https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

GPRChinaTemp1km: a high-resolution monthly air temperature dataset for China (1951–2020) based on machine learning

Qian He[1, 2], Ming Wang[1, 3], Kai Liu[1, 3], Kaiwen Li[1, 2], Ziyu Jiang[1, 2]

[1] Academy of Disaster Reduction and Emergency Management, Beijing Normal University, 100875 Beijing, China
[2] Faculty of Geographical Science, Beijing Normal University, 100875 Beijing, China
[3] The School of National Safety and Emergency Management, Beijing Normal University, 100875 Beijing, China

Correspondence to: Ming Wang (wangming@bnu.edu.cn)

Abstract

An accurate spatially continuous air temperature dataset is crucial for multiple applications in environmental and ecological sciences. Existing spatial interpolation methods have relatively low accuracy and the resolution of available long-term gridded products of air temperature for China is coarse. Point observations from meteorological stations can provide long-term air temperature data series but cannot represent spatially continuous information. Here, we devised a method for spatial interpolation of air temperature data from meteorological stations based on powerful machine learning tools. First, to determine the optimal method for interpolation of air temperature data, we employed three machine learning models: random forest, support vector machine, and Gaussian process regression. Comparison of the mean absolute error, root mean square error, coefficient of determination, and residuals revealed that Gaussian process regression had high accuracy and clearly outperformed the other two models regarding interpolation of monthly maximum, minimum, and mean air temperatures. The machine learning methods were compared with three traditional methods used frequently for spatial interpolation: inverse distance weighting, ordinary kriging, and ANUSPLIN. Results showed that the Gaussian process regression model had higher accuracy and greater robustness than the traditional methods regarding interpolation of monthly maximum, minimum, and mean air temperatures in each month. Comparison with the TerraClimate, FLDAS, and ERA5 datasets revealed that the accuracy of the temperature data generated using the Gaussian process regression model was higher. Finally, using the Gaussian process regression method, we produced a long-term (January 1951 to December 2020) gridded monthly air temperature dataset with 1 km resolution and high accuracy for China, which we named GPRChinaTemp1km. The dataset consists of three variables: monthly mean air temperature, monthly maximum air temperature, and monthly minimum air temperature. The obtained GPRChinaTemp1km data were used to analyse the spatiotemporal variations of air temperature using Theil–Sen median trend analysis in combination with the Mann–Kendall test. It was found that the monthly mean and minimum air temperatures across China were characterized by a significant trend of increase in each month, whereas monthly maximum air temperature showed a more spatially heterogeneous pattern with significant increase, non-significant increase, and non-significant decrease. The GPRChinaTemp1km dataset is publicly available at https://doi.org/10.5281/zenodo.5112122 (He et al., 2021a) for monthly maximum air temperature, at

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

Open Access Earth System Science Data Discussions

https://doi.org/10.5281/zenodo.5111989 (He et al., 2021b) for monthly mean air temperature and at https://doi.org/10.5281/zenodo.5112232 (He et al., 2021c) for monthly minimum air temperature.

1 Introduction

Air temperature is a fundamental variable in various research fields that include the impact of global warming and climate change, ecology, hydrology, agriculture, and human health (Sippel et al., 2020; Abatzoglou et al., 2018; Pathak et al., 2018; Chen et al., 2018). Long-term records of air temperature data with high spatial resolution are necessary for such research. Generally, air temperature data are measured by meteorological station networks or simulated using numerical climate models (dos Santos, 2020; Fu and Weng, 2018). Meteorological stations can provide long-term point-based information of observed air temperature; however, they cannot reflect spatially continuous information regarding regional air temperature.

Various interpolation techniques that include inverse distance weighting (IDW) and ordinary kriging (OK) (Dawood, 2017; Li et al., 2010, 2012; Hadi and Tombul, 2018; Stahl et al., 2006; Benavides et al., 2007; Duhan et al., 2013) are often employed to derive gridded temperature datasets for data-sparse areas. However, the accuracy of the derived results depends on the density of the meteorological stations used for the interpolation (Wang et al., 2017; Peng et al., 2019; Gao et al., 2018; Peng et al., 2014). Using conventional methods for data interpolation in areas with uneven coverage of meteorological stations could diminish the accuracy of the derived data (dos Santos, 2020; Li et al., 2018). The network of meteorological stations in China is characterized by irregular spatial coverage. For example, the observation network has low density in mountain areas (Gao et al., 2018; dos Santos, 2020; Guo et al., 2020), especially on the Tibetan Plateau (Xu et al., 2018; Zhang et al., 2016). Additionally, the number of meteorological stations operational in China in the 1950s was low. Therefore, use of conventional interpolation methods cannot guarantee the accuracy of the derived spatial datasets of air temperature across China. Although various air temperature products are available, e.g., the TerraClimate (Abatzoglou et al., 2018), FLDAS (McNally et al., 2017), and ERA5 (Copernicus Climate Change Service (C3S), 2017) datasets, their spatial resolution is usually coarse (2.5 arc minutes, 0.1 arc degrees, and 0.25 arc degrees, respectively), which restricts their ability to reflect the topographical characteristics and spatial heterogeneity of air temperature across China (Peng et al., 2019; Zhang et al., 2016). Thus, demand remains for a high-resolution long-term spatially continuous dataset of air temperature.

In comparison with traditional techniques, machine learning methods are better able to model nonlinear and highly interactive relationships (Xu et al., 2018). Using mud content samples from the southwest margin of Australia, Li et al. (2011) proved the superior performance of machine learning methods in application to spatial interpolation of environmental variables. Subsequent application of machine learning methods further confirmed their effectiveness as tools for interpolation of environmental variables (Appelhans et al., 2015; Zhu et al., 2018; Alizamir et al., 2020; Kisi et al., 2017). Many previous studies have demonstrated the potential of machine learning techniques in application to estimation of short-term air temperature in small regions, although most such studies interpolated air temperature using satellite-derived predictors such as the Land Surface Temperature and Normalised Difference Vegetation Index based on MODIS products (Appelhans et al.,

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

2015; dos Santos, 2020; Meyer et al., 2016; Xu et al., 2018; Zhang et al., 2016; Yoo et al., 2018). However, MODIS data are only available from 2000, which means that air temperature in earlier years cannot be interpolated using such products. Moreover, optical remote sensing images are easily affected by clouds, limiting the ability of associated models to produce long-term spatially continuous datasets for air temperature across large regions such as China. Therefore, it is necessary to develop a universal model to interpolate long-term air temperature datasets for China. However, how best to design a simple and accurate model for temperature interpolation using machine learning remains unclear.

To interpolate air temperature across China, we employed three machine learning approaches: random forest (RF), support vector machine (SVM), and Gaussian process regression (GPR). Both RF and SVM have been proven effective in previous studies on remote-sensing-based air temperature estimation studies (Yoo et al., 2018; Zhang et al., 2016; Ho et al., 2014; Zeng et al., 2021). GPR is a powerful state-of-the-art probabilistic non-parametric regression method (Calandra et al., 2016; Schulz et al., 2018), which has produced satisfactory results regarding the prediction of daily river temperature (Zhu et al., 2018; Grbić et al., 2013) but has rarely been used for air temperature estimation. In this study, we utilized the RF, SVM, and GPR machine learning methods to develop a model for interpolation of long-term air temperature data for China.

The ultimate objective of the study was production of a long-term high-resolution spatially continuous monthly air temperature product for China, based on meteorological station data and the best-performing model constructed using the machine learning techniques. The specific variables contained in the generated product include monthly mean air temperature (Tmean), monthly maximum air temperature (Tmax), and monthly minimum air temperature (Tmin) from January 1951 to December 2020 across China.

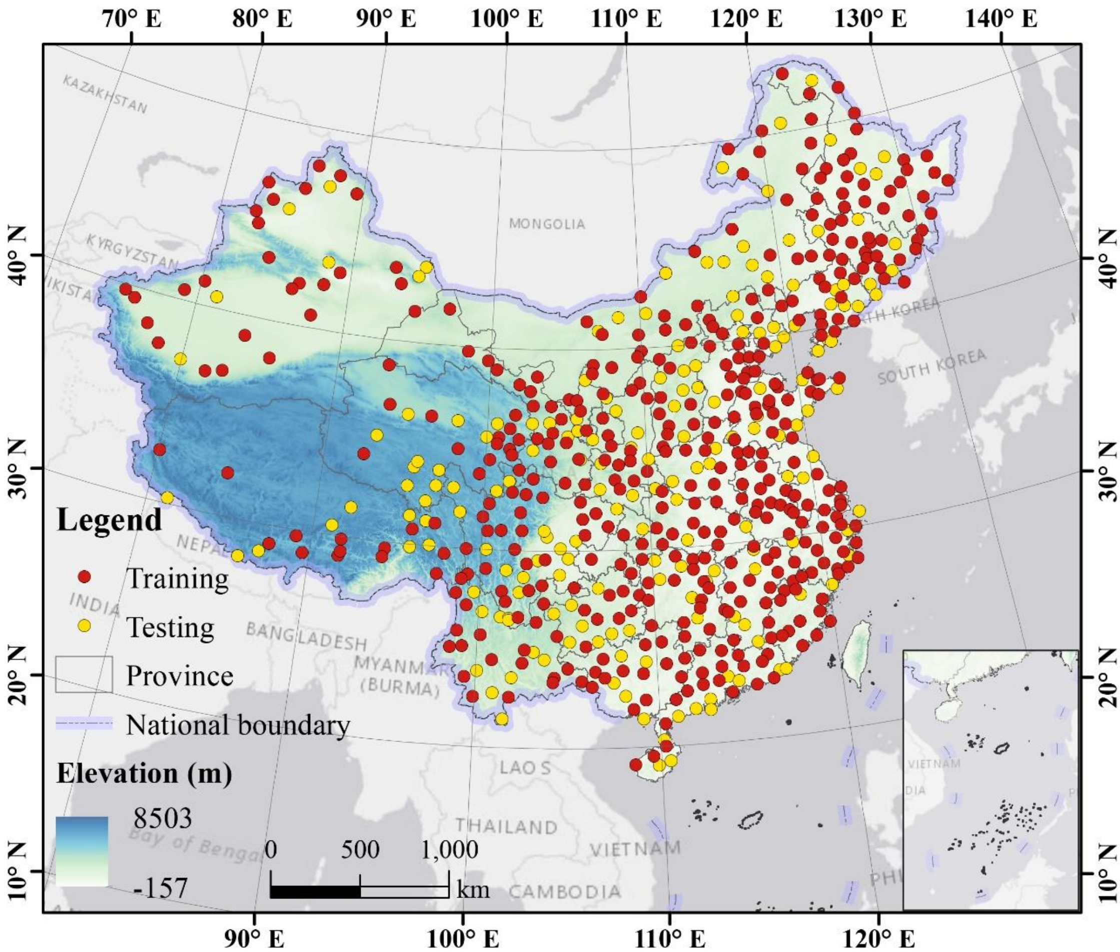
2 Data

2.1 Meteorological station data

Observational data of monthly Tmax, Tmin, and Tmean recorded from January 1951 to December 2020 at meteorological stations distributed across China were downloaded from the China Meteorological Data Service Centre (http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_MUL_MON.html, last access: 15 July 2021). The dataset includes information from 613 stations, which were split randomly into a training set (70%) for model training and a testing set (30%) for model evaluation (Figure 1). We used the “subset features” option of the Geostatistical Analyst Tools in ArcGIS10.8 to divide the original dataset. This tool considers the randomness both in the data and the spatial distritbution of the data.The number of weather stations in different years was not always exactly 613; the early years of the 1950s had notably fewer stations available (See Fig. S1 for further details regarding the number of weather stations in each year and the data records each station contains). Not that we did not impute the missing data to make all then stations have all the monthly temperature data from Januray 1951 to December 2020.

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

Open Access Earth System Science Data Discussions



95 **Figure 1: Elevation and spatial distribution of meteorological stations across China (70% were used for training; 30% were used for testing).**

2.2 Topographic data

The topographic data used in this study comprised a digital elevation model (DEM) obtained from the NASA Shuttle Radar Topographic Mission (SRTM) (https://srtm.csi.cgiar.org/, last access: 15 July 2021). We used STRM version 4, which is the latest SRTM DEM product. The spatial resolution of the DEM is 3 arc seconds (approximately 90 m resolution). The DEM was resampled to 1 km resolution to produce the air temperature dataset with 1 km resolution. Gridded latitudinal and longitudinal coordinates of 1 × 1 km pixels were also used as components. All data used in this study were processed in the WGS84 Geographic Coordinate System (EPSG:4326).

