

RESEARCH ARTICLE

10.1002/2016JD025154

Key Points:

- Dynamically integrating the MODIS LST data from four pass times can greatly reduce the cloud blockage in the final air temperature product
- The differences in model performance are related to the combinations of LST terms and their data quality
- Cubist regression and random forests are the best two statistical models for daily mean air temperature estimation using MODIS LST

Supporting Information:

- Supporting Information S1
- Figure S1
- Figure S2
- Figure S3
- Figure S4
- Figure S5
- Figure S6
- Figure S7
- Figure S8
- Figure S9

Correspondence to:

F. Zhang,
zhangfan@itpcas.ac.cn

Citation:

Zhang, H., F. Zhang, M. Ye, T. Che, and G. Zhang (2016), Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data, *J. Geophys. Res. Atmos.*, 121, 11,425–11,441, doi:10.1002/2016JD025154.

Received 28 MAR 2016

Accepted 16 SEP 2016

Accepted article online 19 SEP 2016

Published online 8 OCT 2016

Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data

Hongbo Zhang^{1,2}, Fan Zhang^{1,2}, Ming Ye³, Tao Che^{2,4}, and Guoqing Zhang^{1,2}

¹Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China, ²CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing, China, ³Department of Scientific Computing, Florida State University, Tallahassee, Florida, USA, ⁴Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Beijing, China

Abstract Recently, remotely sensed land surface temperature (LST) data have been used to estimate air temperatures because of the sparseness of station measurements in remote mountainous areas. Due to the availability and accuracy of Moderate Resolution Imaging Spectroradiometer (MODIS) LST data, the use of a single term or a fixed combination of terms (e.g., Terra/Aqua night and Terra/Aqua day), as used in previous estimation methods, provides only limited practical application. Furthermore, the estimation accuracy may be affected by different combinations and variable data quality among the MODIS LST terms and models. This study presents a method that dynamically integrates the available LST terms to estimate the daily mean air temperature and simultaneously considers model selection, data quality, and estimation accuracy. The results indicate that the differences in model performance are related to the combinations of LST terms and their data quality. The spatially averaged cloud cover of ~14% for the developed product between 2003 and 2010 is much lower than the 35–54% for single LST terms. The average cross-validation root-mean-square difference values are approximately 2°C. This study identifies the best LST combinations and statistical models and provides an efficient method for daily air temperature estimation with low cloud blockage over the Tibetan Plateau (TP). The developed data set and the method proposed in this study can help alleviate the problem of sparse air temperature data over the TP.

1. Introduction

The Tibetan Plateau (TP) is the highest large plateau in the world, with an average elevation over 4000 m and an area of approximately 3×10^6 km² above an elevation of 2500 m (Figure 1) [G Zhang *et al.*, 2013]. Many studies have shown that the TP plays a crucial role in regional and global climate changes due to its extremely complex terrain and physical surface properties [Flohn, 1968; Rohrer *et al.*, 2013; Yasunari *et al.*, 1991]. Air temperature [Liu and Chen, 2000; Thompson *et al.*, 2000; Zhou and Yu, 2006] serves as an important factor for describing the characteristics of terrestrial environmental conditions. In particular, it is a key input for various hydrological, ecological, and environmental models [Daly, 2006; Stahl *et al.*, 2006]. For example, air temperature not only controls the snowmelt processes in mountainous hydrological modeling [Hock, 2003; F Zhang *et al.*, 2015] but also significantly impacts numerous biogeochemical processes [Ninyerola *et al.*, 2007; Spadavecchia and Williams, 2009]. Successfully implementing these models depends on reliable and accurate air temperature data [Daly, 2006; Thornton *et al.*, 1997]. However, air temperature measurements from meteorological stations are usually scarce in remote mountainous areas such as the TP [Green and Hay, 2002; Prihodko and Goward, 1997; Yao and Zhang, 2013; Zakšek and Schroedter-Homscheidt, 2009].

Remotely sensed land surface temperatures (LSTs) have been widely used to estimate daily air temperatures based on the strong correlation between the LST and air temperature resulting from the intense heat exchange and interaction between the land and the atmosphere [Benali *et al.*, 2012; Janatian *et al.*, 2016; Oyler *et al.*, 2016; W Zhang *et al.*, 2011]. Many methods have been developed to estimate air temperature using remotely sensed LST data from various sensors, including Landsat-ETM+ [Wloczyk *et al.*, 2011], the Advanced Very High Resolution Radiometer [Prince *et al.*, 1998], the Spinning Enhanced Visible and Infrared Imager on Meteosat Second Generation [Zakšek and Schroedter-Homscheidt, 2009], and the Moderate Resolution Imaging Spectroradiometer (MODIS) [Benali *et al.*, 2012]. Among the different types of remote sensing data, the most popular and easily obtained data set is the LST product supplied by the MODIS instrument aboard the Terra and Aqua satellites.

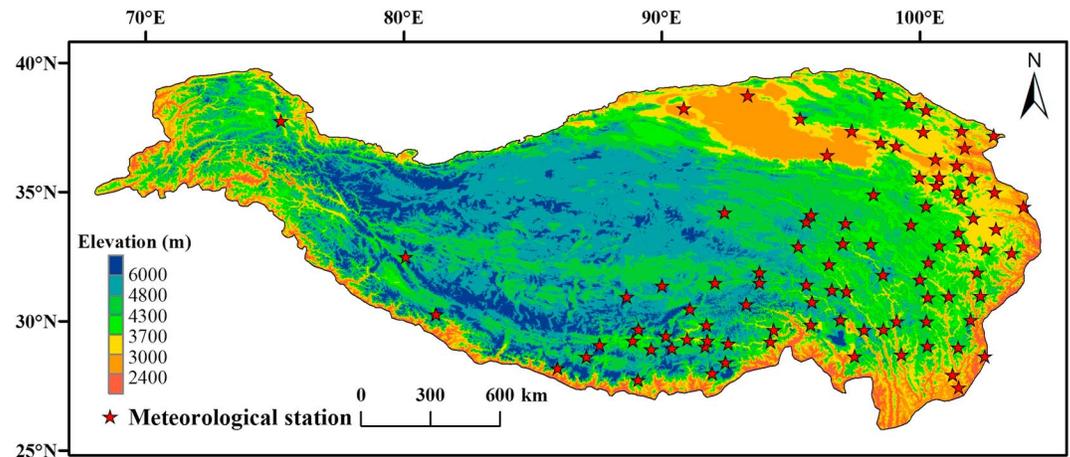


Figure 1. Study region and the locations of the meteorological stations.

However, the cloud blockage in the MODIS LST data set limits its application. When a satellite pixel is covered by clouds, the corresponding MODIS temperature observation for the pixel does not represent the land surface and instead represents the cloud top. As a result, cloud blockage seriously degrades the quantity of available high-quality data in the MODIS LST terms. A single LST term (e.g., Terra Night, Terra Day, Aqua Night, and Aqua Day) for air temperature estimation is subject to serious data gaps caused by high degrees of cloud blockage. Thus, developing a method that can dynamically utilize different combinations of LST terms that are available at different times is critically important to reducing cloud blockage.

In addition, the impact of MODIS LST data quality on the air temperature estimates lacks sufficient attention in the current statistical methods. In addition to the LST value, a MODIS LST product also includes information on data quality stored in built-in quality control (QC) flags, which feature four integer values from 0 to 3. The QC flag of 0 represents the best data quality with an average error of less than 1 K (Kelvin). The average errors for the other QC flags of 1, 2, and 3 are greater than 1 K. Among the MODIS LST data over the TP, the number of data with QC flags of 3 is less than 1%, but the data with QC flags of 1 and 2 comprise 30–50% of the entire data set. Given the large proportion of less accurate data (with QC flags of 1 and 2), simply discarding the data when using statistical methods is not justifiable. Therefore, it is necessary to investigate the negative effects of the less accurate data on the performance of the statistical models.

To improve the estimation accuracy, the multiple linear regression model [Benali *et al.*, 2012; Fu *et al.*, 2011; Good, 2015; Vancutsem *et al.*, 2010] and more complex models have been developed, including neural networks [Zhao *et al.*, 2007], the M5 model tree [Emamifar *et al.*, 2013], and random forests [Y Xu *et al.*, 2014]. Currently, these models have not been compared in the same region, and the best model for air temperature estimation remains unknown.

Due to the scarcity of field observations and the aforementioned unsolved problems associated with air temperature estimations, no accurate data set of daily mean air temperature over the TP is currently available. Such a data set would be of great use for small- and medium-sized watershed modeling because only a few or even no meteorological stations are present in many mountainous watersheds across the TP and because the existing air temperature products feature coarse spatial resolutions (≥ 10 km) that cannot satisfactorily describe the spatial heterogeneity. In this study, we developed a method for air temperature estimation that involves dynamically integrating the available LST terms. Among the various types of air temperature measurements (e.g., maximum, minimum, and average temperatures at different time resolutions such as daily, weekly, and monthly), this study is focused on the daily mean air temperature, which is popularly used for daily-scale models [F Zhang *et al.*, 2015].

2. Data and Methodology

2.1. Meteorological Data

Daily mean air temperatures from 95 meteorological stations (Figure 1) were obtained from the CMA (China Meteorological Administration, <http://cdc.nmc.cn>). The data were measured at 2 m above ground between

2003 and 2010. Most stations are located in the eastern TP. The vast western and central regions have almost no stations. Furthermore, approximately 71% of these stations are at relatively low elevations (<4000 m), and no stations are present at an elevation greater than 5000 m. The sparseness of the stations highlights the necessity of air temperature estimates for areas with few measurements.

2.2. Land Surface Temperature

The primary variable, LST, is from “The MODIS Land Surface Temperature and Emissivity (LST/E) products V005” with product labels of “MOD11A1” and “MYD11A1” for Terra and Aqua, respectively. The “V005” version achieved some refinements compared with previous versions and proved to be accurate in most of clear-sky cases with the errors less than 1 K [Wan, 2008]. The pass times of Terra occur at the approximate local solar times of 10:30 A.M. and 10:30 P.M.; the pass times of Aqua occur at the approximate local solar times of 1:30 P.M. and 1:30 A.M. [Wang et al., 2006]. Thus, within 1 day, MODIS can provide four LST terms. The four LST observations (i.e., “terms”)—Terra Night, Terra Day, Aqua Night, and Aqua Day—are all used in this study.

