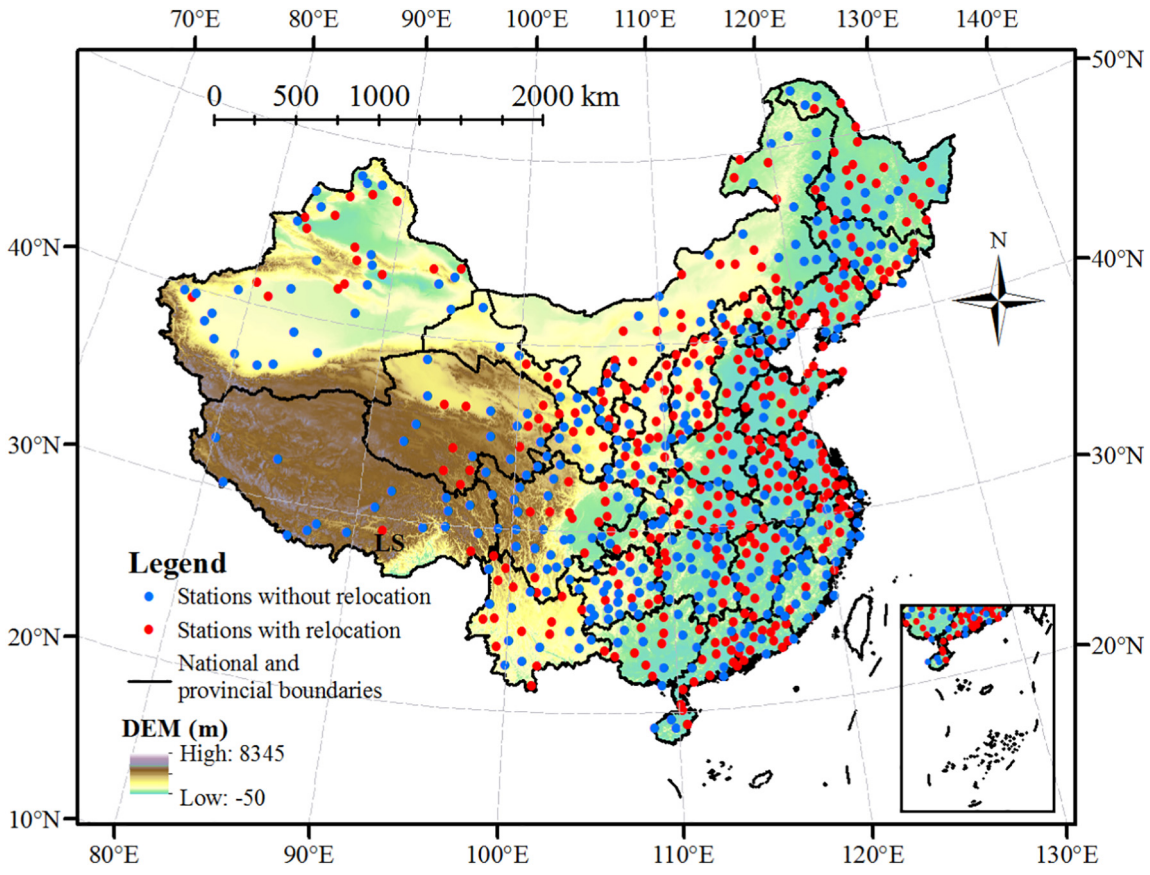


Fig. 1. Study area and locations of the 694 meteorological stations of this study.

2.2. Methods

2.2.1. Variable and model selection

Daytime Ts, nighttime Ts, EVI, solar radiation, topographic index, DEM, latitude and longitude were used as auxiliary variables in this study, since these variables have positive effects in predicting Ta according to the published literatures (Li and Zha, 2019a; Noi et al., 2016; Xu et al., 2018; Yoo et al., 2018). Machine learning methods have been widely used to estimate Ta (Li and Zha, 2019a; Li and Zha, 2019b; Noi et al., 2016; Yoo et al., 2018; Zhang et al., 2016). The Cubist algorithm (Quinlan, 1992) is a rule-based algorithm that is an extension of the M5 model tree. While it is a commercial product, it is also a widely used regression and classification algorithm, as it is made available through R statistical software (Kuhn et al., 2018). Among various machine learning methods, the Cubist algorithm is generally more precise than other methods in terms of predicting Ta (Noi et al., 2017; Xu et al., 2018; Zhang et al., 2016). For instance, Xu et al. (2018) found the accuracy of Cubist algorithm to be more accurate than ten other machine learning models. Zhang et al. (2016) found the Cubist and random forest algorithms to be more accurate than four other algorithms. Thus, in the present study, the Cubist algorithm was used to estimate Ta using the 'Cubist' add-on package available in R. Model parameters were selected using the bootstrap method via the 'caret' add-on package.

Table 1
Data used in this study.

Data	Information	Source
Meteorological data	1 day temporal resolution, 694 stations	http://data.cma.cn/
Ts	8-day composite, 1 km spatial resolution	https://ladsweb.modaps.eosdis.nasa.gov/search/
EVI	Monthly composite, 1 km spatial resolution	https://ladsweb.modaps.eosdis.nasa.gov/search/
Latitude and longitude	1 km spatial resolution	MODIS data
Solar radiation	Monthly composite, 1 km spatial resolution	Calculated in ArcGIS software
Topographic index	15 second spatial resolution	http://eidc.ceh.ac.uk/
DEM	30 m spatial resolution	http://www.gscloud.cn/

Ts: land surface temperature. EVI: enhanced vegetation index. MODIS: Moderate Resolution Imaging Spectroradiometer. DEM: digital elevation model.

2.2.2. Different approaches to input data

The values of pixels positioned closest to each meteorological station were extracted from daytime Ts, nighttime Ts, EVI, solar radiation and topographic index data. DEM, latitude and longitude information obtained from meteorological stations was directly applied. Subsequently, these variables (independent variables) and Ta (dependent variable) were input into the Cubist model. To analyze differences in the accuracy levels of different ways to input data, a total of 5 methods were used in the present study (Table 2). Method 1 involved using data for one month to build the model and predicting Ta for the corresponding month. Method 2 involved using data for the same month from all years to build the model and predicted Ta for the corresponding month from all years. Method 3 is the same as Method 2 but year (2001–2018) was used as an additional variable. Method 4 involved using all data to build the model and predicting Ta for the whole study period. This method has been widely used in previous studies (Li and Zha, 2019a; Lin et al., 2016; Lu et al., 2018; Xu et al., 2018; Yoo et al., 2018; Zhu et al., 2019). Method 5 is the same as Method 4 but year and month (1−12) were utilized as additional variables. It was worth

Table 2
Different ways to input data used in this study.

Method	Description	Variable
Method 1	Data for one month were used to build the model and to predict Ta for the corresponding month	Daytime Ts, nighttime Ts, EVI, DEM, latitude, longitude, topographic index and solar radiation
Method 2	Data for the same month from multiple years were used to build the model and to predict Ta for the corresponding month from all years	Daytime Ts, nighttime Ts, EVI, DEM, latitude, longitude, topographic index and solar radiation
Method 3	Data for the same month from multiple years were used to build the model and to predict Ta for the same month from all years; year information (2001–2018) was used as an auxiliary variable	Daytime Ts, nighttime Ts, EVI, DEM, latitude, longitude, topographic index, solar radiation and year
Method 4	All data were used to build the model and to predict Ta for the whole study period	Daytime Ts, nighttime Ts, EVI, DEM, latitude, longitude, topographic index and solar radiation
Method 5	All data were used to build the model and to predict Ta for the whole study period; year and (2001–2018) month (1–12) information were used as auxiliary variables	Daytime Ts, nighttime Ts, EVI, DEM, latitude, longitude, topographic index, solar radiation, month and year

noting that station relocation (latitude or longitude of the station changed at least one time (Liao et al., 2016)) will not affect Ta estimation. If the location of a meteorological station changes, values of auxiliary variables in new position will be extracted. The relationships between Ta and auxiliary will not be affected by station relocation, thus, the Ta estimation will not be affected.

