

Journal of  
**Applied Remote Sensing**

**Estimation of daily average  
temperature using multisource  
spatial data in data sparse regions of  
Central Asia**

Mingyong Cai  
Shengtian Yang  
Changsen Zhao  
Hongjuan Zeng  
Qiuwen Zhou



# Estimation of daily average temperature using multisource spatial data in data sparse regions of Central Asia

Mingyong Cai,<sup>a</sup> Shengtian Yang,<sup>a</sup> Changsen Zhao,<sup>a</sup> Hongjuan Zeng,<sup>b</sup> and Qiuwen Zhou<sup>a</sup>

<sup>a</sup>Beijing Normal University, School of Geography, State Key Laboratory of Remote Sensing Science, Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, Beijing 100875, China  
[yangshengtian@bnu.edu.cn](mailto:yangshengtian@bnu.edu.cn)

<sup>b</sup>Changjiang Water Resources Commission, Changjiang Soil and Water Conservation Monitoring Center, Wuhan 430010, China

**Abstract.** The distribution of the daily average air temperature with high spatial resolution is vital for hydro-ecological applications. The air temperature usually recorded at fixed-point stations provides little distribution information and easily suffers from the scarce amount and uneven distribution of the stations in the data sparse regions. In this study, a method based on multisource spatial data was developed to estimate the spatial distribution of daily average temperature, especially for data sparse regions. In this method, the instantaneous temperature was retrieved first using the moderate resolution imaging spectroradiometer data, which was then transformed to a daily value using transformation equations. Second, the global land data assimilation system air temperature data were spatially downscaled and used to improve the data accuracy from step 1 at low temperatures. This method was applied in the Ili River basin in Central Asia, and the results were evaluated against data from two stations' observations and *in situ* data from a field test site. The results showed the correlation coefficient varies from 0.90 to 0.94 and the root mean square deviation is  $\sim 3^{\circ}\text{C}$ , indicating the generated temperature matched the observations well. This suggests the method is an alternative for data sparse regions. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.7.073478](https://doi.org/10.1117/1.JRS.7.073478)]

**Keywords:** air temperature; multisource spatial data; data sparse region; temporal scale transformation; spatial downscaling.

Paper 13273 received Jul. 23, 2013; revised manuscript received Sep. 22, 2013; accepted for publication Oct. 18, 2013; published online Nov. 22, 2013.

## 1 Introduction

Air temperature is one of the most important parameters for a wide range of applications in ecology and hydrology. Near surface air temperature with varying spatial and temporal resolution is often required in many hydro-ecological modeling techniques. The values of surface air temperature are now usually recorded at a fixed point at stations, which provide little distribution of air temperature over regions. However, spatial distribution of air temperature is vital for the regional hydro-ecological modeling.

Researchers have explored various methods to generate spatial air temperature. Some researchers like Peterson et al.,<sup>1</sup> Anderson,<sup>2</sup> and Florio et al.<sup>3</sup> used spatial interpolation to estimate spatial distribution of air temperatures. However, the reliability of results from interpolations greatly depends on the density stations and the numbers of observations; thus, these methods are limited for application over large area, especially in data sparse regions.

Other researchers have used remote sensing data to derive spatial distribution of air temperatures because the observations from satellites can provide high temporal-spatial distribution information of the underlying surface, which can be used to derive the air temperature. Land surface temperature (LST), which can be obtained through the split-window technique, is usually used to derive the near surface air temperature.<sup>4-6</sup> Specific algorithms have been

developed for different sensors, such as advanced very high resolution radiometer, moderate resolution imaging spectroradiometer (MODIS), and Meteosat to get LST.<sup>7-10</sup> At present, the methods used to derive near surface air temperature from LST can be divided into three main categories: the statistical approach,<sup>11-14</sup> the temperature-vegetation index (TVX) approach,<sup>15-19</sup> and the energy-balance approach.<sup>20-22</sup>

The statistical approach is usually based on the linear regression between observed air temperature and LST, which is relatively simple and commonly used. Air temperature data can be derived with high accuracy in the regions where the methods or the relationship were established. To enhance the physical basis of the statistical methods, many researchers have attempted to introduce more parameters reflecting the effects of various factors (seasons, vegetation, topography, etc.) when the relationship was built.<sup>23,24</sup> However, these methods suffer from the dependence on the calibration data, which limits the application of these methods, especially in data sparse areas.

The TVX approach is based on the assumptions that a linear and negative relationship exists between LST and normalized difference vegetation index (NDVI) and that the air temperature is equal to the radiometric temperature in a fully vegetated area. This method involves only two factors (LST and NDVI) to retrieve the air temperature and seems easy to implement, but it has limitations as well. Many factors like solar radiation, local topography, and seasonality are overlooked in this approach, and only NDVI is taken into account besides LST, so this approach is sometimes inadequate when the NDVI does not work well.<sup>25</sup>

Energy-balance methods are physically based. In these methods, the physical processes of energy transportation and transformation are considered, instead of using only empirical or statistical relationships. However, energy-balance models require soil, canopy, and atmospheric data that are often unavailable over large areas.

In summary, the methods mentioned above face challenges in deriving spatial distribution of temperature over large areas, especially in data sparse regions either due to the strong dependence on the stations' calibration data, the less consideration of the underlying conditions, or the availability in data acquisition. The temperature derived using remote sensing is often the instantaneous value when the satellite passes by, while the daily value is required in many hydrological models.

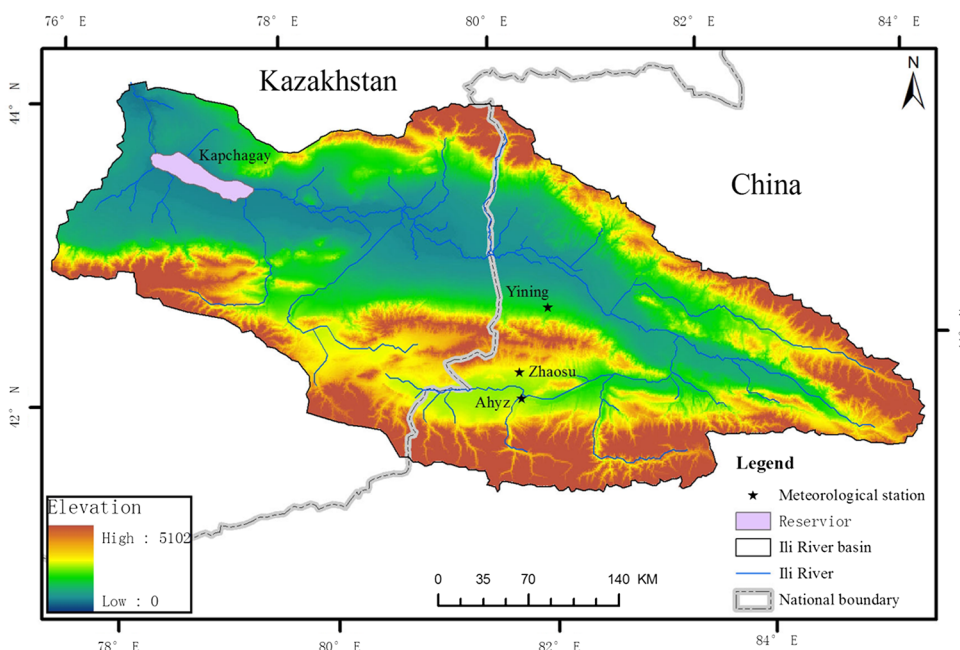
In this study, a new method combining multisource spatial data was developed to estimate the daily mean near surface air temperature at 2 m height ( $T_{2m}$ ), especially for data sparse regions. In this new method, the Klemen algorithm created by Zakšek<sup>24</sup> was used to first obtain the instant  $T_{2m}$  values on the basis of MODIS data products. The Klemen method was adopted for its consideration of various radiation and underlying surface factors and its access to remote sensing data. The instant  $T_{2m}$  from Klemen method was transformed to the daily mean temperature value by introducing the National Center for Atmospheric Research/National Centers for Environmental Prediction (NCAR/NCEP) reanalysis temperature data. The derived daily mean  $T_{2m}$  was finally adjusted with the downscaled temperature data from global land data assimilation system (GLDAS) to achieve higher accuracy at low temperatures.