As previously mentioned, the MODIS LST product also provides information on cloud blockage and data quality in addition to the LST values. If the LST data in a specific pixel are not available because of cloud blockage or other reasons, the data are labeled as “not produced” and can be identified in the LST product. The data quality information is stored in built-in quality control (QC) flags, which have four values from 0 to 3. The average errors associated with the “0,” “1,” “2,” and “3” flags are <1 K, 1–2 K, 2–3 K, and > 3 K, respectively. In practice, the proportion of good data was 50–70% and that of bad data was 30–50% for the four LST terms. The LSTs with average errors of >3 K (QC flag = 3) were removed in this study.

2.3. Auxiliary Variables

Longitude, latitude, Julian day, solar zenith, normalized difference vegetation index (NDVI), and elevation have commonly been used in previous studies as auxiliary variables [Benali et al., 2012; Cristobal et al., 2008; Florio et al., 2004; Jang et al., 2004; Y Xu et al., 2014]. Thus, these six variables were selected as model inputs in our study in addition to the MODIS LST data. Longitude and latitude data were supplied from the meteorological stations. The solar zenith and required spectral information for computing the NDVI were all derived from the MODIS Surface Reflectance products labeled “MOD09GA” and “MYD09GA” for Terra and Aqua, respectively. The daily NDVI was computed in the same manner as in Zhu et al. [2013] and Y Xu et al. [2014] by equation (1):

$$\text{NDVI} = (B_{\text{NIR}} - B_{\text{RED}}) / (B_{\text{NIR}} + B_{\text{RED}}), \quad (1)$$

where B_{NIR} and B_{RED} represent the band 2 (near infrared) and band 1 (red) of the MODIS Surface Reflectance product, respectively. The NDVIs produced at 500 m resolution are converted to 1000 m using nearest neighbor resampling methods to be consistent with MODIS LST products. The elevation information was derived from the Global 30 Arc Second Elevation (GTOPO30) data set, with a spatial resolution similar to that of the MODIS LST data.

2.4. A Method of Integrating the Four MODIS LST Terms

We propose a method that can dynamically make full use of the available LST terms, namely, the LST data among the four passing times to produce air temperature data with minimal cloud blockage.

As shown in Figure 2, a key part of our method is the model ranking table. To generate it, four steps are generally needed:

1. First, we extracted all of the available (i.e., not covered with clouds) LST terms from the pixels where the observation stations were located as “samples.” These data were further classified into two categories based on the built-in quality control flags in the MODIS LST product. The LSTs with average errors of < 1 K (QC flag = 0) were considered to be of good quality, and the LSTs with average errors of > 1 K (QC flag = 1–2) were considered to be of poor quality. To keep things simple, all of the samples in our research are divided into two categories: When all four of the LSTs were of good quality, the combination was labeled an “S1” situation; when at least one LST was of poor quality, the combination was labeled an “S2” situation. This categorization resulted in two groups of samples, namely, “S1 samples” and “S2 samples,” with different data qualities, as shown in Figure 2.
2. Second, after conducting complete permutations and combinations, 15 combinations (i.e., $2^4 - 1$) of the four LST terms were obtained (Figure 3).

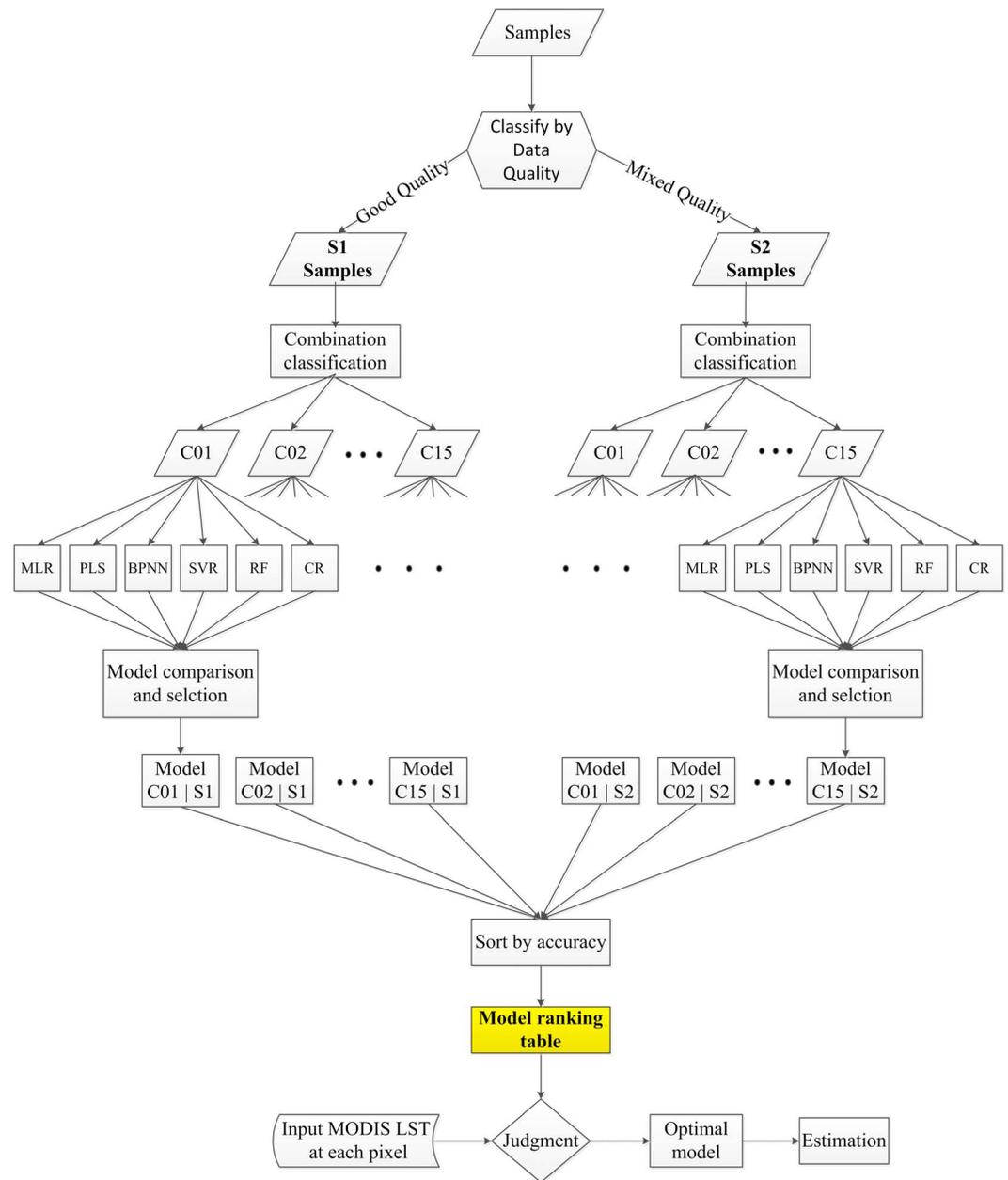


Figure 2. Flow chart of the method for integrating the MODIS LST data from four pass times.

3. Third, for each combination under each quality situation, six statistical models were evaluated including model training and validation. Then, the performances of the six models were compared and the best model with the highest accuracy was chosen for each combination. The models used in this study are described further in section 2.5.
4. Finally, after all of the best models for the 15 combinations in the two quality situations were decided, the 30 (=15 × 2) models were ranked based on their cross-validation results, generating a model ranking table in which models with lower cross-validation root-mean-square difference (RMSD) values were ranked higher.

The ranking table was then used to guide the air temperature product generation process as follows. Every day for each pixel, the data availability status and quality status were identified. Our method automatically selected the best model corresponding to the specific status using the model ranking table.

15 Combinations of MODIS LST

Combination 01: Terra Day
Combination 02: Aqua Day
Combination 03: Terra Night
Combination 04: Aqua Night
Combination 05: Terra Day, Terra Night
Combination 06: Terra Day, Aqua Night
Combination 07: Terra Day, Aqua Day
Combination 08: Terra Night, Aqua Night
Combination 09: Terra Night, Aqua Day
Combination 10: Aqua Day, Aqua Night
Combination 11: Terra Night, Aqua Day, Aqua Night
Combination 12: Terra Day, Aqua Day, Aqua Night
Combination 13: Terra Day, Terra Night, Aqua Night
Combination 14: Terra Day, Terra Night, Aqua Day
Combination 15: Terra Day, Terra Night, Aqua Day, Aqua Night

For example, for a specific pixel on a certain day, the data availability status was as follows. Only Terra day and Aqua night were available because Terra night and Aqua day were covered with clouds (Figure 4a). The quality statuses showed that the Terra day was of good quality and the Aqua night was of poor quality (Figure 4a). Such data conditions result in three possible combinations according to Figure 3 and the quality situations (Figure 4b). In this case, the potential models include Combinations 01 (of the S1 situation), 04 (of the S2 situation), and 06 (of the S2 situation) (Figure 4b). The models for these combinations are sorted using the model ranking table, and the

Figure 3. The compositions of 15 combinations of MODIS LST data.

best one is selected (Figure 4c) as the final model to be used for air temperature estimation for that pixel on that day.