2.2.3. Model validation

Parts of the goals of this study are to evaluate the accuracy of interannual variations and long-term trends of estimated Ta. Meteorological stations with relocation for the period of 2001–2018 should not be used to evaluate the accuracy of interannual variations and long-term trends of estimated Ta. Therefore, a modified 10-fold cross-validation method was employed in this study. There are 357 and 337 stations with and without relocation, respectively, for the period of 2001–2018 (Fig. 1). Samples taken from meteorological stations without relocation were randomly and evenly partitioned into 10 subsets. Subsequently, 9 subsets and samples taken from the meteorological stations with relocation were used to train the model, and the remaining subset was utilized for validation purposes. This step was repeated 10 times. Each subset was used once for verification. That is to say, samples taken from meteorological stations with relocation were only utilized to train the model, and only samples drawn from the meteorological stations without relocation were used to validate. Finally, R^2, RMSE and MAE were calculated to describe the accuracy of estimated monthly Ta. Interannual R^2, RMSE and MAE for the period of 2001–2018 were calculated for each station without relocation, to evaluate the accuracy of interannual variations in estimated monthly Ta. Linear regression was used to calculate the temporal trends of estimated and observed monthly Ta during 2001–2018 for each station without relocation. R^2, RMSE and MAE were then calculated to assess the accuracy of temporal trends of estimated monthly Ta.

3. Results

3.1. The accuracy of predicted monthly Ta

The overall accuracy of predicted monthly Ta is shown in Table 3. The accuracies of predicted monthly mean, maximum and minimum Ta for each month are shown in Figs. 2, S1 and S2, respectively. Five main results were found. First, there were significant differences in the accuracy of predicted monthly Ta derived from the various methods. The MAE, RMSE and R^2 values of estimated monthly mean Ta were found to range from 0.474–0.790 °C, 0.629–1.068 °C and 0.991–0.997, respectively (Table 3). Method 3 was the most accurate method. Second, Method 2 was found to be more accurate than Method 4, and Method 3 was found to be more accurate than Method 5. These results indicate that inputting data for the same month into the model can generate more accurate results than inputting all data into the model. This can be attributed to different relationships between monthly Ta and auxiliary variables across different months, but close relationships for the same month. A model based on 12 months of data may generate a bias for different months (Colombi et al., 2007; Zhang et al., 2011). Third, using temporal variables (i.e., month and year) can significantly increase the accuracy of estimated Ta, as Method 3 is more accurate than Method 2 and Method 5 is more accurate than Method 4. This is the case because the relationships between Ta and auxiliary variables differ by season and year, and using temporal variables can help distinguish between different relationships found at different times and improve accuracy levels (Colombi et al., 2007; Zhang et al., 2011). These results echo those of Zhang et al. (2011), which show that applying solar declination (a variable that is a function of the Julian day) can improve the accuracy of estimated daily Ta. However, to our knowledge, few studies have used temporal variables to predict monthly Ta. Therefore, using temporal variables to predict monthly Ta in future research is highly recommended. Fourth, the accuracy of predicted monthly minimum Ta was stronger than that of maximum Ta. This finding corroborates previous studies and can be attributed to stronger relationships between Ts and minimum Ta than between Ts and maximum Ta (Kloog et al., 2014; Lin et al., 2016; Vancutsem et al., 2010). Fifth, the MAE and RMSE of predicted Ta for warmer months are generally lower than those for colder months while most R^2 values for warmer months are lower than those for colder months (Fig. 2). This result echoes Li and Zha (2019a) and can be attributed to the larger prediction range of colder months than that of warmer months.

Table 3
The accuracy of predicted monthly air temperature (Ta).

	Method 1	Method 2	Method 3	Method 4	Method 5
Monthly mean Ta					
MAE (°C)	0.742	0.730	0.474	0.790	0.633
RMSE (°C)	1.051	0.964	0.629	1.068	0.830
R^2	0.992	0.993	0.997	0.991	0.995
Monthly maximum Ta					
MAE (°C)	0.918	0.948	0.607	0.986	0.811
RMSE (°C)	1.351	1.241	0.798	1.306	1.054
R^2	0.985	0.987	0.995	0.986	0.991
Monthly minimum Ta					
MAE (°C)	0.900	0.784	0.540	0.866	0.686
RMSE (°C)	1.224	1.029	0.719	1.169	0.902
R^2	0.990	0.993	0.997	0.991	0.995

MAE: mean absolute error. RMSE: root mean squared error. R^2: coefficient of determination.

3.2. The accuracy of interannual variations in predicted monthly Ta

The overall accuracy of interannual variations in predicted monthly Ta is shown in Table 4. The accuracy levels of interannual variations in predicted monthly mean, maximum and minimum Ta for each month are shown in Figs. 3, S3 and S4, respectively. Some of the results on the accuracy of interannual variations in predicted monthly Ta are similar to those found on the accuracy of predicted monthly Ta: significant differences in accuracy levels were found across various methods; inputting data for the same month into the model can produce more