## 2 Study Area and Materials

### 2.1 Study Area and In Situ Measurement

The temperature data inversion in this study was a part of the hydrological modeling in ungauged regions in Central Asia. This study was carried out at the region containing runoff of the Ili River (an international river lying between China and Kazakhstan, Fig. 1) covering an area of 115,000 km<sup>2</sup>. The outlet of the study area is the same as the outlet of the Kapchagay Reservoir. The precipitation in this area varies greatly due to the topography. Depending on the variation in topography, the precipitation in the Ili basin can reach up to 1000 mm/year in the mountains compared with 200 mm/year in the valley. The temperature also changes extensively because of the great differences in altitude.

Daily average temperature data from meteorological stations (at Yining and Zhaosu, located in the plain and mountainous areas, respectively) were collected from 2005 to 2009, corresponding to the remote sensing-based temperature simulation and hydrological modeling period.



**Fig. 1** Sketch map of the study area and validation site locations.

These data were collected for the derived daily average temperature validation. It is noticed that the derived daily average temperature using the Klemen method based on multisource spatial data has a height of 2 m above the surface,<sup>24</sup> while the observations from Chinese meteorological stations are recorded at 1.5 m above ground according to the “Technical regulation of meteorological observations” published by the China Meteorological Administration. The derived daily average temperature was transformed to the height of 1.5 m above ground before it was validated using observations.

A field experiment was carried out for a test site (81.10°E, 42.78°N and 2017 m altitude) near Aheyz, a southern study area (Fig. 1). The instantaneous air temperatures are recorded using the automatic meteorological station HOBO and are available at 4 min intervals from August to November in 2009 (see Table 1 for details on the instrumentation for the temperature measurements). The instant temperature recorded by HOBO was used to have the simulated instant temperature value from Klemen method tested, and the daily mean value from the HOBO records were used to validate the derived daily mean values from the Klemen method after temporal transformation and downscaled GLDAS temperature. So the measurements using HOBO were carried out 2 m above the ground to be consistent with the height described in Zakšek.<sup>24</sup>

## 2.2 Spatial Data

Multisource spatial data platform products were used in this study to estimate daily mean values of near surface air temperatures ( $T_{2m}$ ), including the MODIS, GLDAS, Shuttle Radar Topography Mission (SRTM), and NCAR/NCEP products. These data products were selected for their global coverage and common use in various research fields, both of which may offer this study wider application in other regions.

**Table 1** Instrumentation (HOBO) for temperature measurements at the Aheyz.

Parameter	Sensor	Parameters	Maker	Placement height
Air temperature	Temp/RH sensor	(12-bit) w/2m	Onset	2 m

Three MODIS products at 1 km spatial resolution (version 5) were used to derive the instant  $T_{2m}$ : (1) 8-day albedo (MCD43B3), (2) 16-day vegetation index (MOD13A2), and (3) daily land surface temperature and emissivity (MOD11A1).

NCAR/NCEP near surface air temperature data were used in this study to transform the instant  $T_{2m}$  to a daily mean value.<sup>26–28</sup> This NCAR/NCEP dataset has a spatial resolution of 1.875 deg and a temporal resolution of 6 h, so there are four instant values (UTC 00:00, 06:00, 12:00, and 18:00) for each day.

GLDAS provides the near surface air temperature data with a spatial resolution of 0.25 deg and a temporal resolution of 3 h.<sup>29</sup> These data were then spatially downscaled to 1 km using the SRTM elevation data on the basis of an assumed temperature lapse rate. The downscaled results were then used to adjust the MODIS retrieval results.

All of these spatial data can be obtained through the Internet for free (MODIS <http://reverb.echo.nasa.gov/reverb/>; NCAR/NCEP <http://www.esrl.noaa.gov>; GLDAS <http://mirador.gsfc.nasa.gov/>; SRTM <http://datamirror.csdb.cn/>).

### 3 Methodology

Daily average air temperature data were simulated and compared with the observations from two weather stations and a field observation site. For this comparison, two processes were applied: (1) the instant temperature value when the satellite (Terra) passed above was derived first from MODIS products and then was transformed to daily mean value using a statistical method presented in this study; (2) the derived daily average air temperature from step (1) was modified by replacing the simulations at low temperatures with downscaled GLDAS air temperature to achieve better accuracy when simulating low temperatures. The daily average air temperature data from GLDAS with coarse spatial resolution were downscaled to the resolution of 1 km using the relationship between altitude and temperature.

#### 3.1 Klemen Method–Based Daily Average Temperature Retrieval

##### 3.1.1 Instantaneous temperature retrieval using the Klemen method

The method developed by Zakšek<sup>24</sup> was used to obtain instant air temperature values at a height of 2 m. This method [Eq. (1)] was adopted because of its consideration of various effects including season, solar radiation, and underlying surface conditions. The Klemen method accounts for more physical processes compared with the simple statistical methods mentioned above, and all the parameters required for this method can be prepared using multisource spatial data.

$$T_{2m} = \text{LST} + 1.82 - 10.66 \cos z(1 - \text{NDVI}) + 0.566\alpha - 3.72(1 - \text{AL})(\cos i / \cos z + (\pi - s) / \pi)R_s \downarrow - 3.41\Delta h, \quad (1)$$

where  $T_{2m}$  is the instant temperature at 2 m height and the unit is K; LST is the instantaneous daytime land surface temperature from MODIS MOD11A1 in K;  $z$  is the solar zenith angle in rad;  $\alpha$  is the solar azimuth start from the south in rad; AL is the surface albedo;  $i$  is the solar incident angle in rad;  $s$  is the slope in rad;  $R_s \downarrow$  is the down-welling surface short-wave radiation flux in  $\text{w}/\text{m}^2$ ; and  $\Delta h$  is the difference between the pixel elevation and the mean elevation within the vicinity of 20 km in km.

The key parameter in this step is the accuracy and representativeness of the LST. As described in the spatial data section, the LST data used in this study were instant values from the MODIS product (MOD11A1). The LST contained in the MODIS MOD11A1 dataset was generated using the generalized split-window method proposed by Wan and Dozier<sup>30</sup> on the basis of MODIS thermal infrared ray band 31 and band 32. According to Wan and Dozier,<sup>30</sup> the atmospheric effects in the LST retrieving were corrected on the basis of differential absorption in adjacent thermal infrared bands rather than on absolute atmospheric transmission in a single band, so it is less sensitive to the uncertainties in optical properties of the atmosphere and no profiles of atmospheric water vapor and temperature were needed. The accuracy of the

MODIS LST was discussed by many authors and the nominal accuracy was thought to be  $\pm 1\text{K}$ .<sup>30-32</sup> This study was a part of the hydrological modeling in ungauged regions and was designed to provide daily average surface air temperature estimation for distributed hydrological models, especially in data sparse regions, and no further atmospheric correction was applied to the MODIS LST data.

