International Journal of Applied Earth Observations and Geoinformation 97 (2021) 102295

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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

# Creating 1-km long-term (1980–2014) daily average air temperatures over the Tibetan Plateau by integrating eight types of reanalysis and land data assimilation products downscaled with MODIS-estimated temperature lapse rates based on machine learning

Hongbo Zhang [a,b,c,d,e], W.W. Immerzeel [e], Fan Zhang [b,f,g,*], Remco J. de Kok [e], Sally J. Gorrie [h], Ming Ye [h]

[a] College of Water Resources & Civil Engineering, China Agricultural University, Beijing, China
[b] Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences (CAS), 100101 Beijing, China
[c] State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China
[d] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, China
[e] Utrecht University, Department of Physical Geography, PO Box 80115, Utrecht, the Netherlands
[f] CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing, China
[g] University of Chinese Academy of Sciences, Beijing, China
[h] Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, FL, USA

ARTICLE INFO

*Keywords:*
MODIS land surface temperature
Tibetan Plateau
Temperature lapse rate
Reanalysis data
Spatial downscaling

ABSTRACT

Air temperature (*Tair*) is critical to modeling environmental processes (e.g. snow/glacier melting) in high-elevation areas of the Tibetan Plateau (TP). To resolve the issue that *Tair* observations are scarce in the TP western part and at high elevation, many studies have estimated daily air temperatures by using MODIS land surface temperature (LST) and various reanalysis datasets. These estimates are however inadequate for supporting high-resolution long-term hydrological simulations or climate analysis due to the high cloud cover, short time span or low spatial resolution. To improve the *Tair* estimation, this study develops a novel machine-learning based method that uses the Gradient Boosting model to efficiently integrate observations from high-elevation stations with eight widely used air temperature reanalysis and assimilation datasets (i.e., NNRP-2, 20CRV2c, JRA-55, ERA-Interim, MERRA-2, CFSR, ERA5 and GLDAS2) downscaled with remote sensing-based temperature lapse rates (TLR). This method is used to generate a new dataset of daily air temperature with the 1-km resolution for the period of 1980–2014. To overcome the problem that TLR derived from limited stations may be unreliable, a new TLR estimation method is developed to first estimate spatially continuous monthly TLRs from MODIS LST and then downscale daily mean *Tair* from eight reanalysis and assimilation datasets to obtain *Tair* at the 1-km resolution using the MODIS-estimated TLRs. The Gradient Boosting (GB) model is selected for integrating the eight downscaled *Tair* and five other auxiliary variables. The models are trained and validated using observations from 100 common stations (i.e. China Meteorology Administration stations) and 13 independent high-elevation stations (4 on glaciers). The results show that the proposed TLR estimation method can efficiently reduce exceptional TLRs in the meantime keeping acceptable downscaling accuracy. The downscaled *Tair* from JRA-55 is the best among the eight downscaled datasets followed by ERA-Interim, MERRA-2, CFSR and others. Finally, the GB-integrated *Tair* further outperforms the downscaled JRA-55 *Tair* with the mean root-mean-squared-deviation (RMSD) of 1.7 °C versus 2.0 °C, especially in high-elevation stations with mean RMSD of 1.9 °C versus 2.7 °C. Both the MODIS-estimated TLR and the high-elevation training observations are demonstrated to significantly improve the air temperature estimation accuracy of the GB model in high-elevation stations. This study also provides a framework for integrating multiple reanalysis and assimilation temperature data with elevation correction in mountainous regions that is not restricted to the TP.

* Corresponding author at: Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences (CAS), 100101 Beijing, China.
*E-mail address:* zhangfan@itpcas.ac.cn (F. Zhang).

https://doi.org/10.1016/j.jag.2021.102295
Received 16 July 2020; Received in revised form 16 November 2020; Accepted 31 December 2020

## 1. Introduction

The Tibetan Plateau (TP) and its surrounding region are considered as the Third Pole on the Earth with an average elevation of approximately 4000 m. Due to the high elevation, TP has a warming rate higher than the global average, resulting in various severe environmental changes such as glacier retreat (Kraaijenbrink et al., 2017; Yao et al., 2012), permafrost degradation (Cheng and Wu, 2007; Wu and Zhang, 2008), lake expansion (Zhang et al., 2011; Zhang et al., 2017), and increasing river discharge and sediment flux (Zhang et al., 2020). Air temperature is the most direct indicator of global warming (Liu and Chen, 2000; Pepin et al., 2015), and it is also an important variable for investigating the changes and mechanisms of various hydrological and ecological processes such as snow or glacier melting (Immerzeel et al., 2014; Zhang et al., 2015) and photosynthesis (Wu et al., 2012). However, air temperature observations are scarce in TP in particular in high-altitude areas due to harsh terrain and weather conditions. There is an urgent need to estimate air temperature in the high-altitude areas of the TP.

To overcome the problem of air temperature data scarcity, a variety of reanalysis and land data assimilation methods have been developed in the last several decades, such as the NCEP/DOE Reanalysis 2 Project (NNRP-2) (Kanamitsu et al., 2002), the 20th Century Reanalysis Version V2c (20CRV2c) (Compo et al., 2011), the Japanese 55-year Reanalysis Project (JRA-55) (Kobayashi et al., 2015), the ERA-Interim (Dee et al., 2011), the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) (Gelaro et al., 2017), the NCEP Climate Forecast System Reanalysis (CFSR) (Saha et al., 2010), the Global Land Data Assimilation System Version 2 (GLDAS-2) (Rodell et al., 2004) and ERA5 (Service, 2017). These efforts have produced various datasets by using advanced model and assimilation techniques to efficiently incorporate observations from multiple sources with consideration of physical relations between multiple types of the observations. In comparison with the observations at sparse weather stations, the datasets provide spatio-temporally continuous air temperatures estimates at a larger scale.

The air temperature datasets derived from the reanalysis or assimilation data need to be corrected for complex terrains in TP, because spatial resolutions (ranging from 0.25° to 2.5°) of the derived data are too coarse to reflect complex topography in TP. Directly using a block-average value for a point position is subject to a large bias due to the significant elevation variation within the grid block (Zhou et al., 2018). Downscaling is thus needed to achieve a higher horizontal resolution, and the resolution of 1–3 km is considered to be satisfactory for complex terrains (El-Samra et al., 2018; Karger et al., 2017). For downscaling air temperature, the temperature lapse rate (TLR) (i.e., the rate of temperature change with elevation) is the most commonly used variable. A constant value of 0.0065 °C/m has been used for downscaling and distributing the air temperature data of CFSR (Li et al., 2014), GLDAS and JRA-55 (Wang et al., 2011a) from a coarse grid to a finer grid. However, since TLR is not a constant, but varies in space and time in TP and its surrounding areas (Li et al., 2013; Wang et al., 2018; Zhang et al., 2018a), using a constant TLR is unsuitable for downscaling the reanalysis data in TP such as ERA-Interim (Gao et al., 2017; Gerlitz et al., 2014). Methods that estimate more accurate TLRs have been developed, for instance, by calculating local and monthly TLRs based on neighboring stations (Zhang et al., 2015; Zhang et al., 2018b) or using the temperature gradient of synoptic data from different pressure levels (Gao et al., 2017; Karger et al., 2017). It has been shown that the station-based TLR may not be optimal for downscaling when a small number of weather stations are insufficient to represent spatial variability at the scale of a grid block. Recently, based on the strong correlation and consistency between air temperature and remote sensing land surface temperature (LST) (Benali et al., 2012; Mutiibwa et al., 2015; Zhang et al., 2016a), MODIS LST data have been used for estimating TLR for the TP (Wang et al., 2018; Zhang et al., 2018a). The existing methods however have not adequately addressed the problems of missing data (mainly caused by clouds) and unreliable estimates of TLR (mainly resulted from small elevation range among LST pixels), and a more efficient TLR estimation method is needed (Wang et al., 2018).

Another problem related to the reanalysis data is that they lack independent validation data because the reanalysis always assimilates all information embedded in observations from meteorological stations (Wang and Zeng, 2012; Zhou et al., 2018). These stations are unevenly distributed in TP with nearly no stations located in the western part or at high elevation of TP. Although several studies have evaluated accuracy of NNRP-1, NNRP-2, JRA-55, ERA-Interim (Gao et al., 2017), CFSR, GLDAS (Ji et al., 2015; Wang et al., 2016c) or a combination of them in TP (Wang and Zeng, 2012; Zou et al., 2014), the evaluation results may be biased due to the lack of independent validation observations at high elevation. To alleviate this problem, several researchers used observations of low elevation (Wijngaard et al., 2017; Zhou et al., 2015). However, excluding high-elevation data may result in low estimation accuracy. Zhang et al. (2016a) showed that the root mean squared errors of estimated daily air temperature in high-elevation (>4000 m) areas increase significantly from 2.1 °C to 2.7 °C when excluding observational information from high-elevation stations. Given that using multiple reanalysis datasets may improve estimation accuracy due to large spread among different datasets (e.g., their seasonal and regional differences in accuracy) (Mudryk et al., 2015), efforts have been spent to integrate several reanalysis datasets such as the combination of ERA-Interim, MERRA, NNRP-1 and ERA-40 at a global scale based on interpolation (Wang and Zeng, 2013) and the ensembles of NNRP-1, ERA-Interim, CFSR and others in a glacierized region of Peru (Hofer et al., 2012).

In addition to various reanalysis dataset, efforts have been made to use remote sensing land surface temperatures to estimate monthly (Huang et al., 2017; Xu et al., 2018) or daily (Rao et al., 2019; Zhang et al., 2016a; Zhu et al., 2017) air temperatures over the TP in the last decade. The studies employed various types of machine learning models to improve estimation accuracy. For example, based on the Cubist regression model, Rao et al. (2019) recently integrated MODIS daily composite LST data with three cloud-free top-of-atmosphere radiation products to produce all sky daily average air temperatures of 2002–2016 with a relatively low root-mean-squared-deviation of 1.9 °C. However, the short time span of MODIS LST (available after 2000 for TERRA) limits their use in practical applications such as long-term hydrological simulation and climate analysis.

To address the problems discussed above, the objectives of the study are as follows: (1) developing a more efficient method for estimating spatially continuous TLR from MODIS LST; and (2) creating a more accurate 1-km daily air temperature datasets during 1980–2014 for the TP (especially for high-elevation areas) using a machine learning method by integrating eight reanalysis and assimilation datasets downscaled with the newly developed TLRs. In comparison with previous researches on air temperature estimation at relatively high spatial resolutions (e.g., 1-km), our study uses multiple reanalysis air temperature datasets rather than only remote sensing LST. As a result, our product has a larger time span and is intrinsically cloud-free. In addition, using the MODIS-estimated TLR for downscaling can improve accuracy of the input reanalysis temperatures for the machine learning model. Our study also uses the observations from 13 independent high-elevation stations and the observations provide reliable training information from high-elevation areas. The combined use of MODIS-estimated TLR and high-elevation observations improves accuracy of our air temperature estimation. To the best of our knowledge, this study is the first attempt to estimate long-term daily average temperatures at the 1-km resolution using multiple downscaled reanalysis datasets based on machine learning techniques. The rest of the paper is organized as follows: Section 2 describes the data used in this study including the station observations and reanalysis datasets; Section 3 introduces the methods of estimating TLR from MODIS LST, using the estimated TLR

for downscaling the eight reanalysis datasets, and integrating the downscaled air temperature datasets using the machine-learning model; Section 4 presents and evaluates the results with a discussion on uncertainty of the MODIS-estimated TLR and the new air temperature dataset.

## 2. Data

### 2.1. Ground observations

This study uses two kinds of observations, the daily mean air temperature and the sub-daily air temperature. The daily mean air temperature observations were collected during 1980–2014 from 100 meteorological stations (Fig. 1) maintained by China Meteorology Administration (CMA). These stations hereinafter are referred to as the "common stations" (CS), and most of them are located in urban areas within the TP with an average altitude of about 3300 m. The sub-daily air temperature observations were collected from 13 automatic weather stations (Fig. 1) that were set up in the field over different land covers. These stations are referred to as the "high-elevation stations" due to their higher elevations (Zhang et al., 2018a), and the detailed descriptions of the stations are listed in Table 1. All the air temperature measurements mentioned above are made at a height of about 2 m above land surface. It is believed that all or majority of the common stations data have been incorporated into the assimilation of the widely used reanalysis datasets such as ERA-Interim, MERRA and CFSR (Wang et al., 2011b). The data of the 13 high-elevation stations are used in this study. It should be noted that three of the stations are at the elevation higher than 5,000 m, and that four are on glaciers. All the temperatures of high-elevation stations are measured sub-daily with frequencies varying from 10 min to 30 min, and finally averaged to the daily time interval. Although most of high-elevation stations are near to common stations in complex terrains with steep slopes, their elevations are higher than the neighboring common stations. For investigating the consistency between the observation from each high-elevation station and its nearest common station, a correlation analysis of data series is conducted, and the results show that the high-elevation station data are reliable with a correlation coefficient larger than 0.91. Other means of data quality control have been conducted using multiple variables including climatological limits, spatial consistency and temporal coherency by the CMA (for common stations) (Zhou et al., 2017) or data providers (for high-elevation stations), to manually check and remove obvious outliers. Following a number of air temperature estimation studies at the 1-km scale (Xu et al., 2014; Zhang et al., 2016b; Zhu et al., 2013), the station observations are considered as the ground truth for the 1-km pixel, within which the corresponding stations are located. This may introduce certain errors in several situations such as on the glaciers with complex terrains and mixed land covers (Zhang et al., 2018b), and a discussion on the errors is given in Section 4.3.4.

### 2.2. Reanalysis datasets

Eight reanalysis and assimilation datasets are used in this study including NNRP-2, 20CRV2c, JRA-55, ERA-Interim, MERRA-2, CFSR, GLDAS and ERA5. A summary description of them is listed in Table 2. Except for the NNRP-2 dataset that is originally for daily data, all the reanalysis datasets are obtained by averaging sub-daily data to daily data. Table 2 also shows that the eight datasets have different time spans. The time period of 1980–2014 is selected for use in this study, since the datasets of this period give the most overlap during that period. The reanalysis datasets were produced by using advanced data assimilation schemes such as the three-dimensional variational (3DVar) data assimilation (e.g. NNRP-2, CFSR and MERRA-2) or the even better four-dimensional variational (4DVar) data assimilation (e.g. JRA-55, ERA-Interim and ERA5). The aim of the reanalysis is to solve the complex problems of integrating all kinds of information from various sources including the previous forecasts, error uncertainty, physical laws and the observations from both stations and satellites (Barker et al., 2004). For instance, the ERA-Interim dataset incorporates air temperature observations from stations to improve soil moisture estimation (Wang and Zeng, 2012). Due to complexity of the reanalysis, it is difficult to determine whether the observations from the 13 high-elevation stations were used in the reanalysis. However, it is highly likely that the observations have not been used by the reanalysis data.

### 2.3. MODIS LST and other auxiliary data

Four kinds of observations can be obtained from MODIS per day including TERRA daytime (10:30 local time), TERRA nighttime (22:30 LT), AQUA daytime (13:30 LT) and AQUA nighttime (1:30 LT).

