Reconstruction of high spatial resolution surface air temperature data across China: A new geo-intelligent multisource data-based machine learning technique

Xiudi Zhu, Qiang Zhang, Chong-Yu Xu, Peng Sun, Pan Hu

Abstract

Good knowledge of the surface air temperature (SAT) is critical for scientific understanding of ecological environment changes and land-atmosphere thermodynamic interactions. However, sparse and uneven spatial distribution of the temperature gauging stations introduces remarkable uncertainties into analysis of the SAT pattern. From a geo-intelligent perspective, here we proposed a new SAT reconstruction method based on the multisource data and machine learning technique which was developed by considering autocorrelation of the in situ observed SAT in both space and time, or simply STAML, i.e. Geoi-SVM (Geo-Intelligent Support Vector Machine), Geoi-BPNN (Geo-Intelligent Back Propagation Neural Network) and Geoi-RF (Geo-Intelligent Random Forest). The multisource data used in this study include the in situ observed SAT and multisource remotely sensed data such as MODIS land surface temperature, NDVI (Normalized Difference Vegetation Index) data. Intermodel comparisons amidst reconstructed SAT data were done to evaluate reconstructing performance of abovementioned models. Besides, the SAT reconstructed by CART (Classification and Regression Tree) was also included to evaluate the reconstructing performance of the models considered in this study when compared to SAT data by CART algorithm. We found that the estimation error of the reconstructed SAT by the STAML is smaller...
1. Introduction

Surface air temperature (SAT) is a critical meteorological variable describing climate changes and monitoring ecological environment (Chow et al., 1988; Zhang et al., 2015), and is also the key meteorological factor influencing biosphere processes (Prince and Goward, 1995; IPCC, 2013; Schuur et al., 2015; Schleussner et al., 2016; Hughes et al., 2017; Nikulin et al., 2018; Warren et al., 2018), modulating the land-atmosphere exchange of energy and water vapor (Alkama and Cescatti, 2016) and affecting the meteorological and hydrological process (Meehl and Tebaldi, 2004; Zhang et al., 2013; Donnelly et al., 2017; Karmalkar and Bradley, 2017; Nikulin et al., 2018; Warren et al., 2018). In this sense, availability of a high-quality SAT dataset is therefore critical for thorough understanding of spatiotemporal patterns of SAT at regional and even global scale. While, the spatial resolution of the standing SAT datasets cannot satisfy SAT study at regional scale and particularly for those requiring SAT dataset with a finer spatial resolution. This point constitutes the research motivation of this current study.

Detailed and thorough investigation of regional climate changes based on meteorological observations and assimilated meteorological data is critical for human understanding of climate changes at global scale (Zhang et al., 2011). SAT is one of the most critical meteorological data describing global temperature changes and also evaluating climate changes, and nowadays grid air temperature data have been widely used in studies of hydrometeorological processes and also model parameterization (Wang and Zeng, 2013; Zhu et al., 2013; Ge et al., 2014). However, the spatial resolution of the available temperature data products is coarse and hence limits the quality of the study and introduces much uncertainty into findings and conclusions (McCarthy et al., 2010; Oleson, 2012). Particularly, regional climate evaluations require higher spatial resolution of the temperature data and the spatial resolution of the available temperature dataset is not satisfactory (Shi et al., 2015; Wouters et al., 2017). Therefore, reconstruction of temperature data with higher spatial resolution is scientifically, theoretically and practically paramount (José et al., 2016).

In general, SAT data of high spatial resolution were done based on spatial statistical interpolation and remotely sensed temperature data (Nalder and Wein, 1998; Kurtzman and Kadmon, 1999; Shen et al., 2001; Benali et al., 2012; Williamson et al., 2014; Vogt et al., 1997; Gallo et al., 2011; Shen and Leptoukh, 2011; Zhu et al., 2013; Chen et al., 2015; Xu and Liu, 2015). Spatial statistical interpolation is to produce the grid temperature data based on in situ temperature observations (Nalder and Wein, 1998; Kurtzman and Kadmon, 1999; Shen et al., 2001). This method is simple and is easy to use. Therefore, this method was widely used in spatial pattern of temperature changes (e.g. Hofstra et al., 2008; Kilibarda et al., 2014; Stahl et al., 2006). However, the quality of the spatially interpolated temperature data based on the spatial statistical interpolation technique heavily depends on spatial distribution of observational stations and the selection of the spatial statistical interpolation method (Willmott and Robeson, 1995). Therefore, uncertainty and lower accuracy of the spatially interpolated temperature data can be expected. Besides, in the real world, the observational stations are usually sparsely distributed, and the observed temperature only reflects temperature changes in the regions in the vicinity of the observatory stations (Schatz and Kucharik, 2015; Menne et al., 2012).