### 3.1.2 Time-scale transformation to daily average value

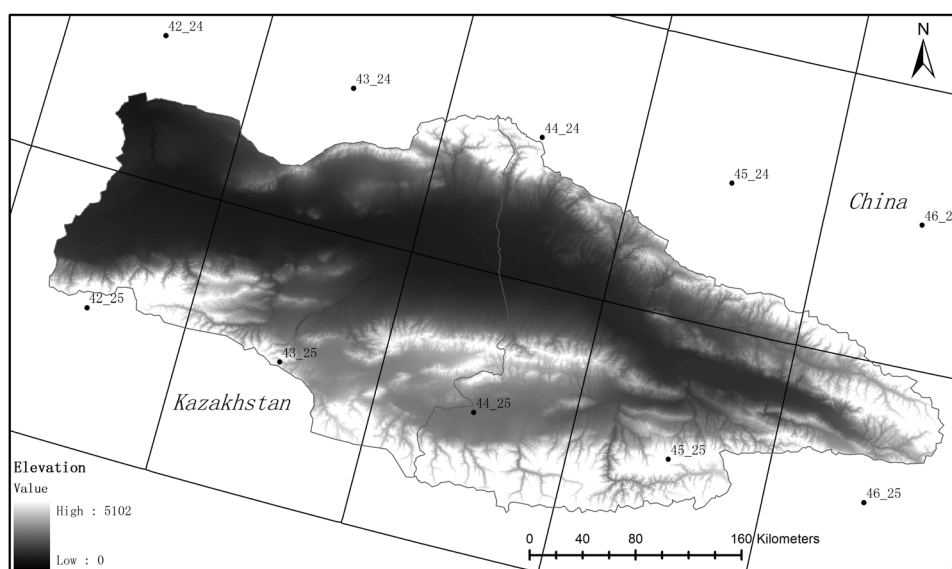
The temperature obtained directly from the Klemen method is the instantaneous value when the satellite (Terra) passes by, while the daily mean value is required in most hydrological models. The temporal scale of the simulated temperature must be transformed before it can be used for hydrological modeling. In this study, the time-scale transformation was based on the statistical relationship between the instant temperature and the daily mean value. Unlike in the research carried out by Colombi, few weather stations exist in the study area, so an insufficient number of observations are available to establish the relationship between instant temperature and daily mean temperature. The NCAR/NCEP reanalysis temperature data were introduced to enhance this relationship.

As described above (in Sec. 2.2), the NCAR/NCEP reanalysis temperature data provide four instant temperature values in the study area (UTC 00:00, 06:00, 12:00, and 18:00) each day. The local time in the study area is  $\sim 5\text{h}$  before the UTC time, which means that the UTC 06:00 is approximately 11:00 in the study area, very close to the time the satellite (Terra) passes this region (the time that Terra passes the study area varies in local time from about 10:00 a.m. to about 12:00 a.m.). If we premise two simple assumptions, the daily average temperature can be derived: (1) a strong relationship exists between the NCAR/NCEP 06:00 temperature and the daily mean NCAR/NCEP temperature and (2) the same relationship exists between the derived instantaneous Klemen  $T_{2m}$  and the simulated daily mean temperature.

Figure 2 shows 10 NCAR/NCEP pixels of the study area and the relationship between the instant temperature value (UTC 06:00) and the daily mean value (the mean of the four NCAR/NCEP instant values) that was established for each pixel [Eqs. (2) to (11)].

$$\text{NCEP42\_24}: y = 0.85x - 1.87 \quad R^2 = 0.96 \quad \text{RSMD} = 1.43^\circ\text{C}, \quad (2)$$

$$\text{NCEP43\_24}: y = 0.84x - 1.96 \quad R^2 = 0.98 \quad \text{RSMD} = 1.59^\circ\text{C}, \quad (3)$$



**Fig. 2** Location of the National Center for Atmospheric Research/National Centers for Environmental Prediction pixels over the study area.

$$\text{NCEP44\_24}: y = 0.82x - 1.81 \quad R^2 = 0.97 \quad \text{RSMD} = 1.68^\circ\text{C}, \quad (4)$$

$$\text{NCEP45\_24}: y = 0.82x - 1.86 \quad R^2 = 0.98 \quad \text{RSMD} = 1.49^\circ\text{C}, \quad (5)$$

$$\text{NCEP46\_24}: y = 0.84x - 2.83 \quad R^2 = 0.98 \quad \text{RSMD} = 1.53^\circ\text{C}, \quad (6)$$

$$\text{NCEP42\_25}: y = 0.92x - 5.72 \quad R^2 = 0.96 \quad \text{RSMD} = 1.57^\circ\text{C}, \quad (7)$$

$$\text{NCEP43\_25}: y = 0.87x - 4.11 \quad R^2 = 0.98 \quad \text{RSMD} = 2.69^\circ\text{C}, \quad (8)$$

$$\text{NCEP44\_25}: y = 0.87x - 4.18 \quad R^2 = 0.97 \quad \text{RSMD} = 1.95^\circ\text{C}, \quad (9)$$

$$\text{NCEP45\_25}: y = 0.85x - 1.92 \quad R^2 = 0.97 \quad \text{RSMD} = 1.88^\circ\text{C}, \quad (10)$$

$$\text{NCEP46\_25}: y = 0.83x - 1.93 \quad R^2 = 0.98 \quad \text{RSMD} = 2.10^\circ\text{C}. \quad (11)$$

Good correlation between the instant NCAR/NCEP value (UTC 06:00) and the daily mean value (obtained from NCAR/NCEP instant values) can be found from Eqs. (2) to (11), each of which stands for the linear relationship between the NCAR/NCEP 06:00 instant temperature and the daily mean NCAR/NCEP temperature within each NCAR/NCEP pixels. The  $R^2$  of these statistical equations vary from 0.96 to 0.98 and most RMSE are  $<2^\circ\text{C}$ , proving the validity of assumption 1 in all 10 NCAR/NCEP pixels.

Though assumption 2 was not directly proven like assumption 1, it is still supported indirectly by the following analysis. The NCAR/NCEP reanalysis data were generated through data assimilation system using multisource observations with advanced quality control and monitoring components.<sup>26–28</sup> It is widely used in the research and climate monitoring communities. So the NCAR/NCEP air temperature values in this study were supposed to be adequate to demonstrate the comprehensive conditions within NCAR/NCEP pixels, and the temporal scale transformation equations [Eqs. (2) to (11)] from the statistical analysis of NCAR/NCEP data are reasonable and representative within NCAR/NCEP pixels. As described above, the time of the derived instantaneous  $T_{2m}$  was close to the time of NCAR/NCEP 06:00 temperature, so the temporal scale transformation equations between NCAR/NCEP 06:00 temperature and daily mean value based on assumption 1 of using reanalysis temperature data can be applied to estimate the daily mean temperature using derived instantaneous  $T_{2m}$ .

In addition, the derived instantaneous  $T_{2m}$  data using Klemen method has higher spatial resolution than the NCAR/NCEP pixels, so there are many instantaneous  $T_{2m}$  pixels within an NCAR/NCEP pixel. In this study, the temporal scale transformation equation of each NCAR/NCEP pixel was applied to the smaller  $T_{2m}$  pixels within its coverage.

### 3.2 New Daily Average Temperature from Downscaled GLDAS Data and Klemen Method

The GLDAS temperature data were used to improve the simulations accuracy at low temperatures from Sec. 3.1. Original GLDAS datasets provided the near surface air temperature with a high temporal resolution of 3 h and a coarse spatial resolution of 0.25 deg. Before adjusting the simulated temperature data based on the MODIS products, the GLDAS temperature data were downsampled to the resolution of the MODIS grid size.

The downscaling process rested on an assumption that the GLDAS temperature value of each grid represents the average condition within this grid, so the GLDAS temperature value of each grid should equal the temperature at the average elevation of the grid. On the basis of this assumption, we introduced the temperature lapse rate to quantify the extent of temperature change with elevation [Eq. (12)]. Along with the prerequisites described above, the core of this method is the introduction of a higher (compared with GLDAS) spatial resolution (1 km) digital elevation model (DEM) refining the spatial distribution of the GLDAS temperature value and proper temperature lapse rate value assignment.

$$\Delta T = \Delta H * \delta, \quad (12)$$

where  $\Delta T$  stands for the temperature difference between two places A and B,  $\Delta H$  is the altitude difference between B and A, and  $\delta$  is the temperature lapse rate.