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

2.3 Existing temperature products for comparison

We used three existing temperature products for comparison: 1) the Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho (TerraClimate) dataset (resolution: 2.5 arc minutes) (https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_TERRACLIMATE, last access: 15 July 2021), 2) the Famine Early Warning Systems Network (FEWS NET) Land Data Assimilation System (FLDAS) dataset (resolution: 0.1 arc degrees) (https://developers.google.com/earth-engine/datasets/catalog/NASA_FLDAS_NOAH01_C_GL_M_V001, last access: 15 July 2021), and 3) the latest climate reanalysis produced by the ECMWF/Copernicus Climate Change Service (ERA5 Monthly aggregates) dataset (resolution: 0.25 arc degrees) (https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_MONTHLY, last access: 15 July 2021). The three datasets were used for comparison with our derived gridded temperature data. TerraClimate was used for comparing Tmax and Tmin using the maximum temperature (tmmx/°C) and the minimum temperature (tmmn/°C) variables, respectively. FLDAS was used for comparing Tmean using the near-surface air temperature variable (Tair_f_tavg/K), and we converted the unit (K) into degrees Celsius. ERA5 was used for comparing Tmax, Tmin, and Tmean using the average air temperature at 2 m height (mean_2m_air_temperature/K), maximum air temperature at 2 m height (maximum_2m_air_temperature/K), and minimum air temperature at 2 m height (minimum_2m_air_temperature/K), respectively, and the unit (K) was converted into degrees Celsius. The available time periods for the TerraClimate, FLDAS, and ERA5 products are: 1958-01-01 to 2020-12-01, 1982-01-01 to 2021-05-01, and 1979-01-01 to 2020-06-01, respectively. Considering the overlapping periods, we chose January 1979 to December 2019 for the comparisons of Tmax and Tmin, and the period January 1982 to December 2019 for the comparisons of Tmean.

3 Methods

3.1 Variable selection

The spatial distribution of air temperature is closely related to latitude, longitude, and elevation (Shao et al., 2012). Use of such auxiliary data can help alleviate to a certain extent the limitation of spatial interpolation associated with the sparse and irregular distribution of meteorological stations and increase estimation accuracy (Chen et al., 2015; Alvarez et al., 2014; Li and Heap, 2011b; Newlands et al., 2011). Figure 2 displays the correlation coefficients between air temperature (i.e., Tmax, Tmin, and Tmean) and the above three geographical variables. Note that the correlation coefficient value for each month represents the average of all years (1951–2020), which was obtained based on all the observed data from meteorological stations. The box plots of the correlation coefficient for each month are provided in Fig. S2. Overall, Tmax, Tmin, and Tmean each have positive (negative) correlation with respect to longitude (latitude and elevation). Longitude and elevation have opposite correlation but a similar trend with Tmax, Tmin, and Tmean, i.e., reasonably high correlation during summer (June–August) and low correlation during winter (December–February). Latitude is correlated negatively with Tmax, Tmin, and

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

Tmean, i.e., strong (weak) correlation in winter (summer). It is evident that strong regularity exists in the relationships between air temperature and longitude, latitude, and elevation. Thus, we chose the three variables as predictor variables for obtaining the gridded temperature raster from the point observations. Owing to the incompleteness of remote sensing data attributable to imaging time constraints and cloud contamination, we did not consider satellite-derived independent variables. We considered only longitude, latitude, and elevation as predictor variables to give the derived model the advantages of ease of use, generalizability, and universality.

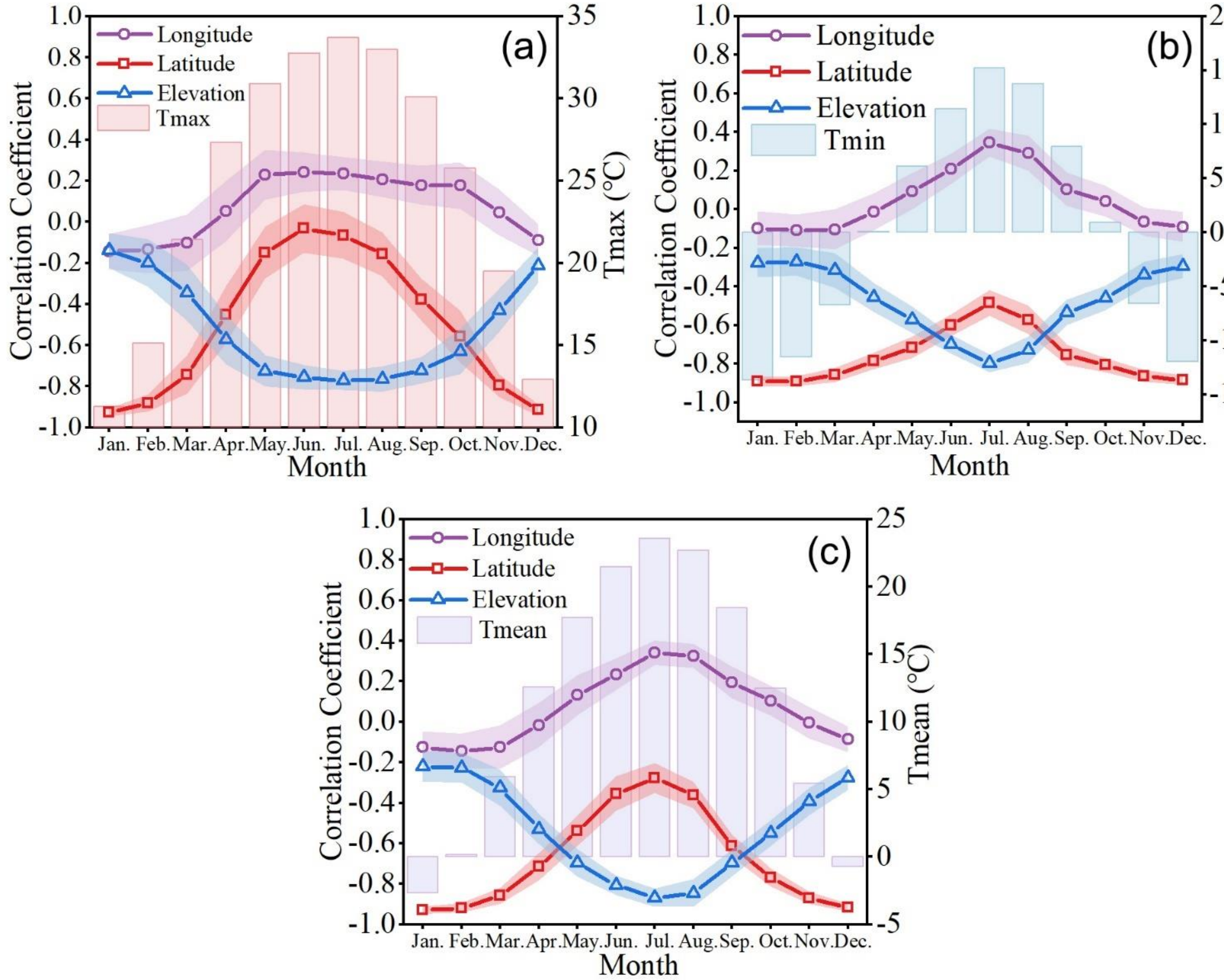


Figure 2: Correlation coefficients between (a) monthly maximum air temperature (Tmax), (b) minimum air temperature (Tmin), and (c) mean air temperature (Tmean) and longitude, latitude, and elevation for each month. Coloured shading indicates the standard deviation. Note that the correlation coefficients are the average values of the correlation coefficients for each month over 70 years (1951–2020).

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

3.2 Machine learning models

3.2.1 Random forest (RF)

RF, proposed by Breiman (2001), has been used widely for regression of geographical variables. RF is an ensemble machining learning method that consists of multiple decision trees. RF can produce high rates of accuracy, and the performance of RF in predicting new data is determined by the aggregation of the results of all the trees (Hengl et al., 2018). The randomization of RF lies in two aspects: the random selection of training samples for a tree through bagging (a form of bootstrapping), and the random selection of predictor variables as the splitting attributes at each node of the tree (Merghadi et al., 2020; Yoo et al., 2018). The randomness of RF makes it resistant to the problem of overfitting. RF, which has been demonstrated promising and flexible in dealing with heterogeneity in the geographical environment, has been applied to prediction of spatial and temporal variables (Hengl et al., 2018; Zeng et al., 2021; Yoo et al., 2018). For further detailed information regarding RF, the reader is referred to Breiman (2001). We used the ensemble algorithm for regression in MATLAB R2020b for the RF implementation. The minimum observations per leaf were set at 8 and the number of ensemble learning cycles was set at 30. The reader is referred to the MATLAB help centre for further details (https://www.mathworks.com/help/stats/fitrensemble.html?searchHighlight=NumLearningCycles&s_tid=srchtitle, last access: 15 July 2021; https://www.mathworks.com/help/stats/ensemble-algorithms.html, last access: 15 July 2021; https://www.mathworks.com/help/stats/fitrensemble.html#bvcj_t2-15, last access: 15 July 2021).

3.2.2 Support vector machine (SVM)

SVM, developed by Vapnik (2013), utilizes the inductive principle of structural risk minimization to obtain the overall optimal response. SVM transforms input data from lower-dimensional into a high-dimension space based on a series of kernel functions (Fan et al., 2018). The input space and the output space are non-linearly related in real applications, and the limitation is solved by mapping the input space on to higher dimension. In regression applications, an optimal hyperplane is constructed that is as close to as many samples as possible. The SVM does not only consider the error approximation to the data but also the model generalization. SVM has been used widely in various fields such as meteorology, hydrology, and agriculture for regression and prediction applications (Ghorbani et al., 2017; Shrestha and Shukla, 2015; Fan et al., 2018). Detailed information regarding SVM can be found in Vapnik (2013). We implemented the SVM algorithm in MATLAB R2020b. The Gaussian kernel was adopted as the kernel function of SVM and the kernel scale parameter was set at 1.7. The box constraint value for the Gaussian kernel function was obtained by dividing the interquartile range of the response variable by 1.349. The predictors were standardized in the SVM model. The reader is referred to the MATLAB help documentation for further technical details (https://www.mathworks.com/help/stats/fitrsvm.html, last access: 15 July 2021 and https://www.mathworks.com/help/stats/understanding-support-vector-machine-regression.html, last access: 15 July 2021).

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

3.2.3 Gaussian process regression (GPR)

GPR is a non-parametric Bayesian technique for solving nonlinear regression problems (Grbić et al., 2013). GPR was originally proposed to provide a "principle, practical, and probabilistic approach to learning in kernel machines" (Rasmussen, 1997, 2004). GPR is based on Bayesian theory and statistical learning theory, which is applicable to regression problems (Zhang et al., 2019). GPR has strength in its seamless combination of several machine learning tasks such as model training, hyperparameter estimation, and uncertainty estimation (Sun et al., 2014; Zhu et al., 2018). GPR has been utilized in diverse applications that include model approximation, experiment design, and multivariate regression (Zhu et al., 2018; Karbasi, 2018); however, previous application of GPR to prediction of air temperature has been limited. For detailed information regarding the GPR model, the reader is referred to Rasmussen (1997, 2004). The explicit basis in the GPR model is "constant" and the kernel function of the GPR algorithm is the exponential kernel. The predictors variables were standardized in the GPR model. GPR was implemented in MATLAB R2020b. The reader is referred to the MATLAB help documentation for further details regarding GPR (https://www.mathworks.com/help/stats/fitrgp.html, last access: 15 July 2021 and https://www.mathworks.com/help/stats/gaussian-process-regression-models.html, last access: 15 July 2021).

We extracted the independent variables (i.e., latitude, longitude, and elevation) relating to the meteorological stations and randomly divided the processed data into a set for model training (70%) and a set for model evaluation and validation (30%). When train the models, the 10-fold cross-validation was used. We constructed a model for each month separately which means we have 840 models for the 840 months from 1951 to 2020. All algorithms were implemented in MATLAB R2020b.

3.3 Model evaluation metrics

We used three metrics to evaluate model performance: mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2), which have all been used widely in previous studies to evaluate model capability in predicting the dependent variable (Graf et al., 2019; Khanal et al., 2018; Peng et al., 2019; Ji et al., 2015). The MAE is the mean value of all the individual errors. The RMSE measures the discrepancy between the observed and predicted values. The MAE and RMSE both summarize the mean difference of the observed and predicted values and are among the best overall measures of model performance (Li and Heap, 2011a). Lower values of MAE and RMSE mean better accuracy. R^2 measures the proportion of variance explained by the model (Sekulić et al., 2021), representing how well the predicted values fit in comparison with the observed values. The higher the R^2 value, the better the model performance:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |P_i - O_i| \,, \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (P_i - O_i)^2} \,, \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \,, \tag{3}$$

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

where P_i is the predicted value in the time series, O_i refers to the observed value from the meteorological stations, n is the number of samples, and $\bar{O}$ represents the average of the observed values from n meteorological stations. All performance measures were calculated using the testing dataset for evaluation purposes.