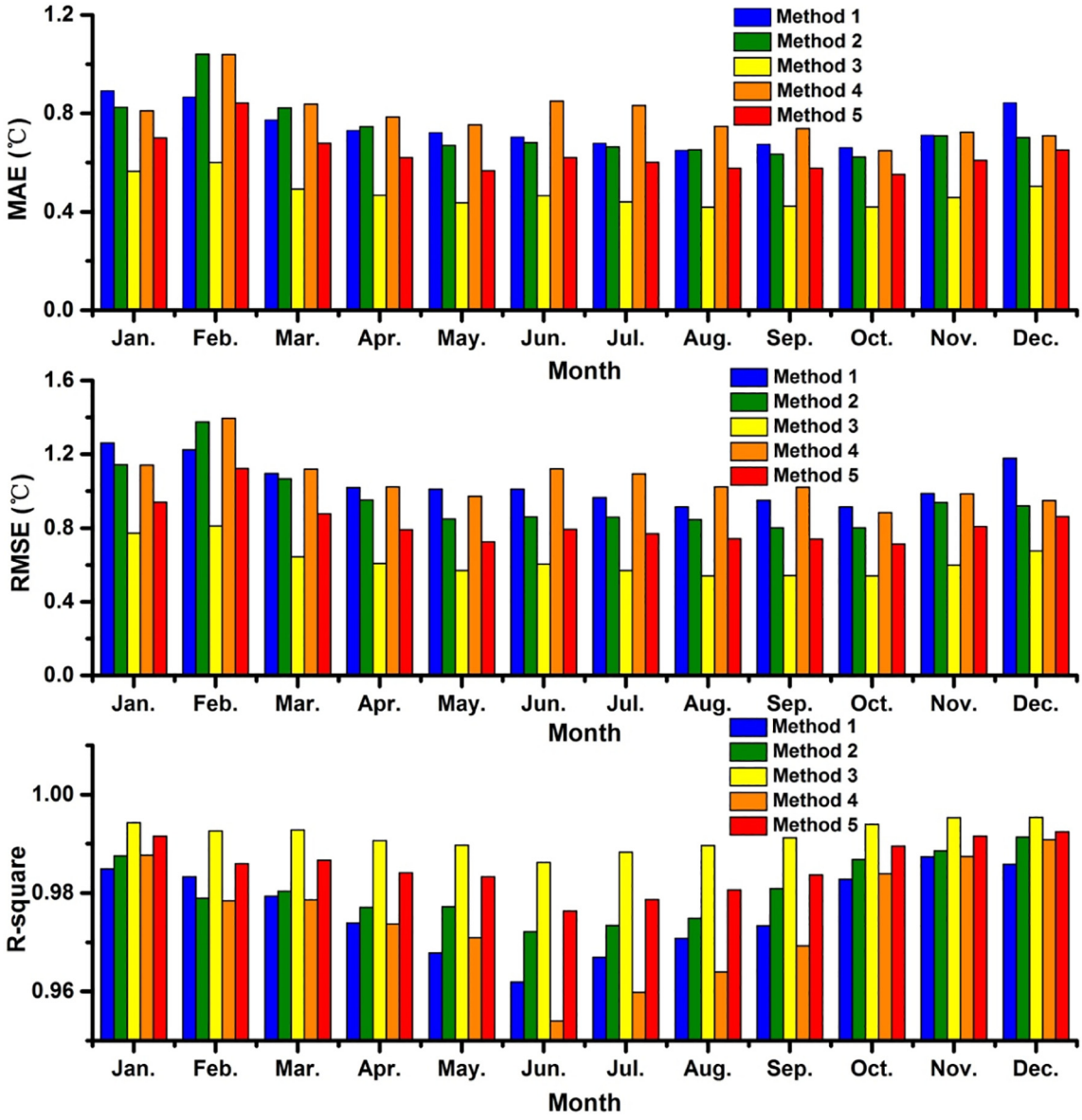


Fig. 2. The accuracy of predicted monthly mean air temperature (Ta) for each month. MAE: mean absolute error. RMSE: root mean squared error. R-square: coefficient of determination.

accurate results than inputting all data into the model; utilizing temporal variables can considerably increase accuracy levels; and the MAE and RMSE of interannual variations in estimated Ta for warmer months are generally lower than those for colder months while most R^2 values for warmer months are lower than those of colder months (Table 4 and Fig. 3).

Table 4
The accuracy of interannual variations in predicted monthly Ta averaged for 337 stations and 12 months.

	Method 1	Method 2	Method 3	Method 4	Method 5
Monthly mean Ta					
MAE (°C)	0.742	0.730	0.474	0.790	0.633
RMSE (°C)	0.866	0.916	0.593	1.004	0.792
R^2	0.774	0.431	0.731	0.404	0.544
Monthly maximum Ta					
MAE (°C)	0.918	0.948	0.607	0.986	0.811
RMSE (°C)	1.066	1.181	0.757	1.235	1.011
R^2	0.787	0.421	0.720	0.390	0.537
Monthly minimum Ta					
MAE (°C)	0.900	0.784	0.540	0.866	0.686
RMSE (°C)	1.040	0.981	0.673	1.104	0.857
R^2	0.700	0.337	0.656	0.324	0.471

However, some different results were found. First, the R^2 of interannual variations in predicted monthly maximum Ta was found to be higher than the mean and minimum Ta while the MAE and RMSE of interannual variations in predicted monthly maximum Ta were found to be higher than the mean and minimum Ta (Table 4). These results can be attributed to the fact the prediction range and standard deviation of interannual variations in monthly maximum Ta were found to be larger than those of mean and minimum Ta (not shown). Second, an interesting finding was that the R^2 of Method 1 was the highest among all methods. This suggests that a highly accurate predicted monthly Ta does not necessarily denote highly accurate interannual variations in predicted monthly Ta, as the accuracies of estimated monthly Ta were found to be similar for Methods 1, 2 and 4 (Table 3). This also suggests that Method 1 can best capture interannual variations in monthly Ta. This phenomenon was observed because the model was built for each year separately for Method 1, whereas one model was built for all years for other

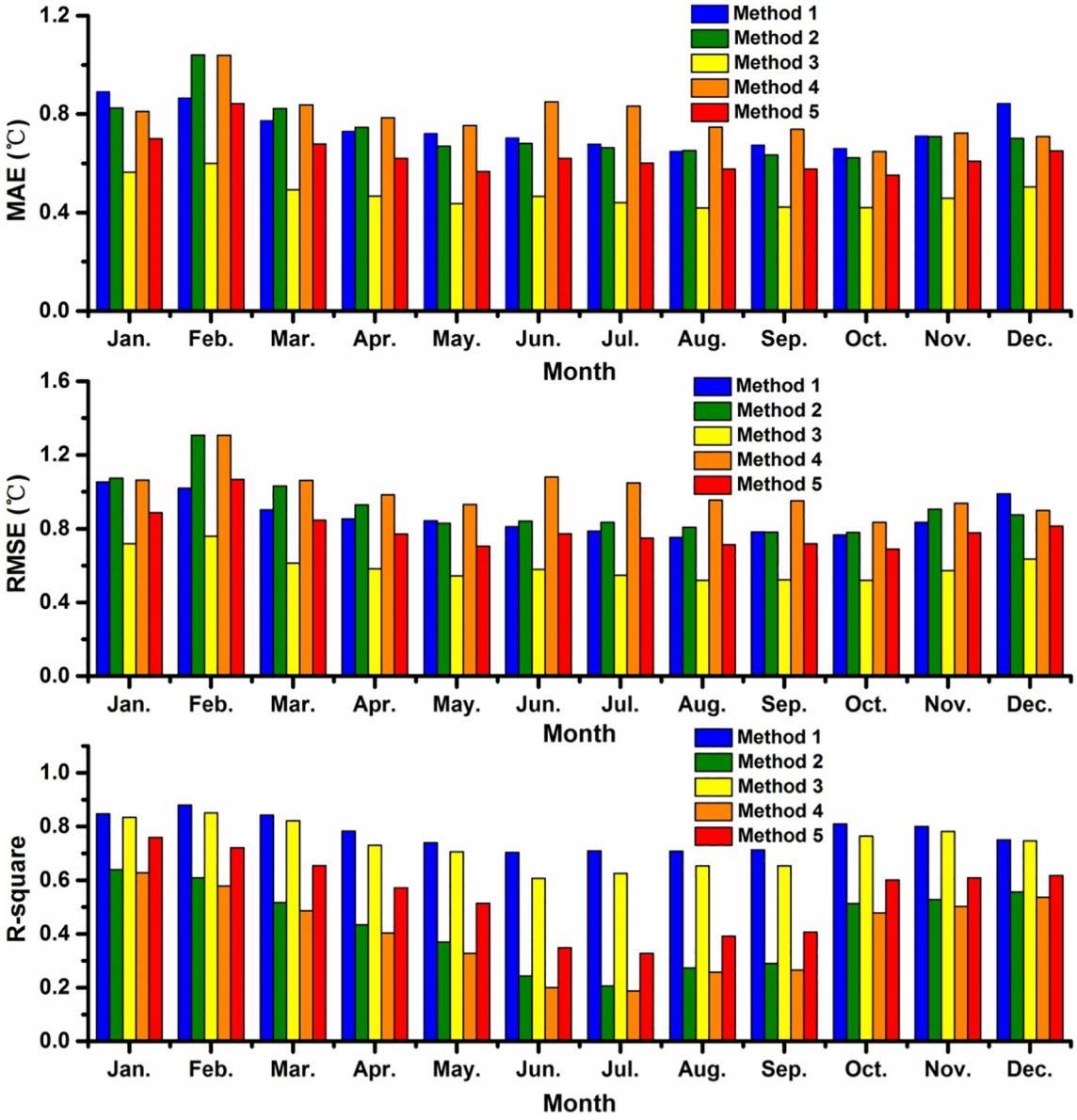


Fig. 3. The accuracy of interannual variations in predicted monthly mean Ta for each month averaged for 337 meteorological stations without relocation.

methods. Thus, differences in relationships between Ta and auxiliary variables observed at different times can be better distinguished using Method 1 than the other methods. In addition, it is difficult to predict extreme values of interannual variations in Ta using a model with data for multiple months or years, as extreme values may be regarded as outliers when data for multiple months or years are used. As indicated by Fig. 4, the interannual R^2 of Method 1 was the highest among all methods. However, Ta was always overestimated when using Method 1, as this model focuses on each separate year, resulting in overestimations for each year. Additionally, although Method 3 generated slightly lower interannual R^2 values than Method 1, it produced significantly lower MAE and RMSE values than Method 1 and did not generate obvious overestimations or underestimations. Furthermore, it should be noted that the magnitude of interannual variations in estimated Ta was found to be lower than that of observed Ta for Methods 2–5 (Fig. 4). The extreme values of interannual variations in observed monthly Ta were clearly more obvious than those of estimated Ta for Methods 2–5, as extreme values may be regarded as outliers when data for multiple years are input into the model.

