

# SCIENTIFIC REPORTS



OPEN

## Investigating spatiotemporal changes of the land-surface processes in Xinjiang using high-resolution CLM3.5 and CLDAS: Soil temperature

Xianyong Meng<sup>1</sup>, Hao Wang<sup>1</sup>, Yiping Wu<sup>2</sup>, Aihua Long<sup>1</sup>, Jianhua Wang<sup>1</sup>, Chunxiang Shi<sup>3</sup> & Xiaonan Ji<sup>4</sup>

Soil temperature plays a key role in the land surface processes because this parameter affects a series of physical, chemical, and biological processes in the soil, such as water and heat fluxes. However, observation of soil temperature is quite limited, especially at the regional scale. Therefore, this study is to investigate the spatiotemporal features of soil temperature in Xinjiang, China, using the Community Land model 3.5 (CLM3.5) with the atmospheric near-surface forcing data of the China Meteorological Administration Land Data Assimilation System (CLDAS). We use the observed soil temperature data collected from 105 national automatic stations during 2009 through 2012 in the study area to verify the simulation capability. The comparison results indicate that the CLM3.5 with the CLDAS driving field could well simulate the spatiotemporal patterns of the soil temperature at hourly, daily, and monthly time scales and at three depths (5 cm, 20 cm, and 80 cm). We also produce a soil temperature database of the region that is continuous both in time and space with high resolution (about 6.25 km). Overall, this study could help understand the regional and vertical characteristics of the soil temperature and provide an important scientific basis for other land-surface processes.

Land surface accounts for one-third of Earth's surface area. Extensive research has shown not only those land-surface conditions either directly or indirectly affect both climate and the atmosphere but also that anomalies in land-surface conditions constantly interact with both climate and the atmosphere<sup>1–10</sup>. It was Charney<sup>11</sup> who, for the very first time, advanced the notion that changes in land surface will result in an abnormal change in the albedo, which in turn may further affect the surface radiation balance and eventually result in climate anomalies. Soil temperature and moisture are particularly important in the land-surface processes because they are key parameters that characterize the thermal properties and soil water content and can affect climate through their impacts on both surface energy and water budget<sup>12</sup>. Soil temperature could also impact the climate and the weather, especially on short-term weather processes<sup>13</sup>. In addition, soil temperature is an important agrometeorological element that affects a series of physical, chemical and biochemical processes in the soil. The most recent studies have shown that soil temperature can affect the short-term precipitation using the Eta model and the Weather Research and Forecasting model<sup>14,15</sup>. Hu and Feng<sup>16</sup> found that the end-of-spring soil temperature is related to summer precipitation in some degree. The anomalies in the temperatures of shallow soil layers will affect short-term weather processes because the thermal anomalies in the surface layer's soil have already been released to the atmosphere before being transferred downward to the deep layers. The anomalies in the soil of the deep layers can affect the regional climate process by gradually releasing energy to the shallow layers. Zhang

<sup>1</sup>State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin & China Institute of Water Resources and Hydropower Research, Beijing, 100038, P. R. China. <sup>2</sup>Department of Earth & Environmental Science, Xi'an Jiaotong University, Xi'an, 710049, P. R. China. <sup>3</sup>National Meteorological Information Center, China Meteorological Administration, Beijing, 100081, P. R. China. <sup>4</sup>Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, 830011, P. R. China. Correspondence and requests for materials should be addressed to H.W. (email: [wanghao@iwhr.com](mailto:wanghao@iwhr.com)) or Y.W. (email: [yipingwu@xjtu.edu.cn](mailto:yipingwu@xjtu.edu.cn)) or A.L. (email: [ahlong@iwhr.com](mailto:ahlong@iwhr.com))

*et al.*<sup>17</sup> found a complex corresponding relationship between changes in the temperature of the atmosphere and that of the soil in Canada during the 20<sup>th</sup> century. Wang *et al.*<sup>18</sup> analyzed long-term trends of Eastern China's soil temperature and moisture and strengthened the understanding of the mechanism and important characteristics of the land-atmosphere interaction in East China. However, measurement of soil temperature is deficient, and even if there are observations, they are spatially non-continuous and discrete.

Simulating the land-surface processes is a useful approach to studying the changes in the land-surface processes and their impacts, especially at the regional scale. Community Land Model (CLM3.5) is a good example, and many scientists pay much attention to the model verification and development, Bonan and Levis<sup>19</sup> simulated the land surfaces of South America's Amazon Plain and the eastern region of the US, respectively, using the CLM series dynamic vegetation model, the CLM3.0, coupled with the Community Atmospheric Model (CAM) (CLM3.0 and CAM3.0). They concluded that it was necessary to further improve the parameterization scheme of the CLM3.0 coupled with the CAM 3.0. CLM3.0 was updated to version 3.5 by modifying a series of land surface processes (e.g., C and N cycles, surface water, and snow accumulation and frozen soil), with a better performance especially in hydrological simulation<sup>20</sup>. Sakaguchi and Zeng<sup>21</sup> proposed a flexible scheme about soil capacitance in CLM4.0, which was used to simulate the impacts of N cycle on the global carbon balance. CLM4.5 is the latest version of CLM, which has revised photosynthesis scheme, improved cold region hydrology and optional VIC-based hydrology, prognostic wetland distribution, a new snow cover fraction parameterization, a new lake model, new crop model features such as fertilization and grain fill, and multiple urban classes<sup>22</sup>, etc. However, we used CLM3.5 because applying CLDAS to drive CLM offline and its validation, while it is quite challenging to do so in CLM4.5, which deserves our future study.