3.4 Methods for spatiotemporal analysis of monthly air temperature

The Theil–Sen slope estimator used in combination with Mann–Kendall (MK) detection, which is an effective approach for trend analysis that reflects the variation in trends of each pixel in a time series, has been used widely in various fields such as hydrology and meteorology (Cai and Yu, 2009; Gocic and Trajkovic, 2013; Jiang et al., 2015). In this study, we used the Theil–Sen estimator coupled with the MK test to detect the trend of the temperature time series.

(1) Theil–Sen estimator

The Theil-Sen estimator, which is a robust non-parametric approach for estimating the slope of a trend, has been used widely in relation to hydrometeorological time series data (Jiang et al., 2015; Gocic and Trajkovic, 2013; Shifteh Some'e et al., 2012; Sayemuzzaman and Jha, 2014). The Theil–Sen slope estimator, which represents the magnitude of a trend, can be expressed as in Eq. (4) (Theil, 1950; Sen, 1968):

$$\beta = \mathrm{Median}\left(\frac{x_j - x_i}{j - i}\right), \forall j > i \,, \tag{4}$$

where β denotes the Theil–Sen median slope, and x_i and x_j refer to the air temperature at time i and j, respectively. The slope derived from the Theil–Sen estimator is a robust estimate of the magnitude of a trend, which can represent an increasing trend ($\beta > 0$) or a decreasing trend ($\beta < 0$) over the study period on the pixel scale. In this study, the Theil–Sen median slope was computed using the MATLAB platform.

(2) Mann–Kendall (MK) test

The MK test quantifies the significance of a trend. It is a non-parametric statistical test, meaning that it does not require samples to follow specific distributions and it is not influenced by outliers. The MK test has frequently been applied to measure the significance of trends in hydrological and meteorological time series data (Jiang et al., 2015; Shifteh Some'e et al., 2012; Da Silva et al., 2015; Gocic and Trajkovic, 2013). The Z statistic is used to evaluate a trend; a positive (negative) value of Z means an increasing (decreasing) trend. Further details regarding the MK test can be found in Jiang et al. (2015) and Shifteh Some'e et al. (2012). In this study, we set the significance level at 5%, similar to many other related studies (Jiang et al., 2015; Shifteh Some'e et al., 2012; Da Silva et al., 2015), which means the variation is significant when |Z| is >1.96; otherwise, the variation is non-significant. The MK test was conducted using MATLAB language.

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

4 Results

4.1 Evaluation of model performance

We used the testing dataset to evaluate the performance of each model. Figure 3(a)–(c) presents the MAE, RMSE, and R^2 values of Tmean, respectively, of the three machine learning models for each month in the time series of 1951–2020. The MAEs of GPR and SVM are close to 1 across the study period (the MAEs are slightly smaller for GPR), while the MAEs of RF are clearly higher than those of both GPR and SVM. The RMSEs have the same order as the MAEs, i.e., GPR outperforms both SVM and RF. The differences in the RMSEs of the three models are evident; GPR has the lowest RMSE in every month throughout the study period (maximum RMSE = 1.35°C, average RMSE = 0.79°C, and Std = 0.15°C). Detailed inspection of the MAEs and RMSEs from January 2015 to December 2020 (Fig. S3 in the Supplementary Material) reveals that the errors are relatively larger in cold months (November–February) and smaller in warmer months. All three models show relatively high values of R^2. GPR and SVM have R^2 values that are very similar, i.e., average R^2 values of 0.97 and 0.96, respectively, while RF has lower values of R^2, especially during the first few years. For Tmean, RF shows distinct fluctuations throughout January 1951 to December 2020, whereas GPR and SVM are relatively stable. The accuracy metrics show that the MAEs and RMSEs fluctuate from month to month, while R^2 remains reasonably constant. The accuracy metrics of GPR averaged over 840 months from January 1951 to December 2020 are as follows: MAE = 0.79°C, RMSE = 0.79°C, and R^2 = 0.97 for Tmean. The three metrics indicate that GPR always has highest accuracy and lowest standard deviation, reflecting the robustness of GPR. For Tmax and Tmin, GPR still performs best according to the evaluation metrics (Figs. S4 and S5). The correlation coefficients of air temperature and the predictor variables (Fig. 2) vary from month to month, which might contribute to the fluctuation in the accuracy of the interpolation with month.

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

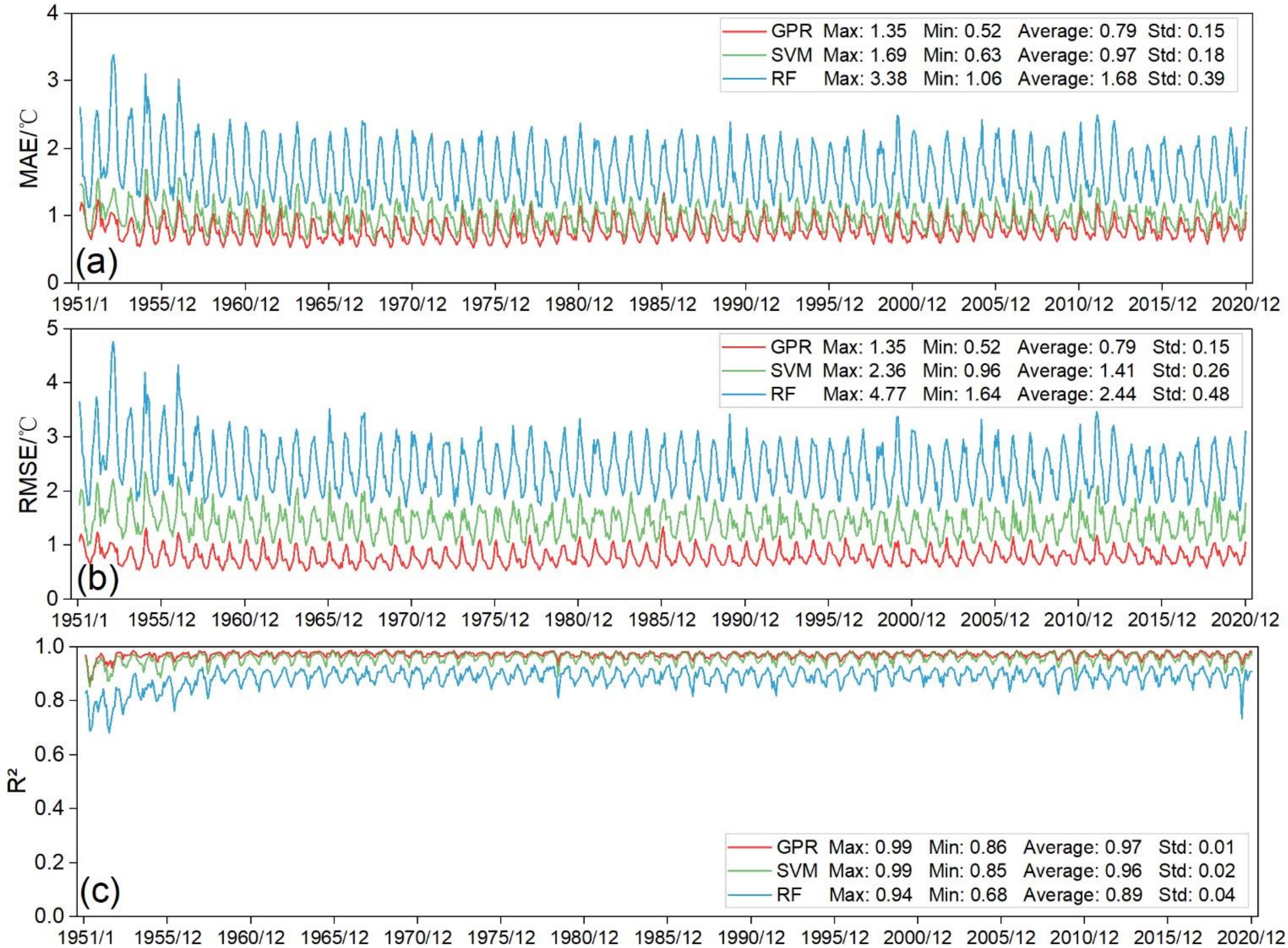


Figure 3: (a) Mean absolute error (MAE), (b) root mean square error (RMSE), and (c) coefficient of determination (R^2) between observed Tmean and predicted Tmean by the three machine learning models (GPR, SVM, RF) of the test meteorological stations over the period from January 1951 to December 2020.

The residuals were obtained as the observed values minus the predicted values. Figure 4 shows box plots of the residuals for Tmean for the test meterorlogical stations in each month during 1951–2020. Overall, the mean residuals of the three models are generally close to 0, and the residuals are smaller during the warm months (June–September) than during the cool/cold months (October–April), particularly for RF and SVM. In comparison with SVM and RF, GPR has the most stable accuracy over the 12 months, i.e., the difference in the residuals among the months is relatively small. GPR also has a quantile range that is narrower than that of the other models. For Tmax and Tmin, the bias of GPR over the 12 months is smaller than that of both RF and SVM (Figs. S6 and S7). Additionally, the accuracy of the estimated Tmax is higher than that of Tmin, consistent with the findings of Tang et al. (2020). The results show that the GPR model could be a better choice than either RF or SVM for estimating Tmean, Tmax, and Tmin for China. The frequency distributions of the residuals of the three machine learning

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

models for Tmean, Tmax, and Tmin are provided in the Supplementary Material (Figs. S8–S10), in which it can be found that GPR generally has the greatset concentration of residuals close to 0.

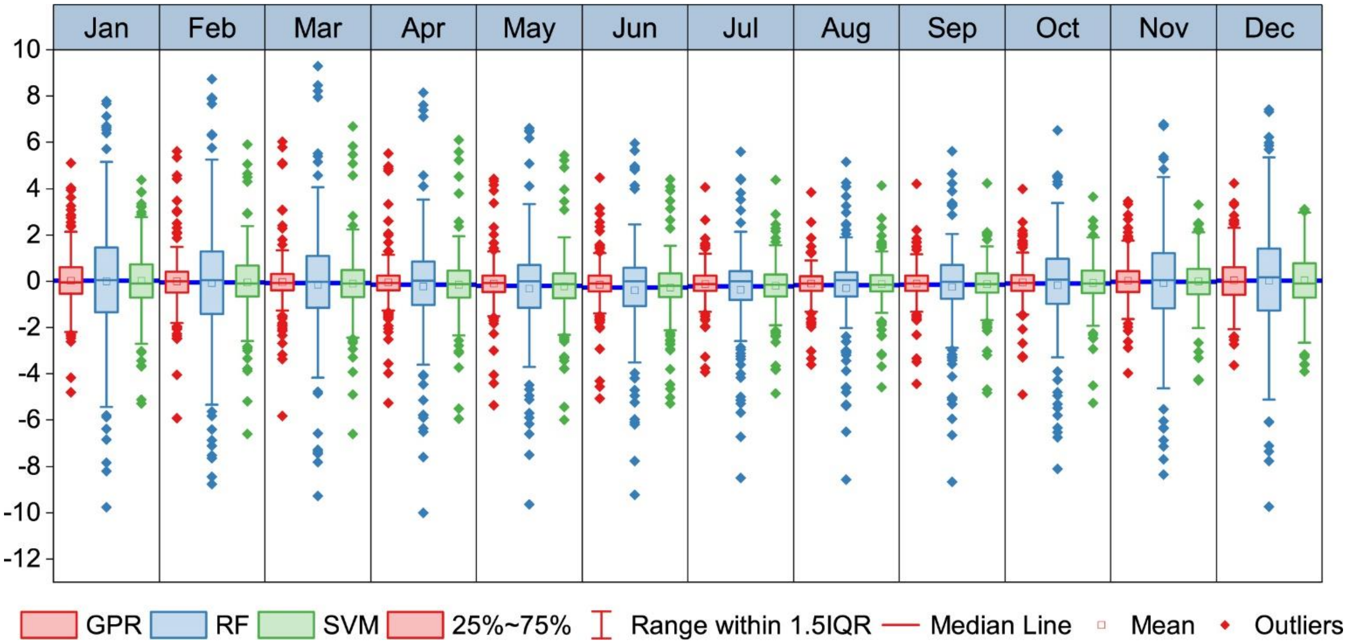


Figure 4: Residuals of the monthly Tmean predicted by the machine learning models with respect to in situ Tmean for the test meteorological stations. Note that the average of the residuals of Tmean from 1951–2020 for each test meteorological station is shown for each month.