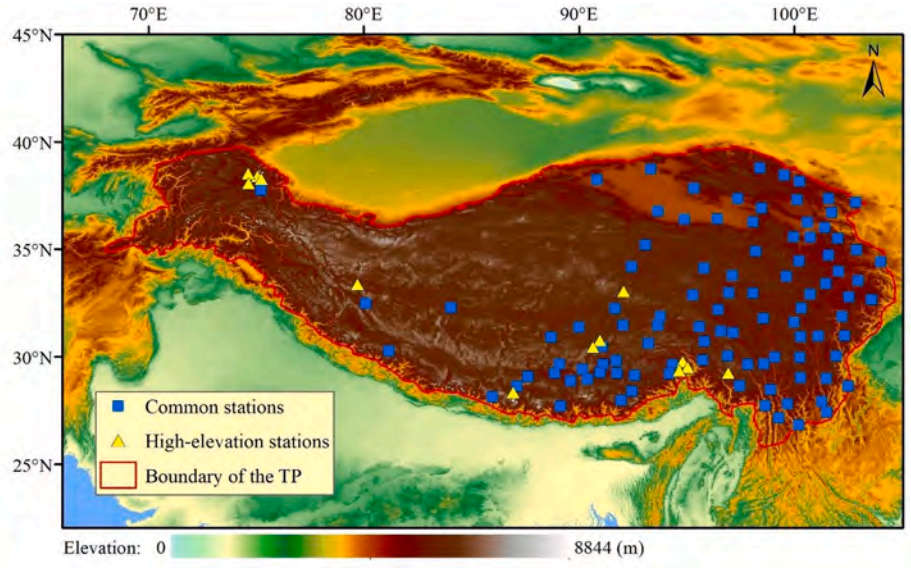


**Fig. 1.** Elevation map of the Tibetan Plateau (TP) and locations of the 100 common stations (blue squares) and 13 high-elevation stations (yellow triangles). The locations of several high-elevation stations are adjusted slightly for a better visualization to avoid overlapping in the plot, and their exact locations are listed in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**
Descriptions of the high-elevation automatic weather stations.

| Station | Longitude (°) | Latitude (°) | Elevation (m) | Land cover | Frequency | Available period |
|---|---|---|---|---|---|---|
| H01 | 86.95 | 28.37 | 4276 | Grasslands | 30 min | 2005.10.1–2008.6.27 |
| H02 | 90.97 | 30.78 | 4730 | Grasslands | 10 min | 2005.10.1–2008.6.26 |
| H03 | 75.06 | 38.41 | 3650 | Grasslands | 30 min | 2009.10.23–2012.7.18 |
| H04 | 75.02 | 38.29 | 4400 | Barren | 30 min | 2005.6.4–2006.9.14, 2009.10.29–2011.4.13 |
| H05 | 75.24 | 38.29 | 4900 | Barren | 30 min | 2003.5.18–2005.6.9 |
| H06 | 94.74 | 29.76 | 3326 | Grasslands | 10 min | 2007.1.1–2009.9.26 |
| H07 | 94.61 | 29.6 | 4390 | Open Shrublands | 30 min | 2005.8.19–2006.7.18, 2007.1.3–2008.5.14 |
| H08 | 94.71 | 29.66 | 4390 | Woody Savannas | 30 min | 2006.11.6–2009.8.1 |
| H09 | 79.7 | 33.39 | 4264 | Barren | 30 min | 2009.10.27–2012.8.22 |
| H10 | 96.93 | 29.25 | 4804 | Glacier | 10 min | 2012.1.5–2012.12.31 |
| H11 | 75.07 | 38.29 | 5900 | Glacier | 10 min | 2011.7.29–2012.7.9 |
| H12 | 92.08 | 33.07 | 5621 | Glacier | 10 min | 2009.1.1–2009.12.31 |
| H13 | 90.65 | 30.47 | 5800 | Glacier | 10 min | 2012.1.1–2012.12.31 |

**Table 2**
Description of eight reanalysis and land data assimilation products.

| Dataset | Full Name | Available time | Spatial resolution (Lon/Lat) | Temporal resolution |
|---|---|---|---|---|
| NNRP-2 | NCEP/DOE Reanalysis 2 Project | 1979–Current | 2.5/2.5 | daily |
| 20CRV2c | 20th Century Reanalysis Version V2c | 1851–2014 | 2.0/2.0 | 6-hourly |
| JRA-55 | Japanese 55-year Reanalysis Project | 1958–Current | 1.25/1.25 | 6-hourly |
| ERA-Interim | Interim version of European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalyses of the global climate | 1979–Current | 0.75/0.75 | 6-hourly |
| MERRA-2 | Modern-Era Retrospective analysis for Research and Applications, Version 2 | 1980–Current | 0.625/0.5 | 3-hourly |
| CFSR | NCEP Climate Forecast System Reanalysis | 1979–2010 | 0.5/0.5 | hourly |
| #CFSR-2 | NCEP Climate Forecast System Version 2 | 2011–Current | 0.5/0.5 | hourly |
| GLDAS-2 | Global Land Data Assimilation System Version 2 | 1948–2010 | 0.25/0.25 | 3-hourly |
| *GLDAS-2.1 | Global Land Data Assimilation System Version 2.1 | 2001–Current | 0.25/0.25 | 3-hourly |
| ERA5 | Fifth generation of ECMWF atmospheric reanalyses of the global climate | 1979–Current | 0.25/0.25 | hourly |

# CFSR-2 is used only for 2011–2014 for extending the CFSR data.
* GLDAS-2.1 is used only for 2011–2014 for extending the GLDAS-2 data.

Comparison of their performances in estimating TLR shows that TERRA nighttime LST is the best for estimating monthly TLR on the TP (Zhang et al., 2018a). This is attributed to the fact that TERRA LST shows higher validation accuracy than AQUA LST over the TP (Min et al., 2015; Yu and Ma, 2011; Zhang et al., 2018b). In addition, comparing with MODIS nighttime LST, MODIS daytime LST is more susceptible to the pixel-scale heterogeneity (Zhang et al., 2016b). MODIS TERRA daily land surface temperature data (MOD11A1) version 6 during 2002–2014 with about 1-km resolution are thus obtained for calculating the TLR at each reanalysis grid block. Considering the data quality and availability, only LST values with quality control (QC) flag <3 (i.e., the average error of LST is ≤3 K) are used. MODIS annual land cover type data (MOD12Q1) version 6 during the same period are also used for creating the water mask and providing the land cover information of station locations. All the elevation information of MODIS pixels and the reanalysis grid blocks are calculated based on the 30 arc second (~1km) resolution elevation data created by averaging the original 3 arc second (~90 m) Shuttle Radar Topography Mission (SRTM) data. The elevation, longitude and latitude information of all the 113 stations are obtained from CMA or the individual data provider. In addition, the gridded specific humidity (SH) derived from CFSR are used for analyzing the spatial and temporal patterns of TLR, which were found to show higher accuracy in the TP than ERA-Interim, MERRA and GLDAS (Wang and Zeng, 2012). The multiyear average snow cover days (SCD) are calculated from a seamless daily snow cover dataset combined from MODIS daily snow cover data and Interactive Multisensor Snow and Ice Mapping System (IMS) that show relatively good accuracy on the TP (Yu et al., 2016).

## 3. Methods

We create a new 1-km daily mean air temperature dataset for the period of 1980–2014 by following the three steps shown in Fig. 2. In the first step, a new method for estimating spatially continuous, multiyear average monthly TLR from MODIS LST is developed with consideration of standard error of elevation (SDE) (Fig. 2a). In the second step, the MODIS-estimated TLRs are used for downscaling air temperature of eight reanalysis datasets, and accuracy of the eight downscaled datasets is also evaluated and compared (Fig. 2b). In the third step, the downscaled 1-km air temperatures from eight reanalysis datasets are further integrated together with five other auxiliary variables (i.e., latitude, longitude, elevation, SCD and Julian day) by the best machine learning model among Cubist, Gradient Boosting and Random Forests (Fig. 2c). The detailed descriptions of the three steps are given below.

### 3.1. Estimating spatially continuous monthly TLR from MODIS LST

Since TLR is only reliable for local areas, several studies have used local samples within a grid block to calculate the grid-based TLR (Wang et al., 2018; Zhang et al., 2018a). However, the average cloud cover of daily MODIS/TERRA nighttime LST is as high as ~40% for the TP (Zhang et al., 2016a). This often renders insufficient valid LST pixels/samples for producing a reliable TLR estimate due to either missing data or small elevation changes within a defined area (Wang et al., 2018). To resolve this problem, a new method for estimating TLR from MODIS LST is developed, and Fig. 3 is the diagram of the method implemented in the following three steps:

(1) Step 1: The multiyear average, monthly MODIS LSTs are calculated by temporally averaging all the valid LST values for each 1-km MODIS LST pixel during 2002–2014. To remove the effects of water, if the 1-km LST pixels contain one or more 500 m pixels

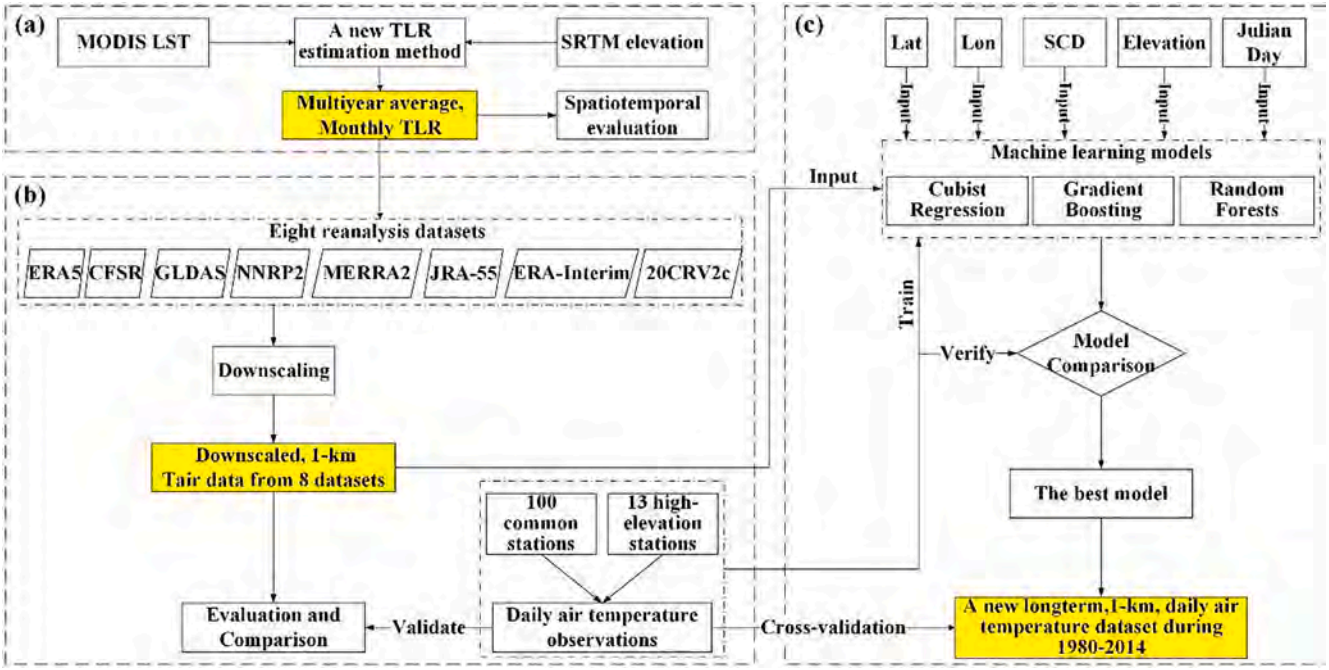


**Fig. 2.** Flow chart describing the three main parts of this study including estimating the multiyear average monthly TLR (a), downscaling 8 types of reanalysis datasets using the MODIS-estimated TLR (b), and integrating the downscaled data to create a new long-term, 1-km, daily air temperature dataset (c). Lat: latitude; Lon: longitude; SCD: multiyear average annual snow cover days.

that are typed as water body from MCD12Q1 dataset, the LST pixels are removed from calculation. For reducing the bias due to high cloud blockage, only the pixels with temporal coverage greater than 1/3 are included in the calculation of monthly TLR. The calculation builds a linear regression between multiyear monthly average LSTs and elevations within each reanalysis grid block via

$$LST_m = -a \times Z + b \quad (1)$$

where, $LST_m$ is the multiyear monthly average LST of each 1-km MODIS pixel, Z is the pixel elevation, and $a$ and $b$ are regression coefficients. The slope, $a$, is taken as the MODIS-estimated TLR, called "original TLR". For ERA5 with the spatial resolution of 0.25°, a reanalysis grid block contains about 900 1-km MODIS LST pixels. There are always many "exceptional TLRs" present in the original TLRs; the "exceptional TLRs" are defined as those with absolute TLR values greater than 0.015 °C/m based on the fact that TLR generally cannot be steeper than 0.015 °C/m according to numerous related studies (Kattel et al., 2015; Li et al., 2013; Rolland, 2003). The exceptional TLRs are dealt with in the next step described below. In addition to exceptional TLRs, there is also a problem of data missing. If the number of valid LST pixels within a reanalysis grid block is less than 4, its TLR is considered missing. The missing data account for about 10% of the original TLRs.