In addition, the spatial statistical interpolation method does not include impacts of topographical features, underlying surface properties and distance to the oceans as well on spatial pattern of temperature changes. Therefore, estimation of SAT and maximum and minimum SAT in particular is not satisfactory in accuracy and uncertainty (Willmott and Robeson, 1995; Yang et al., 2004; Li and Zha, 2018).

Reconstruction of the SAT based on remotely sensed data mainly via three ways: (1) univariate and bivariate regression method (Benali et al., 2012; Williamson et al., 2014; Vogt et al., 1997; Gallo et al., 2011; Shen and Leptoukh, 2011; Zhu et al., 2013; Chen et al., 2015; Xu and Liu, 2015). Estimation of the SAT is based on univariate and bivariate regressive relations between the SAT at the observatory stations and relevant variables (Basist et al., 1998); (2) complex non-linear models (Jang et al., 2004; Jing et al., 2013; Ho et al., 2016; Li et al., 2018). These methods such as machine learning algorithm were usually used to estimate SAT with multiple satellite remotely sensed data; (3) temperature vegetation index (TVX) method. This method is a kind of semi-empirical technique for spatial SAT interpolation based on the statistical relations between the vegetation index and the SAT (Czajkowski et al., 2000; Prihodko and Goward, 1997; Nieto et al., 2011); and (4) thermodynamic balance method. The thermodynamic balance method aims to reconstruct the SAT by coupled relations between the observed SAT and other environmental variables/parameters based on the energy balance equation (Meteotest, 2010; Sun et al., 2005). These aforementioned methods have their own strengths and limitations. However, estimation accuracy of the SAT over a large spatial extent cannot be well guaranteed and there the reconstructed SAT based on abovementioned methods may potentially produce misleading scientific viewpoints and/or findings (Prince and Goward, 1995; Sandholt et al., 2002; Stisen et al., 2007; Vancutsem et al., 2010).

Therefore, the standing spatial interpolation methods do not consider the temporal relationship and the relations amidst SAT and other variables or factors such as topographical features and also interrelations between the SAT changes at neighboring observatory stations. Moreover, the standing methods based on remotely sensed data mainly applied univariate and bivariate regressive methods and these regressive assumptions are sometimes biased. What’s more, the complex non-linear models usually just improve the inversion accuracy of the SAT by training a variety of remotely sensed data without fully considering the in-situ SAT observations. This study proposed a new SAT reconstruction method based on the multisource data and machine learning technique which was developed by considering autocorrelation of SAT in both space and time (STAML). This study adopted and improved a range of machine learning methods by considering spatial and temporal interrelated SAT among different observatory stations. Besides, the newly-reconstructed SAT product was compared with others by different techniques in this study to screen out the right model which performs best in reconstruction of SAT across China. This study helps to provide a new SAT reconstruction technique for SAT reconstruction of other regions over the planet.

2. Data

This study collected and analyzed the following datasets: (1) In-situ SAT observations. In-situ meteorological observations from 2743 stations across China were collected from the China Meteorological
Information Center (e.g., Zhang et al., 2018a, 2018b) (http://data.cma.cn/). (2) Land surface temperature (LST) from MODIS. This LST dataset is the daily LST data product MOD11A1 by Terra MODIS with spatial resolution of 1 km × 1 km (https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a1_v006). The period this LST product covers is 2003–2012. It was well corroborated that the error between the MOD11A1 LST product and observed LST is within ±1 K (Wan, 2014). The MOD11A1 LST product has been widely used in studies of LST (e.g., Huang et al., 2014; Noi et al., 2016). (3) NDVI dataset. The SPOT-NDVI was obtained from the Earth Monitoring System at http://www.vito-eodata.be/collections/srv/eng/main.home which was developed jointly by European Union and France. The spatial resolution is 1 km × 1 km. (4) DEM data. The DEM data are from the version 4 of the STRM with spatial resolution of 90 m (http://srtm.csi.cgiar.org/). The vertical error of the DEM data is within 16 m and is the STRM DEM product of the highest accuracy. (5) Albedo data product. The albedo data are the global 8-day data with spatial resolution of 1 km (http://glcf.umd.edu/data/abd/). (6) Nighttime light data. The nighttime light data are from NOAA and are released annually by DMSP. This dataset can be obtained at https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html. (7) CART (Classification and Regression Tree). The CART is the downscaled SAT data from NCEP/NCAR across China with spatial resolution of 1 km. This dataset was obtained from http://www.geodoi.ac.cn/WebCn/. Considering homogeneity of spatial resolution of multisource datasets as mentioned above, this study processed these datasets on grids at monthly scale. The projection of the data is WGS_1984 and the spatial resolution is set to be 1 km.