Whitfield *et al.*<sup>23</sup> have conducted a comparative study of simulation results for Florida using CLM and the Land Surface Process Model (LSPM), and found these two models were very good at simulating daily changes in soil temperature. Huang *et al.*<sup>24</sup> stated that CLM3.0 could provide a very good simulation not only of various fluxes between the land and atmosphere but also of the characteristics of temporal and spatial distributions of soil temperature. Niu *et al.*<sup>25</sup> conducted verification of three different types of land surface in East Asia (paddy field, sparse vegetation on plateaus and forest) using the CLM3.0. They found that surface temperatures in the sparse vegetation areas on the plateaus simulated using CLM3.0 could fit the observed values relatively well and that CLM3.0 could satisfactorily simulate the characteristics of the change in the soil temperature at different depth. Chen *et al.*<sup>26</sup> investigated China's soil temperature using CLM3.0 driven by Princeton University's global atmospheric forcing field data and found that the coupled model could provide a better simulation of the spatial distribution of soil.

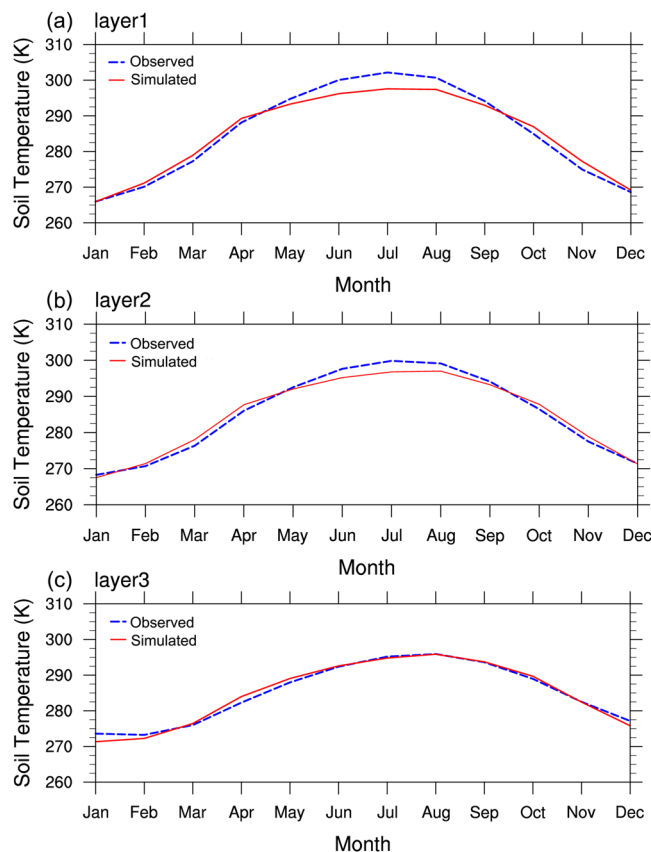
From the above examples, we can see few scientists simulate the spatially continuous distribution of soil temperature at the regional scale. Generally, soil temperatures at a study point are from data collected at a nearby meteorological station, and this could cause large uncertainty especially in an area with complicated terrain<sup>27</sup>.

The Xinjiang Uygur Autonomous Region, spans several latitudes and longitudes and has a very complicated topography. There is a lack of studies that used a land-surface model driven by high-precision atmospheric field data to simulate the continuous temporal and spatial distributions of the soil temperature in an area with a similar size. Therefore, the objective of this study is to 1) simulate soil temperature in the Xinjiang region using the CLM3.5 driven by the high-precision atmospheric driving field of the China Meteorological Administration Land Data Assimilation System (CLDAS), 2) evaluate the mode performance with observed data collected at 105 national automatic soil-temperature stations in the study area, and 3) investigate the characteristics of the spatio-temporal distributions of the soil temperatures at different depths.

## Results

**Monthly soil-temperature validation and features.** Figure 1 shows the simulated and observed monthly mean temperatures of the soil of three layers from 2009 to 2012. The results of the study show that the simulated temperatures of the soil are essentially consistent with the observed temperatures in terms of both the variation trend and the peak-to-valley value. The simulation results reflect the seasonal changes in the temperatures of the soil relatively well, indicating that the CLDAS-driven CLM3.5 can satisfactorily simulate the soil profile temperatures in Xinjiang. In addition, we find that the simulated soil temperatures in spring and autumn are the closest to their observations. From Fig. 1, the differences between the simulated and observed monthly mean temperatures of layer 1 soil in summer are the largest (approximately 8 K), followed by the differences between the simulated and observed monthly mean temperatures of layer 2 soil in summer (close to 5 K). The differences between the simulated and observed monthly mean temperatures of layer 3 soil in summer are the least. However, the simulated monthly mean temperatures of the soil of the two relatively shallow layers (layers 1 and 2) in January and December are relatively poor. In other words, the differences between the simulated and observed monthly mean temperatures of layer 1 and layer 2 soil in January and December are relatively large. We herein analyze the possible reasons for these differences. During the summer, the air temperature in the Xinjiang region changes dramatically, particularly in July, when the air temperature reaches its maximum value, which can be referred to as the extreme value of the air temperature. The surface temperature changes rapidly, but that change does not affect the soil of the deep layers. Therefore, the differences between the simulated and observed monthly mean temperatures of the soil of the shallow layers are larger than the difference between the simulated and observed monthly mean temperature of the soil of the deep layer. In winter, the cold transfers from the shallow layer to the deep layer (layer 3), which may be why the simulated monthly mean temperatures of the soil of the deep layer in winter are relatively poor.