3.3. The accuracy of temporal trends of predicted monthly Ta

The overall accuracy of temporal trends of predicted monthly Ta is shown in Table 5. The accuracy levels of temporal trends of predicted monthly mean, maximum and minimum Ta for each month are shown in Figs. 5, S5 and S6, respectively. Some of the results regarding the accuracy of temporal trends of predicted monthly Ta echo the results on the accuracy of predicted monthly Ta: significant differences in accuracy levels were observed among the various methods, Method 3 was the most accurate; inputting data for the same month into the model can generate more accurate results than inputting all data into the model; and using temporal variables can significantly increase the accuracy of results (Table 5 and Fig. 5). However, the accuracy of temporal trends of estimated monthly Ta varied slightly by season due to insignificant differences in prediction intervals among seasons (not shown). In addition, the estimated monthly minimum Ta was found to be more accurate than maximum Ta (Table 3). However, temporal trends of the estimated monthly maximum Ta were found to be more accurate than those of minimum Ta (Table 5). The cause of this strange phenomenon is not clear and should be investigated further in future

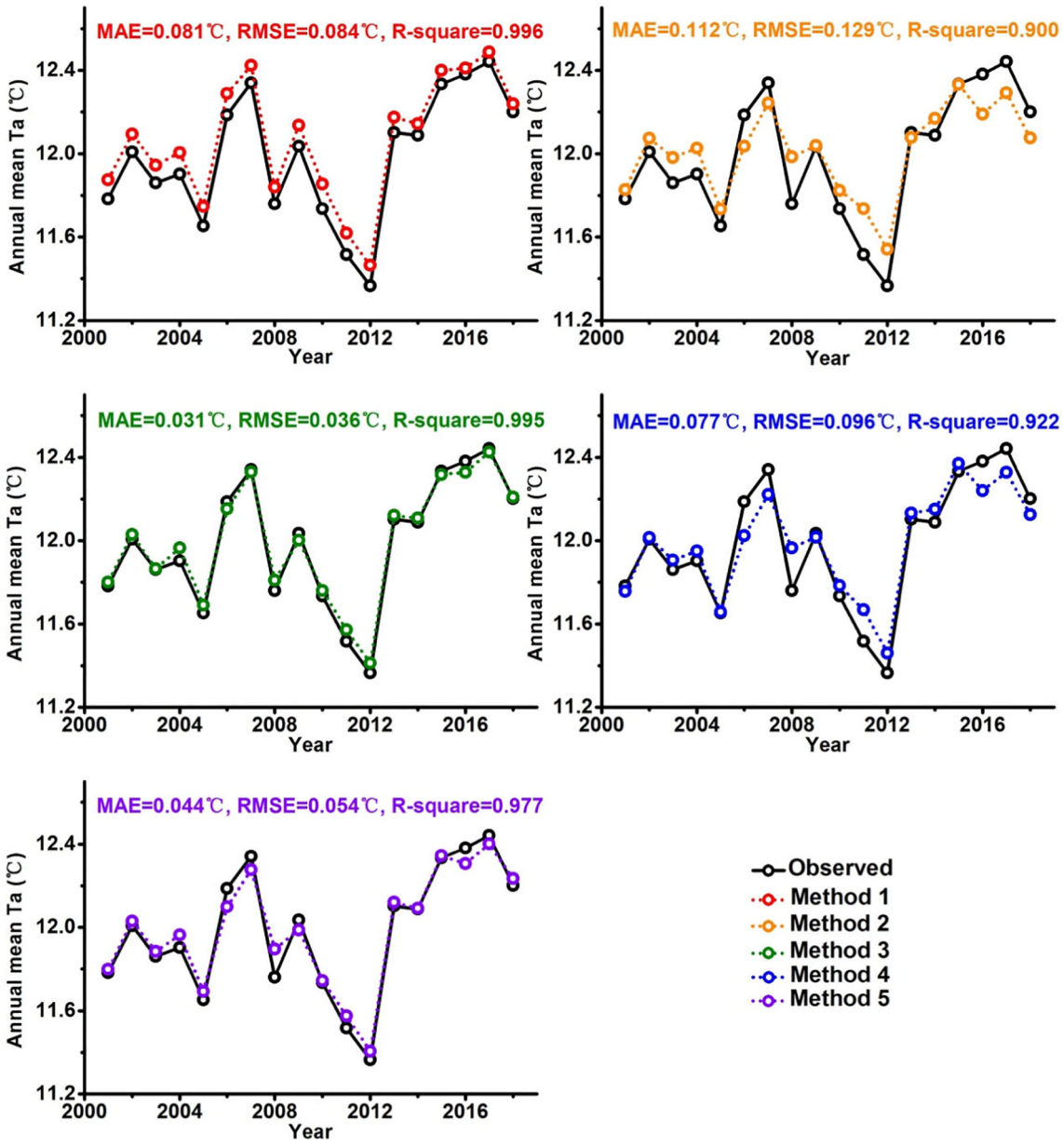


Fig. 4. Interannual variations in observed and estimated mean Ta averaged for 337 stations and 12 months.

studies. Finally, the R^2 values of interannual variations and temporal trends of predicted monthly Ta (always lower than 0.86, Tables 4 and 5) were found to be considerably lower than the R^2 of predicted monthly Ta (always higher than 0.98, Table 3). This was observed because the ranges of interannual variations and temporal trends of Ta were found to be much less pronounced than spatiotemporal variations in Ta (Li and Zha, 2019a). These results show that the temporal accuracy of estimated Ta has considerable room for further improvement.

Table 5
The overall accuracy of temporal trends in predicted monthly Ta.

	Method 1	Method 2	Method 3	Method 4	Method 5
Monthly mean Ta					
MAE (°C/decade)	0.249	0.351	0.155	0.394	0.189
RMSE (°C/decade)	0.322	0.443	0.201	0.509	0.245
R^2	0.649	0.327	0.848	0.258	0.776
Monthly maximum Ta					
MAE (°C/decade)	0.290	0.420	0.182	0.460	0.229
RMSE (°C/decade)	0.385	0.531	0.239	0.586	0.299
R^2	0.671	0.318	0.858	0.270	0.777
Monthly minimum Ta					
MAE (°C/decade)	0.314	0.460	0.185	0.488	0.216
RMSE (°C/decade)	0.413	0.571	0.245	0.625	0.282
R^2	0.521	0.195	0.817	0.141	0.758

3.4. The accuracy of predicted annual Ta

Although many previous works have generated monthly Ta using satellite data (Hooker et al., 2018; Li and Zha, 2019a; Lu et al., 2018; Zhu et al., 2019), seasonal and annual Ta have been more widely used in climate research (Ge et al., 2013; Sun et al., 2016; Wang et al., 2015; Zhang et al., 2017). Thus, changes in accuracy levels when estimated monthly Ta is averaged to coarser temporal resolution Ta were analyzed. Method 3 was employed, as it is the most accurate according to the above results.