Fig. 1. Working flow for SAT estimation in space using machine learning technique considering the adaptively spatial-temporal weighted method. (a) Adaptively spatiotemporal weighted method; (b) Input datasets including the observed SAT data, digital elevation model (DEM), normalized difference vegetation index (NDVI), land surface temperature (LST), albedo, spatial autocorrelation and temporal autocorrelation variable. (c) Training and prediction methods including Back-Propagation neural network (BPNN), random forest (RF), support vector machine (SVM). (d) Output data as the surface air temperature.
3. Methods

3.1. Adaptive spatiotemporal autocorrelation machine learning algorithms

In this study, three original versions of machine learning algorithms were accepted, i.e. Ori-SVM (Support Vector Machine), Ori-BPNN (Back Propagation Neural Network) and Ori-RF (Random Forest) (Breiman, 2001; Ho et al., 2014; Gupta and Christopher, 2009). These original versions of machine learning algorithms were improved by training procedures of multisource remotely sensed datasets and then the improved versions of machine learning algorithms were obtained, i.e. Geoi-SVM (Support Vector Machine), Geoi-BPNN (Back Propagation Neural Network) and Geoi-RF (Random Forest). In addition, the improved machine learning algorithms including adaptive spatiotemporal autocorrelation were introduced here.

The first step is the decision of the spatial/temporal autocorrelation variables. Based on the altitude and distance differences between the target station and its neighboring stations, the weights were obtained by training of the weights for the spatiotemporal weights at pixel by pixel. The obtained weights for each individual pixel will decide that the analysis for each pixel will include spatiotemporal autocorrelation components, i.e. temporal autocorrelation variable (T-T2m), spatial autocorrelation variable (S-T2m) (Fig. 1a). More detailed introduction can be referred to Section 3.2. The second step is the input of the data such as observed SAT, DEM, NDVI, LST, Albedo (Fig. 1b). Training procedure was done on the Ori-SVM, Ori-BPNN, Ori-RF, and the trained original version of the machine learning algorithms were called as Geoi-SVM, Geoi-BPNN, Geoi-RF and more detailed information can be found in Fig. 1c. The last step is the output of the datasets, i.e. the reconstructed SAT in this study (Fig. 1d).

This study accepted the geo-intelligent approach in reconstruction of SAT (X.H. Li et al., 2017; T. Li et al., 2017). With respect to a certain grid, the autocorrelation can be evaluated by:

\[
S_{T2m} = \frac{\sum_{i=1}^{q} w_i T2m_i}{\sum_{i=1}^{q} w_i} \quad w_i = \frac{1}{d_s^2}
\]

\[
T_{T2m} = \frac{\sum_{j=1}^{p} w_j T2m_j}{\sum_{j=1}^{p} w_j} \quad w_j = \frac{1}{d_t^2}
\]

\[
DIS = \min \left( \frac{1}{d_s} \right) \quad i = 1, 2, 3, \ldots, q
\]

wherein \(d_s\) and \(d_t\) denote spatial and temporal distance; \(p\) and \(q\) are respectively 3 and 10. Distance (DIS) denotes the spatial heterogeneity of the spatial distribution of the observatory stations. \(i\) and \(j\) denote respectively the \(i\)th observation station that is close to the pixel spatially and the value of \(j\)th day before the same pixel, respectively. The working flow of this study is shown as Fig. 1. 90% of the observed SAT data were randomly sampled to calculate spatial autocorrelation (S-T2m) and the temporal autocorrelation (T-T2m), and then the autocorrelation variables were taken as the input variable for the multiple machine learning training models. The 10% of the observed SAT were used to verify the accuracy of the model predictions.

However, it should be noted here that spatial autocorrelation is not always statistically significant in regions with evident topographically undulating and also sparse distribution of observatory stations. In this case, threshold values were set to decide whether the autocorrelation analysis will be done or not for each pixel. Given the larger probability of the larger spatial autocorrelation than the threshold value, the
autocorrelation will be done and vice versa. The space for the probability is [0, 1] and two probabilities will be accepted, therefore the median of 0.5 will be accepted for the threshold and the following equation can be obtained (X.H. Li et al., 2017; T. Li et al., 2017):

$$P_{m,n}(s) = \frac{w_{m,n}(s)}{C_{2}} \times P_{m,n}(D) + \frac{1 - w_{m,n}(s)}{C_{0} + C_{1}} \times P_{m,n}(H)$$ \quad (4)

where m and n denote the row and column number of each pixel; $P_{m,n}(s)$ is the probability that the data at a specific pixel will be processed by spatial autocorrelation algorithm; $P_{m,n}(D)$ is the spatial probability obtained using distance information between observatory stations; $P_{m,n}(H)$ is the altitude probability obtained using altitude information of each observatory stations; $W_{m,n}(s)$ is the weight for $P_{m,n}(D)$ in standardization procedure. The computation of $P_{m,n}(D)$, $P_{m,n}(H)$ and $w_{m,n}(s)$ was introduced here.