As shown in Figure 2, our method not only takes into consideration all combinations of the LST terms and the two data quality situations but also conducts model selection among different statistical models based on the estimation accuracy, which depends on the LST combination and data quality. These considerations and processing are deemed valuable for the following reasons:

1. Considering all of the possible combinations of LST can contribute greatly to the accuracy of air temperature estimates. In many cases where multiple LST terms are available, the solution lies in which combination

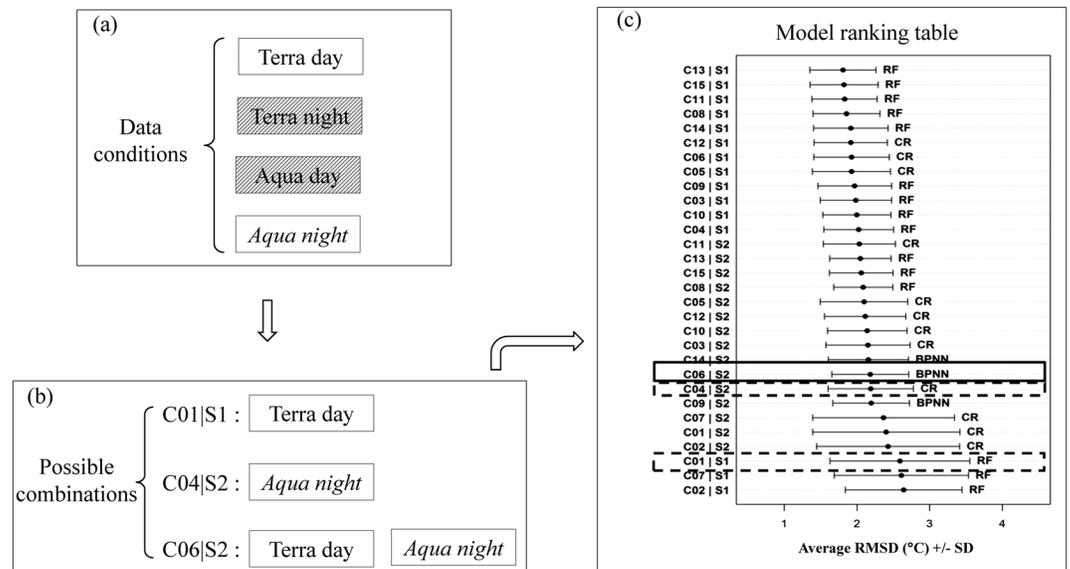


Figure 4. A sketch of the model selection process using a model ranking table. (a) First, the LST data conditions are identified (shaded means unavailable, italics mean poor quality). (b) Second, all the possible combinations are obtained according to Figure 3 and the quality situations. (c) Last, all the corresponding models are sorted based on the model ranking table and the highest one is selected. A dashed-line box indicates a potential model with lower accuracy. A solid line box indicates the selected model with the highest accuracy.

should be selected. This is, in fact, an LST variable selection problem. Our method makes a decision based on estimation performance.

2. Accounting for data quality can help improve product accuracy. As previously stated, the poor quality LST data (QC flags = 1–2) may comprise a large portion of the data, and these “bad” data could have negative effects on model performance. Previous studies just removed the worst data with a QC flag of “3,” which were found to be very limited in our study area. This approach is insufficient. Our method actually considered two situations to resolve two practical problems: (i) when the LST inputs are high quality, our method ensures that the best models for high-quality data are selected, and (ii) when the LST inputs are a mix of both good and poor data, i.e., some of the LST terms are good and others are bad, our method can identify whether the models should use only the “good” LST terms or should include more LST terms that contain some degree of “bad” data.
3. Multiple model comparisons are needed for model selection and for optimizing estimation accuracy. As previously mentioned, different models have been used for air temperature estimation, but whether the selection of different models can improve accuracy is uncertain. Our method can identify which model is best and whether it is necessary to use complex models in place of simple ones, i.e., whether significant differences exist between model performances.

2.5. Models

The six statistical models tested in this study include the multiple linear regression (MLR), the partial least squares (PLS) regression, back propagation neural network (BPNN), support vector regression (SVR), random forests (RFs), and Cubist regression (CR). The first two methods, MLR and PLS, are intrinsically linear. MLR is easy to interpret and is the most common approach in previous studies [Benali *et al.*, 2012; Fu *et al.*, 2011; Good, 2015; Kim and Han, 2013; Lin *et al.*, 2012; Xu *et al.*, 2014; Vancutsem *et al.*, 2010]. The addition of PLS for comparison is mainly to address the colinearity issues of MLR which might degrade its estimation accuracy. However, the actual relationship between the variables and the response may be nonlinear. Although some nonlinear terms can be added to these models, we do not know the real specific form of the nonlinearity. Thus, the methods that are intrinsically nonlinear—machine learning or data mining methods—may help improve the predictions. Therefore, the last four advanced models (BPNN, SVR, RF, and CR) were tested in our research. The six models are included in the supporting information (Text S1).

For the purpose of obtaining more accurate and reliable conclusions for both parameter tuning and model validation, all of the models were built using resampling techniques, including bootstrap [Efron, 1979; Efron and Tibshirani, 1986] or *k*-fold cross-validation [Kohavi, 1995] methods. For efficient computing methods such as MLR and PLS, the bootstrap method with 100 repetitions and the 100-fold cross-validation method were employed, respectively. The remaining four models were tuned using tenfold cross validation due to the heavy computing load.

2.6. Model Validation, Comparison, and Selection

In this study, the RMSD and mean absolute difference (MAD) were used as the performance measurement for tuning, validating, and comparison of the six statistical models used in this study. We validated the modeling results in three ways. In the first validation, all six models were validated using leave-one-out cross validation based on an approach in which every time the samples of one station are left out as validating data, all of the remaining serve as training data. This process was repeated for all stations in turn. The final validation result was the average of the RMSDs validated at all stations, and these results were also used for multiple comparisons discussed later. This process is a traditional and popular validation method.

The second validation involved using the data from stations with longitudes $> 91^\circ$ as training samples (~73%) and using the remainder as testing data (~27%). The third validation involved using the samples from stations with elevations of < 4000 m as training data (~71%) and using the remaining for validation (~29%). The latter two methods can be taken as a stratified validation as described by Daly [2006]. We used the latter two validation methods because the meteorological stations in our study area are not distributed uniformly. According to Figure 1, only a few stations are located on the western TP. With respect to elevation, most of stations are located at relatively low elevations. However, a cross validation does not provide much error information in the unobserved areas, and a stratified validation can be used to test the expandability of the models to the very sparsely observed western region and to areas at higher elevations. The latter two

validation methods were implemented after all of the “optimal” models for each combination and quality situation were determined. Furthermore, the critical 91° longitude and elevation of 4000 m were chosen for splitting all of the samples into appropriate proportions for training and validating.

In addition to the comparison of the six statistical models, different combinations of LST terms and data quality situations needed to be considered. To help confirm the performance differences, statistical tests were performed using multiple comparisons based on a paired unequal variances *t* test with Bonferroni correction for 15 LST combinations. As previously indicated, our first validation method was based on data from 95 different stations, resulting in 95 performance measurements for each combination. Therefore, multiple comparisons can be conducted among different combinations based on these “observations.”

Generally, this multiple comparison approach can yield several “groups” whose members have no significant differences in model performance at the 95 different stations. If listing all the models to be compared are listed in order of average performance measurement (e.g., RMSD), the “groups” are basically arranged with the overlap of certain members. This approach provides a reliable way for us to identify the best group or to determine “several optimal models.” We can then choose the final model that has the least standard error in the “observed” performance measurements (RMSDs). This practice is inspired by the benchmark experiments from *Eugster* [2011]. In our research, the diagnostics of homogeneity indicate that our data are significantly heteroscedastic and that the problem of nonnormality is not severe. Therefore, we selected the paired unequal variances *t* test as our testing method according to *McDonald* [2009]. To alleviate the problem associated with multiple comparisons, we performed simple Bonferroni corrections [*Dunnnett*, 1955] on each test.

2.7. Implementation and Product Generation

Most of the variable samples for model building, including MODIS LST, solar zenith, and NDVI (based on MODIS Surface Reflectance products coded as “MOD09GA” and “MYD09GA”), were extracted or computed directly from the MODIS data files in HDF4 format using standalone software (programmed in C#) developed by the author. The derived records were then imported into R [*Team*, 2012] for statistical analysis and model testing. The final product is in GeoTIFF format with the widely used Albers equal area conic projection based on the WGS84 datum, generated with the R programming software. Various R add-on packages contributed to our work, including *bootStepAIC*, *caret* [*Kuhn and Johnson*, 2013], *pls*, *kernlab*, *nnet*, *randomForest*, *Cubist*, and *rgdal*.

3. Results

3.1. Parameter Tuning Results

Extensive parameter tuning was conducted for each model to assess its full performance. The tuning results are included in the supporting information (Figures S2–S7). For MLR (Figure S2), the number of variables is tuned for each combination of LST using a stepwise method for 100 bootstrap resampling repetitions. The tuning results of PLS (Figure S3), BPNN (Figure S4), and CR (Figure S7) are all scree plots. Using these plots, we easily confirmed the optimal parameter value at which the RMSD no longer significantly declined. SVR (Figure S5) and RF (Figure S6) yielded upward parabola-like graphs; therefore, we straightforwardly and easily determined the final parameter value with the lowest RMSD for each combination. All the tuning results (in the supporting information) are from the S1 quality situation, in which all the LST terms are high quality. The same processes were also performed for the S2 quality situation.

3.2. Different Combinations of MODIS LST Terms for Predicting Performance

Figure 5 plots the RMSDs of each combination for all six statistical models. We also performed multiple comparisons for each model. The results are shown as letters at the top of each combination. In each subplot, the combinations with a same letter at the top exhibit insignificant differences in the RMSD outcomes for the 95 cross validations between these combinations. We can see that for all six models, significant differences generally exist among the different combinations of LST terms, indicating the necessity of considering the combination of LSTs to some degree.