The errors of estimated annual Ta and interannual variations and temporal trends in annual Ta were found to be considerably lower than those of monthly Ta (Table 6). The MAEs and RMSEs of estimated

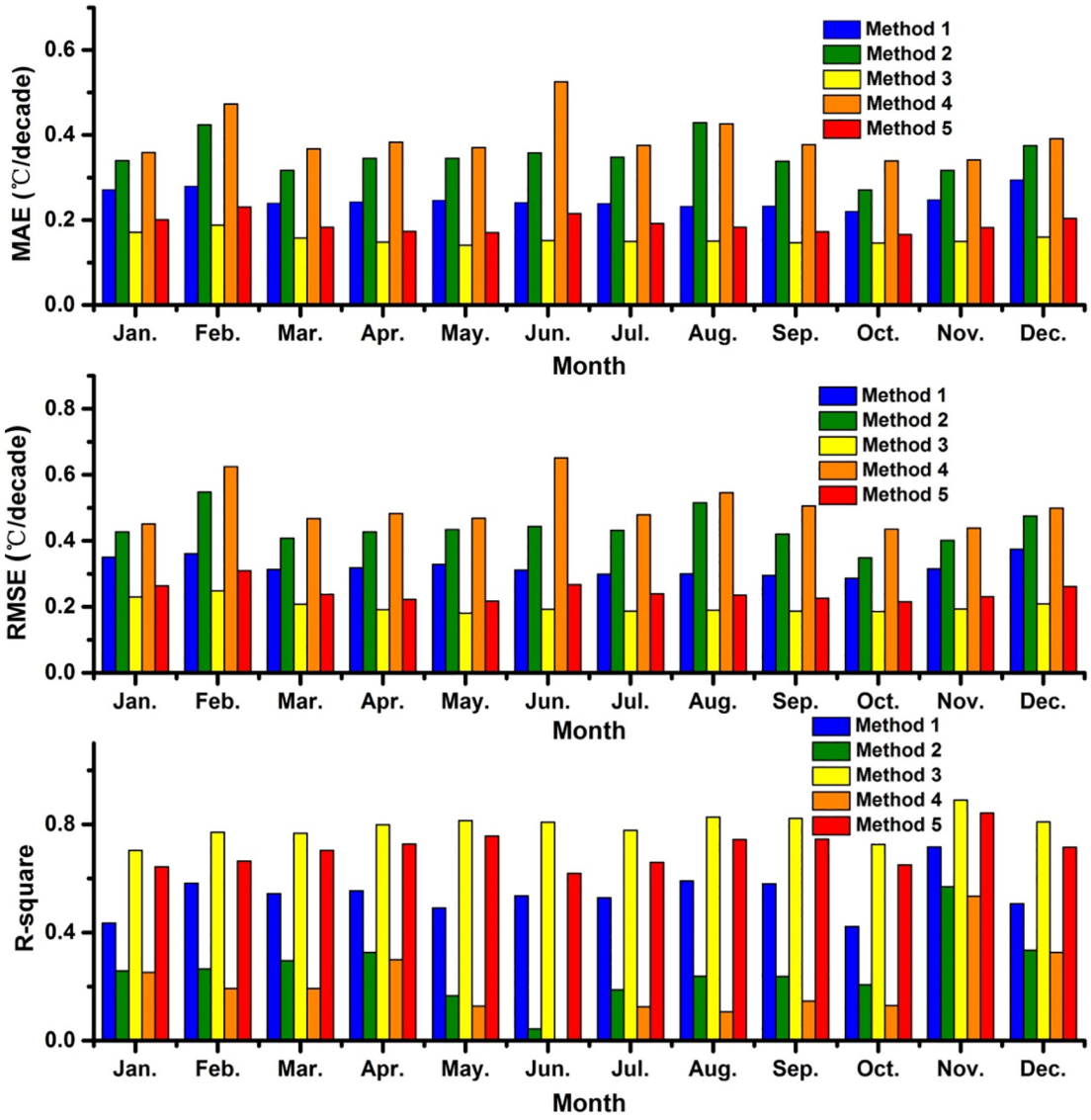


Fig. 5. The accuracy of temporal trends in predicted monthly mean Ta for each month averaged for 337 meteorological stations without relocation.

annual Ta and interannual variations of annual Ta were found to be less than half of monthly Ta (Table 3). There are some outliers in the auxiliary data (e.g., Ts and the EVI) that may affect the accuracy of estimated Ta. The effects of outliers are reduced when monthly Ta is averaged to annual Ta. Thus, the errors of estimated annual Ta were found to be considerably lower than those of monthly Ta. However, the R^2 did not increase significantly and primarily because the prediction interval decreased when estimated monthly Ta was averaged to annual Ta. Interestingly, the observed decrease in error values of temporal trends of estimated annual Ta was found to be less significant than that of annual Ta and interannual variations in annual Ta, as the MAEs and RMSEs of temporal trends of estimated annual Ta were found to be approximately half those of monthly Ta. In addition, the R^2 of predicted annual Ta and interannual variations in annual Ta were found to be slightly higher than those of monthly Ta, while the R^2 of temporal trends of predicted annual Ta was found to be slightly lower than that of monthly Ta (Tables 3 and 6). This is likely the case because temporal trends of Ta were calculated using 18 years of data, while Ta and interannual variations in Ta were not. The effects of outliers may be more effective at predicting monthly Ta and interannual variations in monthly Ta than temporal trends of monthly Ta. Thus, averaging monthly Ta into annual Ta can increase the accuracy of annual Ta and of interannual variations in annual Ta more than temporal trends of annual Ta.

Table 6
The accuracy of predicted annual Ta using Method 3.

	Annual mean Ta	Annual maximum Ta	Annual minimum Ta
The accuracy of predicted annual Ta			
MAE (°C)	0.207	0.263	0.246
RMSE (°C)	0.264	0.334	0.316
R^2	0.998	0.997	0.998
The accuracy of interannual variations in predicted annual Ta			
MAE (°C)	0.207	0.263	0.246
RMSE (°C)	0.251	0.318	0.297
R^2	0.771	0.773	0.722
The accuracy of temporal trends in predicted annual Ta			
MAE (°C/decade)	0.086	0.092	0.122
RMSE (°C/decade)	0.111	0.119	0.162
R^2	0.719	0.744	0.699

3.5. Spatio-temporal variations of the estimated Ta

From the above results it was found that Method 3 was the most accurate with the exception of its R^2 value in interannual variations in