The $P_{m,n}(D)$ is computed as:

$$P_{m,n}(D) = \frac{1}{d_{\text{min}}}$$ \quad (5)

d_{\text{min}} is the minimum distance between the pixel and the observatory stations and the unit is m.

$P_{m,n}(H)$ is the altitude probability obtained by the altitude difference between the pixel and the observatory stations and the unit is m. $P_{m}$,

Table 1
The cross-validation for the modelling performance of the candidate models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sample-based cross-validation</th>
<th>Site-based cross-validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>RMSE</td>
</tr>
<tr>
<td>Ori-SVM</td>
<td>0.994</td>
<td>0.462</td>
</tr>
<tr>
<td>Ori-BPNN</td>
<td>0.958</td>
<td>1.025</td>
</tr>
<tr>
<td>Ori-RF</td>
<td>0.994</td>
<td>0.460</td>
</tr>
<tr>
<td>Geoi-SVM</td>
<td>0.995</td>
<td>0.450</td>
</tr>
<tr>
<td>Geoi-BPNN</td>
<td>0.983</td>
<td>0.724</td>
</tr>
<tr>
<td>Geoi-RF</td>
<td>0.996</td>
<td>0.369</td>
</tr>
</tbody>
</table>

Fig. 3. The cross-validation results for the models considered in this study during different months. Geoi-SVM, Geoi-BPNN, and Geoi-RF denote the developed models of SVM, BPNN, and RF models considering the spatially adaptive autocorrelation analysis, respectively. Ori-SVM, Ori-BPNN, and Ori-RF denote respectively the SVM, BPNN, and RF models without considering the spatially adaptive autocorrelation analysis. Parameters for modelling performance evaluations are respectively Mean Absolute Deviation (MAE), the Root Mean Square Error (RMSE), the coefficient of determination (R²) and the relative prediction error (PRE). Error bar represents the standard deviation of the results by the models considered in this study.
\( n^H \) can be computed by:

\[
p_H^{(m,n)} = \frac{1}{d_{\text{min}}}
\]

\( d_{\text{min}} = \min |h_{(m,n)} - h_{(x,y)}| \) for \( t = 1, 2, 3, \ldots, 10h_{(m,n)} \) denotes the altitude of the pixel which row and column number is \( m \) and \( n \). \( h_{(x,y)} \) denotes the altitude of the nearby observatory station which row and column number is \( x \) and \( y \). \( t \) denotes the \( t \)th observatory stations.

\( w_{(m,n)} \) is the weight for the \( P_{(m,n)}^{D} \) in the standardization procedure. Different temperature data from different data sources are subject to different distribution and hence different \( w_{(m,n)} \). To quantify fractional contribution of \( P_{(m,n)}^{D} \) and \( P_{(m,n)}^{H} \) to \( P_{(m,n)}^{S} \) in an adaptive way, we proposed the adaptive quantification of \( w_{(m,n)} \) as follows.

Based on Eq. (4), if the analysis of the data at a specific pixel does not include autocorrelation analysis, then \( P_{(m,n)}^{S} \leq 0.5 \), i.e.:

\[
\left( p_{(m,n)}^{H} - p_{(m,n)}^{D} \right) \times w_{(m,n)} + p_{(m,n)}^{D} \leq 0.5
\]

In this case, three scenarios can be set:

1. Given \( P_{(m,n)}^{D} > P_{(m,n)}^{H} \), we have \( w_{(m,n)} \geq \frac{0.5 - p_{(m,n)}^{D}}{P_{(m,n)}^{H} - P_{(m,n)}^{D}} \).
2. Given \( P_{(m,n)}^{D} = P_{(m,n)}^{H} \), Eq. (7) has nothing to do with \( w_{(m,n)} \), and therefore we have \( P_{(m,n)}^{H} \leq 0.5 \).
3. Given \( P_{(m,n)}^{D} < P_{(m,n)}^{H} \), we have \( w_{(m,n)} \geq \frac{0.5 - p_{(m,n)}^{D}}{P_{(m,n)}^{H} - P_{(m,n)}^{D}} \).

With respect to the temperature of the same day, given the same condition, if the weights are the same, the weight will be set to be \( w_{H} \). Therefore, the solution of the optimal weight with consideration of spatiotemporal autocorrelation is equivalent to the solution of the optimal weight without consideration of spatiotemporal autocorrelation. Since that \( w_{H} \) is constant and the optimal \( w_{H} \) value can be obtained using the enumeration technique and the detailed procedure was introduced here. Firstly, the spatial probability, \( P_{(m,n)}^{D} \), and altitude probability, \( P_{(m,n)}^{H} \), were computed based on aforementioned algorithms. Then computation of the \( w_{H} \) was done by enumeration with \([0, 1]\) as the value space and the 0.01 as the step. This procedure produced the reclassification results of the pixels with analysis.