Figure 2(a) shows the seasonal changes in the MEs of the soil temperatures at different depths. We can see that most of the simulated monthly mean temperatures of the soil of the three layers in summer (May to September) exhibit a negative deviation, but the simulated monthly mean temperatures of the soil of the three layers at all other times exhibit a positive deviation. Among the simulated monthly mean temperatures of layer 3 soil in all

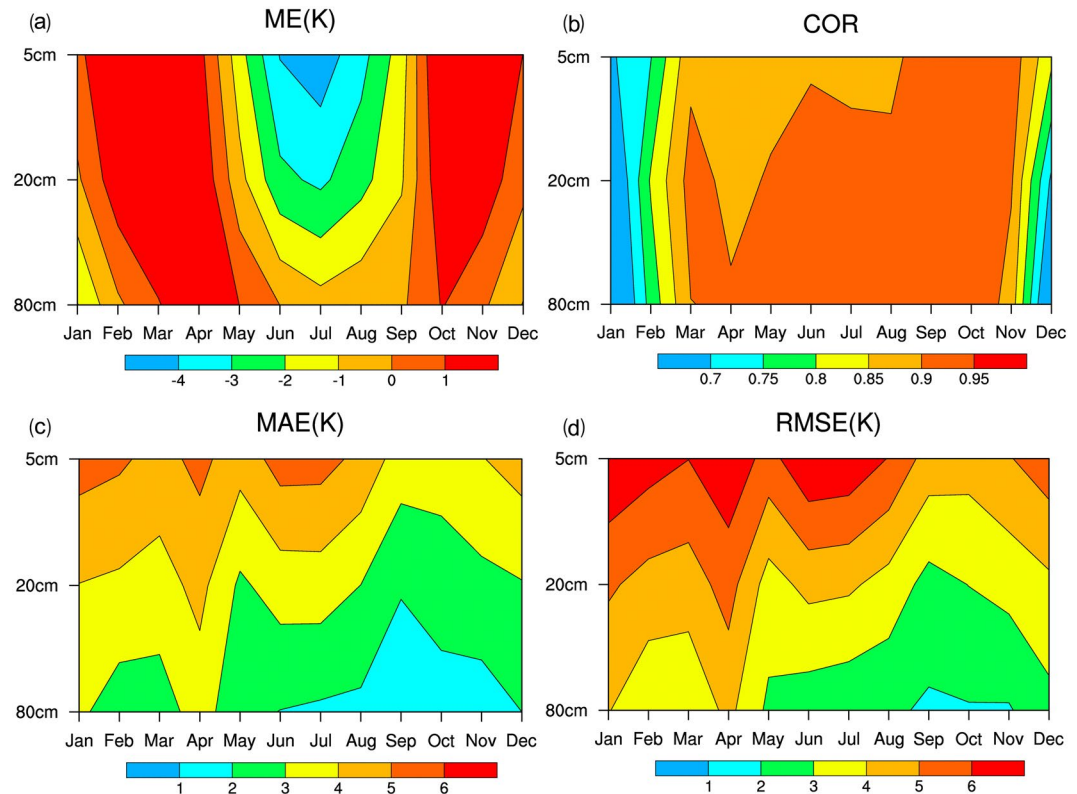


**Figure 1.** Simulation-observation curves of the monthly mean temperatures of the soil at 105 stations in Xinjiang at depths of (a) 5 cm, (b) 20 cm, and (c) 80 cm.

12 months, only the simulated monthly mean temperature in January exhibits a positive deviation. Some of the simulated monthly mean temperatures of the soil of the deeper layer in summer also exhibit a positive deviation. From Fig. 2(b) that most of the anomaly correlation coefficients of the simulated monthly mean temperatures at three layers in all seasons are greater than 0.85, except for early spring and winter, indicating that CLDAS-driven CLM3.5 has a relatively good grasp on the pattern of the seasonal change in the soil temperature. In addition, we find that the anomaly correlation coefficients of layers 1 and 2 are less than those of layer 3. The anomaly correlation coefficients of layers 1 and 2 from May to September range from 0.85 to 0.9, whereas most of the anomaly correlation coefficients of layer 3 are 0.9. From the seasonal angle, the anomaly correlation coefficients of any of the three layers in early spring or late winter are relatively low (most of the anomaly correlation coefficients are below 0.7), which may be caused by the relatively low extreme value of the air temperature. Figure 2(c) and (d) show the seasonal changes in the MAEs and RMSEs, respectively, of the soil temperatures of different layers. We find that the spatial distribution pattern of the MAE is essentially similar to that of the RMSE. In addition, the simulation errors of layer 1 are greater than those of layer 2, and the errors of layer 3 are obviously the least. The errors of layers 1 and 2 in January, February, April, June, July and December are relatively large, whereas the errors of layer 3 are comparatively small.