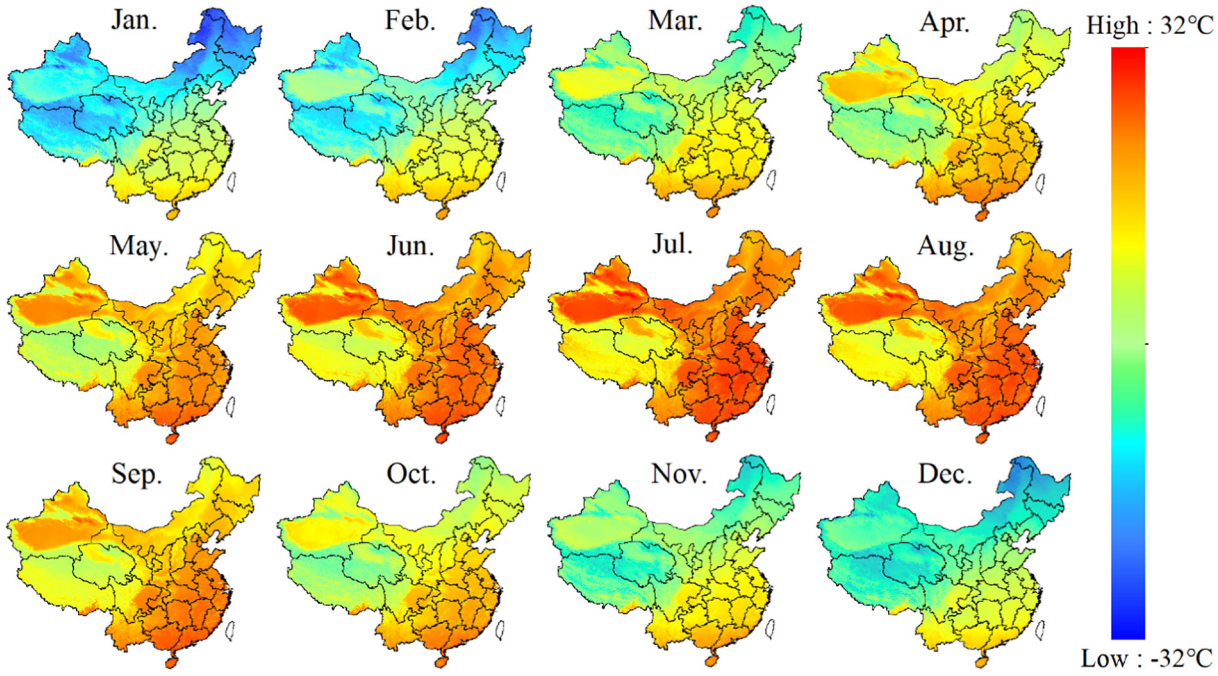


Fig. 6. Spatial and seasonal variations in monthly mean Ta averaged from 2001 to 2018 for Mainland China.

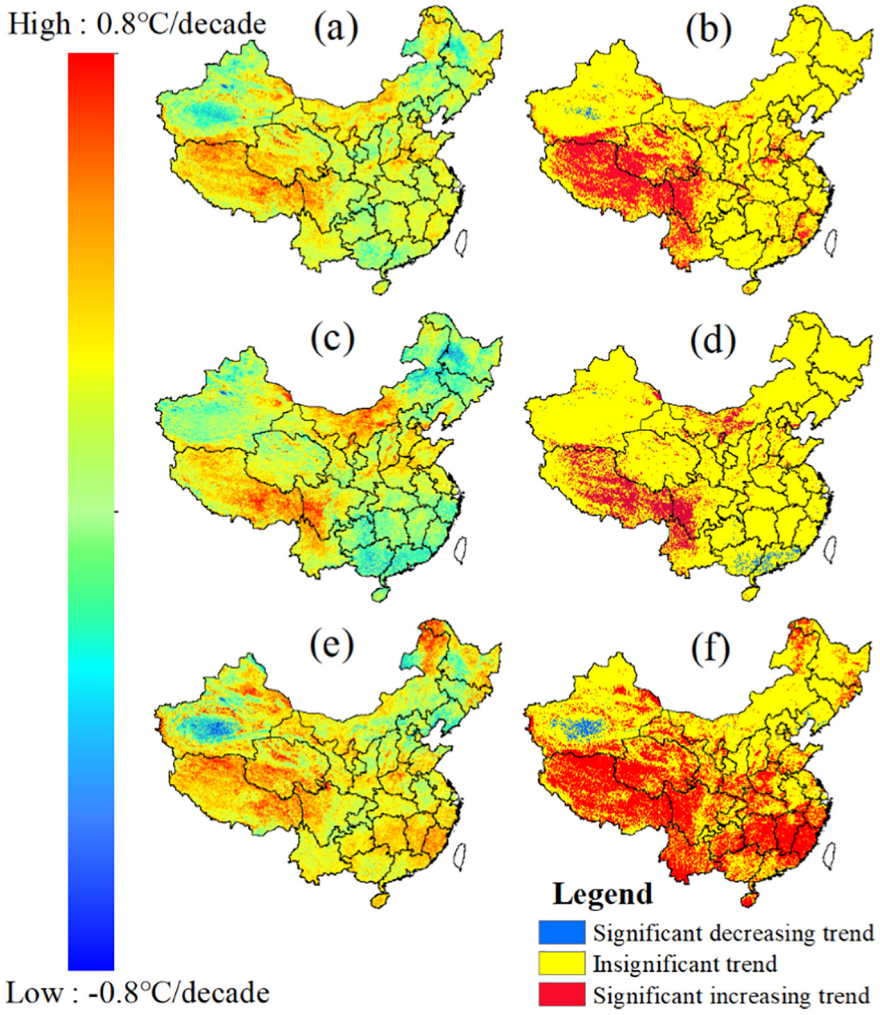


Fig. 7. Long-term trends of Ta for the period of 2001–2018 for Mainland China: (a) slope of annual mean Ta; (b) significance of trend of annual mean Ta; (c) slope of annual maximum Ta; (d) significance of trend of annual maximum Ta; (e) slope of annual minimum Ta; (f) significance of trend of annual minimum Ta.

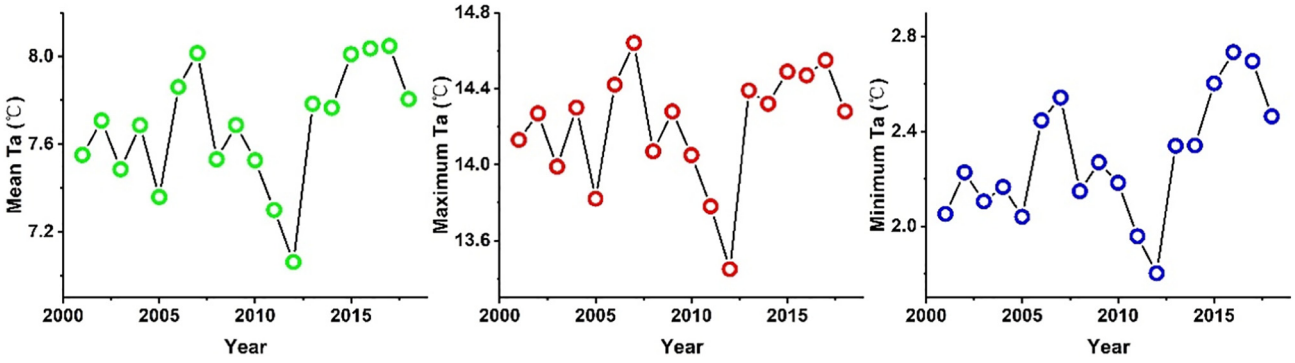


Fig. 8. Interannual variations in annual mean, maximum and minimum Ta for Mainland China during 2001–2018.

estimated Ta. Thus, monthly Ta for 2001–2018 for Mainland China was mapped using Method 3. Spatial and seasonal variations in monthly mean Ta are shown in Fig. 6. Seasonally, summer months (e.g., June to August) generally present higher mean Ta than winter months (e.g., December to February). Spatially, northeastern China and the Qinghai-Tibet Plateau present low mean Ta due to their high latitude and elevation, respectively. The Tarim Basin shows higher mean Ta relative to its surrounding areas, as the area is covered with desert land and is lower in elevation than surrounding areas. The observed spatial and seasonal variations in monthly maximum and minimum Ta echo those of monthly mean Ta (Figs. S7 and S8).

Long-term Ta trends for 2001–2018 for Mainland China are shown in Fig. 7. Annual mean Ta increased at a rate of 0.182 °C/decade ($p = 0.149$). The increasing rate of annual minimum Ta (0.254 °C/decade, $p = 0.025$) was found to be higher than that of maximum Ta (0.127 °C/decade, $p = 0.373$). This is the case because minimum Ta is primarily affected by the greenhouse effect, whereas maximum Ta is also affected by solar radiation and atmospheric moisture (Ren et al., 2017; Wang et al., 2018). Spatially, higher warming rates were found in the Qinghai-Tibet Plateau, corroborating previous studies (Li and Zha, 2019b; Xu et al., 2018). This can primarily be attributed to the increase in heat storage at high altitudes due to changes in snow depth and cloud cover (Yan et al., 2016).