Fig. 4. Fitting behaviors of seven models considered in this study, i.e. Geoi-RF, Geoi-BPNN, Geoi-SVM, Ori-RF, Ori-BPNN, Ori-SVM and Ori-CART. The data by the CART method was by Jing et al. (2013). The same letters in the figure denote no significant differences at 5% level, and different letters represent significant differences at 5% level.
excluding spatiotemporal autocorrelation. The \( w_h \) with the highest accuracy will be accepted for further analysis.

3.2. Verification of the models

In this study, the 10-fold cross-validation technique was accepted for verification of the models and each verification was done for ten times. This procedure used the randomly produced subsamples for training and verification and hence the reliability and the robustness can be well guaranteed (Rodriguez et al., 2010). Besides, the modelling accuracy of the improved models in this study can be well evaluated by coefficient of determination (\( R^2 \)), root mean-square error (RMSE), the mean absolute error (MAE), standard deviation (SD), and the relative prediction error (PRE) (X.H. Li et al., 2017; T. Li et al., 2017).

**Fig. 5.** Frequency distribution of the errors for monthly SAT by seven models considered in this study during a period of 2003–2013. The error is between the estimated SAT and the in situ observed SAT.

**Fig. 6.** Spatial pattern of the error of the SAT by seven models considered in this study. (a)–(g), (h)–(n), (A)–(G), and (H)–(N) represent the spatial pattern of the error during spring (March, April, May), summer (June, July, August), autumn (September, October, November) and winter (December, January, February). The error is the difference between the estimated SAT and the in situ observed SAT.
4. Results

4.1. Autocorrelation analysis

Geophysical variables are usually in spatial and temporal autocorrelation. Modelling results with inclusion of variables may be subject to low quality with large uncertainty if the variables are in significant autocorrelation in both space and time. In this case, autocorrelation was included in the analyses to screen out the right candidate variables for further modelling in SAT reconstruction (Fig. 2). It can be seen from Fig. 2 that S-T2m and T-T2m are in significant relations with the in situ SAT observations and the correlation coefficients are 0.99 respectively. Besides, significant correlation was detected between LST, Albedo, NDVI and T-T2m. Specifically, positive correlation can be found between LST, NDVI and T-T2m with correlation coefficients of 0.93 and 0.72 respectively, implying positive response of LST and NDVI to T-T2m changes. While, negative correlation was found between Albedo and T-T2m with correlation coefficient of −0.72, showing negative response of Albedo to T-T2m changes. In this case, these abovementioned results indicate significant impacts of LST, NDVI and Albedo on SAT variations in both space and time and these factors should be considered in reconstruction of SAT datasets. While, longitude and nighttime lights (NL) are in low correlation with SAT with correlation coefficients of −0.09 and 0.11 respectively. Higher SAT is observed mainly during daytime and therefore nighttime lights were not included in the SAT reconstruction.

4.2. Modelling performance evaluation using the cross-validation technique

With aim to screen out the right model for the reconstruction of SAT, comparison of the modelling performance was done amidst Geoi-SVM, Geoi-BPNN, Geoi-RF, and Ori-SVM, Ori-BPNN, Ori-RF (Table 1, Fig. 2). In general, training and validation results of these models are statistically good with coefficient of determination of R² between 0.968–0.996 and 0.906–0.996, the RMSE between 0.369–1.025 K and 0.367–1.70 K, the MAE between 0.221–0.74 K and 0.221–0.926 K. Wherein, the Geoi-RF model performs best with the lowest RMSE, MAE, and RPE values, and the RMSE and MAE values of the Ori-RF model are one order lower than other models considered in this study. Models usually perform well during training process. While, modelling performance of the models during verification process can convince modelling power of the candidate models. Generally, Geoi-RF and Ori-RF have the reliable modelling performance with minor fluctuations of the MAE, RMSE and RPE. Specifically, RMSE of Geoi-BPNN is 0.95 and 0.046 K larger than that of the Ori-BPNN; MAE of Geoi-BPNN is 0.022 and 0.021 K larger than that of the Ori-BPNN and RPE of Geoi-BPNN is 0.445% and 0.409% larger than that of the Ori-BPNN.