**Daily soil-temperature validation and features.** Figure 3 shows the time-series plots of the 105-sample average daily soil temperature of three layers simulated using CLDAS-driven CLM3.5. The comparison between the simulations and the observations at the 105 national automatic-observation stations shows that the largest difference between is less than 5 K. From Fig. 3(a and b), the differences between the simulated and observed 105-sample average daily soil temperature of layer1 and layer2 in summer and autumn are the largest (around 5 K). The differences between the simulated and observed daily temperatures of layer 3 in January and December are greater than 2 K, whereas the differences in all other time periods are quite small. These results are consistent with the results of the changes in the monthly temperatures discussed in Section 4.1.

Figure 4(a–c) show the seasonal changes in the mean deviations of the daily mean temperatures in different years. It can be seen that the CLM3.5-simulated soil temperatures at the three layers in summer exhibit a deviation from the observed temperatures. The simulated summer temperatures of layer 1 and layer 2 exhibit a more significant deviation. The simulated temperatures of the surface layer from June to September of each year are approximately 2K–4K lower than the observed ones. The simulated soil temperatures of layer 2 from June to September of each year are approximately 1K–3K lower than the observed temperatures. In addition, the simulated temperatures of layer 1 and layer 2 from January to May and from October to November of each year also has a slight deviation (1K–2K). The simulated temperatures of layer 3 in January, February and December exhibit



**Figure 2.** Statistical function graphs of the changes in the temperatures of the soil with season and depth. (a) ME, (b) Anomaly correlation, (c) MAE, and (d) RMSE.

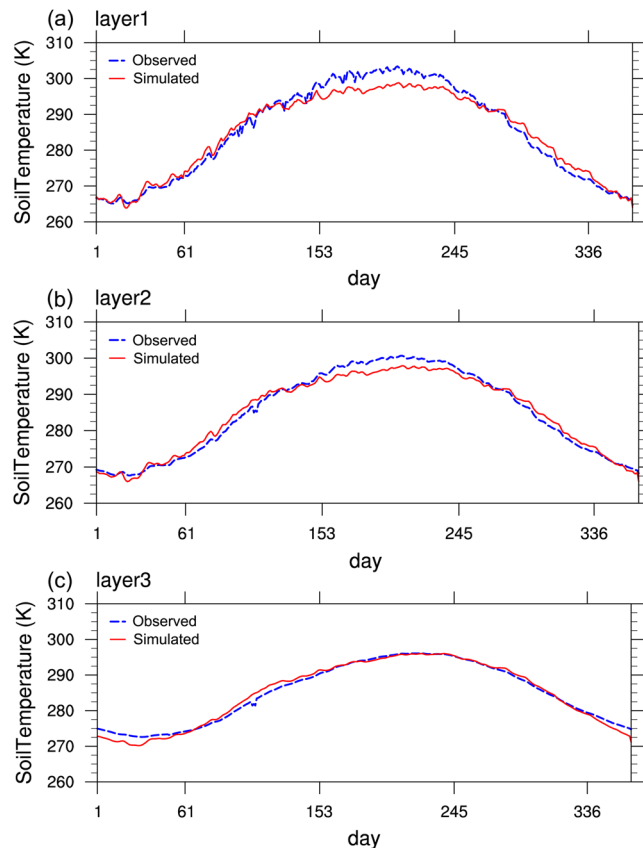
a relatively large deviation (1K–3K lower than the observed temperatures). The seasonal changes in the mean deviations of the simulated temperatures of layer 3 from the observed temperatures at all other times of the year mostly range from  $-1\text{K}$  to  $1\text{K}$ , suggesting that the seasonal fluctuation of the change in the soil temperature of layer 3 is less than that of layers 1 and 2.

**Hourly soil-temperature validation and features.** Figure 5(a1–a3) shows the hourly soil temperatures at the three layers observed at the 105 stations in the Xinjiang region (a1: layer 1, a2: 20 cm, a3: 80 cm). The abscissa axis of each plot represents the coordinated universal time (UTC) and shows both the mean temperature distribution and the mean value of the change in the daily temperature over 24 hours at the 105 soil-temperature stations. The ordinate axis of each plot shows the change in the soil-temperature data from 2009 to 2012. We find that the inter-annual change in the temperature of layer 1 is larger than the inter-annual change in the temperature of layer 2 and the inter-annual change in the temperature of layer 3 on the daily scale. From February to November of each year, the temperature of layer 1 reaches its maximum value (above  $30^\circ\text{C}$ ) at 09UTC each day. In January and December of each year, the daily average temperature of layer 1 decreases to below  $270\text{K}$ . In addition, the largest difference between the daily maximum temperature and the daily minimum temperature of layer 1 exceeds  $30\text{K}$  in all time periods except for January and December. The daily change in the temperature of layer 2 is relatively slow and occurs at a later time. The daily maximum temperature of layer 2 soil (above  $30^\circ\text{C}$ ) occurs at 15UTC, indicating that there is a delay in the transfer of the soil temperature from a shallow layer to a deep layer. Furthermore, the temperature of layer 2 was lower than zero only in early January 2009, at the end of December 2010, in early January 2011, at the end of December 2011 and in January 2012.

Figure 5(a1–a3) also shows that there is essentially no change in the temperature of layer 3 soil on the daily scale (i.e., over each 24 h day). The daily temperatures of layer 3 are between  $290\text{K}$  and  $300\text{K}$  from June to September. The highest daily temperatures of layer 3 are primarily concentrated in July ( $295\text{K}$ – $300\text{K}$ ), whereas the lowest daily temperatures of layer 3 are primarily concentrated in January. The early springs of 2011 and 2012 were relatively cold, resulting in daily temperatures of layer 3 in January and February being between  $270\text{K}$  and  $275\text{K}$ . The differences between the highest and lowest daily temperatures of layer 3 are between  $15\text{K}$  and  $25\text{K}$ , showing a gentler variation than layers 1 and 2. In addition, the energy between the land and atmosphere has a larger impact on the daily temperature of the surface layer than that on the deep layer.