Interannual variations in annual mean Ta for Mainland China during 2001–2018 were similar to annual maximum and minimum Ta (Fig. 8). Annual mean Ta was the lowest in 2012 (7.06 °C) and highest in 2017 (8.05 °C). There were downward trends of Ta for the period of 2001–2012. This is similar to previous studies and can be attributed to the warming hiatus in this period (Du et al., 2019; Li et al., 2015). In addition, there were upward trends of Ta after 2012, indicating the warming hiatus stopped.

The spatially continuous Ta maps developed in this study effectively capture the UHI effect. Therefore, it can be used to investigate UHI effect in future studies. The spatial distributions of Ta for Beijing and its surrounding area averaged for 2001–2018 are shown in Fig. 9. High Ta areas are mainly distributed across the city center of Beijing while low Ta areas are generally distributed across the mountains of northwestern Beijing. In addition, the UHI effect of minimum Ta was found to be stronger than that of maximum Ta. These diurnal variations echo those of previous works, which show that nighttime UHI intensity levels are stronger than daytime levels (Anniballe et al., 2014; Roth et al., 1989; Zhou et al., 2019). Overall, the accurate, national coverage, 1 km spatial resolution and long-term series monthly mean, maximum and minimum Ta dataset developed in the present study can be used as a key variable in many fields.

4. Discussion

The accuracies of the estimated Ta in the present study were generally higher than those of previous studies. First, the MAE and RMSE of monthly mean Ta estimated using Method 3 were recorded as 0.474 and 0.629 °C, respectively. These error values are lower than those recorded in most previous works with RMSE values generally ranging from 1 to 2 °C (Cristóbal et al., 2008; Hooker et al., 2018; Li and Zha, 2019a; Li and Zha, 2019b; Lu et al., 2018; Xu et al., 2018). Second, temporal trends of predicted monthly Ta derived from Method 3 (R^2: 0.697, MAE: 0.00073 °C/month, RMSE: 0.00093 °C/month; note that these data were calculated using different methods from those in Table 5) were found to be significantly more accurate than those generated by a previous study (R^2: 0.630, MAE: 0.00132 °C/month, RMSE: 0.00178 °C/month) (Li and Zha, 2019b). These primarily occurred because temporal information as well as a different method to input the sample data were used. For example, Method 4 was widely used by previous studies. The RMSE value of the estimated Ta of Method 4 was 1.068 °C, which was relatively close to previous studies (Cristóbal et al., 2008; Hooker et al., 2018; Li and Zha, 2019a; Li and Zha, 2019b; Lu et al., 2018; Xu

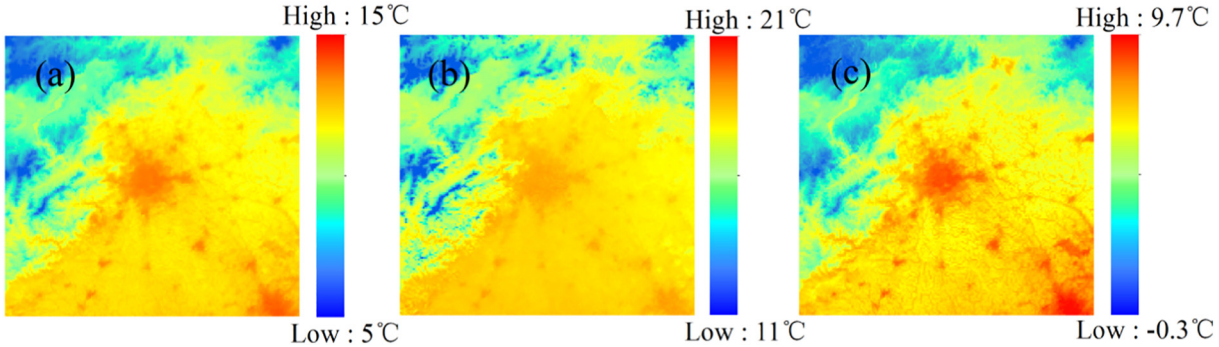


Fig. 9. Ta averaged for the period of 2001–2018 for Beijing: (a) mean Ta; (b) maximum Ta; (c) minimum Ta.

et al., 2018). It is worth noting that Method 4 was found to be the least accurate in measuring interannual variations and temporal trends in predicted monthly Ta (Tables 4 and 5). Therefore, monthly Ta datasets generated in previous studies are not suited to studies on interannual variations and temporal trends in Ta and on their relationships to other variables (e.g., vegetation dynamics, animal growth and other climate variables).

The main contributions of this manuscript include two aspects. First, this study systematically analyzed the temporal accuracy of estimated monthly Ta, which is often ignored in previous studies. Second, this study investigated the differences in the accuracy levels of five ways to input data into the model. The results showed that the method widely used by previous studies (Method 4) was found to be the least accurate. Method 3 was recommended to estimate monthly Ta in future studies.

5. Conclusions

In this study, the temporal accuracy of estimated monthly Ta was comprehensively evaluated. The differences in the accuracy levels of five ways to input data into the model were investigated. Our results are summarized as follows. (1) Inputting data for the same month into the model can generate more accurate results than inputting all data into the model. This can be attributed to different relationships between Ta and auxiliary variables in different months. (2) Using temporal variables (i.e., month and year) can significantly improve the accuracy of estimated Ta. This is because the temporal variables can help distinguish between different relationships between Ta and auxiliary variables and then improve the accuracy of estimated Ta. (3) For the best method (Method 3), the R^2 of the estimated monthly mean Ta was recorded as 0.997, the interannual R^2 of the estimated monthly mean Ta was recorded as 0.731, and the R^2 of temporal trends of estimated monthly mean Ta was recorded as 0.848. (4) An accurate, national coverage, 1 km spatial resolution and long-term series monthly mean, maximum and minimum Ta dataset was ultimately developed and can be used in many fields such as climatology, hydrology and ecology. It was found that annual mean Ta increased at a rate of 0.182 °C/decade ($p = 0.149$) from 2001 to 2018 in Mainland China.

Overall, the present study comprehensively evaluated the temporal accuracy of estimated monthly Ta and developed an accurate monthly Ta dataset for Mainland China. In addition, caution should be exercised when determining how data are to be inputted into the model. Furthermore, the R^2 of interannual variations and temporal trends in predicted monthly Ta were found to be significantly lower than those of predicted monthly Ta, suggesting that the temporal accuracy of estimated Ta has considerable room for further improvement.

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.

Acknowledgements

The daily mean, maximum and minimum air temperature at 699 stations during 2001–2018 were obtained from National Meteorological Information Center of China http://data.cma.cn; MODIS land surface temperature (Ts) and enhanced vegetation index (EVI) products were available from Goddard Space Flight Center Level 2 and Atmosphere Archive Distribution System http://ladsweb.nascom.nasa.gov; Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Digital Elevation Model (DEM) data were derived from Geospatial Data Cloud http://www.gscloud.cn/; Topographic index data were downloaded from the Environmental Information Data Centre http://eidc.ceh.ac.uk/. This work was financially supported by the National Natural Science Foundation of China (No. 41975044, 41771360, 41601044 and 41801021), the Special Fund for Basic Scientific Research of Central Colleges, China University of Geosciences, Wuhan (CUGL170401 and CUGCJ1704).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2019.136037.