(2) Step 2: For automatically removing exceptional TLRs and filling in the missing data, a dynamic checking procedure is conducted by applying two types of statistic index to all the valid 1-km LST pixels within a reanalysis grid block. The statistics are the standard deviation of the elevation change (SDE) and the correlation coefficient (R) between LST and elevation. To examine the reliability of the estimated TLRs, an index, called "α", is firstly defined as the percentage of exceptional TLRs in all the original TLRs. Index α is used here as an error rate and we hypothesize that the exceptional TLRs are resulted from "inapplicable" LST/elevation samples within the reanalysis grid. The "inapplicable" LST/elevation samples may be caused by a low correlation between LST and elevation or by an exceptional TLR due to a small elevation range. For example, both Wang et al. (2018) and Zhang et al. (2018b) indicate that a relatively high correlation coefficient between temperature and elevation should be obtained for a reliable TLR estimation. Huld and Pascua (2015) point that the TLRs are invalid it they are estimated from the reanalysis grid with elevation range less than 200 m. Thus, it is critical to ensure that the samples for TLR calculation are "good" enough, meaning that the correlation between LST and elevation should be adequately strong and that the elevation change should be sufficiently large. Despite that the definition of R is clear, the variation degree of elevation can be measured using either SDE or elevation range (ER) (i.e., the maximum elevation – the minimum elevation). The three statistical parameters, i.e., correlation coefficient (R) between LST and elevation, SDE and ER, are compared in the following way: for each candidate parameter, detailed sensitivity tests are conducted by changing the thresholds of them with the ranges of 0.00–1.00, 0–300 m and 0–300 m for R, SDE and ER, respectively. Within the ranges, the thresholds are changed at the steps of 0.01, 1 m and 1 m for R, SDE and ER, respectively. Selection of the parameter and the corresponding threshold values is determined by checking the response of α. The optimal parameter should give small enough α (which is considered less than 1% here) with minimum loss of samples. The results based on all original TLRs from ERA5 grids (the number of samples is as large as ~220,000) are shown in Supplementary Fig. S1. The results suggest that, although increasing the thresholds of R, SDE or ER decreases α gradually from the initial value of ~14%, the decreases are different for different statistics. If we want to make α less than 1%, R needs to be ≥0.88. This however makes the percentage of remaining samples greatly reduced to less than 16%, which cannot be acceptable (Supplementary Fig. S1a). By contrast, SDE and ER are better in that the remaining samples are still as high as ~63% and ~61%, respectively, when α becomes less than 1% (Supplementary Fig. S1b). Because of the higher percentage of remaining samples and the higher sensitivity to α, SDE is chosen as a checking parameter. It should be

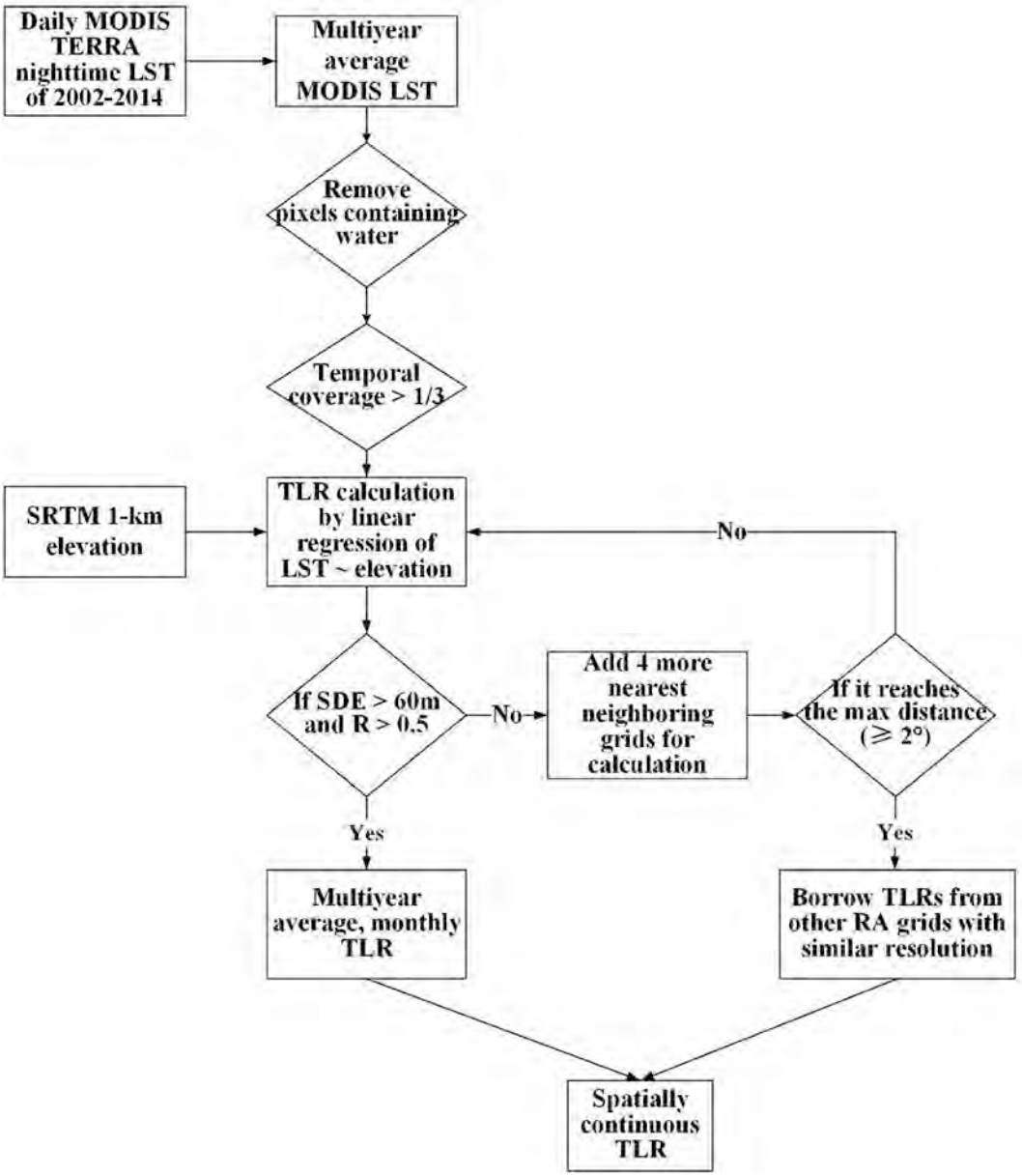


**Fig. 3.** Flow chart describing how TLRs are estimated from MODIS LST based on dynamic SDE checking. SDE means the standard deviation of elevations. R means correlation coefficient. RA means reanalysis.

noted that, since TLR is only reasonable when the correlation between temperature and elevation is relatively high, R is additionally constrained to be larger than 0.5. Supplementary Fig. S1c shows that even when R is constrained to be >0.5, SDE is still very efficient, in that α starts being less than 1% when SDE increases to ~60 m. Thus, the combined thresholds of SDE > 60 m and R > 0.5 are used here. The same tests are conducted for the other seven types of reanalysis datasets, and similar results are obtained, although their optimal SDE thresholds are smaller than 60 m. Then, for both the situations of missing data and those not satisfying the thresholds of SDE and R, more LST/elevation samples from four nearest neighboring reanalysis grid blocks are added for TLR calculation and checking. If the TLR is still missing or the criteria of SDE and R are not met, another four nearest neighboring grid blocks are added. This process continues until a reliable TLR is obtained or the added grid blocks reach the maximum distance to the original grid block, which is defined as 2° based on our previous studies (Zhang et al., 2018a; Zhang et al., 2018b).

(3) Step 3: The final step is to fill the remaining missing TLRs by borrowing values from the nearest valid grid TLR of other reanalysis datasets with similar resolutions which are already developed in Step 2. Basically, this process follows the order of similarity in the grid sizes of reanalysis data. For example, if an ERA5 grid (0.25°) has a missing TLR, it will take the TLR value from the nearest GLDAS grid (0.25°). If the TLR is still missing, then it moves to the nearest CFSR grid (0.5°) to check whether a valid TLR exists. By that analogy, most of the missing data can be filled in with relatively reliable TLRs. After this process, the missing TLRs only account for less than 1% and are further interpolated using the inverse distance weighted (IDW) method based on four nearest valid TLRs. Finally, the spatially continuous TLR dataset is created, and its spatiotemporal evaluation is further conducted by comparing it with multiple meteorological variables and other related studies.

### 3.2. Application of the MODIS-estimated TLR for downscaling reanalysis datasets

Using the MODIS-estimated TLRs, the original gridded air temperature from all the eight reanalysis and assimilation datasets with large resolutions of 0.25–2.5° are downscaled to 0.0083° (~1-km) pixel. This process involves two procedures including the (horizontal) weight calculation and the (vertical) elevation correction (Zhang et al., 2018a). The correction uses the weights estimated using bilinear (BIL) (Bracegirdle and Marshall, 2012; Quiquet et al., 2018; Wang et al., 2016b) or IDW (Jarvis and Stuart, 2001; Stahl et al., 2006; Zhang et al., 2013) interpolation methods. A total of six interpolation schemes are tested here. In the first scheme, called "ORG", the 1-km pixel value is directly assigned with the original grid value without weight calculation or elevation correction. In the second scheme, called "BIL", the original values of the nearest 4 neighboring grids are interpolated to the location of the 1-km pixel using the bilinear method. The third scheme, called "IDW", is similar to BIL but uses the IDW interpolation. In the fourth scheme, called "ORG + TLR", the original grid value is directly corrected from the average elevation of the grid to the 1-km pixel elevation using TLR. In the fifth scheme, called "BIL + TLR", the original values of the 4 nearest neighboring grids are all firstly corrected from their grid average elevations to the elevation of 1-km pixel and then interpolated to the location of 1-km pixel using the bilinear method. The sixth scheme, called "IDW + TLR", is similar to "BIL + TLR" but uses IDW for weight calculation.

The six schemes are then compared with respect to downscaling accuracy for all the eight datasets, and the best scheme is selected. The downscaling accuracy is measured by using the root mean square deviation (RMSD) between downscaled data and station observations according to multiple air temperature evaluations (Benali et al., 2012; Fu et al., 2011; Zhang et al., 2016a). As mentioned before, such evaluations are conducted at both common stations and high-elevation stations for obtaining more reliable results.

### 3.3. Integrating the downscaled air temperature datasets for creating a long-term 1-km data product over the TP using machine learning models

In this section, the downscaled air temperatures are further integrated to create a 1-km daily mean air temperature data product of 1980–2014 over the TP.

#### 3.3.1. Machine learning models

Machine learning methods are used for integrating downscaled air temperatures from the eight reanalysis and assimilation datasets, because of the wide applications and good performances of the methods in air temperature estimations (Noi et al., 2017; Rao et al., 2019; Xu et al., 2018; Zhang et al., 2016a). A total of 13 variables are selected as predictors including the downscaled air temperatures from ERA5, GLDAS, CFSR, MERRA-2, ERA-Interim, JRA-55, 20CRV2c and NNRP-2 as well as longitude, latitude, Julian day (i.e. day of year), SCD and elevation.

We tested and compared three types of machine learning models including Cubist regression (CR), random forests (RF) and gradient boosting (GB), which all use ensemble methods and decision trees. Although CR and RF have been used for daily air temperature estimation and the three models have been all compared for estimating monthly air temperatures (Noi et al., 2017; Rao et al., 2019; Xu et al., 2018; Zhang et al., 2016a), GB has not been used for estimating daily air temperature on the TP. In comparison with monthly air temperature, daily air temperature is more useful to hydrological or climate modeling that needs daily inputs. However, estimating daily air temperature is more difficult than estimating monthly air temperature, which is evidenced by reported estimation accuracy. For instance, the estimation accuracy (RMSD) over the TP for monthly air temperature is mostly clustered around 1.0 °C (Huang et al., 2017; Xu et al., 2018; Zhu et al., 2019) while that for the daily is generally about 2.0 °C (Rao et al., 2019; Zhang et al., 2016a; Zhu et al., 2013). In addition, estimating daily air temperature is more computationally expensive, since the number of samples used for estimating daily air temperature is significantly larger than that for the monthly air temperature estimation.

The three methods all have several parameters that can be tuned. Developed by Quinlan (2001), CR can create a number of committee rule-based models to compensate the predictions from each other or make use of nearest neighbor training points to improve the final predictions. This makes that CR generally has two parameters: *committees* (i.e., the number of committee rule-based models) and *neighbors* (i.e., the number of nearest neighbor training points). RF is proposed by Breiman (2001) and generally creates many random trees, with each tree model trained using bootstrap sampling. To alleviate the correlation problem of tree models with similar structures, RF randomly selects several predictor variables for each splitting of the model tree. Thus, *ntrees* (i.e. number of random trees) and *mtry* (i.e. the number of randomly selected predictors) are the two tuning parameters. GB (Friedman, 2001, 2002) is a framework that can combine a number of "weak prediction models" for an ensemble prediction. It corrects the errors from a prior model through a number of iterations and create a new model combination. The situation for GB is more complex owing to the large number of parameters. The six most widely used parameters are considered here: *eta*, *gamma*, *max_depth*, *min_child_weight*, *subsample* and *colsample_bylevel* which are all booster parameters. These parameters are repeatedly tuned for improving air temperature estimation accuracy in the meantime to avoid the problem of overfitting. The detailed definitions of these parameters can be found in Chen et al. (2015).

All of the three types of models were calibrated in accordance with their different model features and the suggestions given in literatures such as Kuhn and Johnson (2013) and Zhang et al. (2016a). According to the different computing burden, the parameters for CR, RF and GB were finally determined based on the 10-fold, 5-fold and 100-fold cross-validations, respectively. The tuning ranges, steps and the final selection of model parameters are listed in Table 3. Other parameters are determined manually based on our experience. For RF, the parameter of *ntrees* is fixed to 500 which is also used by Zhang et al. (2016a). Using larger values increases computational cost but with negligible improvement in accuracy, For GB, the parameter of *nrounds* (meaning the maximum number of iterations) is fixed to 500 prior to calibration. The three machine learning methods were implemented using the R packages "Cubist", "randomForest" and "xgboost", respectively.

#### 3.3.2. Model comparison and evaluation methods

After the calibration, the performances of the three models are compared by using the widely used leave-one-out cross-validation method. For each station, the machine learning model is trained by the observations from all the other stations and the RMSD and mean absolute deviation (MAD) are calculated using the predictions (from the

**Table 3**
Parameter tuning of Cubist regression, Random forests and Gradient Boosting models.

| Model | Parameter | Tuning Range | Tuning Step | Selected Value |
|---|---|---|---|---|
| Cubist | *committees* | 1–100 | 10 | 80 |
| | *neighbors* | 1–9 | 3 | 9 |
| Random Forests | *mtry* | 1–12 | 1 | 7 |
| Gradient Boosting | *eta* | 0.01–0.1 | 0.01 | 0.03 |
| | *colsample_bylevel* | 0.25–1 | 0.25 | 0.5 |
| | *gamma* | 0–1 | 0.05 | 0.05 |
| | *subsample* | 0.5–1 | 0.1 | 1 |
| | *max_depth* | 2–10 | 2 | 6 |
| | *min_child_weigtht* | 1–30 | 5 | 5 |

trained model) and observations for this station. The final measurements of validation accuracy is averaged from all the station-based RMSDs and MADs. The best model with the highest validation accuracy is selected for producing the final air temperature dataset.

#### 3.3.3. Evaluation of the produced air temperature dataset

The leave-one-out cross-validation accuracy from the best machine learning model is considered as the general accuracy of the estimated air temperature, and this is evaluated by the comparison with the best downscaled reanalysis dataset under the same condition. After the final air temperature dataset is created using the selected machine learning model, the new product is further evaluated by calculating the pixel-based annual temperature trends. The trend of each pixel is calculated based on linear regression between the annual average temperatures of the pixel and the time series (i.e. years 1980–2014) and the significance level (*P*-value) is calculated based on *t*-test. The error distributions are possibly affected by land cover, elevation or data independence, which is discussed below.

## 4. Results and discussions

### 4.1. The spatially continuous monthly TLR

#### 4.1.1. The performance of the proposed method for TLR estimation based on MODIS LST

The performance of the new method is evaluated by estimating TLR of July, when the problem of data missing is the most serious, for the ERA5 grid, and the results are shown step by step in Fig. 4. Fig. 4a shows the TLRs calculated directly from multiyear average monthly MODIS LST without additional procedures. This results in many exceptional values with the maximum absolute TLR reaching 0.68 °C/m which is normally impossible. The calculation also results in a large number of missing data, which could largely limit its applicability. Lots of exceptional values are around large water bodies. After removing the LST pixels with water (Fig. 4b), some exceptional values disappear, with the maximum absolute TLR decreasing to 0.36 °C/m, which is still unreasonable. At the same time, more missing data appear due to the water mask. However, after checking SDE iteratively and including more neighboring grids within the threshold distance (Fig. 4c), the number of exceptional TLRs is greatly reduced with the maximum absolute TLR quickly decreasing to 0.028 °C/m, especially for the Qaidam basin in the northeastern TP where about 1/3 of the basin is covered with deserts and elevation variation is relatively small. It is also noted that the missing data are reduced by ~84%. After filling missing values with those from other reanalysis grids with similar resolutions, all the missing data are eliminated, and the new results have reasonable ranges (Fig. 4d). The performances of the new method in estimating TLRs of all the other 11 months and seven reanalysis datasets are similar, and thus not shown here.