Furthermore, for most of the models, the combinations of nos. 15, 13, 11, and 8 can be taken as a group with lower average RMSD values than the rest, followed by another group consisting of nos. 3, 4, 5, 6, 9, 10, and 14, whose RMSDs are lower than the remaining combinations, including 1, 2, and 7. According to Figure 3, which

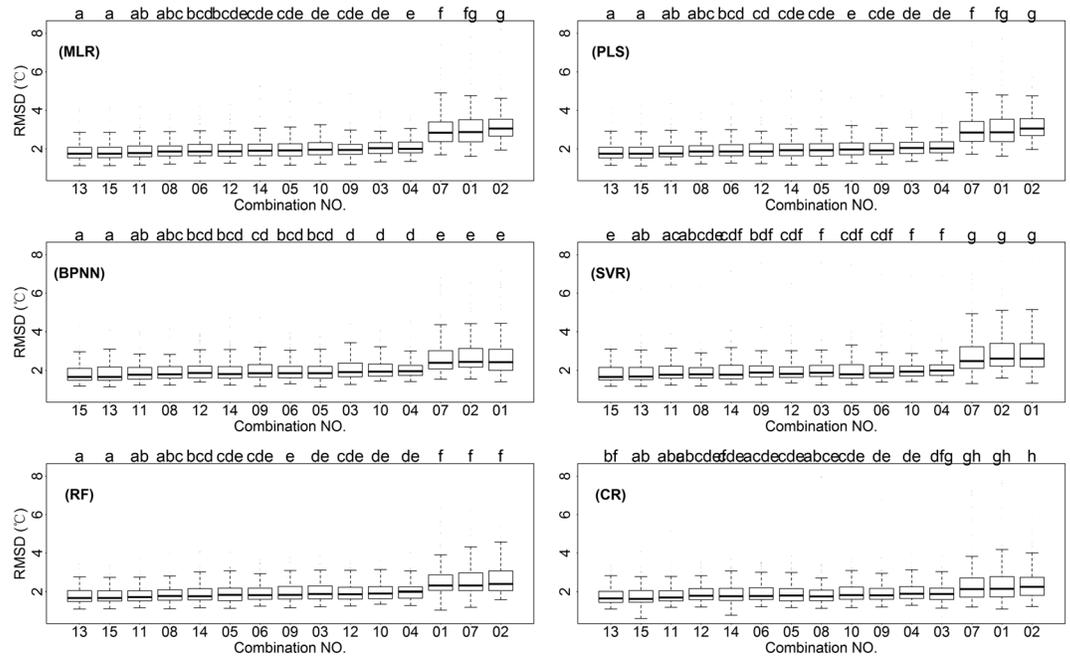


Figure 5. Cross-validation results for every combination of each model and multiple comparisons among different combinations. The x axis is in ascending order of the average RMSD of the cross validation. The box and whiskers show the distributions of RMSDs. Model type indicated in the top left of each panel. See Figure 3 for “combination number.” The same letters at the top mean no statistical significance in difference.

describes the compositions of the LST combinations, we found that the largest difference among the three groups is related to the night LST: the more night LST terms introduced, the higher the accuracy of the model. The number of night LST terms in the combinations of the first, second, and last group is 2, 1, and 0, respectively.

We also found that the LST terms from Terra and Aqua exhibited average mixing in the combinations for each group. This observation indicates that although the overpass times of Terra and Aqua are different, no significant differences exist between their predictive performances with respect to daily mean air temperature estimations. These findings are consistent with those of previous studies [Benali et al., 2012; W Zhang et al., 2011] using MLR.

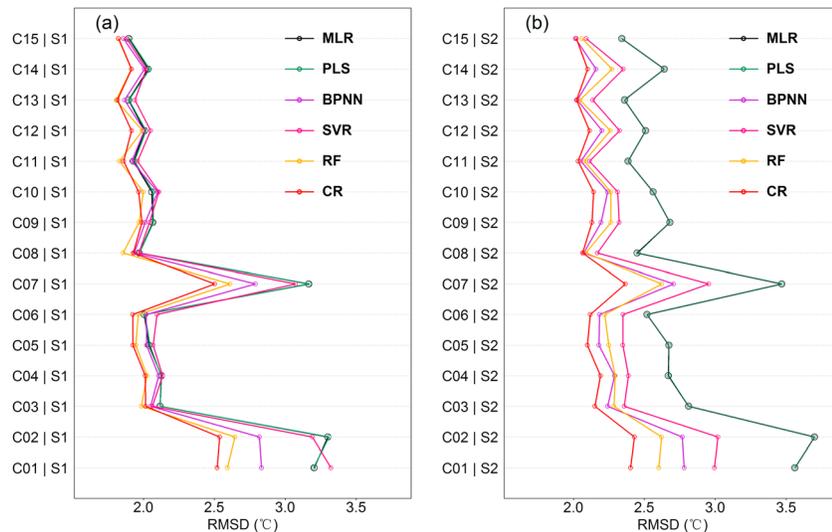


Figure 6. Comparison of different statistical models for each combination under the (a) S1 and (b) S2 quality situations.

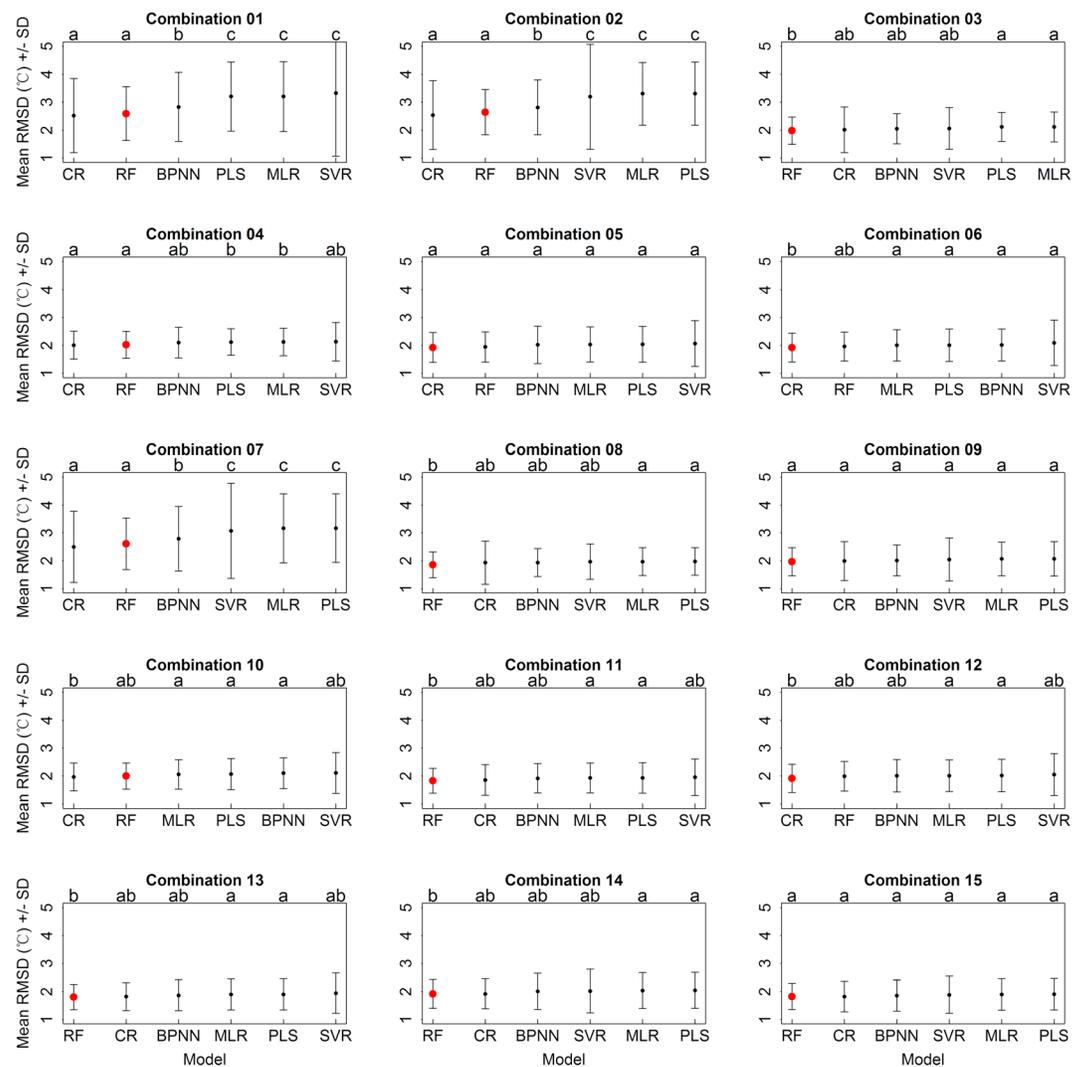


Figure 7. Comparison of different statistical models for each combination under the S1 quality situation based on multiple comparisons. The unique red point represents the final model for each combination. The x axis is in order of the average RMSD of the cross validation. The same letters at the top mean no statistical significance in difference.

However, because we considered all of the possible combinations of LST, we obtained more findings. The letters at the top of the first two groups seem to be rather complicated, indicating that it is hard to identify their differences due to very similar RMSDs. On the contrary, obvious differences are observed between the first two groups and the last one: Among these groups, their top letters are entirely different, and the average RMSDs differ significantly, indicating that if only the day LST term is used, the model performance will be significantly lower.

Please note that all the above results are for the S1 quality situation in which all four LST terms are of good quality. We also conducted the same comparison for the S2 quality situation and achieved similar results (not shown).