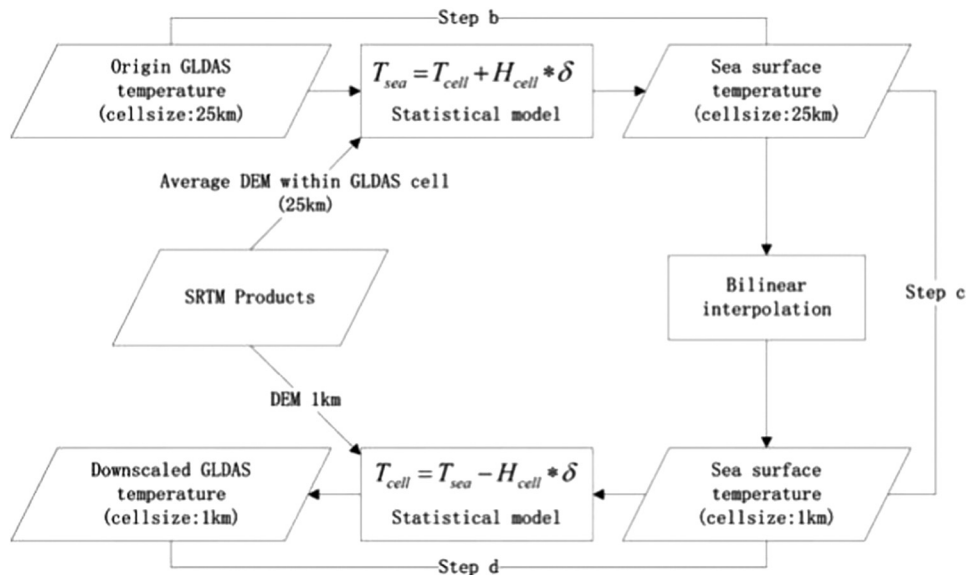
The lapse rate at which air cools with elevation change varies from  $\sim 0.98^\circ\text{C}$  per 100 m for dry air (i.e., the dry-air adiabatic lapse rate) to  $\sim 0.40^\circ\text{C}$  per 100 m (i.e., the saturated adiabatic lapse rate).<sup>33</sup> In this study, the temperature lapse rate was fixed at  $0.56^\circ\text{C}/100$  m according to the previous research work carried out in the study area.<sup>34</sup> The downscaling of GLDAS consists of four steps: (1) the original GLDAS daily mean temperature data (the mean value of eight 3 h instant temperature values) was calculated; (2) the average elevation with the spatial resolution of  $\sim 25$  km matching the original GLDAS grid size was prepared; (3) the temperature at sea level with a spatial resolution of 25 km was calculated according to Eq. (12) using data from (1) and (2) and then bilaterally interpolated to 1 km; (4) the final downscaled temperature result with a spatial resolution of 1 km was calculated by data from (3) and DEM (1 km) using Eq. (12). Figure 3 gives the procedure for downscaling.

Based on the two daily mean air temperature data calculated above, new combined temperature data were generated. According to our simulation results (details in Sec. 4), the daily mean values based on the Klemen method show good heterogeneity of spatial distribution of temperature and perform better over a certain threshold, with these values becoming less accurate below this threshold. The downscaled GLDAS daily mean temperature data were used to replace the values based on the Klemen method under the threshold because of their higher accuracy compared with the station observations at low temperature. However, above the threshold, the values based on the Klemen method were retained because of their ability to represent the spatial heterogeneity of temperatures affected by various underlying parameters, unlike the downscaled temperature data, which are dominated by elevation only. The flow chart of the new daily mean temperature data generation is shown in Fig. 4.

## 4 Results

### 4.1 Klemen Method–Based Daily Average Temperature

As described above, the station observations used in this study were recorded at a height of 1.5 m, which differs from the height of simulated surface daily average air temperature. So,



**Fig. 3** Flow chart of the global land data assimilation system (GLDAS) temperature downscaling.  $T_{sea}$  and  $T_{cell}$  mean the air temperature at sea level and pixel, respectively.  $H_{cell}$  is the pixel elevation and  $\delta$  means the temperature lapse rate.



before the validation was carried out, the simulations were transformed to the height of 1.5 m. The temperature lapse rate in Sec. 3.2 was used for this transformation. The lapse rate at  $0.56^{\circ}\text{C}/100\text{ m}$  should be appropriate and reasonable for the reason that it was from the local study in this region based on the weather stations observations.

#### 4.1.1 Derived instantaneous temperature

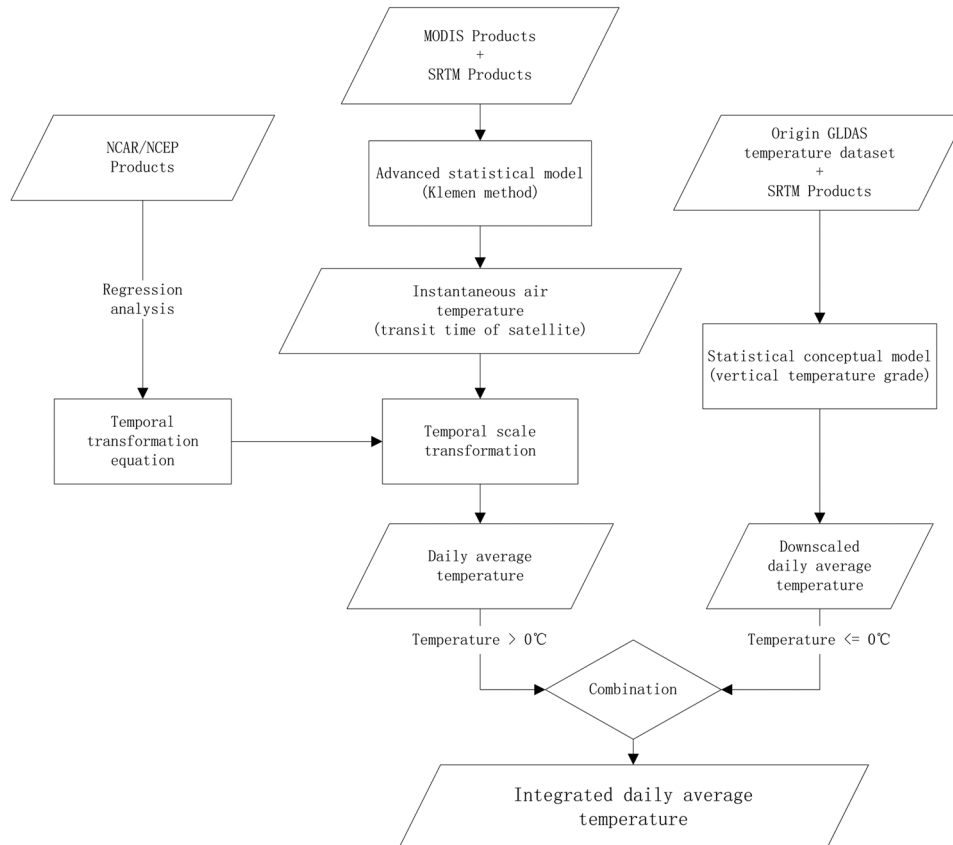
The instant air temperature was simulated and validated with the field observations. The recorded instantaneous air temperature using HOBO at Ahyz was selected according to the transit time of the Terra satellite. The simulation results and HOBO observations with comparable time were then used for the accuracy analysis. Because of the data missing from the MODIS products (no values were available in some image pixels), only part of the observations from HOBO during the observation period have matched simulated values (Fig. 5).

The derived instantaneous temperature has a satisfactory result. The results indicate that the instantaneous air temperature based on multisource spatial data has a high correlation with the observations, and the  $R^2$  can be as high as 0.73 and the RMSE between derived air temperature and observed values is  $2.65^{\circ}\text{C}$ .