#### 4.1.2. Evaluating the spatial and seasonal pattern of MODIS-estimated TLRs across the TP

The spatially continuous TLRs in January for the eight reanalysis and assimilation datasets are shown in Supplementary Fig. S2. The spatial patterns are generally consistent among different grids, e.g., all of them showing steeper TLRs in the southeastern and northwestern parts in January. Similar results are obtained for all the other 11 months, and thus not shown here. However, the influence of grid resolution is observed that the TLR gradually becomes complex as the grid size decreases, with the most obvious evidence of more negative TLRs. In addition, the coarser resolution also shows a smoothing effect that the range (i.e., maximum – minimum) of TLR gradually decreases from the largest 0.033 °C/m (ERA-5) to the smallest 0.006 °C/m (NNRP-2). For simplicity, the following analysis of this section is based on ERA5 grid that has the highest resolution among the eight datasets.

In order to evaluate the spatial patterns of the MODIS-estimated

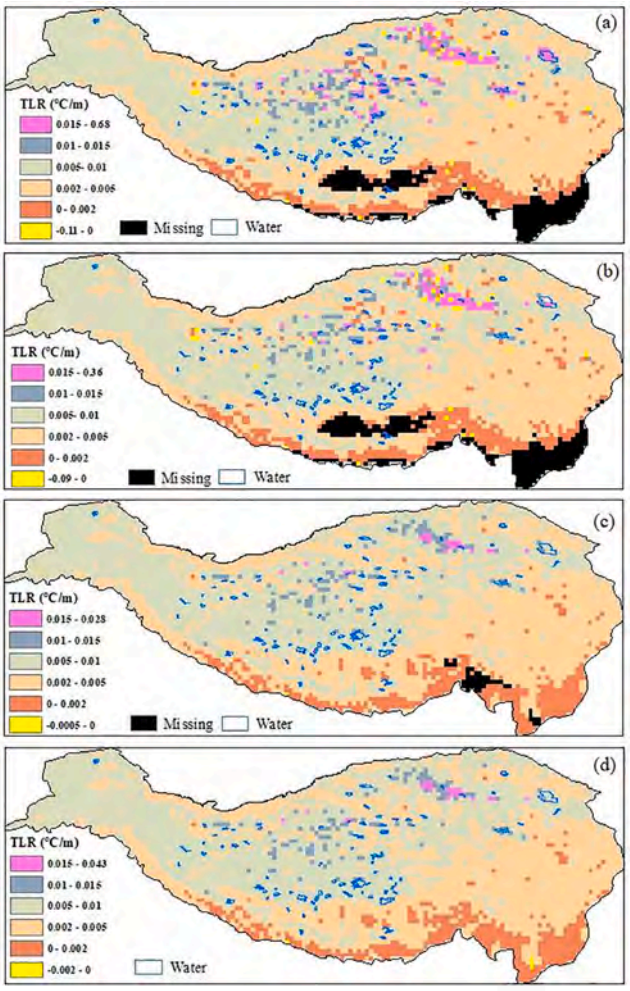


**Fig. 4.** Effects of the proposed method by showing the spatial distribution of the TLRs of July from the original TLRs (a), reprocessing by removing water pixels prior to TLR calculation (b), further processed based on SDE checking (c), and finally filling the missing data by borrowing values from other reanalysis grids with similar resolutions (d).

TLRs, the whole TP is divided into five sub regions including Northwestern TP, Central TP, Northern TP, Southeastern TP and Southern TP (Fig. 5). The spatial pattern of TLRs is very complex, and shows obvious seasonal variations determined by multiple factors. Humidity is a major factor, because the TLR becomes steeper from the northwest to the southeast with a spatial pattern similar to that of specific humidity shown in Fig. 5. This is particularly true for summer and autumn. In summer, the monsoon precipitation is mainly around the southeastern and southern TP and decreases towards the northwest, resulting in a similar humidity pattern. It has been demonstrated that more humid conditions can produce shallower TLRs due to the fact that the rising moist air tends to cool more slowly because of the released latent heat upon condensation (Kattel et al., 2017; Zhang et al., 2018a). Thus, the TLRs of southeastern and southern TP are obviously shallower, and the steepest TLRs are concentrated in the Qaidam basin with a large desert. The spatial pattern of TLR in autumn is similar to that in summer due to the similar pattern of humidity but with smaller differences among different regions caused by the reduced precipitation in the southeastern and southern TP. However, the underlying surface conditions, especially the snow cover and the effect of temperature inversion, can also be important disturbing factors in spring and winter. In spring, both the northwestern and southeastern TP have more snow cover. The high albedo of snow causes lower temperatures at high elevations, leading to

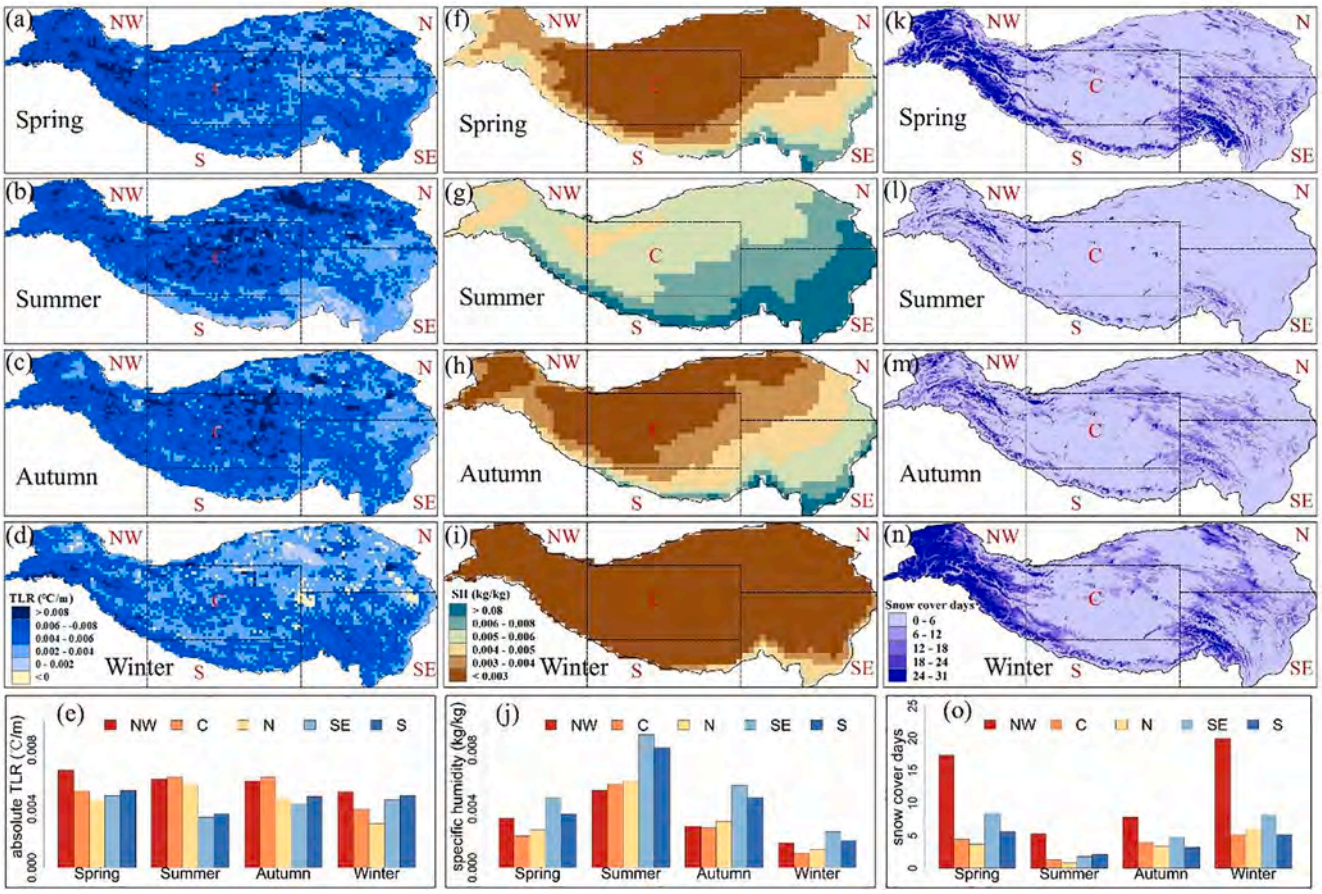


**Fig. 5.** Spatial distribution of seasonal TLRs (left), specific humidity (middle), snow cover days (right) and their averaged values in the 5 regions for 4 seasons (bottom). "NW": Northwestern TP; "C": Central TP; "N": Northern TP; "SE": Southeastern TP; "S": Southern TP. SH: specific humidity.

steeper TLRs. In winter, there are also two steep-TLR regions, including the northwestern and southeastern TP. Again, their higher snow cover frequencies are likely the cause of this. It may be surprising that the southern TP shows steeper TLR than the northern TP (Fig. 5e). Even higher specific humidity and less snow cover are found in the southern part, but it should be noted that there are a number of negative TLRs possibly caused by small elevation ranges or the inversion effect in northern TP, which largely balance the steep positive values. Most of the negative TLRs are found in the grids with a relatively small standard deviation of elevation indicating smaller elevation variation. This was also found by Wang et al. (2018). In addition, temperature inversion is a common phenomenon in winter in high-elevation areas where temperature may increase with elevation when the cold air sinks and pools in the valley, resulting in the coldest locations being the valley bottom (Minder et al., 2010; Rolland, 2003).

The MODIS-estimated TLRs are also compared with those obtained in previous studies around the TP. As for the magnitude of monthly TLRs, few cases with large absolute TLRs greater than 0.015 °C/m and even 0.01 °C/m are generated. This was also found by Jiang et al. (2016) and Li et al. (2013), who used CMA stations all over China including those in the TP. The negative TLRs have also been observed in some areas in TP or its neighboring areas (Jiang et al., 2016; Li et al., 2013), including the Qaidam basin (Wang et al., 2018). This phenomenon was noted to occur mostly in winter, which is consistent with this study. Since most studies use seasonal TLRs for temporal analysis, Supplementary Fig. S3 plots the seasonal patterns of TLRs, taken directly from the literature, reprocessed by averaging monthly values. These are compared with the MODIS-estimated TLRs. For northwestern TP, our results are consistent in both the magnitude and the pattern with a recent study (Sun et al., 2018) that uses not only nine common CMA stations but also nearby six high-elevation field stations. Due to the relatively high snow cover and low humidity (Fig. 5), TLR is the steepest in spring rather than in winter, which is possibly affected by inversion effects. The seasonal patterns of MODIS-estimated TLR are very similar in southeastern and southern TP, and are the steepest in spring and the shallowest in summer. The pattern in southeastern TP is different from that of Jiang et al. (2016), who found that winter has the steepest TLR. This may be due to the unequal definition of southeastern TP, and that there are a number of negative TLRs in winter in this study (Fig. 5d). The situation is a little complex for southern TP, i.e., the Himalayan regions where the southern and northern slopes are found to have different seasonal patterns (Kattel et al., 2013; Kattel et al., 2015). The biggest difference occurs for the winter TLR (Supplementary Fig. S3e), and the true conditions may be between our results and those of Immerzeel et al. (2014). While it is surprising that both the central and northern TP show totally different seasonal patterns with the steepest TLR in summer and the shallowest TLR in winter, these are also reported by Jiang et al. (2016) and Li et al. (2013). However, it should be noted that their humidity and snow cover are relatively low throughout the year indicating much weaker effects of them on the seasonal cycle (Fig. 5). Solar radiation may be responsible for the summer maximum in TLR due to the fact that strong surface heating from increasing solar input can produce steeper lapse rates in dry and warm environments (Blandford et al., 2008; Pepin et al., 1999). The winter minimum TLR is resulted from the combined effects of reduced solar radiation, lower temperature and the frequent cold air drainage, which can also be seen from the negative TLRs in the region (Fig. 5). Several studies have also analyzed the patterns of TLR in China and the TP, but their results are not used for comparison here due to the different definitions of sub-regions (Guo et al., 2016; Li et al., 2015). In summary, the spatial and seasonal patterns of MODIS-estimated TLRs

are considered to be reasonable.

### 4.2. Downscaled reanalysis data and evaluation of the MODIS-estimated TLR

#### 4.2.1. Comparison of different downscaling schemes

Supplementary Fig. S4 shows accuracy comparison of the six downscaling schemes among all the eight reanalysis and assimilation datasets. The three schemes using TLR (with mean RMSD of 3.2 °C) show obviously better accuracy than those without using TLR (with mean RMSD of 5.3 °C), indicating that the use of TLR can significantly improve the downscaling accuracy for all the eight reanalysis datasets. The ORG + TLR method (with mean RMSD of 3.3 °C) is slightly inferior to the other two schemes using TLR, but the BIL + TLR method (with mean RMSD of 3.1 °C) and the IDW + TLR method (with mean RMSD of 3.1 °C) have comparable performance. It indicates that the effect of weight calculation is less important than elevation correction using TLR which is also found by Zhang et al. (2018a). The IDW + TLR method is selected as the downscaling scheme here, because it shows similar accuracy with BIL + TLR and has higher computational efficiency.

#### 4.2.2. Comparison of the downscaled eight reanalysis and assimilation datasets

The accuracy comparison of all the eight downscaled reanalysis and assimilation datasets are shown in Fig. 6a. JRA-55 (mean RMSD: 2.0 °C) shows the highest accuracy followed by ERA-Interim (mean RMSD: 2.1 °C), MERRA-2 (mean RMSD: 2.3 °C), CFSR (mean RMSD: 3.3 °C), ERA5 (mean RMSD: 3.4 °C), 20CRV2c (mean RMSD: 3.5 °C), NNRP-2 (mean RMSD: 4.3 °C) and GLDAS (mean RMSD: 4.6 °C). The results based on all stations (i.e., both common and high-elevation stations) are slightly different from those based on high-elevation stations. For validation based on high-elevation stations, MERRA-2 is the best, and the accuracy from JRA-55 (mean RMSD: 2.68 °C), ERA-Interim (mean RMSD: 2.69 °C) and MERRA-2 (mean RMSD: 2.62 °C) are nearly the same. Due to the negligible accuracy difference among the best three datasets, following Zhang et al. (2016a), we conducted a multiple comparison based on a paired unequal variances *t*-test with Bonferroni correction (Dunnett, 1955). The results indicate that, although no significant difference was observed for high-elevation stations (Fig. 6c), significant differences ($P < 0.05$) are found for common stations (Fig. 6b). Thus, JRA-55 is considered the best among all the eight downscaled datasets, and it probably assimilates more observations from the common stations. It should be noted that the accuracy of the eight datasets significantly decrease when only using high-elevation stations for validation.