The spatial distribution of the average values of the residuals of the GPR results for Tmean throughout the 70 years (1951–2020) at each of the test meteorlogical stations is displayed in Figure 5. Most areas have relatively low absolute residuals, although certain stations in some western areas have relatively high residuals. In January and December, the number of stations with high absolute residuals (>2.5°C) is relatively higher than in other months, i.e., 13 and 12 stations, respectively. Conversely, there are only five, five, and four stations with absolute residuals >2.5°C for June, July, and September, respectively. This might indicate that the GPR model produces better results during warmer months. Furthermore, among the stations with high absolute residuals (>2.5°C), more are positive than negative, indicating that the observed values are higher than the predicted values, i.e., there is slight underestimation by GPR at those stations. Overall, most stations show residuals of between −1°C and 1°C. The maps of the residuals for Tmax and Tmin also display patterns that are spatially similar to the maps of residuals for Tmean; however, the overall residuals of Tmax exhibit better results in comparison with the spatial pattern of the residuals of Tmin (Figs. S11 and S12).

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

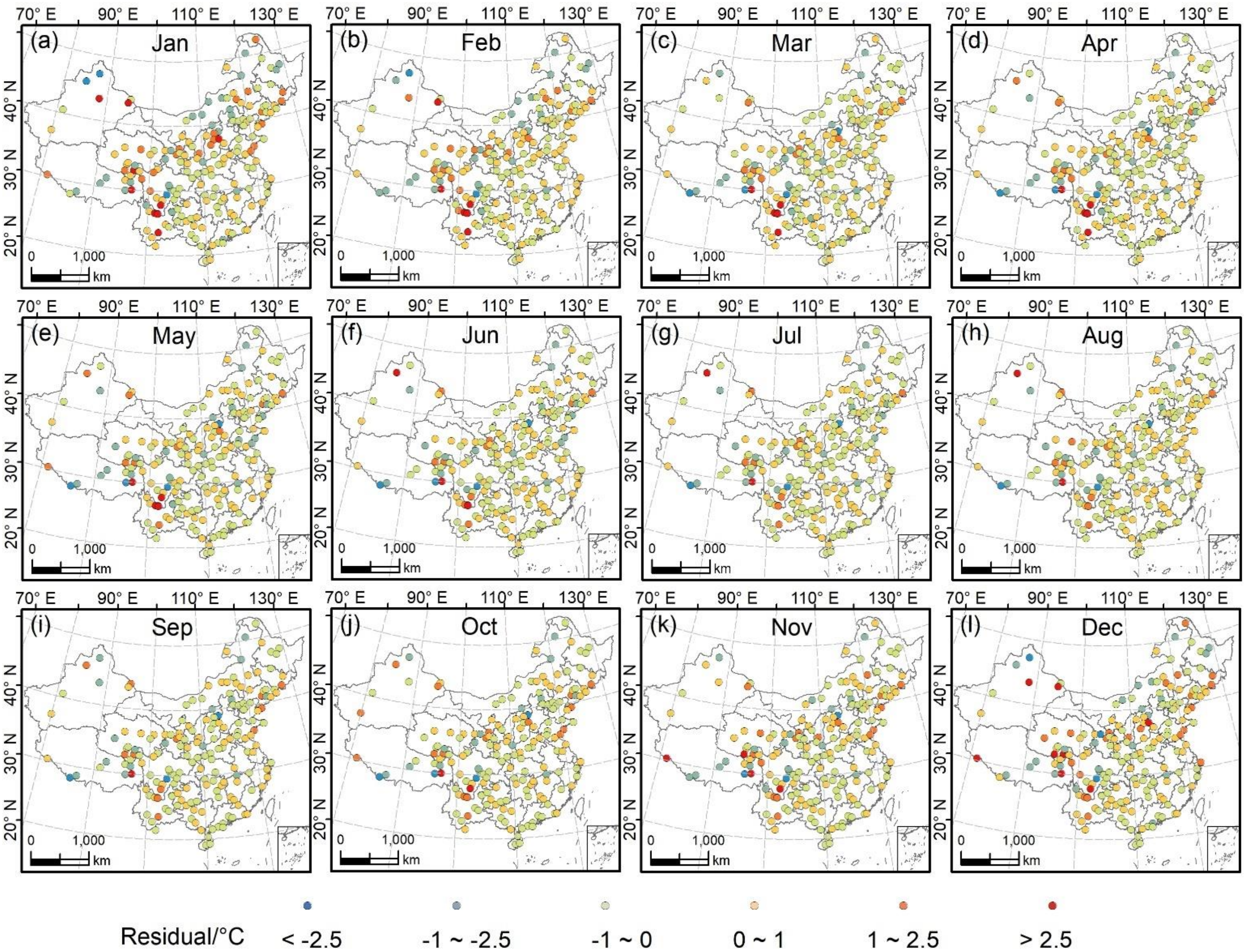


Figure 5: Spatial distribution of residuals between the observed Tmean and the predicted Tmean by GPR for the test meteorological stations for each month. Note that the exhibited residuals are the average residual of 70 years (1951–2020) for each month.

4.2 Spatial distribution of air temperature

According to the model evaluation, we concluded that GPR is the best model for estimating air temperature across China. Therefore, we employed the GPR model to generate the long-term spatial dataset of Tmean, Tmax, and Tmin from January 1951 to December 2020, which we named GPRChinaTemp1km. Figure 6 illustrates the spatial pattern of Tmean estimated by GPR in 2020. The differences between northwestern and southeastern regions are remarkable. Generally, Tmean decreases from the southeast toward the northwest. In winter, the temperature range between northern and southern China is large, whereas the temperature range in summer is relatively small. The lowest Tmean (−27°C) occurs in January and the highest Tmean (34°C) occurs in July, consistent with the fact that January and July are generally the coldest and hottest months, respectively. The maps show reasonable changes as the seasons change, i.e., high temperatures in summer (June–August) and

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

low temperature during winter (December–February). Overall, Tmax and Tmin in China follow a pattern similar to that of Tmean, i.e., decreasing from the south toward the north (Figs. S13 and S14). The highest Tmax of 2020 (44°C) occurs in July (Fig. S13) and the lowest Tmin (−43°C) occurs in December (Fig. S14). The results well describe the spatial heterogeneity of air temperature across China. Additionally, the border of the Tibetan Plateau is evident in the maps of Tmax, Tmin, and Tmean for each month, especially in the winter and summer seasons, further demonstrating the rationality of the derived results.

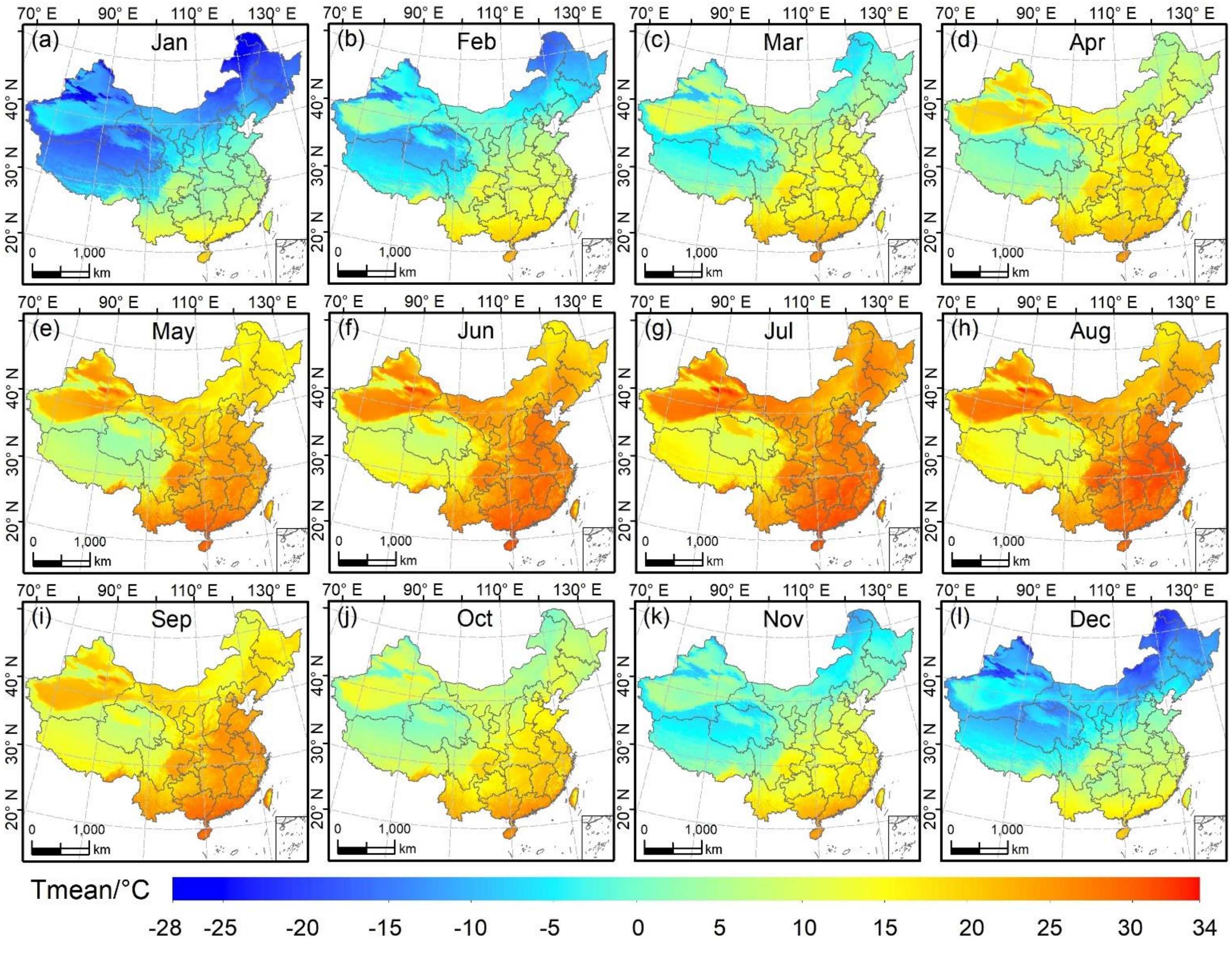


Figure 6: Spatial distribution of monthly Tmean predicted by GPR across China for each month in 2020. Note that only the maps for 2020 are presented as an example (all the data are available in the China GPRChinaTemp1km database).

4.3 Trend analysis of air temperature in China

Theil–Sen median trend analysis was integrated with the MK test and the results were classified into four categories: significant increase, non-significant increase, significant decrease, and non-significant decrease. Figure 7 shows that the trend of the variation of Tmean (1951–2020) in China is dominated by significant increase in each month. There is only a small region in

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

northwestern China that has significant decrease in Tmean in January and December. We found that there is always a small region showing a different trend in comparison with surrounding areas in the Xinjiang Uygur Autonomous Region in northwestern China, which is characterized by a decreasing trend in most months and non-significant increase in the hot months (June–September). This phenomenon could be related to the complex conditions of the region. For example, Bayinbuluke is an intermountain basin surrounded by the Tianshan Mountains with an alpine wetland ecosystem in the arid temperate zone. During summer (June–August), Tmean shows distinct non-significant decrease in central areas of China. In December, the spatial differentiation is the most remarkable, and the increasing trend in most of eastern China is non-significant, which differs from that of other months, and there is a region representing a trend of non-significant decrease on the Yungui Plateau in southwestern China. Overall, the trend of Tmean in China during 1951–2020 shows significant increase in each month, while only a few areas have a trend of decrease. Tmax is characterized by significant increase and non-significant increase, as well as a non-significant decreasing trend (Fig. S21). Tmin exhibits a spatial pattern similar to that of Tmean, showing a significant increasing trend in most areas in each month (Fig. S22).