References

Alkama, R., Cescatti, A., 2016. Biophysical climate impacts of recent changes in global forest cover. Science 351, 600–604.
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.
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.
Colombi, A., Michele, C.D., Pepe, M., Rampini, A., 2007. Estimation of daily mean air temperature from MODIS LST in alpine areas. EARSeL eProceedings 6, 38–46.
Cristóbal, J., Ninyerola, M., Pons, X., 2008. Modeling air temperature through a combination of remote sensing and GIS data. J. Geophys. Res. 113, D13106.
Du, Q., Zhang, M., Wang, S., Che, C., Ma, R., Ma, Z., 2019. Changes in air temperature over China in response to the recent global warming hiatus. J. Geogr. Sci. 29, 496–516.
Ge, Q., Wang, F., Luterbacher, J., 2013. Improved estimation of average warming trend of China from 1951-2010 based on satellite observed land-use data. Clim. Chang. 121, 365–379.
He, C., Gao, B., Huang, Q., Ma, Q., Dou, Y., 2017. Environmental degradation in the urban areas of China: evidence from multi-source remote sensing data. Remote Sens. Environ. 193, 65–75.
Honsey, A.E., Venturelli, P.A., Lester, N.P., 2019. Bioenergetic and limnological foundations for using degree-days derived from air temperatures to describe fish growth. Can. J. Fish. Aquat. Sci. 76, 657–669.
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.
Huang, X., Wang, Y., 2019. Investigating the effects of 3D urban morphology on the surface urban heat island effect in urban functional zones by using high-resolution remote sensing data: a case study of Wuhan, Central China. ISPRS J. Photogramm. 152, 119–131.
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., X, G., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.
Kloog, I., Chudnovsky, A., Koutrakis, P., Schwartz, J., 2012. Temporal and spatial assessments of minimum air temperature using satellite surface temperature measurements in Massachusetts, USA. Sci. Total Environ. 432, 85–92.
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.
Kuhn, M., Weston, S., Keefer, C., Coulter, N., 2018. Cubist: Rule- and Instance-Based Regression Modeling (R Package Version 0.2.2).
Lai, J., Zhan, W., Huang, F., Voogt, J., Bechtel, B., Allen, M., Peng, S., Hong, F., Liu, Y., Du, P., 2018. Identification of typical diurnal patterns for clear-sky climatology of surface urban heat islands. Remote Sens. Environ. 217, 203–220.
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, L., Zha, Y., 2019c. Satellite-based spatiotemporal trends of canopy urban heat islands and associated drivers in China's 32 major cities. Remote Sens. 11, 102.
Li, Q., Yang, S., Xu, W., Wang, X., Jones, P., Parker, D., Zhou, L., Feng, Y., Gao, Y., 2015. China experiencing the recent warming hiatus. Geophys. Res. Lett. 42, 889–898.
Li, X., Zhou, Y., Asrar, G.R., Imhoff, M., Li, X., 2017. The surface urban heat island response to urban expansion: a panel analysis for the conterminous United States. Sci. Total Environ. 605–606, 426–435.
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.
Liao, W., Wang, D., Liu, X., Wang, G., Zhang, J., 2016. Estimated influence of urbanization on surface warming in Eastern China using time-varying land use data. Int. J. Climatol. 37, 3197–3208.
Lin, S., Moore, N.J., Messina, J.P., DeVisser, M.H., Wu, J., 2012. Evaluation of estimating daily maximum and minimum air temperature with MODIS data in east Africa. Int. J. Appl. Earth Obs. 18, 128–140.
Lin, X., Zhang, W., Huang, Y., Sun, W., Han, P., Yu, L., Sun, F., 2016. Empirical estimation of near-surface air temperature in China from MODIS LST data by considering physiographic features. Remote Sens. 8, 629.
Liu, Z., Wu, C., Liu, Y., Wang, X., Fang, B., Yuan, W., Ge, Q., 2017. Spring green-up date derived from GIMMS3g and SPOT-VGT NDVI of winter wheat cropland in the North China Plain. ISPRS J. Photogramm. 130, 81–91.
Liu, Z., Liu, Y., Li, Y., 2018a. Extended warm temperate zone and opportunities for cropping system change in the Loess Plateau of China. Int. J. Climatol. 39, 658–669.

Liu, Z., Liu, Y., Li, Y., 2018b. Anthropogenic contributions dominate trends of vegetation cover change over the farming-pastoral ecotone of northern China. Ecol. Indic. 95, 370–378.
Lu, N., Liang, S., Huang, G., Qin, J., Yao, L., Wang, D., Yang, K., 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.
Medina-Ramon, M., Zanobetti, A., Cavanagh, D.P., Schwartz, J., 2006. Extreme temperatures and mortality: assessing effect modification by personal characteristics and specific cause of death in a multi-city case-only analysis. Environ. Health Perspect. 114, 1331–1336.
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. https://doi.org/10.1002/joc.6188.
Noi, P., Kappas, M., Degener, J., 2016. Estimating daily maximum and minimum land air surface temperature using MODIS land surface temperature data and ground truth data in northern Vietnam. Remote Sens. 8, 1002.
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.
Piao, S., Nan, H., Huntingford, C., Ciais, P., Friedlingstein, P., Sitch, S., Peng, S., Ahlstrom, A., Canadell, J.G., Cong, N., Levis, S., Levy, P.E., Liu, L., Lomas, M.R., Mao, J., Myneni, R.B., Peylin, P., Poulter, B., Shi, X., Yin, G., Viovy, N., Wang, T., Wang, X., Zaehle, S., Zeng, N., Zeng, Z., Chen, A., 2014. Evidence for a weakening relationship between interannual temperature variability and northern vegetation activity. Nat. Commun. 5, 5018.
Quinlan, R., 1992. Learning with continuous classes. Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart, Tasmania, pp. 343–348.
Ren, G., Ding, Y., Tang, G., 2017. An overview of mainland China temperature change research. J. Meteorol. Res. 31, 3–16.
Roth, M., Oke, T.R., Emery, W.J., 1989. Satellite-derived urban heat islands from three coastal cities and the utilization of such data in urban climatology. Int. J. Remote Sens. 10, 1699–1720.
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.
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., Liu, B., Henderson, M., Shen, X., Yang, L., Lian, L., 2018. Warming across decades and deciles: minimum and maximum daily temperatures in China, 1955-2014. Int. J. Climatol. 38, 2325–2332.
Wang, M., Jia, X.J., Ge, J.W., Qian, Q.F., 2019. Changes in the relationship between the interannual variation of Eurasian snow cover and spring SAT over eastern Eurasia. J. Geophys. Res.-Atmos. 124, 468–487.
Warren, R., Price, J., Graham, E., Forstenhaeusler, N., VanDerWal, J., 2018. The projected effect on insects, vertebrates, and plants of limiting global warming to 1.5 °C rather than 2 °C. Science 360, 791–795.
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.
Yan, L., Liu, Z., Chen, G., Kutzbach, J.E., Liu, X., 2016. Mechanisms of elevation dependent warming over the Tibetan plateau in quadrupled CO2 experiments. Clim. Chang. 135, 509–519.
Yang, Q., Huang, X., Tang, Q., 2019. The footprint of urban heat island effect in 302 Chinese cities: temporal trends and associated factors. Sci. Total Environ. 655, 652–662.
Yao, R., Wang, L., Huang, X., Niu, Z., Liu, F., Wang, Q., 2017. Temporal trends of surface urban heat islands and associated determinants in major Chinese cities. Sci. Total Environ. 609, 742–754.
Yao, R., Wang, L., Huang, X., Niu, Y., Chen, Y., Niu, Z., 2018. The influence of different data and method on estimating the surface urban heat island intensity. Ecol. Indic. 89, 45–55.
Yao, R., Wang, L., Huang, X., Gong, W., Xia, X., 2019. Greening in rural areas increases the surface urban heat island intensity. Geophys. Res. Lett. 46, 2204–2212.
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.
Zhang, W., Huang, Y., Yu, Y., Sun, W., 2011. Empirical models for estimating daily maximum, minimum and mean air temperatures with MODIS land surface temperatures. Int. J. Remote Sens. 32, 9415–9440.
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.
Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., Frolking, S., Yao, R., Qiao, Z., Sobrino, J., 2019. Satellite remote sensing of surface urban heat islands: progress, challenges, and perspectives. Remote Sens. 11, 48.
Zhu, W., Lű, A., Jia, S., 2013. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sens. Environ. 130, 62–73.
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.