Another interesting finding is that the higher spatial resolution does not necessarily mean higher evaluation accuracy, because the best three datasets (i.e. JRA-55, ERA-Interim and MERRA-2) are all with medium resolutions ranging from 0.5° to 1.25°. This is true no matter for validation based on the common stations or the high-elevation stations. Although no study has compared the accuracy of air temperature from the eight datasets simultaneously on the TP, our results are partly consistent with other studies evaluating several of the datasets. For example, ERA-Interim is found to have higher accuracy in air

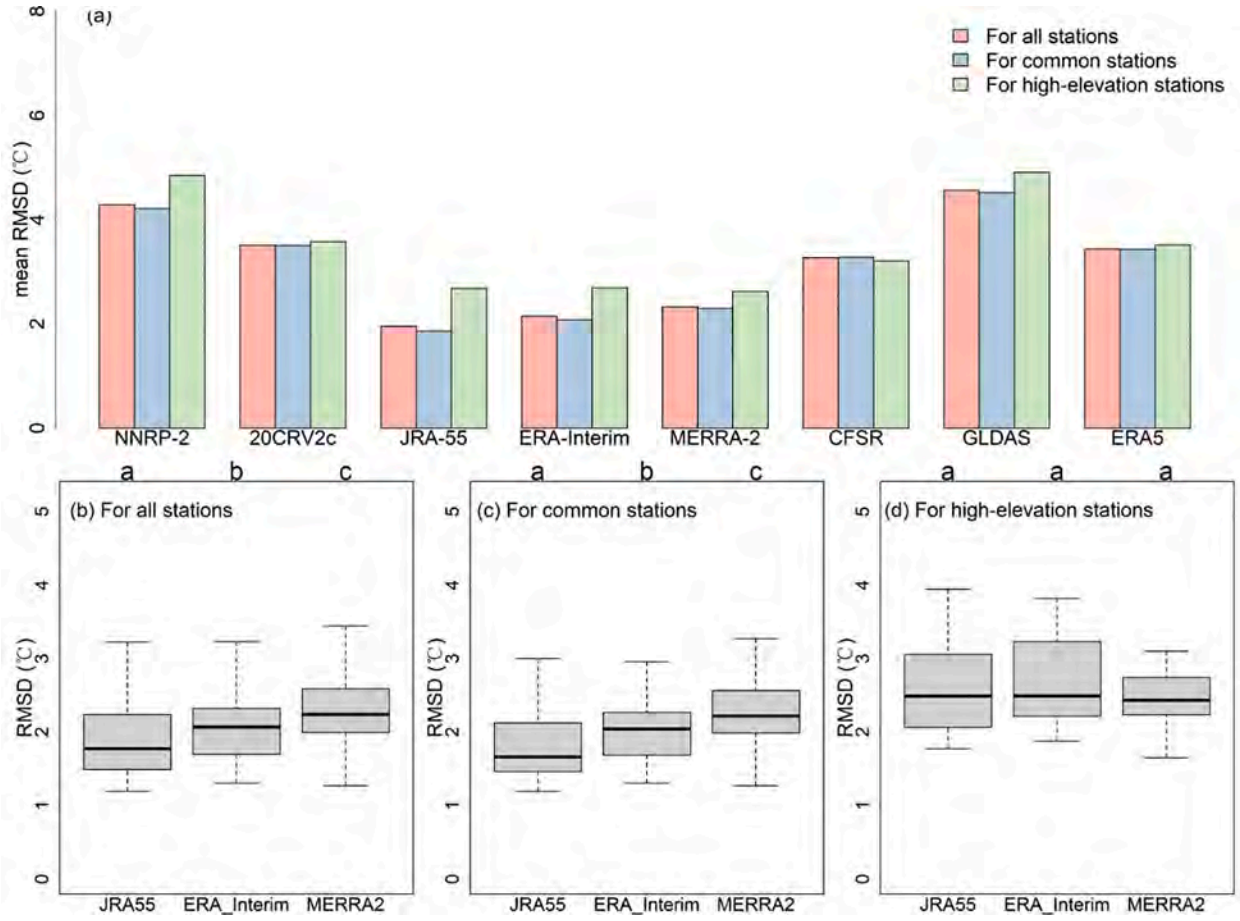


**Fig. 6.** Comparison of the accuracy of eight reanalysis and assimilation datasets downscaled using MODIS-estimated TLR (a), and the comparison between the best three datasets based on all stations (b), 100 common stations (c) and 13 high-elevation stations (d). The box and whiskers in (b), (c), and (d) are the distributions of station-based RMSD. In panels (b), (c) and (d), the letters at the top indicate the significance of the differences: the datasets with a same letter at the top indicate insignificant difference, otherwise indicate significant differences.

temperature than CFSR (Bao and Zhang, 2013), which is better than GLDAS (the second worst) and NNRP-1 (the worst) based on 63 CMA stations across the TP (Wang and Zeng, 2012). Chen et al. (2019) report that for daily air temperature in western China (including TP), MERRA-2 is the best, and that both JRA-55 and ERA-Interim are the second best, followed by CFSR and NNRP-2. Since independent high-elevation observations are used in this study, our comparison results are expected to be more reliable than those only using common station observations.

#### 4.2.3. The reliability of using MODIS-estimated TLR compared with the station-based TLR

In order to better evaluate the reliability of the estimated TLRs, the "station-based TLRs" are also calculated for comparison. For each reanalysis grid block containing more than two stations, the monthly station-based TLRs are calculated based on all the stations located within the corresponding grid block. It should be noted that the TLR may vary among different grid resolutions (Supplementary Fig. S2), and that the station-based TLRs can be sensible to the stations used (Zhang et al., 2018a). Thus, the MODIS-estimated TLRs cannot be directly compared with the station-based TLRs due to the inconsistency between different scales and the insufficiency of stations. We then compare the MODIS-estimated and station-based TLRs based on their performances in downscaling accuracy (Fig. 7). It should be also noted that only two situations of different station numbers are considered here due to the limited number of stations, and that the grid blocks with $\geq$2 stations include those with $\geq$3 stations. In addition, for the grid blocks with $\geq$3 stations, only 3 types of reanalysis datasets that have sufficient grid blocks ($\geq$3) meeting this criteria are considered. For a fair comparison, the ORG + TLR scheme is used here and only common stations are used. It is obvious that the MODIS-estimated TLRs are significantly better than the station-based TLR for all the types of reanalysis and assimilation datasets especially for grid blocks with $\geq$2 stations (Fig. 7a). Although the downscaling accuracy seems to be improved for cases with more than 3 stations (Fig. 7b), the accuracy of the MODIS-estimated TLRs is still obviously better than those using station-based TLR. This is attributed to two reasons: (1) MODIS LST can provide much more temperature information than the sparsely distributed stations within a grid block, and thus results in a more representative TLR; (2) a reliable TLR needs sufficient number of stations/observations, but for the station-based TLR, some grid blocks with very limited number of stations (such as 2 or 3) often lead to abnormal estimates of TLR, which further produces much lower accuracy. The second reason further suggests the importance of developing the MODIS-estimated TLRs.

#### 4.2.4. Uncertainties about using the MODIS-estimated TLR and some limitations

Although the lapse rate of LST is used for downscaling air temperature in this study, it should be noted that LST is not equivalent to air temperature. Because of the intense interaction between land surface and atmosphere, it has been widely observed that air temperature holds a strong correlation with LST (including MODIS LST) (Benali et al., 2012; Fu et al., 2011; Vancutsem et al., 2010). Furthermore, previous studies indicate that the spatial pattern of air temperature is similar to that of MODIS nighttime LST (Mutiibwa et al., 2015; Oyler et al., 2016). Wang et al. (2016a) also demonstrate that, in comparison with the monthly TLRs derived from sparse stations, those estimated from MODIS nighttime LST significantly improve simulation accuracy of snow processes in the upper Yellow River basin. Based on these findings, our previous work further analyzes the feasibility of using MODIS LST to estimate monthly TLR of the TP. We find that the MODIS-estimate TLR can well capture the spatial and temporal variation of station-observed TLR, in that the correlation coefficient between monthly TLRs estimated from TERRA nighttime LST and those from station air temperature observations is 0.88 and that the spatial pattern of them are similar (Zhang et al., 2018a). However, we acknowledge that the fundamental differences between LST and air temperature may pose uncertainties in the proposed TLR estimation method in certain conditions. He and Wang (2020) indicate that LST and air temperature have different sensitivity to surface energy balance that may lead to different lapse rates between them, although they also find that the multiyear average monthly variations of TLRs from LST are very similar to those from air temperature. The differences between their TLRs could be large in areas with complex land covers. Considering that the TP is a large region, it is more realistic to develop the gridded TLRs as in this study rather than a spatially constant one in the entire region or sub-regions (Li et al., 2013; Wang et al., 2016a). The effects of land cover types may gradually diminish with the grid size decreasing from 2.5° (NNRP2) to 0.25° (ERA5), although the sub-grid variability of land covers needs to be addressed in future studies.

Most of the MODIS-estimated TLRs were estimated within the original reanalysis grid, which are considered locally reliable and thus referred to as "local TLR" here. However, the high cloud cover of the TP may impact the proposed TLR estimation method, making some reanalysis grid fail to produce the "local TLR" due to missing data. As discussed in section 3.1, when the conditions of SDE and R cannot be met or LSTs are not available within the original grid block, more neighboring grid blocks are added for calculating TLR. The calculated TLR are here referred to as "extrapolated TLR". In total, the "extrapolated TLRs"

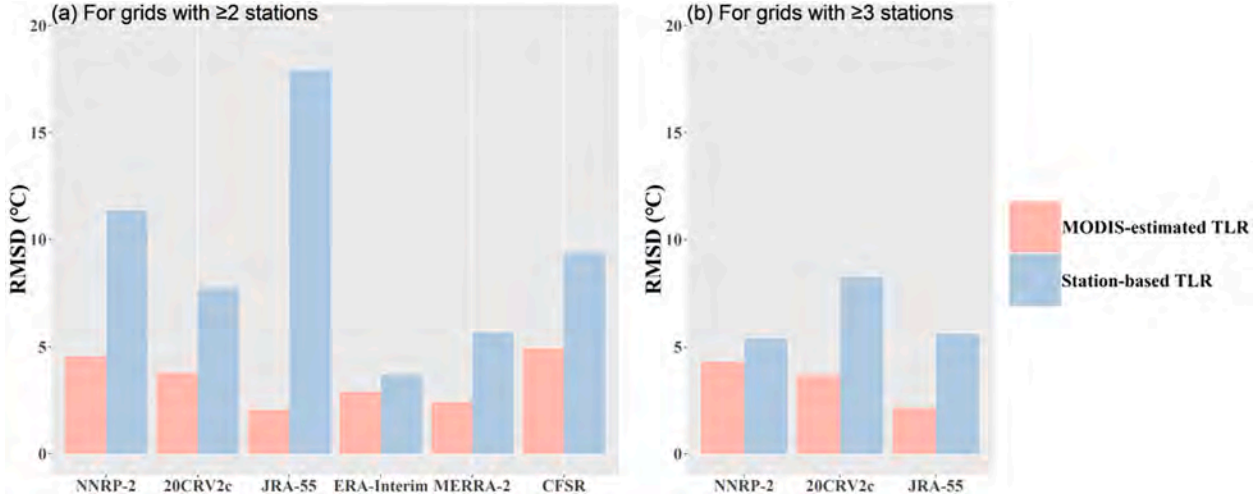


**Fig. 7.** Comparison of downscaling accuracy between from using MODIS-estimated TLR and from using the station-based TLR for two situations including when grids with more than 2 stations are considered (a) and when girds with more than 3 stations are used (b).

caused by high cloud cover or small elevation variation account for about 10% of the developed TLRs for the eight datasets. To evaluate the uncertainty from the extrapolated TLRs, the downscaling accuracy of using the extrapolated TLRs and the local TLRs are compared based on only the stations where both the local and extrapolated TLRs were used for downscaling. The results (Supplementary Fig. S5a) show that the accuracy of using the extrapolated TLRs is indeed lower than that of using the local TLRs for most of the eight datasets expect for NNRP-2 and 20CRV2c of which both the grid resolution and the downscaling errors are already very large. In general, the average increase in RMSD across the 8 datasets is ~ 0.2 °C. Though the maximum increase in RMSD can be as large as 0.7 °C (for CFSR) (Supplementary Fig. S5a), it is still considered highly acceptable in comparison with the poor performance of the station-based TLRs. To provide useful information on such uncertainty, a by-product is created to together with the developed monthly TLR product with the same temporal and spatial resolution. This by-product contained a "quality control code" that records detailed information on whether a "local TLR" is successfully produced for every grid block in each month.

Another uncertainty may be that this study estimates the multi-year average monthly TLRs rather than the year-to-year monthly TLRs which are commonly used in previous studies. The use of the multi-year average monthly TLR is mainly due to the fact that MODIS LSTs are only available after 2001 while the estimated TLRs need to be used for downscaling long-term data (from 1980). One advantage of using multiyear average monthly TLR is the fewer missing data compared with the year-to-year monthly TLRs due to the high cloud cover of MODIS LST. Although the monthly TLRs in TP are reported to show significantly varying trends during 1962–2011 (Li et al., 2013), the averaged (from all common stations) maximum absolute differences between the year-to-year monthly TLRs and multi-year average monthly TLRs are less than 0.0008 °C/m for the four best downscaled datasets indicating limited effects (Supplementary Fig. S5b). We further compared their performances on downscaling the four datasets and found very small accuracy differences with the mean differences in RMSDs across all common stations being less than 0.1 °C. The accuracy of using the multiyear average monthly TLR is even slightly better than that of using year-to-year average one (Supplementary Fig. S5c).

It should be noted that, although previous studies considered latitude or longitude in Eq. (1) for calculating TLR (Li et al., 2013; Rolland, 2003), we did not consider them based on our experience that it may cause the problem of multicollinearity and thus unstable estimates of TLR (Zhang et al., 2018a; Zhang et al., 2018b). In addition, TLR may be affected by other surface characteristics (Firozjaei et al., 2020) such as vegetation, slope aspects and leeward/windward, which will be considered in a future study.

### 4.3. Estimated air temperature by integrating multiple downscaled datasets

#### 4.3.1. Comparison of machine learning methods

The comparison results of accuracy from all the machine learning models indicate that GB performs the best followed by RF and CR (Fig. 8). This is consistent among the three different situations including all stations, common stations and high-elevation stations. For the comparison of accuracy in all stations, the mean RMSD of GB is 1.71 °C, which is only slightly smaller than that of RF (1.74 °C) and CR (1.88 °C). However, for high-elevation stations, GB is obviously superior to RF and CR with the mean RMSDs of 1.88 °C versus 2.10 °C and 2.32 °C. Thus, GB is selected as the final model for integrating the eight downscaled reanalysis and assimilation datasets of air temperature.