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

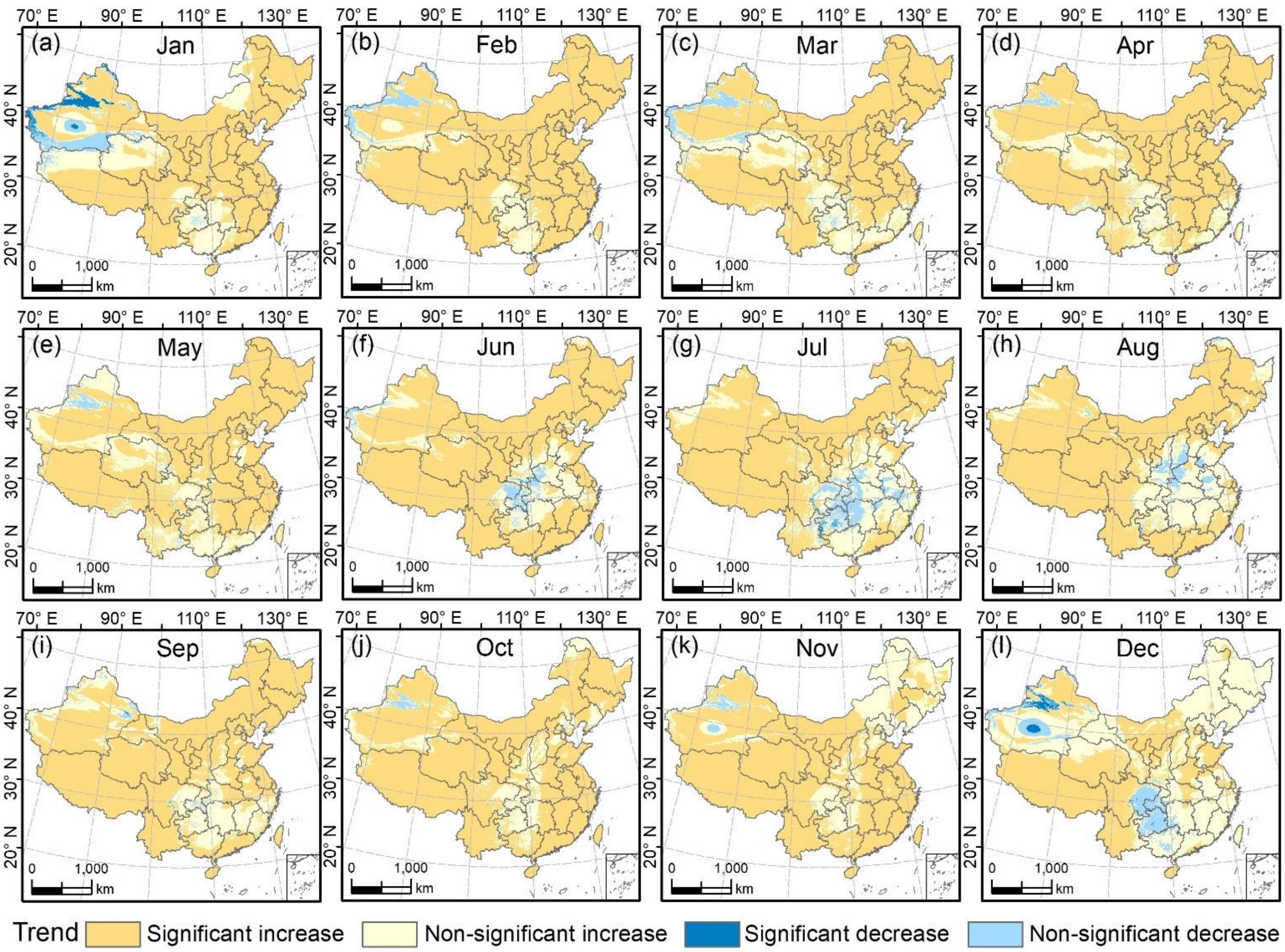


Figure 7: Monthly trends of Tmean change in China during 1951–2020 obtained by Theil–Sen median slope analysis. The significance of the trends is quantified by the Mann–Kendall statistical test at the 95% confidence level. The separate Theil–Sen trend analysis and MK test results for Tmean, Tmax, and Tmin are provided in the Supplementary Material (Figs. S15–S20).

5 Discussion

5.1 Comparison with traditional interpolation methods

Two traditional methods used widely for spatial interpolation are IDW and OK (Li and Heap, 2014, 2011a). In this study, we used ANUSPLIN in addition to IDW and OK for comparison with the machine learning models. ANUSPLIN, which is professional interpolation software that uses the thin-plate smoothing spline algorithm (Hutchinson, 1995, 2004; Xu and Hutchinson, 2013), has been used to create many climatic datasets such as the monthly Climatic Research Unit dataset (New et al., 2000) and the WorldClim dataset (Fick and Hijmans, 2017; Hijmans et al., 2005). We compared the interpolation results derived using the machine learning models with the results obtained using the traditional methods to further assess the

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

interpolation power of the machine learning methods regarding air temperature across China. The accuracy metrics (Fig. 8) show that the performances of GPR, SVM, and ANUSPLIN are of a similar level, while RF, IDW, and OK perform less well. Both IDW and OK have relatively high interpolation errors with higher MAEs and RMSEs than GPR and SVM (Figure 8). Overall, IDW and OK do not perform well in July and January of all the studied years. Figure 9 shows scatter plots of observed monthly Tmean and Tmean estimated by the six models for January and July 2020. It can be seen that OK and IDW both have clear differences between January and July (Figure 9g, h, j, and l), in which the points are relatively widely dispersed in July. GPR, SVM, and ANUSPLIN are slightly affected by the seasonal variation with lower errors (i.e., lower MAEs and RMSEs) in July (Figure 9). As shown in Figure 3, the RMSE, MAE and R2 shows a cycle pattern. In winter (Nov, Dec, Jan and Feb), temperature has relatively lower correlation two variables have relatively lower correlation, i.e longitude and elevation, while in summer, temperature has lower correlation with one variable (latitude) but has high correlations with longitude and elevation. The RMSE and MAE are high for the winter months as shown in the zoomed-in accuracy in Figure S3. This accuracy cycle pattern is probably induced by the correlation difference between summer and winter. GPR has the lowest MAEs and RMSEs, and the highest R^2 values in most months. Considering the proven power of ANUSPLIN in predicting meteorological variables, the GPR yields relatively satisfactory results. Taking the accuracy in 2020 as an example (Figure 9), ANUSPLIN has higher errors and lower R^2 values than GPR, and there are certain points with values estimated by ANUSPLIN that are relatively far away from the observed values in July (Figure 9l). In contrast, the Tmean values estimated by GPR are relatively close to those of the in situ Tmean values (Figure 9b).

Comparison of the performances of the six models for Tmax and Tmin reveals that GPR performs better in terms of Tmax and has the lowest errors (MAEs and RMSEs) in almost all the studied months (Fig. S23). OK and IDW have similar performances, consistent with the findings of previous related studies (Plouffe et al., 2015; Li et al., 2011). It is noticeable that IDW and OK perform relatively poorly. Both IDW and OK depend on the spatial autocorrelation of air temperature and cannot capture the geomorphic characteristics of the interpolation area because neither method includes elevation information (Ozelkan et al., 2015; Wang et al., 2017; Li et al., 2011). Unlike IDW and OK, ANUSPLIN considers longitude, latitude, and elevation (Hijmans et al., 2005). The frequency distributions of the residuals for Tmean, Tmax, and Tmin of the six models for the same months as in Fig. 9 are presented in the Supplementary Material (Figs. S24–S26). The distributions follow a normal distribution, and the residuals of GPR, SVM, and ANUSPLIN are concentrated mainly around 0. Scatter plots of Tmean, Tmax, and Tmin for the same periods as shown in Fig. 9 are provided in the Supplementary Material (Figs. S27–S49), in which the robustness of GPR is clearly demonstrated for Tmean, Tmax, and Tmin in comparison with the other methods. Studies have shown that Gaussian processes are one of the most intuitive techniques for modelling spatial surfaces (Yu et al., 2017; Berger et al., 2001).

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

Open Access Earth System Science Data Discussions

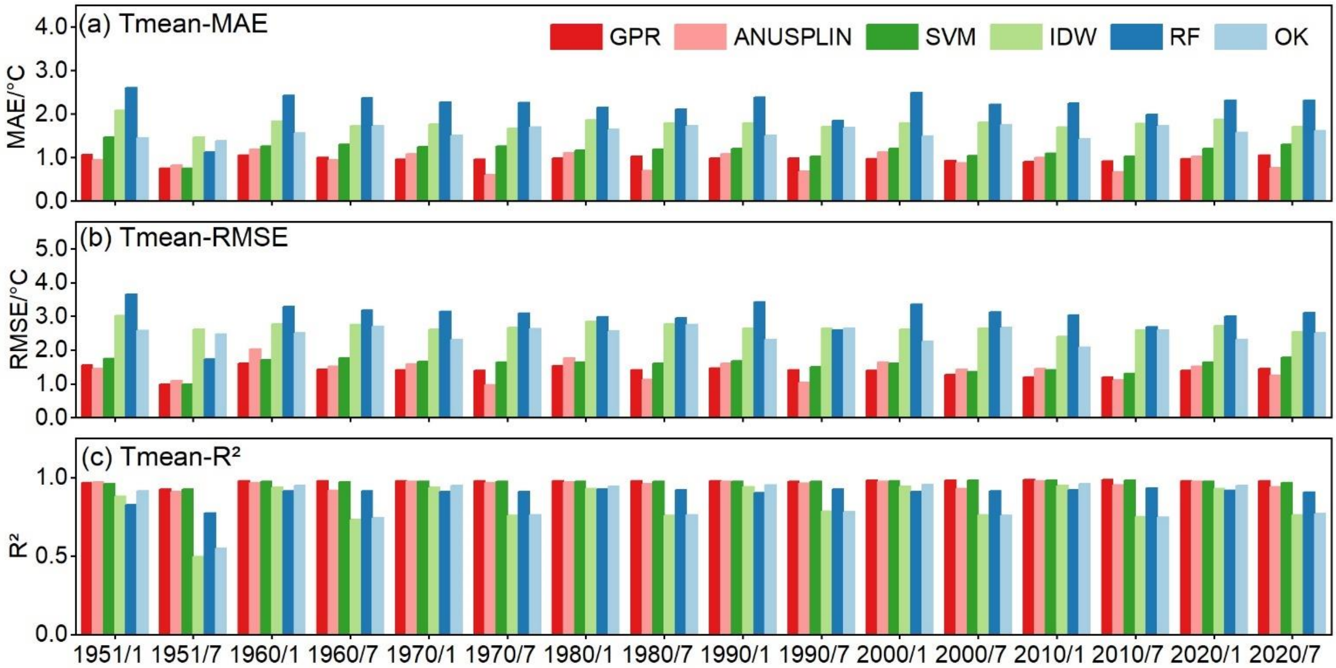


Figure 8: Accuracy of Tmean derived from the machine learning methods and traditional methods for January and July 1951–2020 with an interval of 10 years.

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

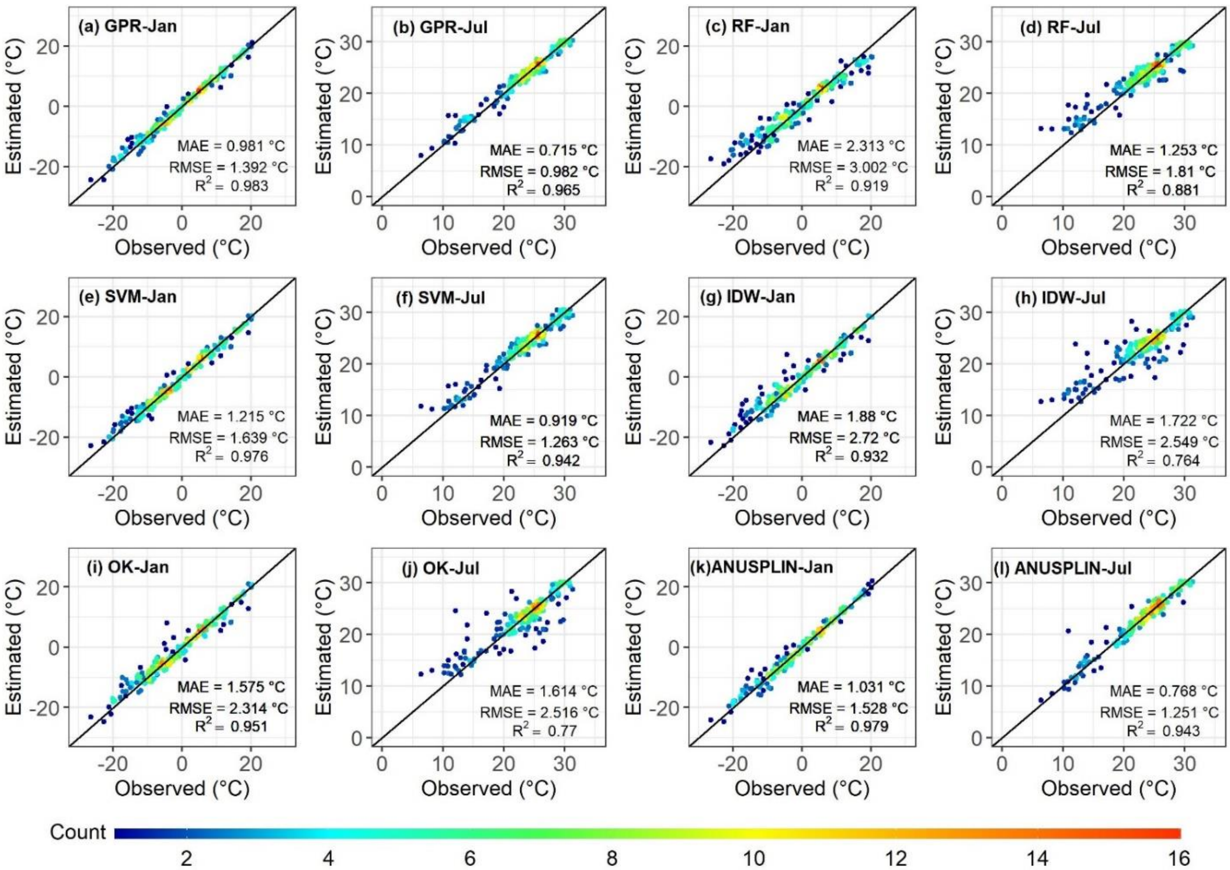


Figure 9: Scatter plots of Tmean estimated by the machine learning models and traditional models against observed monthly mean temperature in January and July 2020.