3.3. The Effects of Different Statistical Models

Figure 6a shows the performances of all six statistical models for each combination in the S1 situation. Basically, their performances can be divided into three grades: RF and CR achieved the first grade because they clearly outperformed the others; BPNN ranked second because the performance was generally slightly better but sometimes (for combination C01, C02, and C07) had a significantly lower RMSD than the other three models; and MLR, PLS, and SVR made up the lowest grade, because their performances were similarly

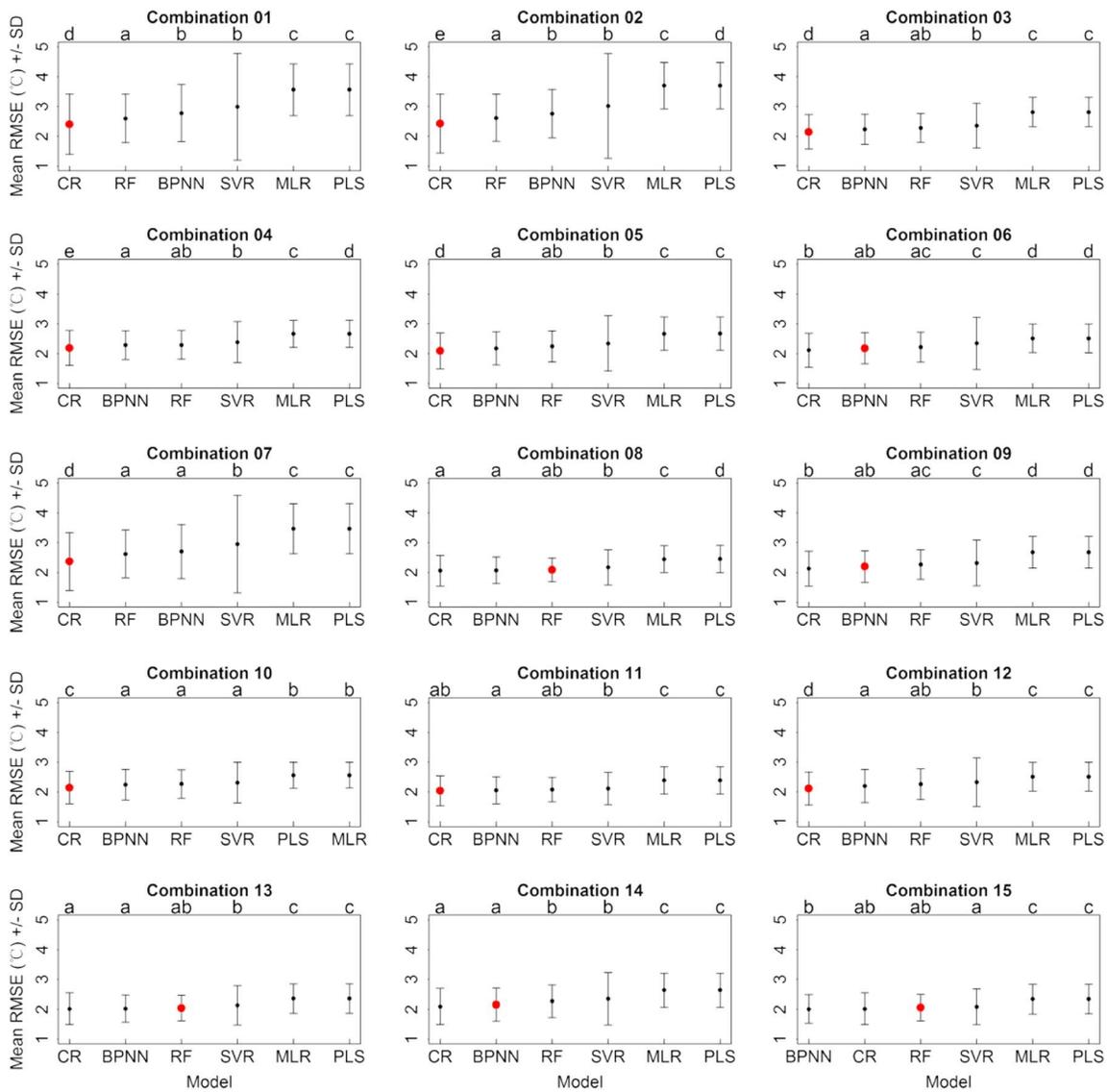


Figure 8. Comparison of different statistical models for each combination under the S2 quality situation based on multiple comparisons. The unique red point represents the final model for each combination. The x axis is in ascending order of the average RMSD of the cross validation. The same letters at the top mean no statistical significance in difference.

poor. To identify the differences between these models for each combination, we performed 15 paired unequal variances *t* tests with Bonferroni correction, and the results are shown in Figure 7. Our study featured not only LST combinations but also multiple models. Thus, their mixed effects are evident.

For combinations that only consider day LST terms, including C01, C02, and C07, the advanced models (especially, CR and RF) perform significantly better than the simple ones, as demonstrated by the clearly different letters at the top (having the same letters means no statistical significance in difference). However, for most combinations that introduce night LST data into the modeling (i.e., all of the combinations except C01, C02, and C07), the performances of the more complex models (e.g., CR, RF, and BPNN) were not significantly better than the MLR or PLS, compared with those only having day LST terms.

From Figures 6a to 7, we conclude that model selection, i.e., choosing the simple or complex model, does have an effect on predicting performance in some cases. When only the day LST terms are available, the advanced models, including the CR, RF, and BPNN, are strongly recommended because they result in obviously lower average RMSDs (Figure 6a), and the differences between these models and the simple models

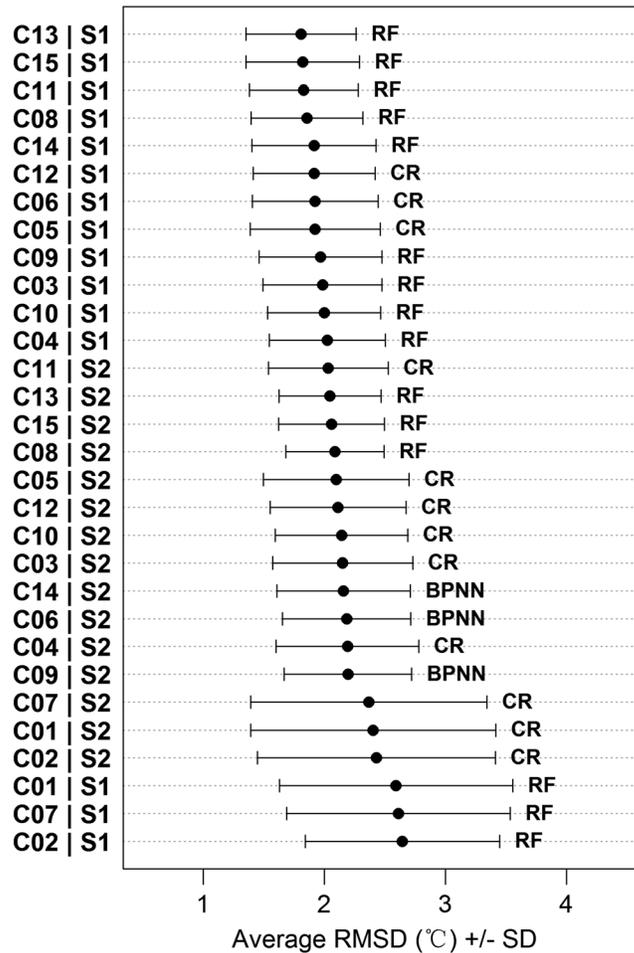


Figure 9. Final model ranking table for each LST combination under both quality situations. The y axis is the number of LST combinations and quality situations; e.g., "C06|S2" indicates a combination of No. 6 in the second quality situation.

LST terms are of good quality. Compared with the results of the S1 situation in Figure 6a, where all the available LST terms are of good quality, certain differences definitely exist: (1) the performances of all models generally decreased to varying degrees; (2) the more complex models performed better than the MLR and PLS, with an obviously larger decrease in RMSD; and (3) CR outperformed all the other models for every combination in the S2 situation. These results imply that no significant performance differences exist among the candidate models for most combinations in the S1 quality situation; however, when the LSTs include poor quality data, the advanced or complex models obviously outperform the simple ones. Therefore, it is necessary to replace the simple methods such as MLR or PLS with the more complex models.

To confirm the final model selection for all combinations in the S2 situation, we performed the same multiple comparisons among models for each combination plotted in Figure 8. As in the S1 situations, we selected the model with the lowest standard RMSD value from the best group of models, i.e., those both at the front and with the same letter at the top. The comparison results based on MAD are also included in the supporting information (Figure S9).

3.5. Ranking Table

Having determined the best model for each combination under both quality situations, we generated a ranking table (Figure 9) of all of the selected models in ascending order of validation performance as measured by average RMSD. When the night LST is considered, the performances of all corresponding combinations in the

are significant (Figure 7). However, in cases where the night LST terms are available, the selection of models generally does not have a significant effect on the estimation performance.

It should be noted that to consider the robustness of the model performance measurement, the same comparisons are also conducted based on MAD. The results are included in the supporting information (Figure S8), and similar conclusions are achieved.

In our research, to explicitly choose a target model for each combination (see Figure 2 and section 2.4), the selection process was as follows: For each combination, the best model group was first determined by listing all six models in order of their average RMSD values, which are representative of model performance, and choosing the foremost group that had the same letter at the top, which indicates that the models in this group had insignificant differences in model performance. Then, the model with the lowest standard error of validation RMSDs in the chosen group was selected as the final model.

3.4. The Effects of LST Quality

Figure 6b shows the general performance information of the S2 quality situation when not all of the available

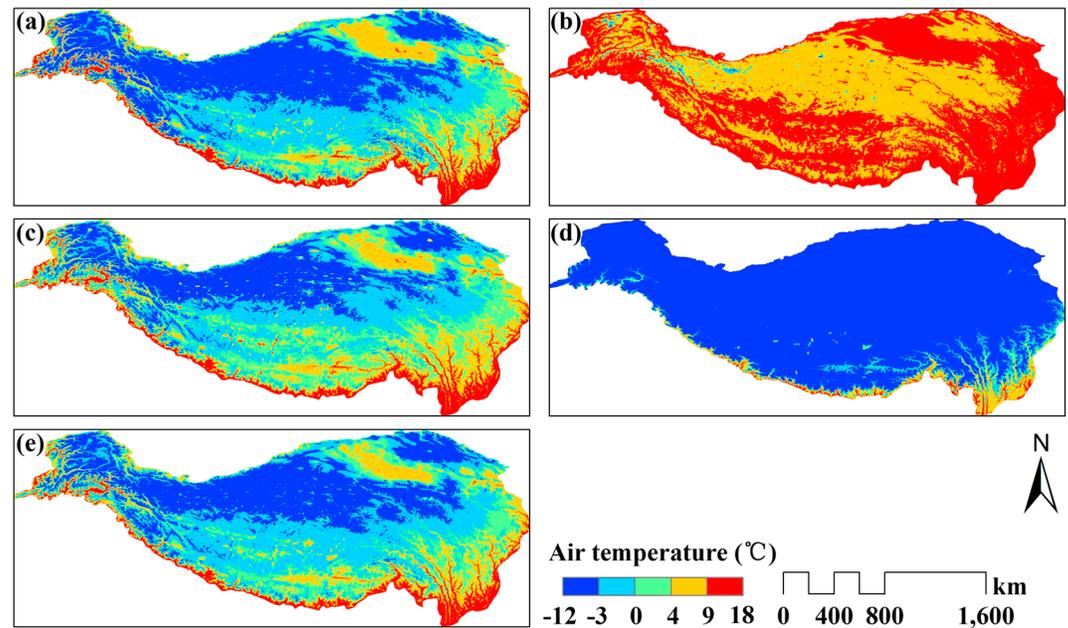


Figure 10. Spatial distribution of the seasonally averaged daily mean air temperatures for 2003–2010 in (a) spring, (b) summer, (c) autumn, (d) winter, and (e) the full year.