#### 4.3.2. Evaluating the integrated air temperature of 8 downscaled reanalysis datasets using GB

Comparison between the further integrated air temperatures using GB model and those directly from the downscaled JRA-55 shows that the

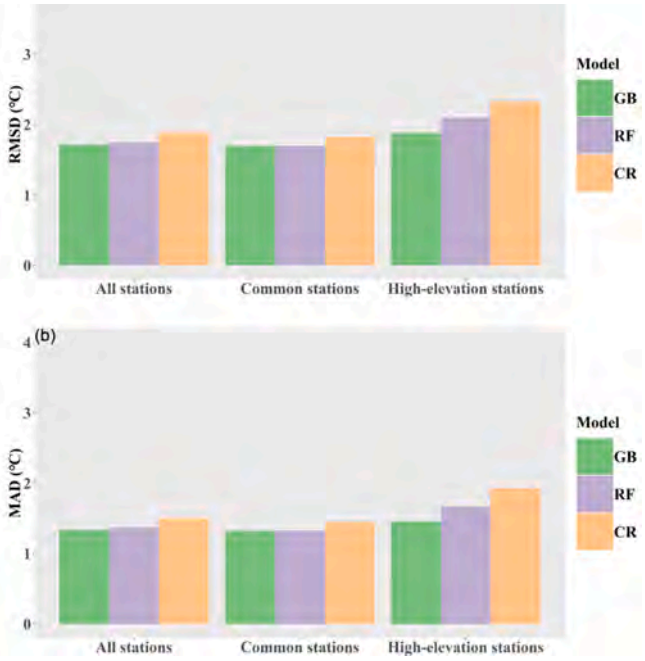


**Fig. 8.** Comparison of mean RMSD (a) and MAD (b) between GB, RF and CR models. The RMSD and MAD are averaged from all stations.

GB-integrated are better in both common stations and high-elevation stations (Fig. 9a). The validation RMSD (MAD) of the GB-integrated is 1.7 (1.3) ± 0.3 (0.3) °C in common stations (Number = 100) and 1.9 (1.5) ± 0.5 (0.4) °C in high-elevation stations (N = 13). Though the reduction in RMSD is only 0.17 °C for common stations, the GB-integrated air temperatures show obviously higher accuracy than the downscaled JRA-55 for high-elevation stations with the reduction in RMSD as large as 0.81 °C making its RMSD for high-elevation stations even less than 2.0 °C. This conclusion is also supported by the scatter plots of the GB-integrated and the downscaled JRA-55 (Supplementary Fig. S6) showing that the GB-integrated significantly improve the estimation accuracy in high-elevation stations compared with the downscaled JRA-55. The spatial distributions of RMSDs are also plotted in Supplementary Fig. S7 showing that for most stations the GB-integrated are truly better than the downscaled JRA-55, especially for high-elevation stations (Fig. 9b). We further evaluated the accuracy on different land cover types and elevation ranges. The GB-integrated air temperature shows better accuracy than the downscaled JRA-55 in the two elevation situations including both above and below 3555 m (the median of all elevations of stations), and its superiority is much more evident for stations higher than 3555 m (Fig. 9c). It should also be noted that the accuracy difference between the two different elevation ranges is small (~0.1 °C) for the GB-integrated. The correlation analysis between accuracy and elevations also shows that there is no significant correlation for the GB-integrated data whereas the accuracy of the downscaled JRA-55 truly shows significant ($P < 0.05$) positive correlation with elevation. All analyses indicate that the GB-integrated air temperature may be more reliable than the downscaled JRA-55 in high-elevation areas. Similar comparison results are also found on different land cover types. Due to the limited number of stations, the original land cover types of the 113 stations obtained from MODIS land cover product are reclassified into three categories including "urban and barren", "vegetated" and "glacier". The GB-integrated air temperature performs consistently better than the downscaled JRA-55 on all the three land cover types (Fig. 9d), particularly for glacier surfaces with a sharp decrease in RMSD of ~ 0.8 °C.

Fig. 10 compares the temporal trends calculated from the new product and those gathered directly from the eight reanalysis and

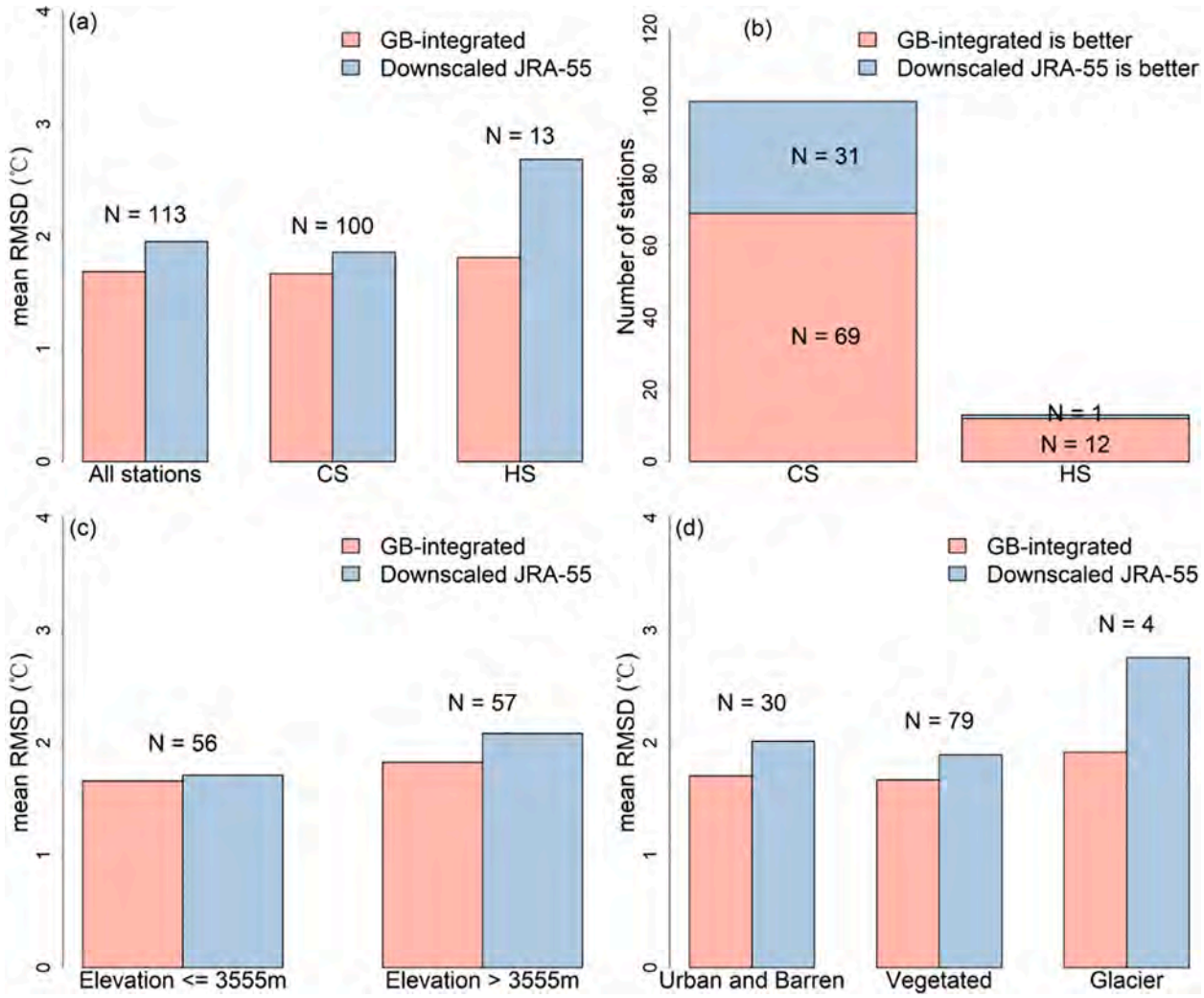


**Fig. 9.** Comparison of the GB-integrated and downscaled JRA-55 air temperature grouped by stations (a and b), elevations (c) and land cover types (d). CS: common stations; HS: high-elevation stations. N: number of stations.

assimilation datasets. The trends produced by the GB-integrated are more similar with JRA-55 and ERA-Interim both of which have higher validation accuracy than the other six datasets. This is consistent with an evaluation study of the suitability of 12 types of reanalysis datasets for investigating regional warming trends in China, which also find that ERA-Interim and JRA-55 show the best performance in simulating air temperature climatology (Zhou et al., 2018). However, the downscaling using TLR enables the GB-integrated to obtain more detailed air temperature trends at higher spatial resolution and the accuracy improvement in air temperature data through machine learning is expected to produce more accurate temperature trends at higher resolutions. It should be noted that though the new product generally has relatively good accuracy in daily air temperature estimation, it may not meet the strict accuracy requirements for practical warming trend analysis. Continued efforts should be made in future to collect more observations to further evaluate the suitability of our new product for investigating the warming trends over the TP.

*4.3.3. The importance of the use of TLR and high-elevation observations*

Both the use of TLR and high-elevation observations may have important effects on the air temperature estimation in the present study. In order to distinguish their effects, another three GB models are re-trained. The first is trained using only observations from commons stations and with the same predictor variables as the original GB model, called "GB_No_HS". The second one is trained using observations from both common and high-elevation stations but replacing the downscaled air temperature of eight datasets with their original grid values from the predictors of GB model, called "GB_No_TLR". The last GB model is trained both using only common station observations and taking original reanalysis data as predictors, called "GB_No_Both". The same leave-one-out cross-validation results are shown in Fig. 11. The strong influence of TLR can be seen from the obvious accuracy decline from GB to GB_No_TLR with the relatively large increase in RMSD of 0.6 °C and 1.1 °C for both common and high-elevation stations, respectively. It is interesting that though elevation has been considered in GB_No_TLR model as a predictor variable, its accuracy is still much lower than the original GB model and even lower than the downscaled JRA-55 indicating that the use of TLR is much more efficient than simply incorporating elevation into a statistical model. The important influence of high-elevation observations mainly lies in high-elevation areas that can be seen from the clearly lower accuracy of GB_No_HS than the original GB model in high-elevation stations. It is very interesting that for high-elevation stations, the accuracy reduction of GB_No_Both seems to be greatly exaggerated compared with both GB_No_TLR and GB_No_HS considering that for common stations they are much closer. It implies the great importance of both high-elevation observations and TLR. When reanalysis air temperature data need to be used in high-elevation areas, either reliable TLR or independent high-elevation observations for correction can produce a much better accuracy, however, without using either of them, the estimated air temperature would be

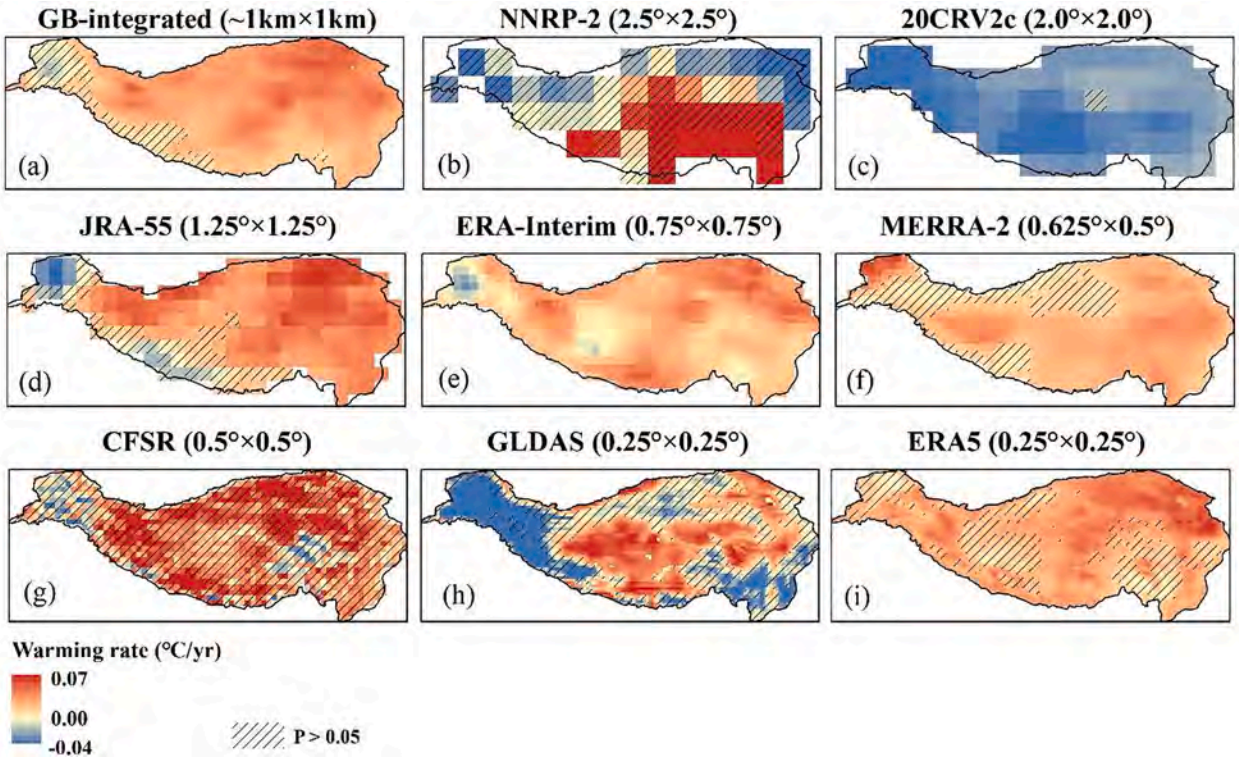


**Fig. 10.** Comparison of the temperature trends between the GB-integrated (a), NNRP-2 (b), 20CRV2c (c), JRA-55 (d), ERA-Interim (e), MERRA-2 (f), CFSR (g), GLDAS (h) and ERA5 (i). Shadowed area means insignificant trends.

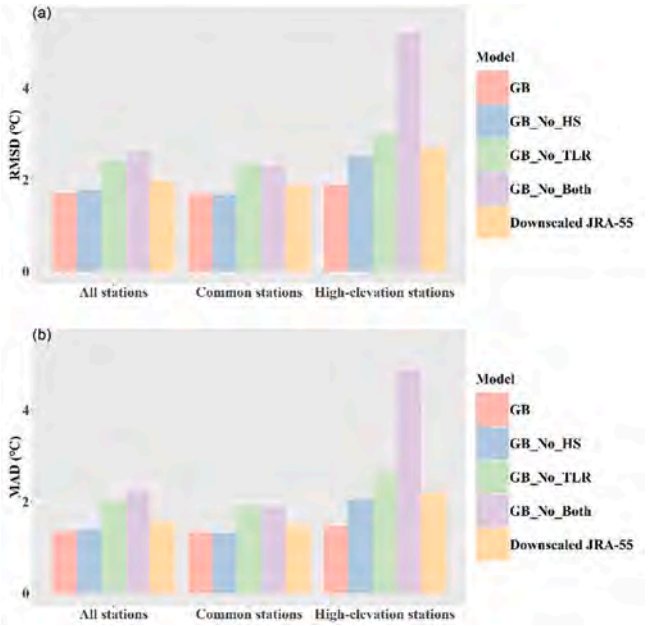


**Fig. 11.** Comparison of accuracy among GB, GB_No_HS, GB_No_TLR, GB_No_-Both and the downscaled JRA-55 based on RMSD (a) and MAD (b).

very unreliable. This finding is important because accurate air temperature is urgently needed in high-elevation areas where many environmental processes such as snow or glacier melting are very sensitive to air temperature variation (Immerzeel et al., 2014; Zhang et al., 2015). The results also show that simply integrating multiple reanalysis air temperatures cannot significantly improve the accuracy especially considering the very similar accuracy between GB_No_HS and the downscaled JRA-55 though the former is truly slightly better.