In terms of the seasonal change in the soil temperature, the temperatures at all three layers of the soil reach their maximum values in summer and their minimum values in winter, suggesting clear seasonal changes.

Fig. 5(b1–b3) shows the distributions of the soil temperatures at the three layers ((b1), (b2) and (b3)) simulated using the CLDAS-driven CLM3.5. The simulated temperatures are interpolated to the matching latitudes and longitudes and depths of the 105 national automatic stations for statistical analysis. It is found that both the



**Figure 3.** Simulation-observation curves of the daily mean temperatures of the soil at the 105 stations in Xinjiang (a) 5 cm; (b) 20 cm; (c) 80 cm.

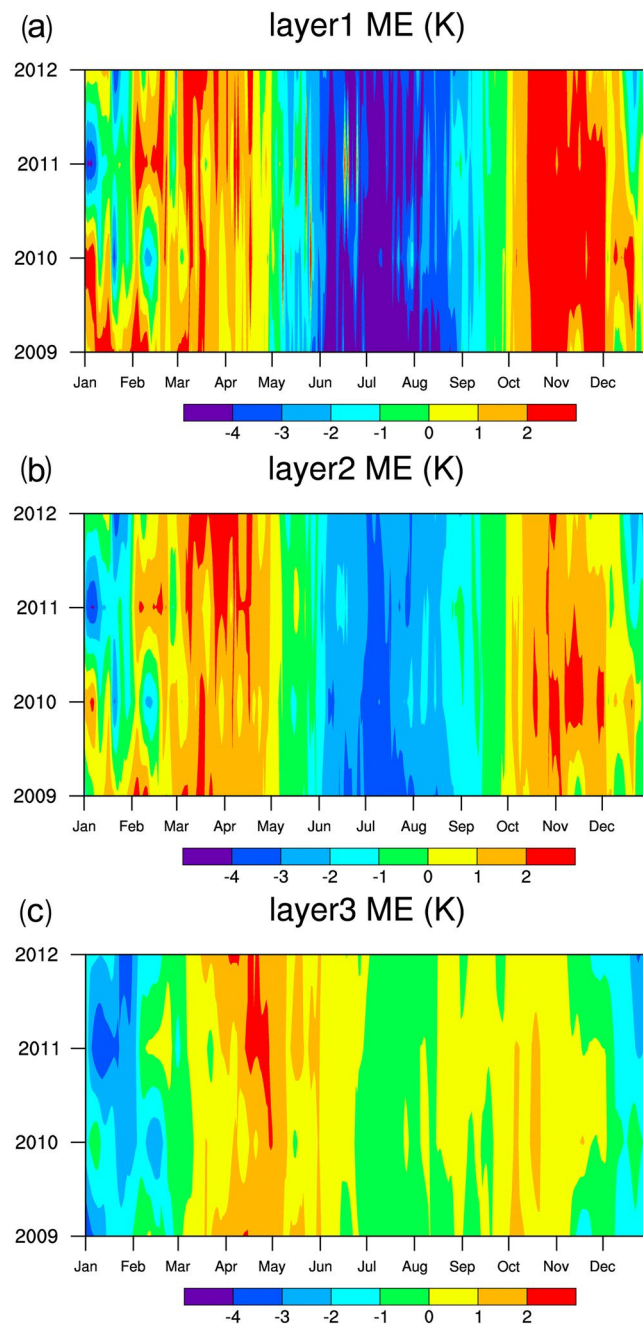
inter-annual change and the hourly change of each day in the soil temperature of each layer simulated using CLDAS-driven CLM3.5 fit the observed results very well, except for the simulated maximum temperature of layer 2 (15UTC). Overall, we could say the simulations of soil temperature at three soil depths by CLM3.5 in the Xinjiang region is acceptable.

Figure 5(c1–c3) shows a statistical analysis of the differences of the soil temperatures at the three layers between simulation and observation at the national automatic-soil temperature stations. Figure 5 tells that the simulation of soil temperatures is generally consistent with the observed ones. From the interpolation graph of the observed and simulated soil temperatures of layer1, we can see that the simulated temperatures of layer 1 between 03UTC and 21UTC from January to April and from September to November of each year are higher than their respective observed temperatures (0K–2K), whereas the simulated temperatures of layer 1 soil between 21UTC and 00UTC in the same time periods are lower than the observed temperatures (–1K–0K). In addition, the simulated temperatures of layer 1 throughout the day from May to August of each year are lower than the observed temperatures. The simulated maximum temperature of layer 1 occurs at 09UTC and is approximately 4 K lower than the observed maximum temperature. The reason for the relatively poor simulated values may be because the change in temperature in the Xinjiang region is the largest between 03UTC and 14UTC, and extreme values also occur in this time period. It is found that the differences between the simulated and observed temperatures of layer-1 soil are between –1 K and 1 K in all other non-extreme value time periods. The simulation results of the temperatures in non-extreme value time periods are ideal. The differences between the simulated and observed temperatures of layer 2 from January to April and from September to November of each year are mostly between –1 K and 1 K. Simulated temperatures of layer 2 at 12UTC reaches the maximum value, which is consistent with the information shown in Fig. 5(a2). We also find that the characteristics of the daily change in the soil temperature of layer 3 are not prominent. The simulated temperatures of layer 3 in late spring, summer and early autumn are 0–1 K higher than observations. The simulated temperatures of layer 3 in all other seasons are lower than the respective observed temperatures, and in winter the difference reach the largest (around 1–4 K) in winter.