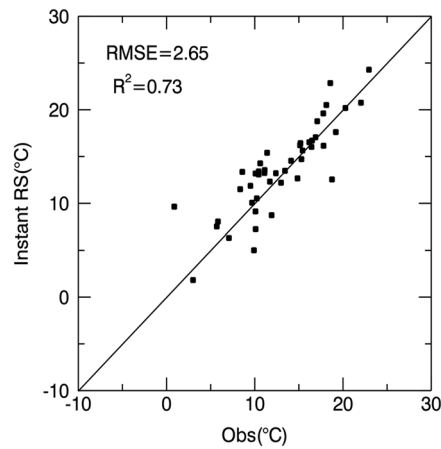
#### 4.1.2 Daily average temperature

The derived instantaneous air temperature was transformed to a daily average value and was then validated using the observed daily average temperature from 2005 to 2009 at the Zhaosu and Yining stations; the daily mean temperature observations from HOBO equipment were also used.

Figure 6 shows that the daily average temperature after temporal transformation was less accurate compared to the instantaneous values (Fig. 5). The RMSE and  $R^2$  at the Zhaosu station during the simulation period are  $4.9^{\circ}\text{C}$  and 0.91, respectively. These two indexes performed



**Fig. 4** The flow chart of daily average temperature estimation using the combined method.



**Fig. 5** Derived instantaneous temperature validation use HOBO observations.

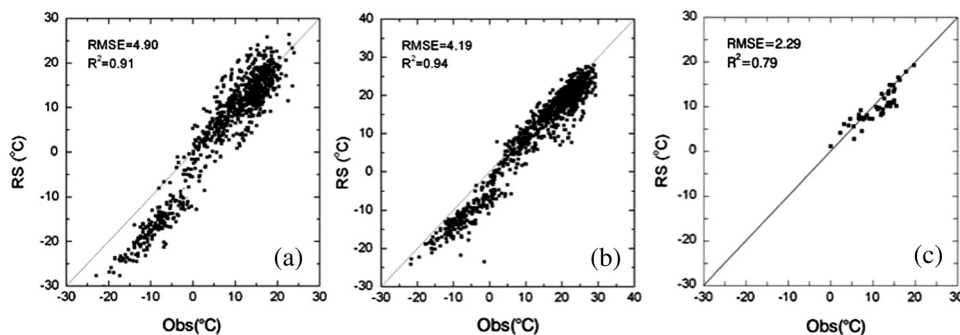
better at the Yining station where the RMSE is  $4.19^{\circ}\text{C}$  and the  $R^2$  is 0.94. When validated using the observed HOBO daily average air temperature, the results based on the Klemen method have a good performance, where the  $R^2$  is 0.79 and the RMSE is  $2.29^{\circ}\text{C}$ . According to Fig. 6, the derived daily average temperature was more consistent with the observations at high temperatures. The simulated daily average temperature was lower than the actual value when the temperature was low, and the difference increases as the temperature decreases. The threshold is  $\sim 0^{\circ}\text{C}$ .

#### 4.2 New Daily Average Air Temperature

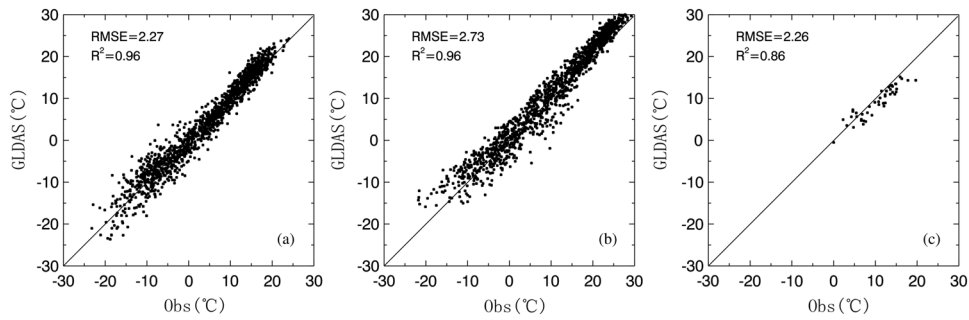
Similarly, the observed daily average temperature from 2005 to 2009 at the Zhaosu and Yining stations, and the observations from HOBO were also used for downscaled GLDAS daily average temperature validation. The results are shown in Fig. 7.

As shown in Fig. 7, the downscaled GLDAS daily average temperature performed well at both weather stations. According to the regression analysis, the  $R^2$  at the Zhaosu station is 0.96 and the RMSE is  $2.27^{\circ}\text{C}$  during the simulation period, while the  $R^2$  and RMSE at the Yining station are 0.96 and  $2.73^{\circ}\text{C}$ , respectively. According to the validation at HOBO location, the downscaled GLDAS data have a similar accuracy as the data from the Klemen method. The  $R^2$  and RMSE are 0.86 and  $2.26^{\circ}\text{C}$ , respectively.

The new daily average temperature data comprised a combination of the derived daily mean temperature on the basis of remote sensing and the downscaled GLDAS data. Through the analysis of Fig. 6,  $0^{\circ}\text{C}$  was selected as the threshold value mentioned in Sec. 3.3 to determine if the simulated daily mean temperature using the Klemen method should be replaced with the GLDAS downscaled value.



**Fig. 6** Derived daily average temperature validation based on Klemen's method using observations from the Zhaosu (a) and Yining (b) stations during 2005 to 2009 and HOBO (c) from August to November in 2009.



**Fig. 7** Validation of the GLDAS temperature downscaling data at Zhaosu (a) and Yining (b) stations during 2005 to 2009 and HOBO (c) from August to November in 2009.

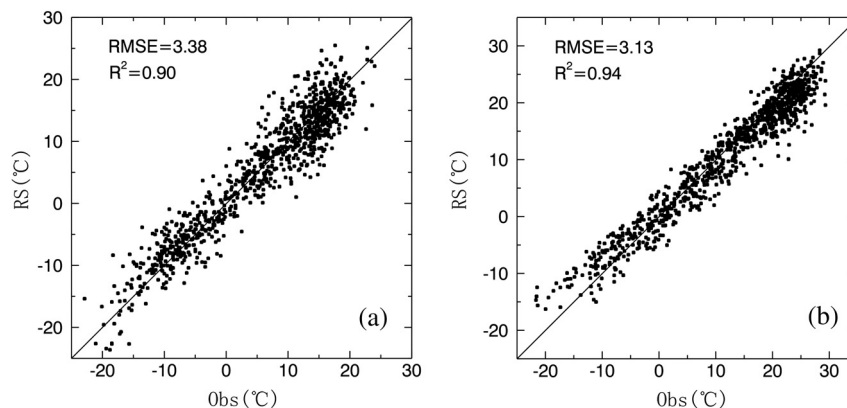
The new daily average temperature shows significant improvement at low temperatures compared with data obtained using the Klemen method (Fig. 8). During the simulation period, the  $R^2$  between the combined daily average temperature and observations from the Zhaosu station was 0.90, and this value was 0.94 at the Yining station. The RMSE at the Zhaosu station is  $3.38^\circ\text{C}$ , while the RMSE is  $3.13^\circ\text{C}$  at the Yining station.

## 5 Discussion

### 5.1 Klemen Method–Based Daily Average Temperature

The instantaneous temperature simulated on the basis of multisource spatial data using the Klemen method performs better than the daily average value derived by temporal transformation. Two factors can possibly explain the obvious decrease in accuracy: temporal transformation and the characteristics of the Klemen method.

The relationship between the instantaneous temperature at the satellite transit time and the daily mean value was assumed to be linear in this study. This simplification will certainly introduce uncertainties. The relationship established using NCAR/NCEP data stands for the comprehensive level within each NCAR/NCEP grid. Error will also be introduced when this relationship was used in pixels with higher spatial resolution. However, considering the high correlation between the temperature at the transit time and the daily average value described in Sec. 3.1.2, the high accuracy of the simulations validated using HOBO observations at both instant and daily time scale, and the obvious underestimation shown in Fig. 6 for low temperatures, the temporal transformation should not be primarily responsible for the decrease in accuracy in the daily average temperature simulation.