In addition, Fig. 3 illustrates seasonal shifts of fitting performance and modelling reliability of the models considered in this study. Expected is the reliable modelling performance of the Geoi-RF and Ori-RF models with R² > 0.992, and moderate fluctuations of MAE, RMSE and RPE when compared to other models considered in this study. Meanwhile, fitting performance of Geoi-RF and Geoi-BPNN was greatly improved when compared to Ori-RF and Ori-BPNN. Therefore, considering spatiotemporal autocorrelation of SAT can greatly improve modelling practice of the RF and BPNN, implying feasibility and rationality of the model improvement in this study. It should be noted here that spatiotemporal autocorrelation of SAT is physically and theoretically correct and should be considered in reconstruction of SAT in both space and time. However, consideration of the spatiotemporal autocorrelation of SAT does not necessarily improve the modelling performance of all machine learning algorithms such as SVM model in this study. Therefore, selection of machine learning algorithms for reconstruction of meteor-hydrological variables, e.g. SAT in this study, should be prudent and should be convinced with more state-of-the-art analysis procedures.

4.3. Evaluation of the general modelling performance

In this study, Tukey’s range test was used to compare modelling performances of the models considered in this study. Fig. 4 compared MAE, RMSE, SD and R² of 7 models considered in this study. Fig. 4 illustrated relatively high modelling efficiency of Ori-CART with MAE, RMSE and SD values of ~2 K and R² of 0.89. However, 6 models considered in this study still have better modelling performance or higher modelling efficiency than Ori-CART. The MAE value of the Ori-CART is 6.96, 4.17, 1.92, 3.10, 1.79, and 2.39 times larger than that of Geoi-RF, Geoi-BPNN, Geoi-SVM, Ori-RF, Ori-BPNN and Ori-SVM. The RMSE value of the Ori-CART is 5.71, 3.62, 1.34, 2.83, 1.81, and 2.25 times larger than that of Geoi-RF, Geoi-BPNN, Geoi-SVM, Ori-RF, Ori-BPNN and Ori-SVM. Meanwhile, SD
value of the Ori-CART is 5.46, 3.45, 1.27, 2.68, 1.71, and 2.13 times larger than that of Geoi-RF, Geoi-BPNN, Geoi-SVM, Ori-RF, Ori-BPNN and Ori-SVM. The models of Ori-RF, Ori-BPNN and Ori-SVM that included multisource remotely sensed datasets have MAE, RMSE and SD values of <1 K and R² of >0.95. While, the improved versions of Ori-RF and Ori-BPNN, i.e. Geoi-RF and Geoi-BPNN models, that included multisource remotely sensed datasets and spatiotemporal autocorrelation properties of the SAT have greatly improved modelling performance with MAE, RMSE and SD values of <0.5 K and R² of >0.99. In this case, inclusion of multisource remotely sensed datasets and adaptive spatiotemporal autocorrelation algorithm can significantly improve modelling efficiency of the models considered in this study. However, SVM is an exception, i.e. Geoi-SVM has degraded modelling performance (Fig. 4).

4.4. Prediction accuracy of the models

Evaluation of prediction accuracy of the models considered in this study was based on rates of the accurate prediction (RAP) which was defined as the percentage of the practices with prediction error of <1.5 K and/or 0.5 K to the total prediction practices. The prediction error was defined as predicted values minus observed values. Fig. 5 showed frequency of the prediction error of the SAT for 7 models considered in this study during 2003–2012. It can be seen that higher prediction accuracy of the models can be observed for SAT reconstruction during March to August (warm season) than during September to February (cold season) of the subsequent year. Specifically, the RAP of the Geoi-RF, Ori-RF, Geoi-BPNN, Ori-BPNN, Geoi-SVM, and Ori-SVM is respectively 91.82%, 73.01%, 83.00%, 47.17%, 63.21%, and 59.70%. When compared to the dataset by the CART technique (Jing et al., 2013), the RAP of the models considered in this study was greatly improved with magnitude of 63.17%, 44.36%, 54.35%, 18.51%, 34.56%, and 31.05%, implying that inclusion of multisource data into analyses and consideration of the spatially adaptive autocorrelation did benefit improvement of the prediction performance of the models.

In addition, the RAP os Geoi-RF is 18.82% higher than that of the Ori-RF and the RAP of the Geoi-BPNN is 35.84% higher than that of the Ori-BPNN. However, the RAP of Geoi-SVM is 4.90% lower than that of the Ori-SVM. The RAP of the Geoi-RF is 3.16% higher than that of the Ori-RF, Ori-BPNN, Geoi-BPNN, Ori-BPNN, Geoi-SVM, and Ori-SVM is respectively 99.40%, 96.24%, 97.98%, 89.55%, 91.01%, and 95.01%, being 30.25%, 27.09%, 28.84%, 20.40%, 21.87%, and 26.77% higher than that of the CART method. All these results further corroborated roles of the multisource data and the spatially adaptive autocorrelation analysis in improvement of the prediction accuracy of the models considered in this study. The same results were found for RF and BPNN models. The RAP of the Geoi-RF is 3.16% higher than that of the Ori-RF, and the RAP of the Geoi-BPNN is 8.43% higher than that of the Ori-BPNN. However, the RAP of the Geoi-SVM is 4.90% lower than that of the Ori-SVM. In this case, when considering multisource data and spatially adaptive autocorrelation, selection of the right machine learning models is the critical step to be done in reconstruction of the SAT.