## Discussion

The simulation of soil temperature is the key component of a land-surface model. The accuracy of the calculation of soil temperature using a land-surface model directly affects the material and energy exchanges between the land and the atmosphere in the model, which in turn affects the accuracy of the numerical model. This study uses the CLDAS data developed by the National Meteorological Information Center of China to force CLM3.5 to simulate the soil temperature in the Xinjiang region. This study describes both data integration and the core assimilation algorithm used in the preparation process of the CLDAS in details. This study validates and analyzes

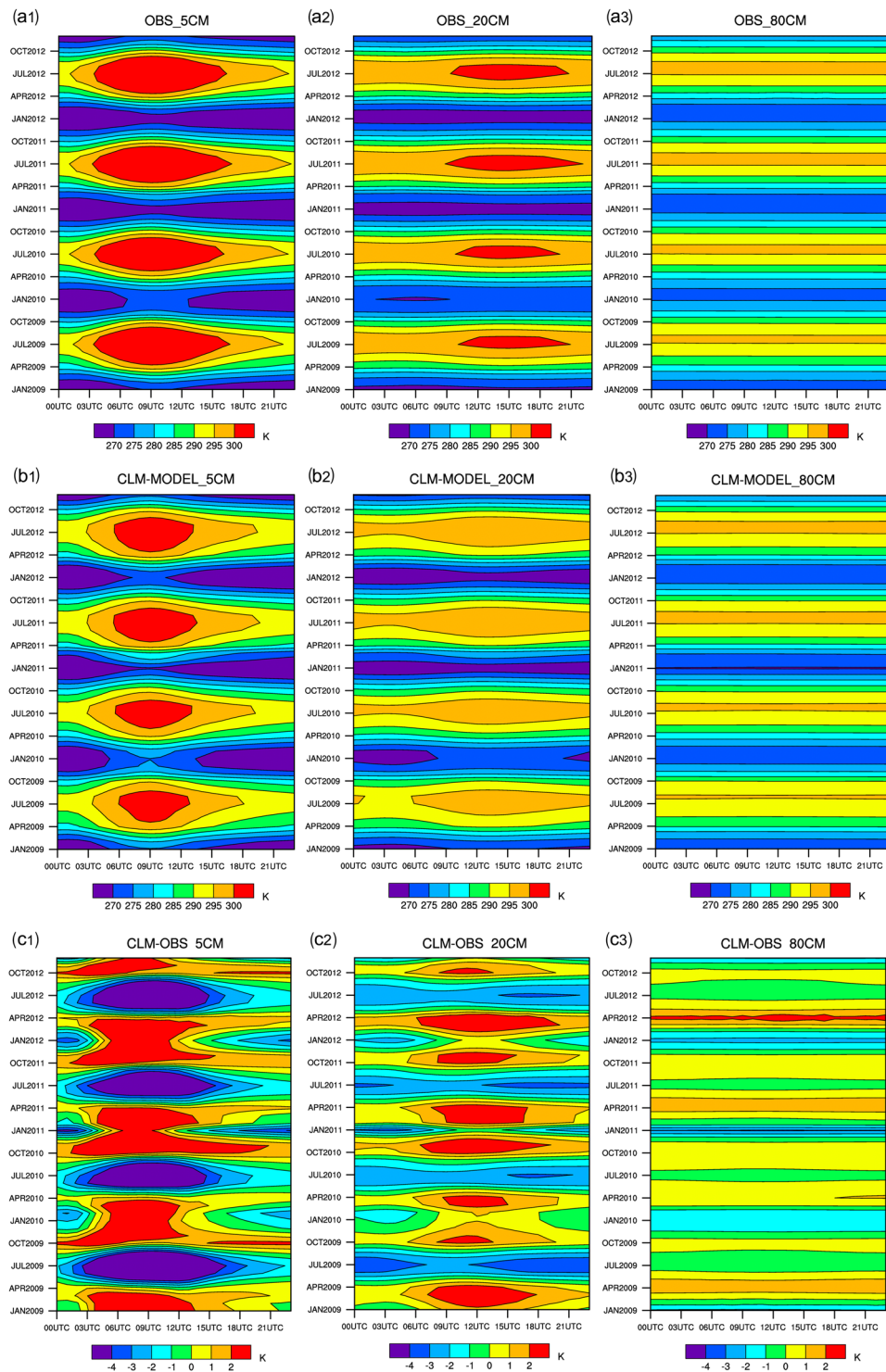




**Figure 4.** Seasonal changes in the mean deviations of the simulated daily mean temperatures from the observations in different years.