S1 situation were better than those of S2. However, counterintuitively, the performances of C01, C02, and C07 under the S1 situation are worse than those in the S2 situations. We attribute this “abnormality” to the relatively weak correlation between the day LST and daily mean air temperature; the errors in the day LSTs may add significant randomness to the final results.

3.6. Produced Air Temperature

Based on the ranking table (Figure 9), the scheme described in sections 2.4 and 2.7 produced the final product of daily mean air temperature over the TP with a high spatial resolution of 1 km. Furthermore, to provide potentially useful information on errors, a by-product was created to accompany the daily mean air temperature product with the same temporal and spatial resolution. This by-product contained a “model code” that records specific information on which combination and quality situation were used in the estimation for every pixel on each day.

To provide the overall perspective of the final product, the spatial patterns of the generated air temperatures were evaluated. Figure 10 plots the spatial distribution of the seasonally averaged daily mean air temperatures from 2003 to 2010. In spring (Figure 10a) and autumn (Figure 10c), the mean air temperatures for most areas are below 0°C. Only very few areas are above 0°C in winter (Figure 10d) and below 4°C in summer (Figure 10b). According to *Li et al.* [2003], the spatial distribution of air temperature over the TP is determined by vertical and latitudinal zonality: (1) the higher the elevation, the lower the air temperature and (2) the higher the latitude, the lower the air temperature. Generally, these trends can be reflected in the spatial pattern as in Figure 10, with the latitude, longitude, and elevation information indicated in Figure 1.

4. Discussion

4.1. Selection of Variables

Removing problematic variables or adding informative ones may improve model performance. In our research, we obtained six auxiliary predictors that were most commonly used in former studies. Previously reported results do not indicate whether adding certain auxiliary variable actually results in significant improvement [*Benali et al.*, 2012]. Other potential variables, including the albedo [*Y Xu et al.*, 2014], solar radiation [*Emamifar et al.*, 2013], and surface moisture conditions [*Kim and Han*, 2013], may also be significant; however, it is difficult to consider every possible factor. For the MLR, we used the popular stepwise method to remove confounding variables for each combination, but the other five models did not take noise variable

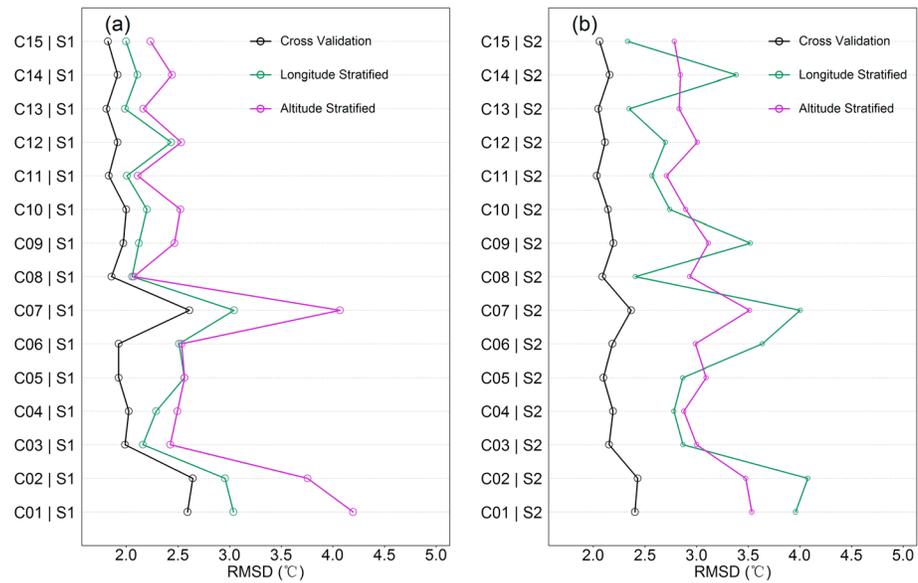


Figure 11. Cross validation and longitude- and altitude-stratified validations of the final 30 models under the (a) S1 and (b) S2 quality situations.

removal into account. Variable selection in the machine learning methods, including BPNN, SVR, RF, and CR, may make some difference. However, the much heavier computational burden of the machine learning methods (relative to the simple MLR) would take too much time to perform detailed predictor selection for the four advanced models. Therefore, we focused on different combinations and quality situations associated with the LST data because they are inevitably the most important factors in this study.

4.2. Model Accuracy and Validation

Typically, the errors in air temperature estimation are larger in association with high temporal resolution than with lower temporal resolution. For example, *Benali et al.* [2012] reported an RMSD of 1.33°C for an 8 day averaged air temperature estimation in Portugal with relatively uniformly distributed and dense stations, whereas *Emamifar et al.* [2013] observed an RMSD of 2.3°C for the daily mean air temperature estimation in southwest Iran. In previous studies, errors in daily air temperature estimation using statistical methods generally fall in the range of 2–3°C [*Benali et al.*, 2012; *W Zhang et al.*, 2011]. It should be noted that the accuracies reported in these studies are actually incomparable with those of this research, owing to the different study areas, LST accuracies, and other factors. However, all the statistical models used in previous studies mentioned above are tested and compared under the same conditions in our study. The accuracies of our product using the proposed method are considered to be optimal among the selected six models and different LST conditions, as described in section 2.4.

In this study, the average RMSDs based on leave-one-out cross validation ranges from 1.81°C to 2.64°C for the S1 situation and from 2.03°C to 2.43°C for the S2 situation (Figure 9). Except for the combinations containing only the day LST, the average RMSDs are all approximately 2°C. Furthermore, the air temperatures produced using the models containing night LST terms account for 87% of the total LSTs estimated. Thus, in most cases, the accuracy of our product is relatively high according to the validation results of the final models listed in Figure 9. Given the complex terrain, climate conditions, and high elevations on the TP, we considered the validation results to be highly acceptable.

The results of the stratified validations using the method described in section 2.5 are displayed in Figure 11. Compared with cross validation, the performances of all of the models of decreased to varying degrees under stratified validation. In the S1 situation, for combinations using the night LST, the mean RMSDs of the longitude- and altitude-stratified validations were 2.20°C and 2.38°C, respectively (Figure 11a). For combinations using the night LST in the S2 situation, the mean RMSDs of the longitude- and altitude-stratified validations were 2.66°C and 2.70°C, respectively (Figure 11b). In our study, the air temperatures

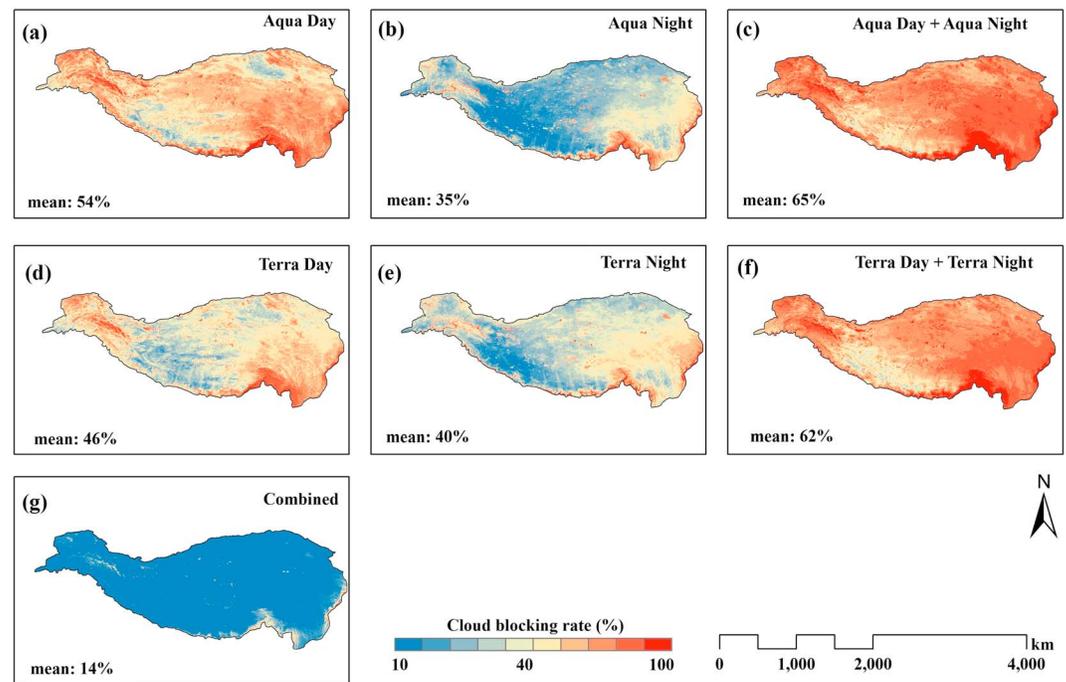


Figure 12. Spatial distribution of the cloud-blocking rate for the LSTs of (a) Aqua day, (b) Aqua night, (c) Aqua day and night, (d) Terra day, (d) Terra night, (e) Terra day and night, and (g) the developed product. The “cloud-blocking rate” represents the percentage of days for which the LST data were not available during the investigation period.

produced using the models containing night LST terms accounted for 87%; therefore, in most cases, the accuracies in the west and high-elevation areas with few gauging stations are still highly acceptable. The errors in the daily air temperature estimation generally fall in the range of 2–3°C [Benali *et al.*, 2012]. The decreased performance in both the S1 and S2 situations in the west and high-elevation areas may necessitate establishing more stations in the ungauged areas for future research.