Figure 10 presents maps of the residuals (observed values minus estimated values) for Tmean in January and July 2020 estimated using the six methods. Bias is apparent in RF (Figure 10c and d), IDW (Figure 10g and h), and OK (Figure 10i and j). Comparison of the maps of the residuals reveals that Tmean estimated by GPR generally agrees well with the in situ data, with large bias at only a few stations that are distributed mainly in western and northern China, which might be related to the scarcity of meteorological stations and the complex regional topography (Ji et al., 2015). It is also evident that the absolute residuals in July are generally lower than in January (Figure 10). For China, the spatial homogeneity of temperature in summer is stronger than that in winter, which might be one reason for the lower bias observed in July. We note that RF has poor performance in comparison with the other machine learning methods. Although we do not have sufficient evidence to deduce the causes for the lower accuracy of RF, the small number of meteorological stations might be a major reason. Additionally, RF regression has a limitation regarding the conditions beyond the range of the training dataset because only the values included in the training data are used for splitting the trees (Mutanga et al., 2012; Jeong et al., 2016). Among the three machine learning algorithms, GPR and SVM both perform relatively well, although the performance of GPR is better. Note that we

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

used the medium Gaussian SVM and exponential GPR in MATLAB R2020b. GPR and SVM are both non-parametric kernel-based models that rely on the Gaussian principle. The Gaussian function has the desired characteristics of being an inverse-distance algorithm and a smoothing filter (Thornton et al., 1997), which might explain the better performances of GPR and SVM.

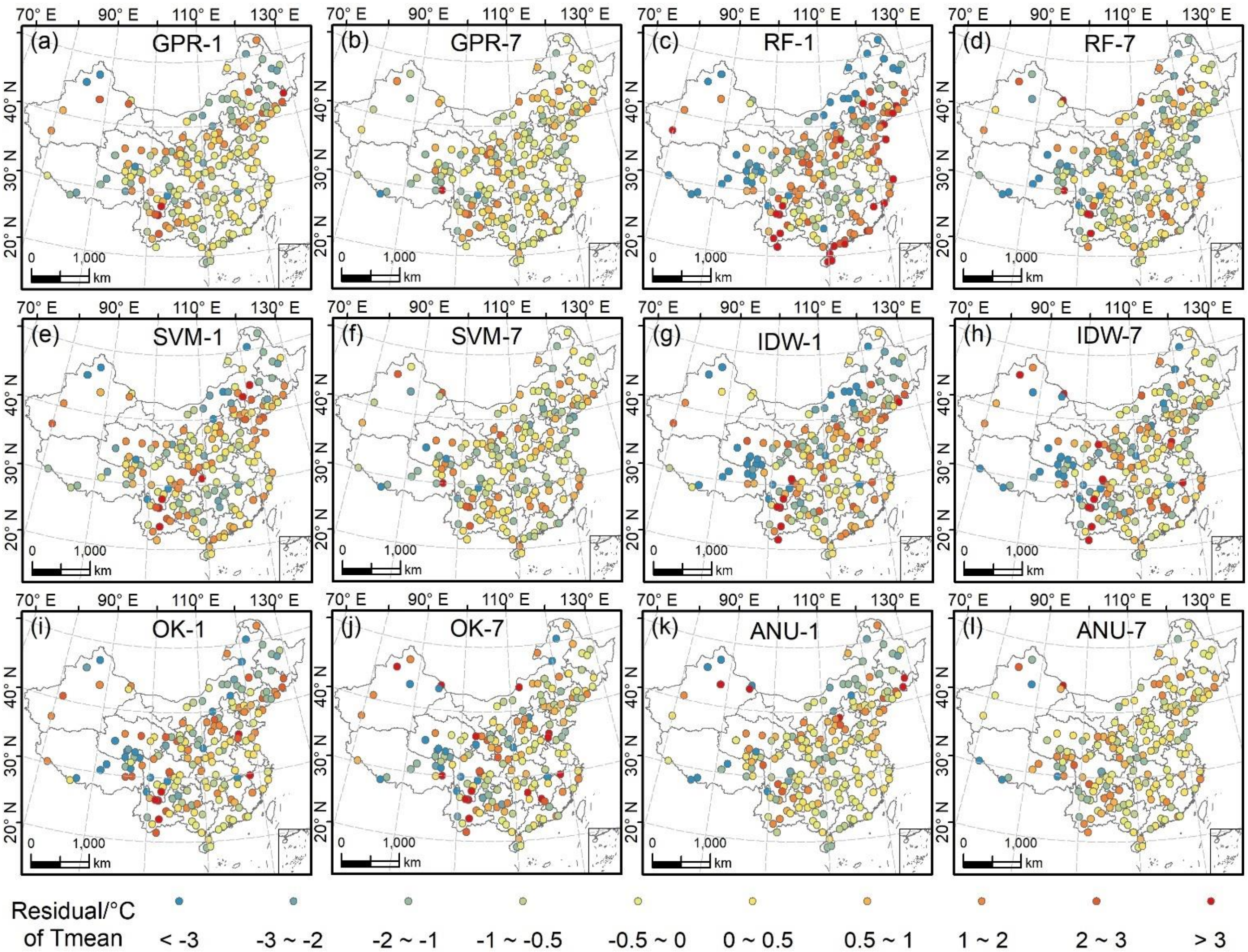


Figure 10: Comparison of the spatial distribution of the residuals between the machine learning methods and the traditional methods for Tmean in January and July 2020 (similar figures for Tmax and Tmin are provided in Figs. S50 and S51).

In summary, ANUSPLIN is an interpolation method that is better than IDW and OK in modelling air temperature over complex terrain (Plouffe et al., 2015; Newlands et al., 2011); however, the robustness of ANUSPLIN is no better than that of GPR. Moreover, ANUSPLIN is based on the principle of thin-plate splines, the skill of which can be limited in regions with high elevations and sparse observations, i.e., areas such as the Tibetan Plateau (Jobst et al., 2017). Furthermore, in our study, running ANUSPLIN was more time consuming in comparison with running the GPR model, making it difficult to generate

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

long-term monthly datasets for all 12 months over 70 years. The spatial maps of temperature generated by the six models (Figs. S52–S54) reveal that GPR obtained reasonable results for Tmean, Tmax, and Tmin. However, in the case of Tmax and Tmin, ANUSPLIN does not appear to have a rational range for Tmin (Fig. S54k and l). Therefore, in production of long-term high-resolution datasets over large land areas such as China, it is more feasible, efficient, and accurate to use the GPR model.

5.2 Comparison with other products

We used the TerraClimate, ERA5, and FLDAS temperature datasets for comparison with our dataset generated using the GPR model. Taylor diagrams were constructed to compare the accuracy between our data and that of the other products for Tmax, Tmin, and Tmean (Figure 11). For Tmax, it can be seen that the GPR-simulated air temperature best matches the observations, with the closer standard deviation to the observed variability, lower centred RMSE, and higher correlation than both ERA5 and TerraClimate. For Tmin, the standard deviation and RMSE values of ERA5 are clearly greater than those of both TerraClimate and GPR. GPR has the almost same standard deviation as the observations with the lowest RMSE and highest correlation, whereas TerraClimate has slightly less spatial variability (lower standard deviation) with a higher RMSE value and lower correlation. In the case of Tmean, GPR and FLDAS have almost the same variability (with a standard deviation close to the observed variability), while GPR has the highest correlation and lowest RMSE. Generally, the GPR-derived dataset is better in terms of Tmax, Tmin, and Tmean than the datasets obtained using the other products. The better outcome using the GPR model is characterized by the closest distance in terms of the variability compared with the observations, the lowest RMSE, and the highest correlation for all three temperature variables. The Taylor diagrams also show that the GPR model leads in terms of the reliability of the gridded temperature datasets and has greatest potential regarding spatial interpolation of air temperature.

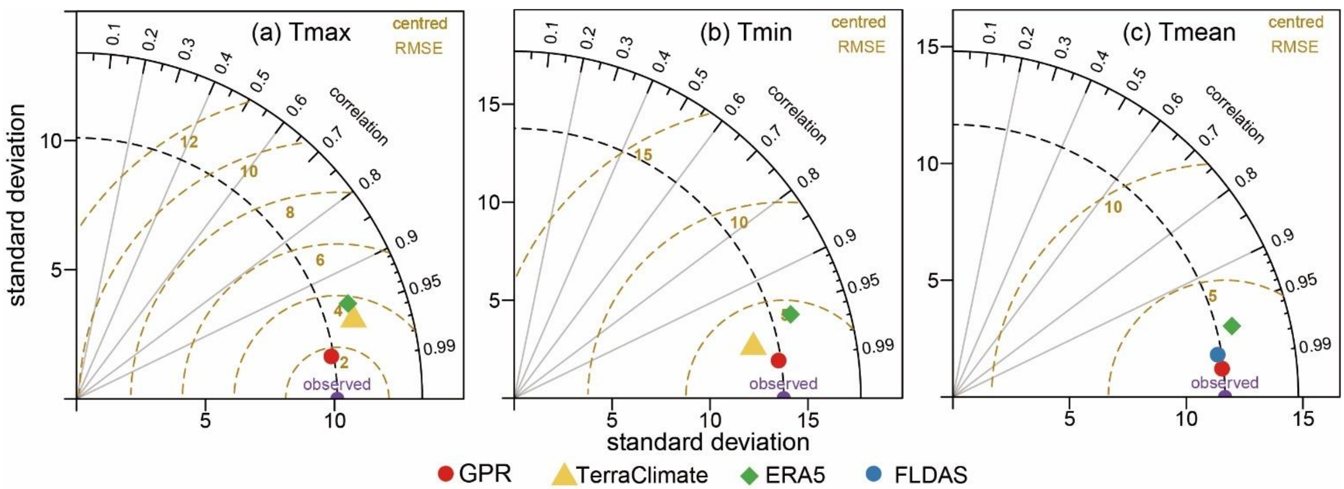


Figure 11: Taylor diagrams displaying a statistical comparison with observations between our products generated using the GPR model and the other products. Given the overlapping time of the datasets, January 1979 to December 2019 was used for comparing Tmax and Tmin and January 1982 to December 2019 was used for comparing Tmean. Comparisons for each month are presented in the Supplementary Material (Figs. S55–S57).

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

5.3 Limitations

China covers a vast territory with complex topography and diverse climate, meaning that auxiliary data such as elevation are particularly important regarding temperature interpolation (Appelhans et al., 2015; Vicente-Serrano et al., 2003). Air temperature is strongly impacted by topography, and the DEM represents a fundamental variable for interpolating air temperature in our methodology. The quality of auxiliary environmental predictors is vital and an appropriate DEM is crucial for accurate interpolation (Li and Heap, 2011b; Diodato, 2005). The DEM data adopted in this study were from the SRTM Version 4, which represents a substantial improvement on previous versions. Although the updated version is promoted as the highest quality SRTM dataset available (https://srtm.csi.cgiar.org/, last access: 15 July 2021), certain limitations remain. For example, Mukul et al. (2017) reported that the accuracy of the SRTM product in the region of the Himalayas decreases as elevation rises. Additionally, only a limited number of external studies validating the SRTM version 4 product have been reported (Tan et al., 2015), and the uncertainty of the data in our application to air temperature interpolation should be assessed in future work.