**Fig. 8** Validation of the new derived daily average temperature using observations from the Zhaosu (a) and Yining (b) stations during 2005 and 2009.

Equation (1) in Sec. 3.1.1 indicates that many parameters were included in the Klemen method to depict the influence of various factors, such as radiation, season, vegetation, and topography. The air temperature derived on the basis of this method shows sufficient spatial distribution details, and the spatial pattern of the air temperature depending on the underlying surface conditions is more reasonable (see Sec. 5.3). The Klemen method was proposed by Zakšek<sup>24</sup> using observations (from May to December 2005) from dozens of meteorological stations in Slovenia, Germany, and France. Due to the warm and humid climate in Europe, most of the observations from these stations in 2005 fell into the temperature span above 0°C. Thus, Eq. (1), which is based on these observations, may be innately more suitable for use at high temperatures. The simulated instantaneous air temperature and daily average air temperature validated using HOBO observations [Figs. 5 and 6(c)] have high accuracy and may profit from the field observation time between August and September, when the temperature is high. The climate conditions in the study area are different from where Eq. (1) was established; the temperature span is wider and more low temperatures (<0°C) appear, which may make the regional LST and NDVI in the study area no longer fit the relationship calibrated in Europe at low temperatures.

In this study, no further atmospheric correction was applied to the LST data from MODIS products. The atmospheric effects such as air moisture were partly removed or corrected when the LST was generated using the generalized split-window method; the accuracy of MODIS LST is  $\sim \pm 1\text{K}$ . The air moisture, which will become lower and drier at low temperatures, will certainly influence the simulation accuracy, but it should not mainly account for the low simulation accuracy of this method (the overall RMSE of daily average simulation is  $>4\text{K}$ , Fig. 6) at low temperatures.

In addition, in this study, the transformation equations obtained at coarse spatial resolution based on NCAR/NCEP data were applied to the smaller  $T_{2m}$  pixels within each NCAR/NCEP pixel with MODIS grid size. Uncertainties will be introduced in this procedure because the statistical relationship between instant air temperature and the daily average values vary when the spatial resolutions change. This method was carried out for its simplicity and feasibility in the data sparse area where the transformation equations with high spatial resolution are difficult to establish and the computational burden will be heavy. In further study, more spatial resolution independent transformation equations are needed.

## 5.2 Downscaled GLDAS Daily Average Temperature Data

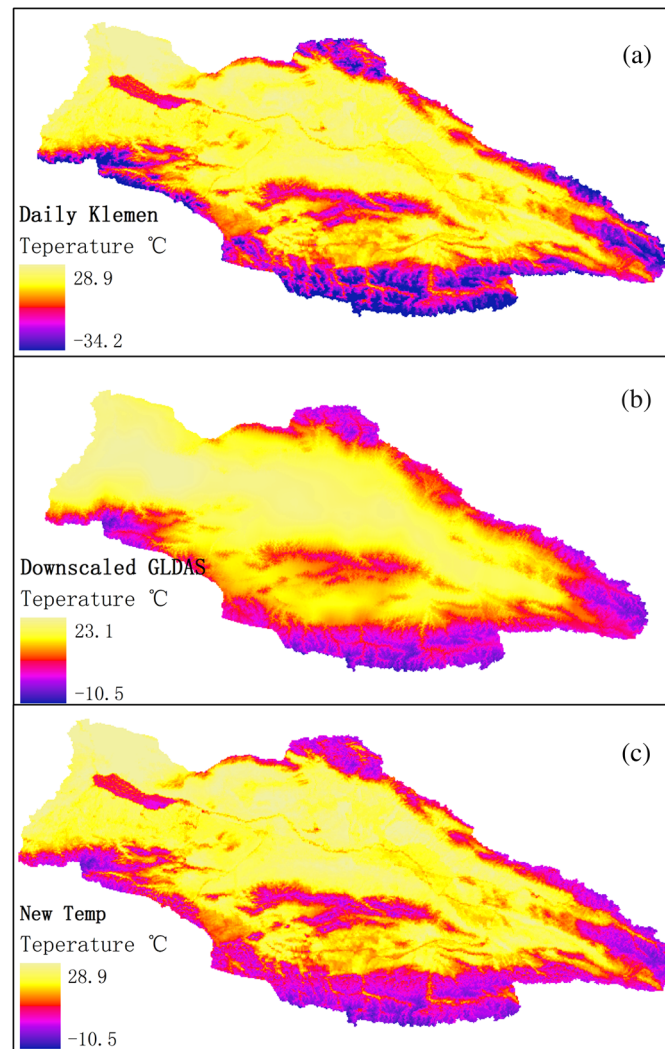
The downscaled GLDAS daily average temperature performed better at two validated weather stations than the data based on the Klemen method [Figs. 7(a) and 7(b)]. The RMSE between the downscaled data and the observation is within 3 K, while the RMSE is  $>4\text{K}$  in Figs. 6(a) and 6(b). Though the downscaled data have better accuracy at the stations, this improved accuracy did not equate to higher data quality. The downscaling process was based on the temperature lapse rate and the assumption of a relationship between the air temperature and the elevation (see Sec. 3.2). During this process, no factors other than elevation in this area were taken into account to downscale the GLDAS air temperature data. Without considering the underlying surface complexity, the downscaled GLDAS temperature data may face challenges where the underlying surface conditions (water surface, dense vegetation, aspect, and so on) have significant influences. The comparison between Figs. 6(c) and 7(c) shows that the data from downscaled GLDAS and the Klemen method have similar accuracy or the data from the Klemen method perform a little better at high temperatures using HOBO observations during the observation period. It is reasonable to believe that the downscaled daily average air temperature may not be more accurate than the data from the Klemen method for other locations (such as the valley in Fig. 9) when the temperature is high.

The reasonable regional spatial distribution and interrelation of parameters are important in the hydrological or ecological applications. The spatial pattern of the downscaled air temperature data was determined only by the elevation, following the variation of DEM. Due to more physically based parameters being introduced into the simulation using the Klemen method, the derived daily average temperature has a better spatial pattern and a better depiction of the distribution, especially at high temperatures.

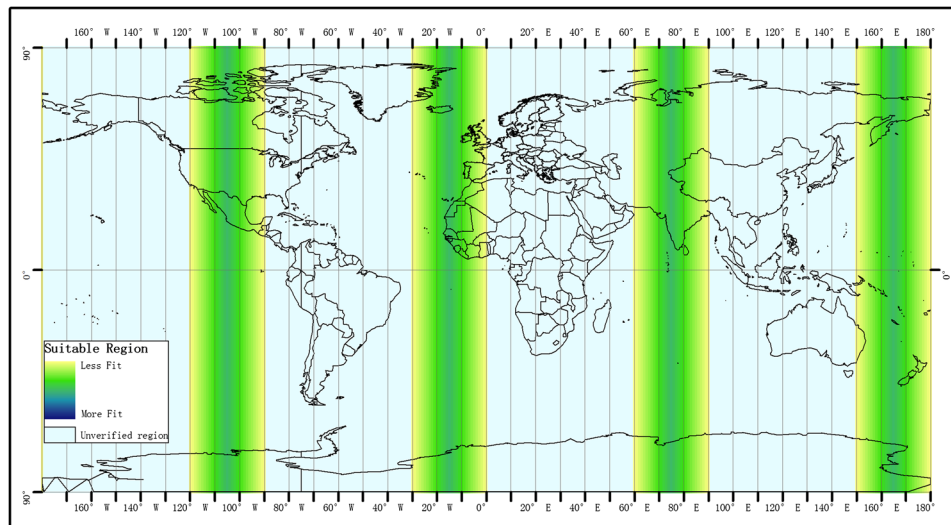
In addition, though the temperature lapse rate used in the spatial downscaling of GLDAS data and in the process of translating the derived daily air temperature data to the height of 1.5 m above the ground was based on previous study in this area using meteorological observations, problems still exist. The temperature lapse rate in this study was fixed to the constant for the whole study area while values with spatial variation are more reasonable, though it is difficult to obtain the spatial distribution of lapse rate in the data sparse region. The temperature lapse rate is more suitable to demonstrate the variation of temperature with elevation in large scale, so it may be coarse to use the temperature lapse rate for the temperature transformation with only 0.5 m height differences. A more accurate algorithm should be discussed in further study to eliminate the height differences in the simulated daily average temperature and observations.