It should be noted that the accuracies of the data set produced using the proposed approach can be very heterogeneous. The spatial and temporal properties of the errors can be evaluated using the by-product mentioned in section 3.6 which record the “model code” for every pixel on each day. The model code values of 1–30 correspond to models listed in Figure 9 (from top to bottom).

4.3. Data Availability

One of the most important objectives of the proposed method was to reduce the cloud blockage in the final air temperature product. Figures 12a–12f show the cloud-blocking rate in our study region in six common cases that only use single LST terms or a simple combination of LST terms from 2008 to 2010. The cloud-blocking rate indicates the percentage of days when the LST data were not available during the investigation period. The clearly higher cloud-blocking rate of night LST than that of day LST may be due to the fact that the daytime cloud algorithm of MODIS is expected to present more confidence than that for nighttime due to the availability of visible channel data in daytime [Ackerman *et al.*, 1998]. All six cases suffered from the serious problem of high cloud blockage, with a mean cloud-blocking rate of 35–65%. However, the final product using our method features much less cloud blockage than the spatially averaged cloud-blocking rate, with a value of only 14%, as shown in Figure 12g.

We found that most of the areas unobserved in situ in the center and western TP have very low cloud blockage. Furthermore, the areas with low MODIS data availability are mostly located along the southeastern boundary of the TP, where the meteorological stations are relatively dense, making it feasible to fill in the missing data through interpolation of observation data using geostatistical methods [Kilibarda *et al.*, 2014] in future research.

4.4. Implications for the Selection of LST Terms and Statistical Models in Practical Applications

Our study indicates that in practical applications, the accuracies of daily air temperature estimation using MODIS LST may be influenced by the combinations and qualities of LST and the selection of statistical models. It should be noted that the findings obtained in this study only apply to the estimations of daily mean air temperatures, and the results for other types of air temperatures (e.g., daily maximum) may be well different.

When multiple LST terms are available, it may be important to determine which LST term or which combination of LST terms should be used. Based on the results shown in section 3.2 (Figure 1), the night LST term is a guarantee of high accuracy and should always be used regardless of the model employed or the quality situation. It is also useful to understand that in practice, when night LST is available, adding the day LST term will achieve little improvement (for example, compare the combinations C03 and C05 or the combinations C08 and C13). This finding is particularly useful in cases when good night LST and bad day LST are available. In fact, in these cases, we do not need to use additional day LST data for air temperature estimations because they introduce the uncertainty associated with poor quality day LST data.

The selection of models depends on the LST combinations and quality situations. When the night LST is available and the model is trained under good quality situation (S1), it is not necessary to use complex methods because these methods impose a much heavier computing burden and may introduce greater uncertainty compared to MLR or PLS with little improvement in the prediction performance, as indicated in section 3.3 (Figures 6a and 7).

However, in cases where only the day LST is available, it is essential to replace the simple methods with advanced models according to our results (Figures 6a and 7) regardless of the quality situation. This finding can be explained to some degree by the differences in the relationships between the LST terms of different pass times and the daily mean air temperature. In practice, the night LST has a very high correlation (up to 0.93) with daily mean air temperature, indicating that the linear relationship between these variables is so strong that using a typical linear model MLR or PLS is sufficient. However, the correlation between the day LST and daily mean air temperature is much lower than the correlation with the night LST. The relationship between them features increases nonlinearly, which can be simulated more accurately using complex models.

When models are trained under the mixed quality situation (S2), in which not all the LST data are of good quality, especially when the good LST samples alone are not sufficient to build a reliable model fit, complex models are always recommended regardless of the LST combination, according to the results shown in section 3.4 (Figures 6b and 8). The obvious differences between complex and simple models under this situation are considered to be largely attributable to the inherent characteristics of these models. Generally, small errors in the variables do not greatly affect the model fit. However, LST inputs of extremely poor quality are bound to introduce more "outliers." Typical linear regression models, such as MLR or PLS, are more sensitive to outliers than complex machine learning methods, such as RF or SVR. For example, tree-based models, such as RF and CR, can alleviate the exceptional effects of outliers by splitting the training samples. It should be noted that LST outliers can be identified using MLR to some degree, so detecting and removing the outliers may greatly improve the accuracy in some cases. However, it is not conducted in this study given the great difficulty of detecting all the possible outliers present in as many as four LST terms and six auxiliary variables.

5. Conclusions

This study proves that reasonably integrating multiple MODIS LST terms can greatly reduce cloud blockage and simultaneously retain relatively high accuracies. Instead of simply combining the four LST terms, our proposed method optimizes the estimation accuracy depending on data quality, available LST combinations, and model selection. This method greatly reduced the cloud blockage in the final air temperature product to a spatial average of only 14% while maintaining relatively high accuracy in most cases, with cross-validation RMSDs values of approximately 2°C. In addition, it accounted for the effects of LST data quality by considering different quality situations in a model ranking process and identified the best model for air temperature estimation under various conditions among the six popular statistical models tested.

Some important implications for practical application can be concluded: (1) night LST is a guarantee of high accuracy and should always be used, regardless of the model employed or the quality situation; (2) when the

night LST is available and the model is trained under a good quality situation, simple models (typical linear regression models, such as MLR or PLS) are recommended; (3) when only the day LST is available, advanced models (intrinsically nonlinear models, such as BPNN, SVR, RF, or CR) are recommended, regardless of the quality situation; and (4) when models are trained under a mixed quality situation in which not all the LST data are of good quality, complex models are always recommended.

Acknowledgments

This work was supported by the "Strategic Priority Research Program (B)" of the Chinese Academy of Sciences (grant XDB03030300) and the National Natural Science Foundation of China (grants 41422101 and 41371087). We are grateful to the Chinese Meteorology Administration for providing the air temperature data. Daily temperature data were collected from the Climatic Data Center, National Meteorological Information Center, China Meteorological Administration at <http://data.cma.cn>. The MODIS LST and reflectance products can be downloaded freely from the Reverb website (<http://reverb.echo.nasa.gov/>). The GTOPO30 data set can be obtained from <https://lta.cr.usgs.gov/GTOPO30>. The produced daily air temperature data set using the method proposed in present study can be obtained at http://hongbozhang.net/wph/?page_id=357.

References

- Ackerman, S. A., K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley (1998), Discriminating clear sky from clouds with MODIS, *J. Geophys. Res.*, *103*(D24), 32,141–32,157, doi:10.1029/1998JD200032.
- Austin, P. C., and J. V. Tu (2004), Bootstrap methods for developing predictive models, *Am. Stat.*, *58*(2), 131–137.
- Benali, A., A. C. Carvalho, J. P. Nunes, N. Carvalhais, and A. Santos (2012), Estimating air surface temperature in Portugal using MODIS LST data, *Remote Sens. Environ.*, *124*, 108–121, doi:10.1016/j.rse.2012.04.024.
- Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Oxford Univ. Press, Inc, New York.
- Breiman, L. (1996), Bagging predictors, *Mach. Learn.*, *24*(2), 123–140.
- Breiman, L. (2001), Random forests, *Mach. Learn.*, *45*(1), 5–32.
- Cristobal, J., M. Ninyerola, and X. Pons (2008), Modeling air temperature through a combination of remote sensing and GIS data, *J. Geophys. Res.*, *113*, D13106, doi:10.1029/2007JD009318.
- Daly, C. (2006), Guidelines for assessing the suitability of spatial climate data sets, *Int. J. Climatol.*, *26*(6), 707–721.
- Drucker, H., C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik (1997), Support vector regression machines, *Adv. Neural Inf. Process. Syst.*, *9*, 155–161.
- Dunnett, C. W. (1955), A multiple comparison procedure for comparing several treatments with a control, *J. Am. Stat. Assoc.*, *50*(272), 1096–1121.
- Efron, B. (1979), Bootstrap methods: Another look at the jackknife, *Ann. Stat.*, *7*, 1–26.
- Efron, B., and R. Tibshirani (1986), Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Stat. Sci.*, *1*, 54–75.
- Emamifar, S., A. Rahimikhoob, and A. A. Noroozi (2013), Daily mean air temperature estimation from MODIS land surface temperature products based on M5 model tree, *Int. J. Climatol.*, *33*(15), 3174–3181, doi:10.1002/joc.3655.
- Eugster, M. J. (2011), Benchmark experiments: A tool for analyzing statistical learning algorithms, München, Univ., Diss., 2011.
- Faraway, J. J. (2014), *Linear Models with R*, CRC Press.
- Flohn, H. (1968), *Contributions to a Meteorology of the Tibetan Highlands*, Department of Atmospheric Science, Colorado State University Fort Collins, Colorado.
- Florio, E., S. Lele, Y. Chi Chang, R. Sterner, and G. Glass (2004), Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: A statistical approach, *Int. J. Remote Sens.*, *25*(15), 2979–2994.
- Fu, G., Z. Shen, X. Zhang, P. Shi, Y. Zhang, and J. Wu (2011), Estimating air temperature of an alpine meadow on the Northern Tibetan Plateau using MODIS land surface temperature, *Acta Ecol. Sin.*, *31*(1), 8–13, doi:10.1016/j.chnaes.2010.11.002.
- Geladi, P., and B. R. Kowalski (1986), Partial least-squares regression: A tutorial, *Anal. Chim. Acta*, *185*, 1–17.
- Gleason, C. J., and J. Im (2012), Forest biomass estimation from airborne LiDAR data using machine learning approaches, *Remote Sens. Environ.*, *125*, 80–91, doi:10.1016/j.rse.2012.07.006.
- Good, E. (2015), Daily minimum and maximum surface air temperatures from geostationary satellite data, *J. Geophys. Res. Atmos.*, *120*, 2306–2324, doi:10.1002/2014JD022438.
- Green, R. M., and S. I. Hay (2002), The potential of Pathfinder AVHRR data for providing surrogate climatic variables across Africa and Europe for epidemiological applications, *Remote Sens. Environ.*, *79*(2–3), 166–175, doi:10.1016/S0034-4257(01)00270-X.
- Hock, R. (2003), Temperature index melt modelling in mountain areas, *J. Hydrol.*, *282*(1–4), 104–115, doi:10.1016/S0022-1694(03)00257-9.
- Hsu, C.-W., C.-C. Chang, and C.-J. Lin (2003), *A Practical Guide to Support Vector Classification*.
- Janatian, N., M. Sadeghi, S. H. Sanaeinejad, E. Bakhshian, A. Farid, S. M. Hashemina, and S. Ghazanfari (2016), A statistical framework for estimating air temperature using MODIS land surface temperature data, *Int. J. Climatol.*, doi:10.1002/joc.4766.
- Jang, J.-D., A. Viau, and F. Anctil (2004), Neural network estimation of air temperatures from AVHRR data, *Int. J. Remote Sens.*, *25*(21), 4541–4554.
- Karatzoglou, A., A. Smola, K. Hornik, and A. Zeileis (2004), Kernlab—an S4 package for kernel methods in R, *J. Stat. Software*, *11*(9), 1–20.
- Kilibarda, M., T. Hengl, G. B. M. Heuvelink, B. Gräler, E. Pebesma, M. Perčec Tadić, and B. Bajat (2014), Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution, *J. Geophys. Res. Atmos.*, *119*, 2294–2313, doi:10.1002/2013JD020803.
- Kim, D. Y., and K. S. Han (2013), Remotely sensed retrieval of midday air temperature considering atmospheric and surface moisture conditions, *Int. J. Remote Sens.*, *34*(1), 247–263, doi:10.1080/01431161.2012.712235.
- Kohavi, R. (1995), A study of cross-validation and bootstrap for accuracy estimation and model selection, in *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1137–1145, Morgan Kaufmann Publishers Inc, San Francisco, Calif.
- Kuhn, M., and K. Johnson (2013), *Applied Predictive Modeling*, Springer, New York.
- Li, X., G. Cheng, and L. Lu (2003), Comparison study of spatial interpolation methods of air temperature over Qinghai-Xizang Plateau, *Plateau Meteorol.*, *22*(6), 565–573.
- Lin, S. P., N. J. Moore, J. P. Messina, M. H. DeVisser, and J. P. Wu (2012), Evaluation of estimating daily maximum and minimum air temperature with MODIS data in east Africa, *Int. J. Appl. Earth Obs. Geoinf.*, *18*, 128–140, doi:10.1016/j.jag.2012.01.004.
- Liu, X. D., and B. D. Chen (2000), Climatic warming in the Tibetan Plateau during recent decades, *Int. J. Climatol.*, *20*(14), 1729–1742, doi:10.1002/1097-0088(20001130)20:14<1729::aid-joc556>3.0.co;2-y.
- McDonald, J. H. (2009), *Handbook of Biological Statistics*, Sparky House Publishing, Baltimore, Md.
- Neaupane, K. M., and S. H. Achet (2004), Use of backpropagation neural network for landslide monitoring: A case study in the higher Himalaya, *Eng. Geol.*, *74*(3–4), 213–226, doi:10.1016/j.enggeo.2004.03.010.
- Ninyerola, M., X. Pons, and J. M. Roure (2007), Objective air temperature mapping for the Iberian Peninsula using spatial interpolation and GIS, *Int. J. Climatol.*, *27*(9), 1231–1242, doi:10.1002/joc.1462.
- Oyler, J. W., S. Z. Dobrowski, Z. A. Holden, and S. W. Running (2016), Remotely sensed land skin temperature as a spatial predictor of air temperature across the conterminous United States, *J. Appl. Meteorol. Climatol.*, doi:10.1175/JAMC-D-15-0276.1.