Some studies use the remote sensing data to generate the air temperature dataset, such as land surface temperature, normalized differential vegetation index (NDVI), land-use (Hooker et al., 2018; Li and Zha, 2019; Li et al., 2018). Although these variables are correlated with the air temperature, these remote sensing data are usually not available before 2000 since our goal is to generate long term data series from 1951 to 2020. Furthermore, the MODIS data are not available for each month from January 2000 to December 2020. As shown in Figure S58, the percentage of the available MODIS images are low in northeast China and southern regions. Thus, the remote sensing data are not appropriate for generating long-term temperature data in our study. Furthermore, there is inherent data uncertainty in the remote sensing data itself, such as the land use data.

It should be noted that in July 1951, there were only 38 samples available for testing and 96 samples available for training. The scarcity of meteorological stations in the early years of the 1950s represents one of the major limitations regarding the use of the machine learning methods. Generally, this study found that GPR estimates Tmean and Tmax better than Tmin. The average MAEs and RMSEs of the GPR model for Tmean are both 0.79°C, i.e., smaller than 1°C (Figure 3), whereas the average MAEs and RMSEs for Tmax and Tmin are >1°C (Tmax: average MAE = 1.20, average RMSE = 1.70; Tmin: average MAE = 1.41, average RMSE = 1.92) (Figs. S4 and S5). Therefore, the GPR model requires further improvement regarding interpolation of Tmax and Tmin.

6 Data availability

The GPRChinaTemp1km dataset includes monthly maximum air temperature, minimum air temperature, and mean air temperature at 1 km spatial resolution over China from January 1951 to December 2020. The datasets are publicly available in GeoTIFF format on Zenodo at https://doi.org/10.5281/zenodo.5112122 (He et al., 2021a) for monthly maximum air temperature, at https://doi.org/10.5281/zenodo.5111989 (He et al., 2021b) for monthly mean air temperature, and at https://doi.org/10.5281/zenodo.5112232 (He et al., 2021c) for monthly minimum air temperature. The unit of the data is °C.

7 Conclusions

A long-term, high-resolution, current, and spatially continuous dataset of air temperature over China is fundamental for understanding climatic dynamics and conducting related scientific research. We used meteorological station data available from January 1951 to December 2020 throughout China as the dependent variable, and longitude, latitude, and elevation were considered as independent variables for interpolation. We used three machine learning models (i.e., RF, SVM, and GPR) to investigate the potential of machine learning techniques regarding interpolation of air temperature over China. Results showed that GPR performed best, followed by SVM and RF. The machine learning models were also compared with conventional interpolation methods (i.e., IDW, OK, and ANUSPLIN), and the results showed that GPR was generally superior power for interpolating Tmax, Tmin, and Tmean for each month over China. Comparison of the GPR-derived results with existing products (i.e., TerraClimate, FLDAS, and ERA5) revealed that GPR outperformed the three products with regard to Tmax, Tmin, and Tmean. We constructed a new 1 km resolution monthly maximum, minimum, and mean air temperature dataset (named GPRChinaTemp1km) for China from 1951 to 2020 using the advanced GPR machine learning method. Most regions of China display significant increases for Tmean and Tmin in each month, while the trends of significant increase, non-significant increase, and non-significant decrease are prominent for Tmax. More profound analysis can be conducted based on our temperature datasets, which could help further understanding regarding global warming and climate change.

Author contributions

QH and MW designed the research and developed the methodology; QH wrote the manuscript; all other authors reviewed and revised the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We thank the China Meteorological Data Service Centre for providing the required temperature data of the meteorological stations. We also gratefully acknowledge other institutions that have contributed to the study. This work was supported by the Research on Key Technologies of Multi-level Accurate Rescue for Major Natural Disasters (grant no. 2017YFC1502902).

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C.: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015, 5, 170191, https://doi.org/10.1038/sdata.2017.191, 2018.

Alizamir, M., Kisi, O., Ahmed, A. N., Mert, C., Fai, C. M., Kim, S., Kim, N. W., and El-Shafie, A.: Advanced machine learning model for better prediction accuracy of soil temperature at different depths, PLOS ONE, 15, e0231055, https://doi.org/10.1371/journal.pone.0231055, 2020.

Appelhans, T., Mwangomo, E., Hardy, D. R., Hemp, A., and Nauss, T.: Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania, Spatial Statistics, 14, 91–113, https://doi.org/10.1016/j.spasta.2015.05.008, 2015.

Benavides, R., Montes, F., Rubio, A., and Osoro, K.: Geostatistical modelling of air temperature in a mountainous region of Northern Spain, Agricultural and Forest Meteorology, 146, 173–188, https://doi.org/10.1016/j.agrformet.2007.05.014, 2007.

Berger, J. O., De Oliveira, V., and Sansó, B.: Objective Bayesian analysis of spatially correlated data, Journal of the American Statistical Association, 96, 1361–1374, 2001.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Cai, B. and Yu, R.: Advance and evaluation in the long time series vegetation trends research based on remote sensing, J. Remote Sens, 13, 1170–1186, 2009.

Calandra, R., Peters, J., Rasmussen, C. E., and Deisenroth, M. P.: Manifold Gaussian Processes for regression, in: 2016 International Joint Conference on Neural Networks (IJCNN), 2016 International Joint Conference on Neural Networks (IJCNN), 146 citations (Semantic Scholar/DOI) [2021-07-06], 3338–3345, https://doi.org/10.1109/IJCNN.2016.7727626, 2016.

Chen, R., Yin, P., Wang, L., Liu, C., Niu, Y., Wang, W., Jiang, Y., Liu, Y., Liu, J., Qi, J., You, J., Kan, H., and Zhou, M.: Association between ambient temperature and mortality risk and burden: time series study in 272 main Chinese cities, BMJ, 363, k4306, https://doi.org/10.1136/bmj.k4306, 2018.

Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, Copernicus Climate Change Service Climate Data Store (CDS), 2017.

Da Silva, R. M., Santos, C. A., Moreira, M., Corte-Real, J., Silva, V. C., and Medeiros, I. C.: Rainfall and river flow trends using Mann–Kendall and Sen's slope estimator statistical tests in the Cobres River basin, Natural Hazards, 77, 1205–1221, 2015.

Dawood, M.: Spatio-statistical analysis of temperature fluctuation using Mann–Kendall and Sen's slope approach, Climate dynamics, 48, 783–797, 2017.

Diodato, N.: The influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain, 25, 351–363, https://doi.org/10.1002/joc.1131, 2005.

Duhan, D., Pandey, A., Gahalaut, K. P. S., and Pandey, R. P.: Spatial and temporal variability in maximum, minimum and mean air temperatures at Madhya Pradesh in central India, Comptes Rendus Geoscience, 345, 3–21, https://doi.org/10.1016/j.crte.2012.10.016, 2013.

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

Open Access Earth System Science Data Discussions

Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., and Xiang, Y.: Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China, Agricultural and Forest Meteorology, 263, 225–241, https://doi.org/10.1016/j.agrformet.2018.08.019, 2018.

Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, Int. J. Climatol, 37, 4302–4315, https://doi.org/10.1002/joc.5086, 2017.

Fu, P. and Weng, Q.: Variability in annual temperature cycle in the urban areas of the United States as revealed by MODIS imagery, ISPRS Journal of Photogrammetry and Remote Sensing, 146, 65–73, https://doi.org/10.1016/j.isprsjprs.2018.09.003, 2018.

Gao, L., Wei, J., Wang, L., Bernhardt, M., Schulz, K., and Chen, X.: A high-resolution air temperature data set for the Chinese Tian Shan in 1979–2016, Earth Syst. Sci. Data, 10, 2097–2114, https://doi.org/10.5194/essd-10-2097-2018, 2018.

Ghorbani, M. A., Shamshirband, S., Zare Haghi, D., Azani, A., Bonakdari, H., and Ebtehaj, I.: Application of firefly algorithm-based support vector machines for prediction of field capacity and permanent wilting point, Soil and Tillage Research, 172, 32–38, https://doi.org/10.1016/j.still.2017.04.009, 2017.

Gocic, M. and Trajkovic, S.: Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia, Global and Planetary Change, 100, 172–182, https://doi.org/10.1016/j.gloplacha.2012.10.014, 2013.

Graf, R., Zhu, S., and Sivakumar, B.: Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach, Journal of Hydrology, 578, 124115, https://doi.org/10.1016/j.jhydrol.2019.124115, 2019.

Grbić, R., Kurtagić, D., and Slišković, D.: Stream water temperature prediction based on Gaussian process regression, Expert Systems with Applications, 40, 7407–7414, https://doi.org/10.1016/j.eswa.2013.06.077, 2013.

Guo, B., Zhang, J., Meng, X., Xu, T., and Song, Y.: Long-term spatio-temporal precipitation variations in China with precipitation surface interpolated by ANUSPLIN, Sci Rep, 10, 81, https://doi.org/10.1038/s41598-019-57078-3, 2020.

Hadi, S. J. and Tombul, M.: Comparison of Spatial Interpolation Methods of Precipitation and Temperature Using Multiple Integration Periods, J Indian Soc Remote Sens, 46, 1187–1199, https://doi.org/10.1007/s12524-018-0783-1, 2018.

He, Q., Wang, M., Liu, K., Li, K., and Jiang, Z.: GPRChinaTemp1km: 1 km monthly maximum air temperature for China from January 1951 to December 2020, https://doi.org/10.5281/zenodo.5112122, 2021a.

He, Q., Wang, M., Liu, K., Li, K., and Jiang, Z.: GPRChinaTemp1km: 1 km monthly mean air temperature for China from January 1951 to December 2020, https://doi.org/10.5281/zenodo.5111989, 2021b.

He, Q., Wang, M., Liu, K., Li, K., and Jiang, Z.: GPRChinaTemp1km: 1 km monthly minimum air temperature for China from January 1951 to December 2020, https://doi.org/10.5281/zenodo.5112232, 2021c.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., and Gräler, B.: Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables, PeerJ, 6, e5518, https://doi.org/10.7717/peerj.5518, 2018.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, Int. J. Climatol., 25, 1965–1978, https://doi.org/10.1002/joc.1276, 2005.

Ho, H. C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., and Henderson, S. B.: Mapping maximum urban air temperature on hot summer days, Remote Sensing of Environment, 154, 38–45, https://doi.org/10.1016/j.rse.2014.08.012, 2014.

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

Hooker, J., Duveiller, G., and Cescatti, A.: A global dataset of air temperature derived from satellite remote sensing and weather stations, Sci Data, 5, 180246, https://doi.org/10.1038/sdata.2018.246, 2018.

Hutchinson, M.: ANUSPLIN Version 4.3, Centre for Resource and Environmental Studies. The Australian National University, Canberra, Australia, 2004.

Hutchinson, M. F.: Interpolating mean rainfall using thin plate smoothing splines, null, 9, 385–403, https://doi.org/10.1080/02693799508902045, 1995.

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K.-M., Gerber, J. S., Reddy, V. R., and Kim, S.-H.: Random Forests for Global and Regional Crop Yield Predictions, PLOS ONE, 11, e0156571, https://doi.org/10.1371/journal.pone.0156571, 2016.

Ji, L., Senay, G. B., and Verdin, J. P.: Evaluation of the Global Land Data Assimilation System (GLDAS) Air Temperature Data Products, 16, 2463–2480, https://doi.org/10.1175/JHM-D-14-0230.1, 2015.

Jiang, W., Yuan, L., Wang, W., Cao, R., Zhang, Y., and Shen, W.: Spatio-temporal analysis of vegetation variation in the Yellow River Basin, Ecological Indicators, 51, 117–126, https://doi.org/10.1016/j.ecolind.2014.07.031, 2015.

Jobst, A. M., Kingston, D. G., Cullen, N. J., and Sirguey, P.: Combining thin-plate spline interpolation with a lapse rate model to produce daily air temperature estimates in a data-sparse alpine catchment: COMBINING THIN-PLATE SPLINE INTERPOLATION WITH A LAPSE RATE MODEL, Int. J. Climatol., 37, 214–229, https://doi.org/10.1002/joc.4699, 2017.

Karbasi, M.: Forecasting of Multi-Step Ahead Reference Evapotranspiration Using Wavelet- Gaussian Process Regression Model, Water Resour Manage, 32, 1035–1052, https://doi.org/10.1007/s11269-017-1853-9, 2018.

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., and Shearer, S.: Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield, Computers and Electronics in Agriculture, 153, 213–225, https://doi.org/10.1016/j.compag.2018.07.016, 2018.