### 5.3 New Daily Average Temperature Data

The new data were generated from the combination of the two daily average air temperatures discussed previously. Due to the discussion in Sec. 5.2, the daily average temperature based on the Klemen method was adopted at high temperatures ( $>0^{\circ}\text{C}$ ) to keep its reasonable spatial pattern and was replaced by the downscaled GLDAS temperature data (see Sec. 3.2) to increase the accuracy at low temperatures ( $<0^{\circ}\text{C}$ ).



**Fig. 9** Comparison of the results from three daily average temperature retrieval methods (take the 275th day in 2005, for example). (a) Results from the Klemen method after temporal transformation. (b) Downscaled GLDAS data. (c) Integrated daily average temperature data.



**Fig. 10** Regions (four gradient color regions) fit for the daily average air temperature retrieving method in this study. The applicability of this method in other regions (with background blue) is unverified.

At high temperatures, more details about the air temperature distribution are shown in Figs. 9(a) and 9(c), reflecting the influence of various underlying surface conditions. A large reservoir named Kapchagay exists in the study area downstream of the Ili River. The water surface evaporation cools the air above the reservoir leading to temperatures (dark gray area under the red rectangle) lower than the surrounding temperatures. A large area of the Gobi Desert lies within the study area (the bright parts), where the temperature is obviously higher. Most of the light gray areas in Figs. 9(a) and 9(c) cover well-developed vegetation. Many of these spatial pattern details cannot be seen in the downscaled air temperature distribution [Fig. 9(b)].

At the low temperatures [the black areas in Fig. 9(a)], most of which appear on the top of the mountain, the data from the Klemen method become less accurate. A useful spatial texture of the air temperature can no longer be observed, and the accuracy of the simulated temperatures obviously decreases. The downscaled GLDAS air temperature data with the DEM texture were used.

In this study, the derived daily average air temperature based on the Klemen method was simply replaced with the downscaled GLDAS data to achieve overall simulation accuracy. In the future, weights may be introduced to integrate these two temperature data. Different weight values according to certain rules will be assigned to the Klemen method data and the downscaled GLDAS data to reflect the contributions from each, instead of simply replacing one with another. In order to achieve better overall simulation accuracy, the coefficients used in the Klemen method can also be recalibrated in a wider temperature span to make it more suitable in different regions.

The temporal scale transformation equations used in this study were based on the NCAR/NCEP reanalysis temperature data (details in Sec. 3.1.2). NCAR/NCEP provides four instant values (UTC 00:00, 06:00, 12:00, and 18:00) each day. If the Terra passing time in certain regions is close to any of the four times, the temporal scale transformation equations and the daily average air temperature can be made in the same way. In this study, the time that satellite passes the study area varies from about local 10:00 a.m. to about local 12:00 a.m., is about an hour before or after the UTC 06:00 (local time 11:00 a.m.). So that the local satellite passing time is close to any of the four NCAR/NCEP time means the difference between them is <1h. Figure 10 shows the regions around the world where the daily average air temperature deriving method in this study can be used.

## 6 Conclusions

This study shows how the daily average air temperature at 2 m height can be obtained using multisource spatial datasets over a large area, especially in data sparse regions. An integrated

method using the combination of an advanced statistical model (the Klemen method) with temporal transformation and spatial downscaling was proposed. Using the integrated method, air temperature was estimated during the study period with an accuracy of  $\sim 3\text{K}$  for two stations in the Ili River basin in Central Asia; the portability of this method is also discussed.

The Klemen method provides a depiction of the spatial pattern and distribution of the air temperature by incorporating parameters such as radiation, season, and underlying surface conditions. At high temperatures, this method has high simulation accuracy. One disadvantage of this approach is its less accurate simulation at low temperatures. The Klemen method was developed and validated using observations from meteorological stations in Germany, Slovenia, and France between May and December 2005, when low temperatures ( $<0^\circ\text{C}$ ) seldom appear. The statistical relationship between the air temperature and parameters developed under such conditions does not fit for application at low temperatures. Further research can be carried out to have the coefficients in the Klemen equations recalibrated in a wider temperature span to make this method more suitable at low temperatures.

The downscaling procedure used in this study is based on the vertical temperature gradient. The origin GLDAS temperature data with coarse spatial resolution were spatially detailed by introducing DEM with higher resolution. Though this procedure does not suffer from inaccuracy at low temperatures, it does not account for other factors influencing the air temperature. For further applications, additional factors related to the season, radiation, and underlying surface conditions should be taken into consideration during the downscaling.

The daily average temperature data obtained by the integrated method provided both sufficient spatial distribution details and high simulation accuracy, which were especially valuable because no field observations were used in the simulation. In this study, the integrated data comprised data from the Klemen method after temporal transformation for high temperatures and downscaled GLDAS data for low temperatures. More sophisticated methods can be developed by combining these two datasets in future applications.

According to the discussion, the regions where the satellite (Terra) passing time is close to any of the four UTC times provided by the NCAR/NCEP, which is about one third of the total global area, can use the method proposed in this study to estimate the daily average air temperature. The feasibility of this method in other regions in the world needs further study.

## Acknowledgments

The authors would like to thank all the researchers who participated in the project. The authors would also like to thank the Ministry of Water Resources of China (Grant Nos. 201101037 and 201201092), the General Program of National Natural Science Foundation of China (Grant No. 41271414), and the Fundamental Research Funds for the Central Universities for financial support. Furthermore, the authors would like to thank the anonymous reviewers for their helpful and constructive comments.

## References

1. T. C. Peterson et al., "A blended satellite-in situ near-global surface temperature dataset," *Bull. Am. Meteorol. Soc.* **81**(9), 2157–2164 (2000), [http://dx.doi.org/10.1175/1520-0477\(2000\)081<2157:ABSSNS>2.3.CO;2](http://dx.doi.org/10.1175/1520-0477(2000)081<2157:ABSSNS>2.3.CO;2).
2. S. Anderson, "An evaluation of spatial interpolation methods on air temperature in Phoenix, AZ," <http://www.cobblestoneconcepts.com/ucgis2summer/anderson/anderson.htm> (2002).
3. E. Florio et al., "Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: a statistical approach," *Int. J. Remote Sens.* **25**(15), 2979–2994 (2004), <http://dx.doi.org/10.1080/01431160310001624593>.
4. B. Tang et al., "Generalized split-window algorithm for estimate of land surface temperature from Chinese geostationary FengYun meteorological satellite (FY-2C) data," *Sensors* **8**(2), 933–951 (2008), <http://dx.doi.org/10.3390/s8020933>.
5. K. Mao et al., "A practical split-window algorithm for retrieving land-surface temperature from MODIS data," *Int. J. Remote Sens.* **26**(15), 3181–3204 (2005), <http://dx.doi.org/10.1080/01431160500044713>.