#### 4.3.4. *Factors affecting the air temperature estimation accuracy and some limitations*

The first important factor of the estimation accuracy may be the reliability of the downscaled reanalysis datasets. Among the eight datasets, the downscaled JRA-55 is possibly the most important as it shows the best downscaling accuracy. The relative importance of variables automatically calculated by the GB model based on the improvement in squared error due to different variables (Kuhn and Johnson, 2013) also indicates that JRA-55 is the most important input data for our machine learning model (see Supplementary Fig. S8). The importance ranking is also highly consistent with the comparison results in Section 4.2.2 in that the best three datasets and their relative orders are the same. The important effects of JRA-55 can also be seen from its highly consistent spatial patters of errors compared with that of the GB-integrated (Supplementary Fig. S7a and b). The correlation analysis shows that the RMSDs of the GB-integrated and the downscaled JRA-55 hold a significant ($P < 0.01$) positive correlation of 0.65.

Some special land covers, such as glaciers, may also be an important uncertainty source. Though the accuracy of the GB-integrated is largely improved on glacier surfaces compared with the downscaled JRA-55 (Fig. 9d), its RMSD (2.0 °C) is obviously higher than that on the vegetated (1.7 °C) and urban and barren surfaces (1.7 °C). The land cover effects could be exaggerated together with the scale problem about the difference between 1-km pixel and point position, especially for glacier surfaces (Qie et al., 2020; Wu et al., 2015; Zhang et al., 2018b). Though no significant correlation between elevation and RMSD is observed, the effects of elevation remain uncertain because the highest altitude of our stations is only 5900 m and the number of higher-than-5000 m stations is only 3. Such uneven distribution of stations might also make the predictions from GB model less reliable in areas with no stations, though the GB model shows better performance than CR and RF models. Future study may collect more field stations covering more land cover types and higher elevations for training a more reliable GB model.

It should also be noted that a fair comparison between multiple machine learning models requires carefully designed experiments that

could be affected by different configurations of hyperparameters and validation strategies. Though all the three models tested in this study have been tuned following common practice as described in Section 3.3.1 and are compared based on the same station-based leave-one-out cross-validations, more efforts could be made in future to increase the reliability of the comparison results. For instance, due to the huge computational burden, only 10-fold and 5-fold cross-validations are currently used in parameter tuning for CR and RF models, respectively, while the GB model uses 100-fold cross-validations.

Selection of ancillary variables also has effects on the estimation accuracy. Elevation, longitude, latitude and Julian day are the four most commonly used ancillary variables in air temperature estimation based on statistical models (Benali et al., 2012; Cristobal et al., 2008; Xu et al., 2014). Thus, these four variables are firstly included as ancillary predictors of our machine learning models. Considering that snow or glacier surface may have special effects on the air temperature estimation (Zhang et al., 2018b), the multiyear average snow cover days are also used as an ancillary variable. It is difficult to test every possible ancillary variable for a machine learning model due to the heavy computing burden. Including more ancillary variables such as albedo (Xu et al., 2014), radiation (Emamifar et al., 2013) and soil moisture (Kim and Han, 2013) may have the potential to further improve the estimation accuracy. Future studies may also consider using the new MOD21 land surface temperature data for developing TLR to replace the MOD11 product used in present study, since MOD21 is claimed to have addressed the problem of frequent cold bias in MOD11 product (Hulley et al., 2016). The high cloud cover of MODIS LST data may also affect the air temperature estimation accuracy by producing less reliable TLRs as discussed in Section 4.2.4.

### 4.4. Comparison with other air temperature estimation studies on the TP

Previous studies mainly estimate high resolution air temperature of the TP by downscaling reanalysis data or by incorporating MODIS LST. As for downscaling air temperature, most studies use TLR or a similarly simple empirical method (e.g., the linear regression) (Gao et al., 2017; Gerlitz et al., 2014; Huld and Pascua, 2015; Kazmi et al., 2015; Wang et al., 2011a; Wang et al., 2016b). For instance, Gerlitz et al. (2014) made corrections on both 1-km grid elevation and monthly bias to downscale daily air temperatures from ERA-Interim during 1989–2010 and finally get a high accuracy with a general average RMSD of 2.1 °C in the TP, though the monthly bias correction is highly dependent on local observations. Some studies also employ more advanced downscaling methods such as the machine learning models. For example, Pérez et al. (2015) develop a high-resolution air temperature dataset of east Iberian Peninsula based on NCEP/NCAR reanalysis data downscaled using a hybrid artificial neural network. Cao et al. (2017) use the temperature difference between near surface and pressure-level from ERA-Interim as a predictor and a surface correction parameter to improve upon downscaling results. This resulted in very high accuracy in Qilian mountains at a fine scale, but the model parameters need careful local calibration and are hard to transfer to other regions. This study selects TLR method as the downscaling scheme mainly due to two reasons: (1) TLR is easy to apply in mountainous regions and does not need dense stations, whereas more advanced models generally need clearly more ground observations for calibration; (2) TLR is also an important parameter in hydrological or ecological models in alpine areas, so the MODIS-estimated TLRs that are developed only using remote sensing temperatures can contribute to TLR parameterization in areas with no stations.

A number of existing studies have been using various advanced statistical methods by taking MODIS LSTs as the most important predictors to estimate 1-km air temperatures over the TP obtaining higher and higher accuracy with the RMSD reduced from 0.5 to 1.2 °C (Huang et al., 2017), ~1.0 °C (Xu et al., 2018) to ~0.7 °C (Zhu et al., 2019). However, they all have limitations for application because of their mostly estimating monthly temperatures and inability to produce long-term historical air temperature due to the short time span of MODIS LST. By contrast, the proposed method is not restricted to the time span of MODIS LST and is intrinsically free of clouds due to the sole dependence of space–time continuous reanalysis data rather than remote sensing products. It is also inspiring that compared with the limited researches targeting at estimating daily mean air temperature from MODIS LST, the accuracy from this study is relatively high in the TP with the mean RMSD of 1.7 °C versus 2.4–3.0 °C (Zhu et al., 2013), 2–4 °C (Kilibarda et al., 2014), ~2.0 °C (Zhang et al., 2016a) and 1.6–2.8 °C (Rao et al., 2019).

## 5. Conclusions

This study creates a novel 1-km daily average air temperature data product of 1980–2014 by integrating observation data of 113 stations and eight reanalysis datasets using machine learning. We demonstrate that integrating both the MODIS-estimated TLR and high-elevation observations can contribute greatly to the optimal use of multiple reanalysis and assimilation data for estimating high-resolution daily air temperatures of the Tibetan Plateau. Three main findings are concluded.

First, dynamically checking elevation variation of neighboring LST pixels can efficiently detect exceptional TLRs existing in the MODIS-estimated TLRs. A new TLR estimation method is thus proposed for estimating spatially continuous monthly TLRs from MODIS LST. The MODIS-estimated TLR shows reasonable spatial and seasonal patterns and is more reliable than the station-based one on the Tibetan Plateau. Second, all the eight reanalysis and assimilation air temperature data benefit a lot from being spatially downscaled with the MODIS-estimated TLR. The downscaled air temperature of JRA-55 shows the best accuracy among the eight downscaled datasets followed by ERA-Interim, MERRA-2, CFSR and others. Third, the GB model shows better performance than Random Forests and Cubist Regression in integrating multiple downscaled reanalysis and assimilation data for daily average air temperature estimation. The new air temperature data estimated by GB model are better than the downscaled JRA-55 with the mean RMSDs of 1.7 °C versus 2.0 °C in general and especially the distinct mean RMSDs of 1.9 °C versus 2.7 °C in high-elevation stations. Both independent high-altitude training data and reliable TLRs are crucial for machine-learning based integration of multiple reanalysis and assimilation air temperatures in the Tibetan Plateau, especially for high-altitude areas, however, simply integrating multiple reanalysis data cannot produce significant improvement in estimation accuracy.

Although the new 1-km daily average temperature product has a relatively high accuracy and a long time span, its uncertainties and limitations need further investigation in future. The MODIS-estimated TLR could be improved by considering more terrain and vegetation factors. Especially, the extrapolated TLRs due to insufficient LST samples may have lower reliability and could be significantly improved by introducing more temperature information from other data sources such as MODIS AQUA LST. Despite 13 high-elevation stations' observations used for model training, obviously more high-elevation observations should be collected in future to help improve the estimation accuracy through machine learning. It should be noted that the framework proposed for creating high-resolution air temperature data by integrating multiple reanalysis data with elevation correction is globally applicable and is not restricted to the Tibetan Plateau.

## Author contributions

All authors made significant contributions to this study. W.W.I., H.Z. and F.Z. designed this study; H.Z. performed all analyses; H.Z., W.W.I. and F.Z. wrote the manuscript; R.J.d.K. and M.Y. discussed the results, commented on the manuscript and contributed to writing. S.G. contributed to writing.

## CRediT authorship contribution statement

**Hongbo Zhang:** Writing - original draft, Conceptualization, Methodology, Software. **W.W. Immerzeel:** Conceptualization, Writing - review & editing. **Fan Zhang:** Software, Writing - review & editing. **Remco J. de Kok:** Writing - review & editing. **Sally J. Gorrie:** Writing - review & editing. **Ming Ye:** Writing - review & editing.

## Declaration of Competing Interest

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

## Acknowledgements

The produced 1-km daily air temperature of the Tibetan Plateau during 1980-2014 can be obtained at http://data.tpdc.ac.cn/en/data/62234872-c39c-4614-a7ac-348c9437e7d5/, with DOI numbers 10.11888/Meteoro.tpdc.270377. The developed monthly TLRs with eight types of resolutions and the corresponding quality control files can be obtained at http://hongbozhang.net/wph/?page_id=504. This study was supported by State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences (Grant Number: SKLCS-OP-2020-13), the Second Tibetan Plateau Scientific Expedition and Research Program (Grant No. 2019QZKK0203), State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute (Grant No. 2019nkms02), the National Natural Science Foundation of China (Grant No. 41701079), the "Strategic Priority Research Program" of the Chinese Academy of Sciences (Grant No. XDA20100300 and XDA20060202), the European Research Council (ERC) under the European Union Horizon 2020 Research and Innovation Program (Grant Agreement No. 676819), Netherlands Organization for Scientific Research (Grant No. 016.181.308 and ALWOP.467) and the China Scholarship Council. We appreciate China Meteorological Administration and multiple field stations of the Institute of Tibetan Plateau Research, Chinese Academy of Sciences for providing daily or sub-daily air temperature observations. Thanks to Dr. Philip Kraaijenbrink for helping obtain the daily mean temperature data from ERA5.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2021.102295.

## References

Bao, X., Zhang, F., 2013. Evaluation of NCEP–CFSR, NCEP–NCAR, ERA-Interim, and ERA-40 reanalysis datasets against independent sounding observations over the Tibetan Plateau. J. Clim. 26, 206–214.
Barker, D.M., Huang, W., Guo, Y.-R., Bourgeois, A., Xiao, Q., 2004. A three-dimensional variational data assimilation system for MM5: Implementation and initial results. Mon. Weather Rev. 132, 897–914.
Benali, A., Carvalho, A.C., Nunes, J.P., Carvalhais, N., Santos, A., 2012. Estimating air surface temperature in Portugal using MODIS LST data. Remote Sens. Environ. 124, 108–121.
Blandford, T.R., Humes, K.S., Harshburger, B.J., Moore, B.C., Walden, V.P., Ye, H., 2008. Seasonal and synoptic variations in near-surface air temperature lapse rates in a mountainous basin. J. Appl. Meteorol. Climatol. 47, 249–261.
Bracegirdle, T.J., Marshall, G.J., 2012. The reliability of Antarctic tropospheric pressure and temperature in the latest global reanalyses. J. Clim. 25, 7138–7146.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Cao, B., Gruber, S., Zhang, T., 2017. REDCAPP (v1. 0): parameterizing valley inversions in air temperature data downscaled from reanalyses. Geosci. Model Dev. 10, 2905–2923.
Chen, S., Gan, T.Y., Tan, X., Shao, D., Zhu, J., 2019. Assessment of CFSR, ERA-Interim, JRA-55, MERRA-2, NCEP-2 reanalysis data for drought analysis over China. Clim. Dyn. 1–21.
Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., 2015. Xgboost: extreme gradient boosting. R package version 0.4-2, pp. 1–4.
Cheng, G., Wu, T., 2007. Responses of permafrost to climate change and their environmental significance, Qinghai-Tibet Plateau. J. Geophys. Res. Earth Surf. 112.
Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., 2011. The twentieth century reanalysis project. Q. J. R. Meteorolog. Soc. 137, 1–28.
Cristobal, J., Ninyerola, M., Pons, X., 2008. Modeling air temperature through a combination of remote sensing and GIS data. J. Geophys. Res.-Atmos. 113.
Dee, D.P., Uppala, S.M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, D.P., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorolog. Soc. 137, 553–597.
Dunnett, C.W., 1955. A multiple comparison procedure for comparing several treatments with a control. J. Am. Stat. Assoc. 50, 1096–1121.
El-Samra, R., Bou-Zeid, E., El-Fadel, M., 2018. What model resolution is required in climatological downscaling over complex terrain? Atmos. Res. 203, 68–82.
Emamifar, S., Rahimikhoob, A., Noroozi, A.A., 2013. Daily mean air temperature estimation from MODIS land surface temperature products based on M5 model tree. Int. J. Climatol. 33, 3174–3181.
Firozjaei, M.K., Fathololoumi, S., Alavipanah, S.K., Kiavarz, M., Vaezi, A.R., Biswas, A., 2020. A new approach for modeling near surface temperature lapse rate based on normalized land surface temperature data. Remote Sens. Environ. 242, 111746.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38, 367–378.
Fu, G., Shen, Z., Zhang, X., Shi, P., Zhang, Y., Wu, J., 2011. Estimating air temperature of an alpine meadow on the Northern Tibetan Plateau using MODIS land surface temperature. Acta Ecol. Sin. 31, 8–13.
Gao, L., Bernhardt, M., Schulz, K., Chen, X., 2017. Elevation correction of ERA-Interim temperature data in the Tibetan Plateau. Int. J. Climatol. 37, 3540–3552.
Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Silva, A.M.D., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J. Clim. 30, 5419–5454.
Gerlitz, L., Conrad, O., Thomas, A., Böhner, J., 2014. Warming patterns over the Tibetan Plateau and adjacent lowlands derived from elevation-and bias-corrected ERA-Interim data. Clim. Res. 58, 235–246.
Guo, X., Wang, L., Tian, L., 2016. Spatio-temporal variability of vertical gradients of major meteorological observations around the Tibetan Plateau. Int. J. Climatol. 36, 1901–1916.
He, Y., Wang, K., 2020. Contrast patterns and trends of lapse rates calculated from near-surface air and land surface temperatures in mainland China from 1961 to 2014. Sci. Bull.
Hofer, M., Marzeion, B., Mölg, T., 2012. Comparing the skill of different reanalyses and their ensembles as predictors for daily air temperature on a glaciated mountain (Peru). Clim. Dyn. 39, 1969–1980.
Huang, F., Ma, W., Wang, B., Hu, Z., Ma, Y., Sun, G., Xie, Z., Lin, Y., 2017. Air temperature estimation with MODIS data over the Northern Tibetan Plateau. Adv. Atmos. Sci. 34, 650–662.
Huld, T., Pascua, I., 2015. Spatial downscaling of 2-meter air temperature using operational forecast data. Energies 8, 2381–2411.
Hulley, G., Malakar, N., Freepartner, R., 2016. Moderate Resolution Imaging Spectroradiometer (MODIS) Land Surface Temperature and Emissivity Product (MxD21) Algorithm Theoretical Basis Document Collection-6. JPL Publication, pp. 12–17.
Immerzeel, W., Petersen, L., Ragettli, S., Pellicciotti, F., 2014. The importance of observed gradients of air temperature and precipitation for modeling runoff from a glacierized watershed in the Nepalese Himalayas. Water Resour. Res. 50, 2212–2226.
Jarvis, C.H., Stuart, N., 2001. A comparison among strategies for interpolating maximum and minimum daily air temperatures. Part II: The interaction between number of guiding variables and the type of interpolation method. J. Appl. Meteorol. 40, 1075–1084.
Ji, L., Senay, G.B., Verdin, J.P., 2015. Evaluation of the global land data assimilation system (GLDAS) air temperature data products. J. Hydrometeorol. 16, 2463–2480.
Jiang, J., Liu, J., Qin, C., Miao, Y., Zhu, A.X., 2016. Near-surface air temperature lapse rates and seasonal and type differences in China. Progr. Geogr. 35.
Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J., Fiorino, M., Potter, G., 2002. Ncep–doe amip-ii reanalysis (r-2). Bull. Am. Meteorol. Soc. 83, 1631–1644.
Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's land surface areas. Sci. Data 4, 170122.
Kattel, D.B., Yao, T., Panday, P.K., 2017. Near-surface air temperature lapse rate in a humid mountainous terrain on the southern slopes of the eastern Himalayas. Theor. Appl. Climatol. 1–13.
Kattel, D.B., Yao, T., Yang, K., Tian, L., Yang, G., Joswiak, D., 2013. Temperature lapse rate in complex mountain terrain on the southern slope of the central Himalayas. Theor. Appl. Climatol. 113, 671–682.
Kattel, D.B., Yao, T., Yang, W., Gao, Y., Tian, L., 2015. Comparison of temperature lapse rates from the northern to the southern slopes of the Himalayas. Int. J. Climatol. 35, 4431–4443.