Beside evaluation of the RAP of the models in time, evaluation of the RAP of the models in space is also necessary. Fig. 6 illustrated spatial pattern of the errors between the predicted SAT and the in situ observed SAT across China. Fig. 6 indicated increasing errors from southeastern China to northwestern China. It can be observed from Fig. 6 that spatial distribution of the in-situ observational stations in northwestern China is relatively sparse when compared to southeastern China. Besides, the in-situ observational stations for SAT in northwestern China are distributed along the transient zones from mountains to deserts and also plains. The observed SAT in northwestern China was heavily influenced by albedo and solar radiation from the deserts and so on. When compared to the SAT datasets by the CART method, the RAP of these six models considered in this study is greatly improved. However, the prediction performance of these six models over space is varying from one region to another. The relative errors of the SAT by the Geoi-RF and Geoi-BPNN models are greatly lower than those by the CART technique. Specifically, the Geoi-RF and BPNN models evidently reduced the overestimation (underestimation) of the SAT by the CART method in northern China (southwestern China) and the errors of the estimation SAT are mostly <0.5 K. Comparatively, the prediction errors of the SAT by the Geoi-RF are smaller than those of the SAT by the Geoi-BPNN in the Tibet Plateau and are relative stable seasonally. Table 3 showed that the Geoi-RF performs the best among the seven models considered in this study in the Tibetan Plateau. Compared with the Ori-CART model, the RMSE, MAE and PRE of the Geoi-RF reduced by 2.279 K, 1.695 K and 27.430%, respectively. Compared with the other five models, the accuracy of the Geoi-RF model is also significantly improved. This result showed that our method
can effectively improve the accuracy of SAT prediction even in sparsely populated areas. Although compared with the China, the model efficiency is reduced, our model does show obvious advantages in the area with sparse meteorological stations. In this sense, the Geo-i RF model is the best one in SAT prediction amidst all the models considered in this study. Fig. 7 showed relations between the in situ observed monthly average SAT and the predicted monthly average SAT by seven models considered in this study. Table 2 displayed comparisons between the in situ observed monthly average SAT and the predicted monthly average SAT by seven models considered in this study. In general, the estimated monthly SAT over regions dominated by different land use and land cover changes was in good agreement with the in situ observed monthly SAT with $R^2 > 0.95$. The errors of the estimated monthly SAT should also be attributed to some other driving factors such as the land surface temperature, wind velocity and direction, wind and also heat waves due to human activities such as urbanization-induced heat waves (Chen et al., 2015; X.H. Li et al., 2017; T. Li et al., 2017). Besides, the modelling performance of CART technique in describing lower temperature changes is not statistically good. However, machine learning techniques involving spatially adaptive autocorrelation and multisource datasets as well can greatly improve the modelling performance of the models in describing changing pattern of the lower air temperature.

4.5. Spatial pattern of the SAT

Figs. 8–11 illustrated spatial pattern of the in situ observed seasonal SAT and the estimated seasonal SAT by seven models considered in this study. It can be seen from Figs. 8–11 that seven models have statistically
satisfactory modelling performance for the SAT changes in the southeastern China. The seasonal SAT is decreasing from southeastern China to the inland which should be attributed to latitudinal distribution of the SAT. However, significantly different altitudes of the Tibet Plateau trigger evident difference of the SAT and hence different modelling performance of the models considered in this study can be expected. Even so, the Geoi-RF model still has the statistically satisfactory modelling performance for the SAT change in both space and time. During spring season, seven models considered in this study can generally describe the spatial patterns of spring SAT across China. However, SVM cannot model SAT changes in the right way. Therefore, the Geoi-SVM and Ori-SVM tended to overestimate the SAT when compared to other alternative models considered in this study. However, the Geoi-SVM model can well describe SAT changes in the Sichuan region given consideration of the multisource datasets and spatially adaptive autocorrelation relations between SAT of different regions. During summer, autumn and winter seasons, seven models considered in this study have the similar modelling performance for the SAT over the eastern China in comparison with the in situ observed SAT. However, the models have different modelling performance for the SAT over the Tibet Plateau. Geoi-SVM and Ori-SVM tended to overestimate the SAT when compared to the estimated SAT by other models. While, Geoi-BPNN, Ori-BPNN and CART techniques tended to underestimate the SAT. It is surprising to find that the Geoi-RF still has the stable modelling performance for the SAT over the Tibet Plateau, further corroborating the remarkable modelling performance of this model for the SAT across China. This finding also...
provides good reference information for the reconstruction of the SAT in other regions of the world.