6. C. Coll, V. Caselles, and T. J. Schmugge, "Imation of land surface emissivity differences in the split-window channels of AVHRR," *Remote Sens. Environ.* **48**(2), 127–134 (1994), [http://dx.doi.org/10.1016/0034-4257\(94\)90135-X](http://dx.doi.org/10.1016/0034-4257(94)90135-X).
7. J. C. Price, "Land surface temperature measurements from the split window channels of the NOAA 7 advanced very high resolution radiometer," *J. Geophys. Res.: Atmos.* **89**(D5), 7231–7237 (1984), <http://dx.doi.org/10.1029/JD089iD05p07231>.
8. A. Pinheiro et al., "Development of a daily long term record of NOAA-14 AVHRR land surface temperature over Africa," *Remote Sens. Environ.* **103**(2), 153–164 (2006), <http://dx.doi.org/10.1016/j.rse.2006.03.009>.
9. Z. Wan et al., "Validation of the land-surface temperature products retrieved from terra moderate resolution imaging spectroradiometer data," *Remote Sens. Environ.* **83**(1), 163–180 (2002), [http://dx.doi.org/10.1016/S0034-4257\(02\)00093-7](http://dx.doi.org/10.1016/S0034-4257(02)00093-7).
10. M. Atitar and J. A. Sobrino, "A split-window algorithm for estimating LST from Meteosat 9 data: test and comparison with in situ data and MODIS LSTs," *IEEE Geosci. Remote Sens. Lett.* **6**(1), 122–126 (2009), <http://dx.doi.org/10.1109/LGRS.2008.2006410>.
11. E. Chen et al., "Comparison of winter-nocturnal geostationary satellite infrared-surface temperature with shelter—height temperature in Florida," *Remote Sens. Environ.* **13**(4), 313–327 (1983), [http://dx.doi.org/10.1016/0034-4257\(83\)90033-0](http://dx.doi.org/10.1016/0034-4257(83)90033-0).
12. R. M. Green and S. I. Hay, "The potential of Pathfinder AVHRR data for providing surrogate climatic variables across Africa and Europe for epidemiological applications," *Remote Sens. Environ.* **79**(2), 166–175 (2002), [http://dx.doi.org/10.1016/S0034-4257\(01\)00270-X](http://dx.doi.org/10.1016/S0034-4257(01)00270-X).
13. G. V. Mostovoy et al., "Statistical estimation of daily maximum and minimum air temperatures from MODIS LST data over the state of Mississippi," *GIsci. Remote Sens.* **43**(1), 78–110 (2006), <http://dx.doi.org/10.2747/1548-1603.43.1.78>.
14. A. Colombi et al., "Estimation of daily mean air temperature from MODIS LST in Alpine areas," *EARSeL eProc.* **6**(1), 38–46 (2007).
15. R. R. Nemani and S. W. Running, "Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data," *J. Appl. Meteorol.* **28**(4), 276–284 (1989), [http://dx.doi.org/10.1175/1520-0450\(1989\)028<0276:EORSRT>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1989)028<0276:EORSRT>2.0.CO;2).
16. S. N. Goward et al., "Ecological remote sensing at OTTER: satellite macroscale observations," *Ecol. Appl.* **4**(2), 322–343 (1994), <http://dx.doi.org/10.2307/1941937>.
17. S. N. Goward, Y. Xue, and K. P. Czajkowski, "Evaluating land surface moisture conditions from the remotely sensed temperature/vegetation index measurements: an exploration with the simplified simple biosphere model," *Remote Sens. Environ.* **79**(2), 225–242 (2002), [http://dx.doi.org/10.1016/S0034-4257\(01\)00275-9](http://dx.doi.org/10.1016/S0034-4257(01)00275-9).
18. L. Pihodko and S. N. Goward, "Estimation of air temperature from remotely sensed surface observations," *Remote Sens. Environ.* **60**(3), 335–346 (1997), [http://dx.doi.org/10.1016/S0034-4257\(96\)00216-7](http://dx.doi.org/10.1016/S0034-4257(96)00216-7).
19. S. Stisen et al., "Estimation of diurnal air temperature using MSG SEVIRI data in West Africa," *Remote Sens. Environ.* **110**(2), 262–274 (2007), <http://dx.doi.org/10.1016/j.rse.2007.02.025>.
20. R. Pape and J. Löffler, "Modelling spatio-temporal near-surface temperature variation in high mountain landscapes," *Ecol. Modell.* **178**(3), 483–501 (2004), <http://dx.doi.org/10.1016/j.ecolmodel.2004.02.019>.
21. J. Remund, "Quality of Meteornorm Version 6.0," [http://www.meteotest.ch/fileadmin/user\\_upload/Sonnenenergie/pdf/2008WREC\\_mn6\\_qual\\_paper.pdf](http://www.meteotest.ch/fileadmin/user_upload/Sonnenenergie/pdf/2008WREC_mn6_qual_paper.pdf) (2008).
22. Y. J. Sun et al., "Air temperature retrieval from remote sensing data based on thermodynamics," *Theor. Appl. Climatol.* **80**(1), 37–48 (2005), <http://dx.doi.org/10.1007/s00704-004-0079-y>.
23. J. D. Jang, A. Viau, and F. Anctil, "Neural network estimation of air temperatures from AVHRR data," *Int. J. Remote Sens.* **25**(21), 4541–4554 (2004), <http://dx.doi.org/10.1080/01431160310001657533>.
24. K. Zakšek and M. Schroedter-Homscheidt, "Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments," *ISPRS J. Photogramm. Remote Sens.* **64**(4), 414–421 (2009), <http://dx.doi.org/10.1016/j.isprsjprs.2009.02.006>.



25. K. P. Czajkowski et al., "Thermal remote sensing of near surface environmental variables: application over the Oklahoma Mesonet," *Prof. Geogr.* **52**(2), 345–357 (2000), <http://dx.doi.org/10.1111/0033-0124.00230>.
26. E. Kalnay et al., "The NCEP/NCAR 40-year reanalysis project," *Bull. Am. Meteorol. Soc.* **77**(3), 437–471 (1996), [http://dx.doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](http://dx.doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
27. S. Saha et al., "The NCEP climate forecast system reanalysis," *Bull. Am. Meteorol. Soc.* **91**(8), 1015–1057 (2010), <http://dx.doi.org/10.1175/2010BAMS3001.1>.
28. R. Kistler et al., "The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation," *Bull. Am. Meteorol. Soc.* **82**(2), 247–267 (2001), [http://dx.doi.org/10.1175/1520-0477\(2001\)082<0247:TNNYRM>2.3.CO;2](http://dx.doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2).
29. M. Rodell et al., "The global land data assimilation system," *Bull. Am. Meteorol. Soc.* **85**(3), 381–394 (2004), <http://dx.doi.org/10.1175/BAMS-85-3-381>.
30. Z. Wan and J. Dozier, "A generalized split-window algorithm for retrieving land-surface temperature from space," *IEEE Trans. Geosci. Remote Sens.* **34**(4), 892–905 (1996), <http://dx.doi.org/10.1109/36.508406>.
31. G. V. Mostovoy et al., "Statistical estimation of daily maximum and minimum air temperatures from MODIS LST data over the state of Mississippi," *GIs. Remote Sens.* **43**(1), 78–110 (2006), <http://dx.doi.org/10.2747/1548-1603.43.1.78>.
32. Z. Wan et al., "Validation of the land-surface temperature products retrieved from terra moderate resolution imaging spectroradiometer data," *Remote Sens. Environ.* **83**(1), 163–180 (2002), [http://dx.doi.org/10.1016/S0034-4257\(02\)00093-7](http://dx.doi.org/10.1016/S0034-4257(02)00093-7).
33. R. Dodson and D. Marks, "Daily air temperature interpolated at high spatial resolution over a large mountainous region," *Clim. Res.* **8**(1), 1–20 (1997), <http://dx.doi.org/10.3354/cr008001>.
34. D. H. Chen et al., "Study on spatial interpolation of the average temperature in the Yili River Valley based on DEM," *Spectrosc. Spectral Anal.* **31**(7), 1925–1929 (2011). (in Chinese)

Biographies and photographs of the authors are not available.