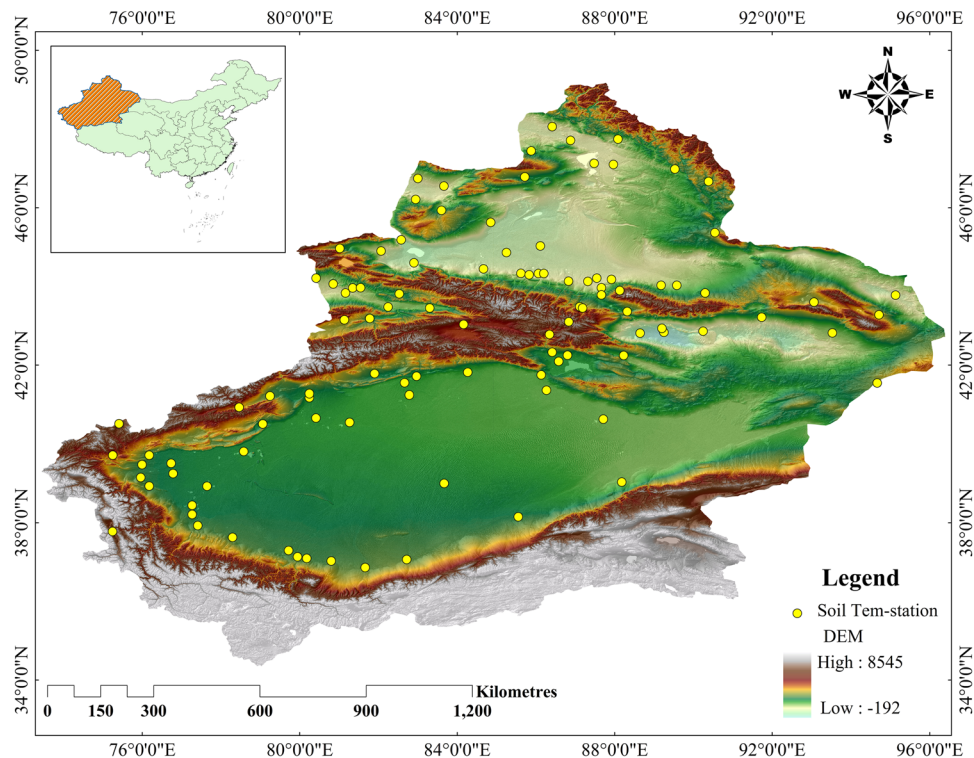
the soil temperatures at 105 national automatic stations in the Xinjiang region. The validation of soil temperature simulation using 105 national automatic stations indicated that CLM3.5 driven by CLDAS can simulate the changes of the soil temperatures at multiple layers in the Xinjiang region quite well, with the correlation coefficients being greater than 0.85. The deviations of the simulated values from the observed values are relatively small (no more than 2K–4K). The errors in the simulated soil temperatures of the shallow layers are relatively large, whereas they are relatively small for deep layers. Therefore, future studies should focus on further adjusting the parameters related to soil heat transfer and the parameterization schemes for other processes in the CLM to reduce the uncertainty of the model and obtain the optimal simulation results.

We admitted that there are still some drawbacks or limitations in our study. (1) From Fig. 2(a) and (b), we can see soil temperature in late winter and early spring was underestimated. Further, we also found some spikes (relatively high/low values) in the measurement data in this period, and this might be caused by the physical malfunction of probes during the melting period. The model did not perform well either in capturing this phenomenon. Actually, soil temperature variation involves water and energy fluxes and air temperature changes. The poor performance of soil temperature simulation from the late winter through the early spring due to the snow cover



**Figure 5.** Annual and hourly changes in the observed (a) and simulated (b) soil temperatures at three depths and their differences (c).

and air temperature were reported previously<sup>28–31</sup>. This complicated interaction complicated and abrupt change of soil temperature and poor model behavior deserves further investigation in future studies. (2) The soil temperature for the 105 stations were measured at a series depth of 5 cm, 10 cm, 15 cm, 20 cm, 40 cm, 80 cm, 160 cm, and 320 cm. The simulated soil temperature were for 10 soil layers with depths of 0.007 m, 0.0279 m, 0.0623 m, 0.1188 m, 0.212 m, 0.366 m, 0.619 m, 1.038 m, 1.727 m, 2.846 m. Our study was designed to take simulations from depths 0.0623 m, 0.366 m, and 1.038 m over the Xinjiang region for comparison with measurements from depths of 5 cm, 20 cm, and 80 cm. We understand this slight mismatch in depths could result in some biases including overestimation and underestimation, but it is unavoidable due to the model setting.



**Figure 6.** Location of the Xinjiang region and 105 automatic soil temperature observation stations. The map is generated with ArcMap Version 10.1 (<http://www.esri.com/en/arcgis/arcgis-for-desktop/>).

Application of numerical models usually involves input parameters uncertainty and its effects on model output, and this topic is quite attractive but challenging especially for large-scale studies. Many studies indicate that using an envelope of models could help assess the uncertainties and might reduce the systematic biases. Therefore, we are considering to investigate this issue in our future studies (1) using multiple models including CLM, Noah Land Surface Model, and/or the Common Land Model (CoLM) together with post statistical approaches such as Bayesian Model Averaging, and (2) using different climate and land-use scenarios to analyze the variabilities/uncertainties of model output<sup>32</sup>.

In spite of the above limitations and there was a national assessment of soil temperature of China using CLM3.0 with coarse resolutions ( $1^\circ$  and 3-hour)<sup>26</sup>. In contrast, our model simulations can be more informative and useful because it investigates the spatial-temporal patterns of soil profile (at three depths) temperature in Xinjiang using CLM3.5 with high spatial ( $1/16^\circ$ ) and temporal (hourly) resolutions. The high-resolution input and output can help evaluate the spatial-temporal features of the land surface processes especially in local environment.