Kazmi, D.H., Li, J., Rasul, G., Tong, J., Ali, G., Cheema, S.B., Liu, L., Gemmer, M., Fischer, T., 2015. Statistical downscaling and future scenario generation of temperatures for Pakistan Region. Theor. Appl. Climatol. 120, 341–350.
Kilibarda, M., Hengl, T., Heuvelink, G.B.M., Gräler, B., Pebesma, E., Perčec Tadić, M., Bajat, B., 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J. Geophys. Res.: Atmos. 119, 2294–2313.
Kim, D.Y., Han, K.S., 2013. Remotely sensed retrieval of midday air temperature considering atmospheric and surface moisture conditions. Int. J. Remote Sens. 34, 247–263.
Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., 2015. The JRA-55 reanalysis: General specifications and basic characteristics. J. Meteorol. Soc. Jpn. Ser. II 93, 5–48.
Kraaijenbrink, P., Bierkens, M., Lutz, A., Immerzeel, W., 2017. Impact of a global temperature rise of 1.5 degrees Celsius on Asia's glaciers. Nature 549, 257.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer.
Li, T., Zheng, X., Dai, Y., Yang, C., Chen, Z., Zhang, S., Wu, G., Wang, Z., Huang, C., Shen, Y., 2014. Mapping near-surface air temperature, pressure, relative humidity and wind speed over Mainland China with high spatiotemporal resolution. Adv. Atmos. Sci. 31, 1127–1135.
Li, X., Wang, L., Chen, D., Yang, K., Xue, B., Sun, L., 2013. Near-surface air temperature lapse rates in the mainland China during 1962–2011. J. Geophys. Res.: Atmos. 118, 7505–7515.
Li, Y., Zeng, Z., Zhao, L., Piao, S., 2015. Spatial patterns of climatological temperature lapse rate in mainland China: A multi–time scale investigation. J. Geophys. Res.: Atmos. 120, 2661–2675.
Liu, X., Chen, B., 2000. Climatic warming in the Tibetan Plateau during recent decades. Int. J. Climatol. 20, 1729–1742.
Min, W., Yueqing, L., Zhou, J., 2015. Validation of MODIS land surface temperature products in east of the Qinghai-Xizang Plateau. Plateau Meteorol. 34, 1511–1516.
Minder, J.R., Mote, P.W., Lundquist, J.D., 2010. Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains. J. Geophys. Res.: Atmos. 115.
Mudryk, L.R., Derksen, C., Kushner, P.J., Brown, R., 2015. Characterization of Northern Hemisphere snow water equivalent datasets, 1981–2010. J. Clim. 28, 8037–8051.
Mutiibwa, D., Strachan, S., Albright, T., 2015. Land surface temperature and surface air temperature in complex terrain. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 4762–4774.
Noi, P., Degener, J., Kappas, M., 2017. Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. Remote Sens. 9, 398.
Oyler, J.W., Dobrowski, S.Z., Holden, Z.A., Running, S.W., 2016. Remotely sensed land skin temperature as a spatial predictor of air temperature across the conterminous United States. J. Appl. Meteorol. Climatol. 55, 1441–1457.
Pepin, N., Benham, D., Taylor, K., 1999. Modeling lapse rates in the maritime uplands of northern England: implications for climate change. Arct. Antarct. Alp. Res. 31, 151–164.
Pepin, N., Bradley, R., Diaz, H., Baraër, M., Caceres, E., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M., Liu, X., 2015. Elevation-dependent warming in mountain regions of the world. Nat. Clim. Change 5, 424–430.
Pérez, J.J.M., Navarro, M.J.E., Cantos, J.O., 2015. Statistical downscaling and attribution of air temperature change patterns in the Valencia region (1948–2011). Atmos. Res. 156, 189–212.
Qie, Y., Wang, N., Wu, Y., Chen, A.A., 2020. Variations in winter surface temperature of the Purog Kangri Ice Field, Qinghai-Tibetan Plateau, 2001–2018. Using MODIS data. Remote Sens. 12, 1133.
Quinlan, J., 2001. Rulequest. www.rulequest.com/see5-info.html.
Quiquet, A., Roche, D.M., Dumas, C., Paillard, D., 2018. Online dynamical downscaling of temperature and precipitation within the iLOVECLIM model (version 1.1). Geosci. Model Dev. 11, 453–466.
Rao, Y., Liang, S., Wang, D., Yu, Y., Song, Z., Zhou, Y., Shen, M., Xu, B., 2019. Estimating daily average surface air temperature using satellite land surface temperature and top-of-atmosphere radiation products over the Tibetan Plateau. Remote Sens. Environ. 234, 111462.
Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The global land data assimilation system. Bull. Am. Meteorol. Soc. 85, 381–394.
Rolland, C., 2003. Spatial and seasonal variations of air temperature lapse rates in Alpine regions. J. Clim. 16, 1032–1046.
Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Delst, P.V., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Dool, H.V.D., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R.W., Rutledge, G., Goldberg, M., 2010. The NCEP climate forecast system reanalysis. Bull. Am. Meteorol. Soc. 91, 1015–1058.
Service, C.C.C., 2017. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.
Stahl, K., Moore, R.D., Floyer, J.A., Asplin, M.G., McKendry, I.G., 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, 224–236.
Sun, C.-J., Li, W., Li, X.-G., Zhang, Z.-Y., Chen, R.-X., Chen, W., 2018. Spatio-temporal variation of near-surface temperature lapse rates over the Northwestern Tibetan Plateau. J. Natural Resour. 33, 1270–1282.
Vancutsem, C., Ceccato, P., Dinku, T., Connor, S.J., 2010. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Remote Sens. Environ. 114, 449–465.
Wang, A., Zeng, X., 2012. Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau. J. Geophys. Res.: Atmos. 117.
Wang, A., Zeng, X., 2013. Development of global hourly 0.5 land surface air temperature datasets. J. Clim. 26, 7676–7691.
Wang, F., Wang, L., Koike, T., Zhou, H., Yang, K., Wang, A., Li, W., 2011a. Evaluation and application of a fine-resolution global data set in a semiarid mesoscale river basin with a distributed biosphere hydrological model. J. Geophys. Res.: Atmos. 116.
Wang, L., Sun, L., Shrestha, M., Li, X., Liu, W., Zhou, J., Yang, K., Lu, H., Chen, D., 2016a. Improving snow process modeling with satellite-based estimation of near-surface-air-temperature lapse rate. J. Geophys. Res. Atmos. 121.
Wang, T., Hamann, A., Spittlehouse, D., Carroll, C., 2016b. Locally downscaled and spatially customizable climate data for historical and future periods for North America. PLoS ONE 11, e0156720.
Wang, W., Cui, W., Wang, X., Chen, X., 2016c. Evaluation of GLDAS-1 and GLDAS-2 forcing data and Noah model simulations over China at the monthly scale. J. Hydrometeorol. 17, 2815–2833.
Wang, W., Xie, P., Yoo, S.-H., Xue, Y., Kumar, A., Wu, X., 2011b. An assessment of the surface climate in the NCEP climate forecast system reanalysis. Clim. Dyn. 37, 1601–1620.
Wang, Y., Wang, L., Li, X., Chen, D., 2018. Temporal and spatial changes in estimated near-surface air temperature lapse rates on Tibetan Plateau. Int. J. Climatol. 38, 2907–2921.
Wijngaard, R.R., Lutz, A.F., Nepal, S., Khanal, S., Pradhananga, S., Shrestha, A.B., Immerzeel, W.W., 2017. Future changes in hydro-climatic extremes in the Upper Indus, Ganges, and Brahmaputra River basins. PLoS ONE 12, e0190224.
Wu, Q., Zhang, T., 2008. Recent permafrost warming on the Qinghai-Tibetan Plateau. J. Geophys. Res.: Atmos. 113.
Wu, S.H., Jansson, P.-E., Kolari, P., 2012. The role of air and soil temperature in the seasonality of photosynthesis and transpiration in a boreal Scots pine ecosystem. Agric. For. Meteorol. 156, 85–103.
Wu, Y., Wang, N., He, J., Jiang, X., 2015. Estimating mountain glacier surface temperatures from Landsat-ETM+ thermal infrared data: A case study of Qiyi glacier, China. Remote Sens. Environ. 163, 286–295.
Xu, Y., Knudby, A., Ho, H.C., 2014. Estimating daily maximum air temperature from MODIS in British Columbia, Canada. Int. J. Remote Sens. 35, 8108–8121.
Xu, Y., Knudby, A., Shen, Y., Liu, Y., 2018. Mapping monthly air temperature in the Tibetan Plateau from MODIS data based on machine learning methods. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 345–354.
Yao, T., Thompson, L., Yang, W., Yu, W., Gao, Y., Guo, X., Yang, X., Duan, K., Zhao, H., Xu, B., 2012. Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings. Nat. Clim. Change 2, 663–667.
Yu, J., Zhang, G., Yao, T., Xie, H., Zhang, H., Ke, C., Yao, R., 2016. Developing Daily Cloud-Free Snow Composite Products From MODIS Terra-Aqua and IMS for the Tibetan Plateau. IEEE Trans. Geosci. Remote Sensing 54, 2171–2180.
Yu, W., Ma, M., 2011. Validation of the MODIS land surface temperature products—A case study of the Heihe River Basin. Remote Sens. Technol. Appl. 26, 705–712.
Zhang, F., Shi, X., Zeng, C., Wang, L., Xiao, X., Wang, G., Chen, Y., Zhang, H., Lu, X., Immerzeel, W., 2020. Recent stepwise sediment flux increase with climate change in the Tuotuo River in the central Tibetan Plateau. Sci. Bull. 65, 410–418.
Zhang, F., Zhang, H., Hagen, S.C., Ye, M., Wang, D., Gui, D., Zeng, C., Tian, L., Liu, J., 2015. Snow cover and runoff modelling in a high mountain catchment with scarce data: effects of temperature and precipitation parameters. Hydrol. Process. 29, 52–65.
Zhang, G., Xie, H., Kang, S., Yi, D., Ackley, S.F., 2011. Monitoring lake level changes on the Tibetan Plateau using ICESat altimetry data (2003–2009). Remote Sens. Environ. 115, 1733–1742.
Zhang, G., Yao, T., Shum, C., Yi, S., Yang, K., Xie, H., Feng, W., Bolch, T., Wang, L., Behrangi, A., 2017. Lake volume and groundwater storage variations in Tibetan Plateau's endorheic basin. Geophys. Res. Lett. 44, 5550–5560.
Zhang, H., Zhang, F., Ye, M., Che, T., Zhang, G., 2016a. Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data. J. Geophys. Res.: Atmos. 121, 11,425–411,441.
Zhang, H., Zhang, F., Zhang, G., Che, T., Yan, W., 2018a. How accurately can the air temperature lapse rate over the Tibetan Plateau be estimated from MODIS LSTs? J. Geophys. Res.: Atmos. 123, 3943–3960.
Zhang, H., Zhang, F., Zhang, G., He, X., Tian, L., 2016b. Evaluation of cloud effects on air temperature estimation using MODIS LST based on ground measurements over the Tibetan Plateau. Atmos. Chem. Phys. 16, 13681–13696.
Zhang, H., Zhang, F., Zhang, G., Ma, Y., Yang, K.U.N., Ye, M., 2018b. Daily air temperature estimation on glacier surfaces in the Tibetan Plateau using MODIS LST data. J. Glaciol. 1–16.
Zhang, L., Su, F., Yang, D., Hao, Z., Tong, K., 2013. Discharge regime and simulation for the upstream of major rivers over Tibetan Plateau. J. Geophys. Res.: Atmos. 118, 8500–8518.
Zhou, C., He, Y., Wang, K., 2018. On the suitability of current atmospheric reanalyses for regional warming studies over China. Atmos. Chem. Phys. 18, 8113–8136.
Zhou, C., Wang, K., Ma, Q., 2017. Evaluation of eight current reanalyses in simulating land surface temperature from 1979 to 2003 in China. J. Clim. 30, 7379–7398.
Zhou, J., Wang, L., Zhang, Y., Guo, Y., Li, X., Liu, W., 2015. Exploring the water storage changes in the largest lake (Selin Co) over the Tibetan Plateau during 2003–2012 from a basin-wide hydrological modeling. Water Resour. Res. 51, 8060–8086.

Zhu, W., Lű, A., Jia, S., 2013. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. Remote Sens. Environ. 130, 62–73.
Zhu, W., Lű, A., Jia, S., Yan, J., Mahmood, R., 2017. Retrievals of all-weather daytime air temperature from MODIS products. Remote Sens. Environ. 189, 152–163.
Zhu, X., Zhang, Q., Xu, C.-Y., Sun, P., Hu, P., 2019. Reconstruction of high spatial resolution surface air temperature data across China: A new geo-intelligent multisource data-based machine learning technique. Sci. Total Environ. 665, 300–313.
Zou, H., Zhu, J., Zhou, L., Li, P., Ma, S., 2014. Validation and application of reanalysis temperature data over the Tibetan Plateau. J. Meteorolog. Res. 28, 139–149.