5. Discussions and closing remarks

In this study, 3 models were considered in reconstruction of the SAT dataset and three models were developed with consideration of the autocorrelation of SAT in both space and time and also multisource data. The R² values of the models considering autocorrelation of the in situ observed SAT in both space and time are 11.9–39.6% higher than those original version of the models (Vogt et al., 1997; Shen and Leptoukh, 2011; Gallo et al., 2011; Benali et al., 2012; Evrendilek et al., 2012; Williamson et al., 2014; Zhu et al., 2013; Xu and Liu, 2015; Chen et al., 2015). Previous studies mostly emphasized linear relations between LST and SAT and regressive relations were considered between SAT and LST in the study of the spatial pattern of the SAT (Vogt et al., 1997; Gallo et al., 2011; Shen and Leptoukh, 2011; Benali et al., 2012; Williamson et al., 2014; Zhu et al., 2013; Chen et al., 2015; Xu and Liu, 2015). However, relations between LST and SAT were heavily modulated by underlying surface properties, weather conditions, and even cloud coverage. Simple linear relations cannot fully describe physical relations between LST and SAT and hence the reconstructed SAT in the regions without in situ observatory stations was often estimated with considerable errors (Ho et al., 2014). Besides, autocorrelations were observed between the SAT and other variables. Therefore, variables in good autocorrelations with SAT should be excluded from the analyses. More variable inputs may introduce much more uncertainties into the reconstructed SAT (Ho et al., 2014; Li et al., 2018). In addition, when compared to previous studies, this current study included multisource datasets and multiple algorithms. In so doing considers fully the shifting relations between LST and SAT. Introduction of multisource datasets also helped to reduce uncertainty and increase prediction accuracy of the reconstructed SAT (Li et al., 2018). This study combined spatially adaptive autocorrelation algorithm and machine learning models, and multisource datasets as well in the reconstruction of the SAT, which greatly improved the prediction accuracy of the SAT and the newly-developed models have improved modelling performance when compared to standing models.

This study reconstructed the SAT over continuous spatial pattern across China with spatial resolution of 1 km. In this sense, the reconstructed SAT over each data grid represents the average temperature conditions within 1 km x 1 km regions. Therefore, temperature analysis over a finer scale of space is not good due to uneven spatial pattern of the SAT as a result of different altitudes and intermittent landscape pattern. Besides, the remotely sensed datasets have been widely used in hydrometeorological study. However, influencing factors such as underlying surface properties, regional climates, topographies, cloud coverage and algorithms as well can cause systematic errors. These standing errors can also potentially introduce much uncertainty and also heavily influenced estimation accuracy of the reconstructed datasets. Moreover, due to cloud coverage, remote sensing images also contain missing data (Crosson et al., 2012; Kloog et al., 2014; Li et al., 2018; Zhang et al., 2016). Therefore, how to process missing data due to cloud coverage for the remote sensing images is still another scientific issue to be addressed in the ongoing work.

In summary, this study screened out the variables as the optimal input variable based on spatial and temporal autocorrelation algorithms. Besides, the reconstructed SAT datasets were done for the period of 2003–2012 using spatially adaptive spatial autocorrelation algorithm and machine learning models. Evaluations of the models were done based on comparison between the reconstructed SAT data, the in situ observed SAT and the data by the CART method. Our findings indicated that the Geo-i-RF and Geo-i-BPNN models can well reconstructed the SAT in the Sichuan province and the Tibet Plateau where the SAT was badly reconstructed using other alternative models and it is particularly true for the maximum and minimum SAT. Combination of the multisource remotely sensed datasets and the spatially adaptive spatial autocorrelation algorithm can greatly improve the modelling performance of the candidate models. The modelling accuracy of the developed models in this study increased as much as 18.5%–63.17%. Wherein, the Geo-i-RF model has the best modelling performance when compared to other models considered in this study in terms of modelling accuracy and the modelling stability. This study provides reference information in terms of selection of models for reconstruction of the SAT in other regions of the globe.

Acknowledgements

This work is financially supported by the National Science Foundation for Distinguished Young Scholars of China (Grant No.: 51425903), the National Natural Science Foundation of China (No. 41771536), and the Fund for Creative Research Groups of National Natural Science Foundation of China (Grant No.: 41621061). Our cordial gratitude should be extended to the editor, Prof. Dr. Pingqing Fu and anonymous reviewers for their professional and pertinent revision suggestions and comments which are greatly helpful for further quality improvement of this current manuscript.

References