- Prihodko, L., and S. N. Goward (1997), Estimation of air temperature from remotely sensed surface observations, *Remote Sens. Environ.*, *60*(3), 335–346, doi:10.1016/S0034-4257(96)00216-7.
- Prince, S. D., S. J. Goetz, R. O. Dubayah, K. P. Czajkowski, and M. Thawley (1998), Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using Advanced Very High-Resolution Radiometer satellite observations: Comparison with field observations, *J. Hydrol.*, *212–213*, 230–249, doi:10.1016/S0022-1694(98)00210-8.
- Quinlan, J. (2001), Rulequest: Data mining tools See5 and C5.0.
- Quinlan, J. R. (1992), Learning with continuous classes, in *5th Australian Joint Conference on Artificial Intelligence*, edited by N. Adams and L. Sterling, pp. 343–348, World Scientific, Singapore.
- Quinlan, J. R. (1993), Combining instance-based and model-based learning, in *Proceedings of the Tenth International Conference on Machine Learning*, pp. 236–243, Morgan Kaufmann, Amherst, Mass.
- Rohrer, M., N. Salzmann, M. Stoffel, and A. V. Kulkarni (2013), Missing (in-situ) snow cover data hampers climate change and runoff studies in the Greater Himalayas, *Sci. Total Environ.*, *468*, S60–S70.
- Spadavecchia, L., and M. Williams (2009), Can spatio-temporal geostatistical methods improve high resolution regionalisation of meteorological variables?, *Agric. For. Meteorol.*, *149*(6–7), 1105–1117, doi:10.1016/j.agrformet.2009.01.008.
- Stahl, K., R. D. Moore, J. A. Floyer, M. G. Asplin, and I. G. McKendry (2006), Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density, *Agric. For. Meteorol.*, *139*(3–4), 224–236, doi:10.1016/j.agrformet.2006.07.004.
- Team, R. C. (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna.
- Thompson, L. G., T. Yao, E. Mosley-Thompson, M. E. Davis, K. A. Henderson, and P. N. Lin (2000), A high-resolution millennial record of the South Asian Monsoon from Himalayan ice cores, *Science*, *289*(5486), 1916–1919, doi:10.1126/science.289.5486.1916.
- Thornton, P. E., S. W. Running, and M. A. White (1997), Generating surfaces of daily meteorological variables over large regions of complex terrain, *J. Hydrol.*, *190*(3–4), 214–251, doi:10.1016/S0022-1694(96)03128-9.
- Vancutsem, C., P. Ceccato, T. Dinku, and S. J. Connor (2010), Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa, *Remote Sens. Environ.*, *114*(2), 449–465, doi:10.1016/j.rse.2009.10.002.
- Vapnik, V. (1995), *The Nature of Statistical Learning Theory*, Springer, Berlin.
- Venables, W. N., and B. D. Ripley (2002), *Modern Applied Statistics with S*, Springer Science & Business Media.
- Walton, J. T. (2008), Subpixel urban land cover estimation, *Photogramm. Eng. Remote Sens.*, *74*(10), 1213–1222.
- Wan, Z. (2008), New refinements and validation of the MODIS land-surface temperature/emissivity products, *Remote Sens. Environ.*, *112*(1), 59–74, doi:10.1016/j.rse.2006.06.026.
- Wang, K., Z. Li, and M. Cribb (2006), Estimation of evaporative fraction from a combination of day and night land surface temperatures and NDVI: A new method to determine the Priestley–Taylor parameter, *Remote Sens. Environ.*, *102*(3–4), 293–305, doi:10.1016/j.rse.2006.02.007.
- Wang, Y., G. Li, W. Min, and Y. Zhang (2014), Representativeness analysis of meteorological stations based on temperature estimated from MODIS data, *Meteorol. Monthly*, *40*(3), 373–380.
- Wloczyk, C., E. Borg, R. Richter, and K. Miegel (2011), Estimation of instantaneous air temperature above vegetation and soil surfaces from Landsat 7 ETM+ data in northern Germany, *Int. J. Remote Sens.*, *32*(24), 9119–9136.
- Wold, S., A. Ruhe, H. Wold, and W. J. Dunn III (1984), The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, *SIAM J. Sci. Stat. Comput.*, *5*(3), 735–743.
- Xu, Y., A. Knudby, and H. C. Ho (2014), Estimating daily maximum air temperature from MODIS in British Columbia, Canada, *Int. J. Remote Sens.*, *35*(24), 8108–8121, doi:10.1080/01431161.2014.978957.
- Yao, Y., and B. Zhang (2013), MODIS-based estimation of air temperature and heating-up effect of the Tibetan Plateau, *Acta Geogr. Sin.*, *68*(1), 95–107.
- Yasunari, T., A. Kitoh, and T. Tokioka (1991), Local and remote responses to excessive snow mass over Eurasia appearing in the northern spring and summer climate—A study with the MRI GCM, *J. Meteorol. Soc. Jpn.*, *69*(4), 473–487.
- Zakšek, K., and M. Schroedter-Homscheidt (2009), Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments, *ISPRS J. Photogramm. Remote Sens.*, *64*(4), 414–421, doi:10.1016/j.isprsjprs.2009.02.006.
- Zhang, F., H. Zhang, S. C. Hagen, M. Ye, D. Wang, D. Gui, C. Zeng, L. Tian, and J. Liu (2015), Snow cover and runoff modelling in a high mountain catchment with scarce data: Effects of temperature and precipitation parameters, *Hydrol. Processes*, *29*(1), 52–65, doi:10.1002/hyp.10125.
- Zhang, G., T. Yao, H. Xie, S. Kang, and Y. Lei (2013), Increased mass over the Tibetan Plateau: From lakes or glaciers?, *Geophys. Res. Lett.*, *40*, 2125–2130, doi:10.1002/grl.50462.
- Zhang, W., Y. Huang, Y. Q. Yu, and W. J. Sun (2011), Empirical models for estimating daily maximum, minimum and mean air temperatures with MODIS land surface temperatures, *Int. J. Remote Sens.*, *32*(24), 9415–9440, doi:10.1080/01431161.2011.560622.
- Zhao, D., W. Zhang, and S. Xu (2007), A neural network algorithm to retrieve nearsurface air temperature from landsat ETM+ imagery over the Hanjiang River Basin, China, in *IEEE International Geoscience and Remote Sensing Symposium*, edited, pp. 1705–1708, IEEE, Barcelona, Spain, doi:10.1109/IGARSS.2007.4423146.
- Zhou, T., and R. Yu (2006), Twentieth-century surface air temperature over China and the globe simulated by coupled climate models, *J. Clim.*, *19*(22), 5843–5858, doi:10.1175/jcli3952.1.
- Zhu, W., A. Lú, and S. Jia (2013), Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products, *Remote Sens. Environ.*, *130*, 62–73, doi:10.1016/j.rse.2012.10.034.