Kisi, O., Sanikhani, H., and Cobaner, M.: Soil temperature modeling at different depths using neuro-fuzzy, neural network, and genetic programming techniques, Theor Appl Climatol, 129, 833–848, https://doi.org/10.1007/s00704-016-1810-1, 2017.

Li, J. and Heap, A. D.: A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors, Ecological Informatics, 6, 228–241, https://doi.org/10.1016/j.ecoinf.2010.12.003, 2011a.

Li, J. and Heap, A. D.: A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors, Ecological Informatics, 6, 228–241, https://doi.org/10.1016/j.ecoinf.2010.12.003, 2011b.

Li, J. and Heap, A. D.: Spatial interpolation methods applied in the environmental sciences: A review, Environmental Modelling & Software, 53, 173–189, https://doi.org/10.1016/j.envsoft.2013.12.008, 2014.

Li, J., Heap, A. D., Potter, A., and Daniell, J. J.: Application of machine learning methods to spatial interpolation of environmental variables, Environmental Modelling & Software, 26, 1647–1659, https://doi.org/10.1016/j.envsoft.2011.07.004, 2011.

Li, L. and Zha, Y.: Estimating monthly average temperature by remote sensing in China, Advances in Space Research, 63, 2345–2357, https://doi.org/10.1016/j.asr.2018.12.039, 2019.

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

Li, X., Zhou, Y., Asrar, G. R., and Zhu, Z.: Developing a 1 km resolution daily air temperature dataset for urban and surrounding areas in the conterminous United States, Remote Sensing of Environment, 215, 74–84, https://doi.org/10.1016/j.rse.2018.05.034, 2018.

Li, Y., Zhao, C., Zhang, T., Wang, W., Duan, H., Liu, Y., Ren, Y., and Pu, Z.: Impacts of Land-Use Data on the Simulation of Surface Air Temperature in Northwest China, J Meteorol Res, 32, 896–908, https://doi.org/10.1007/s13351-018-7151-5, 2018.

Li, Z., Zheng, F., Liu, W., and Flanagan, D. C.: Spatial distribution and temporal trends of extreme temperature and precipitation events on the Loess Plateau of China during 1961–2007, Quaternary International, 226, 92–100, https://doi.org/10.1016/j.quaint.2010.03.003, 2010.

Li, Z., Zheng, F.-L., and Liu, W.-Z.: Spatiotemporal characteristics of reference evapotranspiration during 1961–2009 and its projected changes during 2011–2099 on the Loess Plateau of China, Agricultural and Forest Meteorology, 154–155, 147–155, https://doi.org/10.1016/j.agrformet.2011.10.019, 2012.

McNally, A., Arsenault, K., Kumar, S., Shukla, S., Peterson, P., Wang, S., Funk, C., Peters-Lidard, C. D., and Verdin, J. P.: A land data assimilation system for sub-Saharan Africa food and water security applications, Sci Data, 4, 170012, https://doi.org/10.1038/sdata.2017.12, 2017.

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., and Abderrahmane, B.: Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance, Earth-Science Reviews, 207, 103225, https://doi.org/10.1016/j.earscirev.2020.103225, 2020.

Meyer, H., Katurji, M., Appelhans, T., Müller, M. U., Nauss, T., Roudier, P., and Zawar-Reza, P.: Mapping Daily Air Temperature for Antarctica Based on MODIS LST, 8, 732, https://doi.org/10.3390/rs8090732, 2016.

Mukul, M., Srivastava, V., Jade, S., and Mukul, M.: Uncertainties in the Shuttle Radar Topography Mission (SRTM) Heights: Insights from the Indian Himalaya and Peninsula, Sci Rep, 7, 41672, https://doi.org/10.1038/srep41672, 2017.

Mutanga, O., Adam, E., and Cho, M. A.: High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm, International Journal of Applied Earth Observation and Geoinformation, 18, 399–406, https://doi.org/10.1016/j.jag.2012.03.012, 2012.

New, M., Hulme, M., and Jones, P.: Representing Twentieth-Century Space–Time Climate Variability. Part II: Development of 1901–96 Monthly Grids of Terrestrial Surface Climate, 13, 2217–2238, https://doi.org/10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2, 2000.

Newlands, N. K., Davidson, A., Howard, A., and Hill, H.: Validation and inter-comparison of three methodologies for interpolating daily precipitation and temperature across Canada, 22, 205–223, https://doi.org/10.1002/env.1044, 2011.

Ozelkan, E., Bagis, S., Ozelkan, E. C., Ustundag, B. B., Yucel, M., and Ormeci, C.: Spatial interpolation of climatic variables using land surface temperature and modified inverse distance weighting, 36, 1000–1025, https://doi.org/10.1080/01431161.2015.1007248, 2015.

Pathak, T. B., Maskey, M. L., Dahlberg, J. A., Kearns, F., Bali, K. M., and Zaccaria, D.: Climate Change Trends and Impacts on California Agriculture: A Detailed Review, 8, 25, https://doi.org/10.3390/agronomy8030025, 2018.

Peng, S., Zhao, C., Wang, X., Xu, Z., Liu, X., Hao, H., and Yang, S.: Mapping daily temperature and precipitation in the Qilian Mountains of northwest China, J. Mt. Sci., 11, 896–905, https://doi.org/10.1007/s11629-013-2613-9, 2014.

Peng, S., Ding, Y., Liu, W., and Li, Z.: 1 km monthly temperature and precipitation dataset for China from 1901 to 2017, Earth Syst. Sci. Data, 11, 1931–1946, https://doi.org/10.5194/essd-11-1931-2019, 2019.

Plouffe, C. C. F., Robertson, C., and Chandrapala, L.: Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: A case study of Sri Lanka, Environmental Modelling & Software, 67, 57–71, https://doi.org/10.1016/j.envsoft.2015.01.011, 2015.

Rasmussen, C. E.: Evaluation of Gaussian processes and other methods for non-linear regression, 1997.

Rasmussen, C. E.: Gaussian Processes in Machine Learning, in: Advanced Lectures on Machine Learning, vol. 3176, edited by: Bousquet, O., von Luxburg, U., and Rätsch, G., Springer Berlin Heidelberg, Berlin, Heidelberg, 63–71, https://doi.org/10.1007/978-3-540-28650-9_4, 2004.

dos Santos, R. S.: Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data, International Journal of Applied Earth Observation and Geoinformation, 88, 102066, https://doi.org/10.1016/j.jag.2020.102066, 2020.

Sayemuzzaman, M. and Jha, M. K.: Seasonal and annual precipitation time series trend analysis in North Carolina, United States, Atmospheric Research, 137, 183–194, https://doi.org/10.1016/j.atmosres.2013.10.012, 2014.

Schulz, E., Speekenbrink, M., and Krause, A.: A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions, Journal of Mathematical Psychology, 85, 1–16, https://doi.org/10.1016/j.jmp.2018.03.001, 2018.

Sekulić, A., Kilibarda, M., Protić, D., and Bajat, B.: A high-resolution daily gridded meteorological dataset for Serbia made by Random Forest Spatial Interpolation, 8, 123, https://doi.org/10.1038/s41597-021-00901-2, 2021.

Sen, P. K.: Estimates of the regression coefficient based on Kendall's tau, Journal of the American statistical association, 63, 1379–1389, 1968.

Shao, J., Li, Y., and Ni, J.: The characteristics of temperature variability with terrain, latitude and longitude in Sichuan-Chongqing Region, J. Geogr. Sci., 22, 223–244, https://doi.org/10.1007/s11442-012-0923-4, 2012.

Shifteh Some'e, B., Ezani, A., and Tabari, H.: Spatiotemporal trends and change point of precipitation in Iran, Atmospheric Research, 113, 1–12, https://doi.org/10.1016/j.atmosres.2012.04.016, 2012.

Shrestha, N. K. and Shukla, S.: Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment, Agricultural and Forest Meteorology, 200, 172–184, https://doi.org/10.1016/j.agrformet.2014.09.025, 2015.

Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E., and Knutti, R.: Climate change now detectable from any single day of weather at global scale, Nat. Clim. Chang., 10, 35–41, https://doi.org/10.1038/s41558-019-0666-7, 2020.

Stahl, K., Moore, R. D., Floyer, J. A., Asplin, M. G., and McKendry, I. G.: Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density, Agricultural and Forest Meteorology, 139, 224–236, https://doi.org/10.1016/j.agrformet.2006.07.004, 2006.

Sun, A. Y., Wang, D., and Xu, X.: Monthly streamflow forecasting using Gaussian Process Regression, Journal of Hydrology, 511, 72–81, https://doi.org/10.1016/j.jhydrol.2014.01.023, 2014.

https://doi.org/10.5194/essd-2021-442
Preprint. Discussion started: 9 December 2021

Open Access Earth System Science Data Discussions

Tan, M. L., Ficklin, D. L., Dixon, B., Ibrahim, A. L., Yusop, Z., and Chaplot, V.: Impacts of DEM resolution, source, and resampling technique on SWAT-simulated streamflow, Applied Geography, 63, 357–368, https://doi.org/10.1016/j.apgeog.2015.07.014, 2015.

Tang, G., Clark, M. P., Newman, A. J., Wood, A. W., Papalexiou, S. M., Vionnet, V., and Whitfield, P. H.: SCDNA: a serially complete precipitation and temperature dataset for North America from 1979 to 2018, Earth Syst. Sci. Data, 12, 2381–2409, https://doi.org/10.5194/essd-12-2381-2020, 2020.

Theil, H.: A rank-invariant method of linear and polynomial regression analysis, Indagationes mathematicae, 12, 173, 1950.

Thornton, P. E., Running, S. W., and White, M. A.: Generating surfaces of daily meteorological variables over large regions of complex terrain, Journal of Hydrology, 190, 214–251, https://doi.org/10.1016/S0022-1694(96)03128-9, 1997.

Vapnik, V.: The nature of statistical learning theory, Springer science & business media, 2013.

Vicente-Serrano, S. M., Saz-Sánchez, M. A., and Cuadrat, J. M.: Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature, 24, 161–180, https://doi.org/10.3354/cr024161, 2003.

Wang, M., He, G., Zhang, Z., Wang, G., Zhang, Z., Cao, X., Wu, Z., and Liu, X.: Comparison of Spatial Interpolation and Regression Analysis Models for an Estimation of Monthly Near Surface Air Temperature in China, Remote Sensing, 9, 1278, https://doi.org/10.3390/rs9121278, 2017.

Xu, T. and Hutchinson, M. F.: New developments and applications in the ANUCLIM spatial climatic and bioclimatic modelling package, Environmental Modelling & Software, 40, 267–279, https://doi.org/10.1016/j.envsoft.2012.10.003, 2013.

Xu, Y., Knudby, A., Shen, Y., and Liu, Y.: Mapping Monthly Air Temperature in the Tibetan Plateau From MODIS Data Based on Machine Learning Methods, 11, 345–354, https://doi.org/10.1109/JSTARS.2017.2787191, 2018.

Yoo, C., Im, J., Park, S., and Quackenbush, L. J.: Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data, ISPRS Journal of Photogrammetry and Remote Sensing, 137, 149–162, https://doi.org/10.1016/j.isprsjprs.2018.01.018, 2018.

Yu, W., Liu, Y., Ma, Z., and Bi, J.: Improving satellite-based PM 2.5 estimates in China using Gaussian processes modeling in a Bayesian hierarchical setting, Sci Rep, 7, 7048, https://doi.org/10.1038/s41598-017-07478-0, 2017.

Zeng, L., Hu, Y., Wang, R., Zhang, X., Peng, G., Huang, Z., Zhou, G., Xiang, D., Meng, R., Wu, W., and Hu, S.: 8-Day and Daily Maximum and Minimum Air Temperature Estimation via Machine Learning Method on a Climate Zone to Global Scale, 13, 2355, https://doi.org/10.3390/rs13122355, 2021.

Zhang, H., Zhang, F., Ye, M., Che, T., and Zhang, G.: Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data, 121, 11,425-11,441, https://doi.org/10.1002/2016JD025154, 2016.

Zhang, Z., Ye, L., Qin, H., Liu, Y., Wang, C., Yu, X., Yin, X., and Li, J.: Wind speed prediction method using Shared Weight Long Short-Term Memory Network and Gaussian Process Regression, Applied Energy, 247, 270–284, https://doi.org/10.1016/j.apenergy.2019.04.047, 2019.

Zhu, S., Nyarko, E. K., and Hadzima-Nyarko, M.: Modelling daily water temperature from air temperature for the Missouri River, PeerJ, 6, e4894, https://doi.org/10.7717/peerj.4894, 2018.