## Methods

**Study area.** As shown in Fig. 6, Xinjiang Uygur Autonomous Region is located in the hinterland of the Eurasian continent between  $73^\circ 40'E$  and  $96^\circ 18'E$  and between  $34^\circ 25'N$  and  $48^\circ 10'N$ . With a total area of  $1.6649 \times 10^6$  km<sup>2</sup>, Xinjiang accounts for nearly one-sixth of China's total area and is the largest among the provincial-level administrative regions in China. With the closest coastline 2,648 km (straight-line distance) away, the Gurbantünggüt Desert in Xinjiang ( $46^\circ 16.8'N$ ,  $86^\circ 40.2'E$ ) is the remotest point of land from any sea. There are alternating mountain ranges and basins across Xinjiang—basins surrounded by high mountains and high mountains surrounded by basins. The Altai Mountains are located in the north of Xinjiang, and the Kunlun Mountain system is located in the south of Xinjiang. The Tianshan Mountains span the center of Xinjiang and divide Xinjiang into southern and northern halves. The Tarim Basin is situated in the southern half, and the Dzungarian Basin is located in the northern half. Xinjiang's unique landform—two basins sandwiched among three mountains, resulting in complicated underlying surface conditions and a huge spatial-temporal difference in climate distribution. Unfortunately, there are a deficient number of meteorological stations in the Xinjiang region, and literature review showed that most scientists just used reanalyzed data and/or some single-point data to study this region. Therefore, the existing studies on the land-surface processes in the Xinjiang region may be inaccurate or unsystematic.

**Models.** The CLM developed by the US National Center for Atmospheric Research (NCAR) is used as the land-surface model in this study. This model was developed based on extensive, painstaking research on the numerical simulation of climate, vegetation ecology and watershed hydrology<sup>33</sup>. Possessing the advantages of existing land-surface models worldwide CLM is one of the most reliable land-surface process models in the



world. CLM is the land-surface model component of the Community Earth System Model developed by the US NCAR<sup>28</sup>, and was primarily developed to couple with the earth-system model, the CLM can be operated offline (i.e., independently). When the CLM is operated offline, users need to provide the atmospheric forcing data, such as solar radiation, temperature, pressure, near-surface wind speed, specific humidity and precipitation rate. The CLM's physical processes includes the heat and water exchange between the land and atmosphere, the dynamic growth processes of vegetation, the thermodynamic processes of soil and the hydrologic processes. These processes are primarily expressed using parameterization schemes.

The CLM series adopted the sub-grid variability of each model—each grid contains multiple types of land unit, such as glacier, lake, wetland, city and plant. The plant unit can also be divided into multiple plant function types (PFTs, i.e., tiles) (Fig. 6), and energy and water conservation must be maintained on each tile, which has its own diagnostic variable. Each tile in a lattice point obtains mean-state atmospheric forcing from the lattice point of the atmosphere corresponding to that lattice point. The tiles in a lattice point contribute their water and heat fluxes to the lattice point based on the proportion of the area. There is no direct interaction among the tiles in each lattice point. The CLM is divided into one plant layer, 10 soil layers of different thicknesses and a maximum of five snow layers (the number of snow layers is determined by the thickness of the snow) in the vertical direction. This layered structure is beneficial to obtain relatively high-resolution daily and seasonal changes in each layer's temperature.

CLM3.5 was released by the US NCAR in 2007, and its modifications to CLM3.0 consist of an improved hydrologic process and the introduction of new surface datasets<sup>34–36</sup>. The introduction of new surface datasets simultaneously improves the description of the land surface and the simulation of the surface albedo, surface temperature and precipitation. Niu *et al.*<sup>35</sup> introduced a new frozen soil scheme. Oleson *et al.*<sup>20</sup> and Lawrence *et al.*<sup>34</sup> provided a detailed description of the physical processes and performance of CLM3.5.

CLM3.5 calculates soil temperatures using soil heat transfer equations:

The first law of heat transfer can be expressed as follows:

$$F = -\lambda \Delta T \quad (1)$$

where  $F$  represents the total amount of heat that transfers through a unit cross section area per unit of time ( $Wm^2$ );  $\lambda$  represents thermal conductivity ( $Wm^{-1}K^{-1}$ ); and  $\Delta T$  represents the spatial gradient of temperature ( $Km^{-1}$ ). The one-dimensional (1D) form of the first law of heat transfer is expressed as follows:

$$F_z = -\lambda \frac{\partial T}{\partial z} \quad (2)$$

where  $z$  represents the vertical direction (m) and is positive downward; and  $F_z$  is positive upward. To illustrate unstable or transient conditions, the principle of conservation of energy is called using the following continuity equation:

$$c \frac{\partial T}{\partial t} = -\frac{\partial F_z}{\partial z} \quad (3)$$

where  $c$  represents the heat capacity of the snow/soil of the equivalent body ( $Jm^{-3}K^{-1}$ ); and  $t$  represents time (s).

By simultaneously solving equations (2) and (3), the 1D form of the second law of heat transfer can be obtained:

$$c \frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \left[ \lambda \frac{\partial T}{\partial z} \right] \quad (4)$$

The soil and snow temperatures can be calculated by numerically solving the above equations.

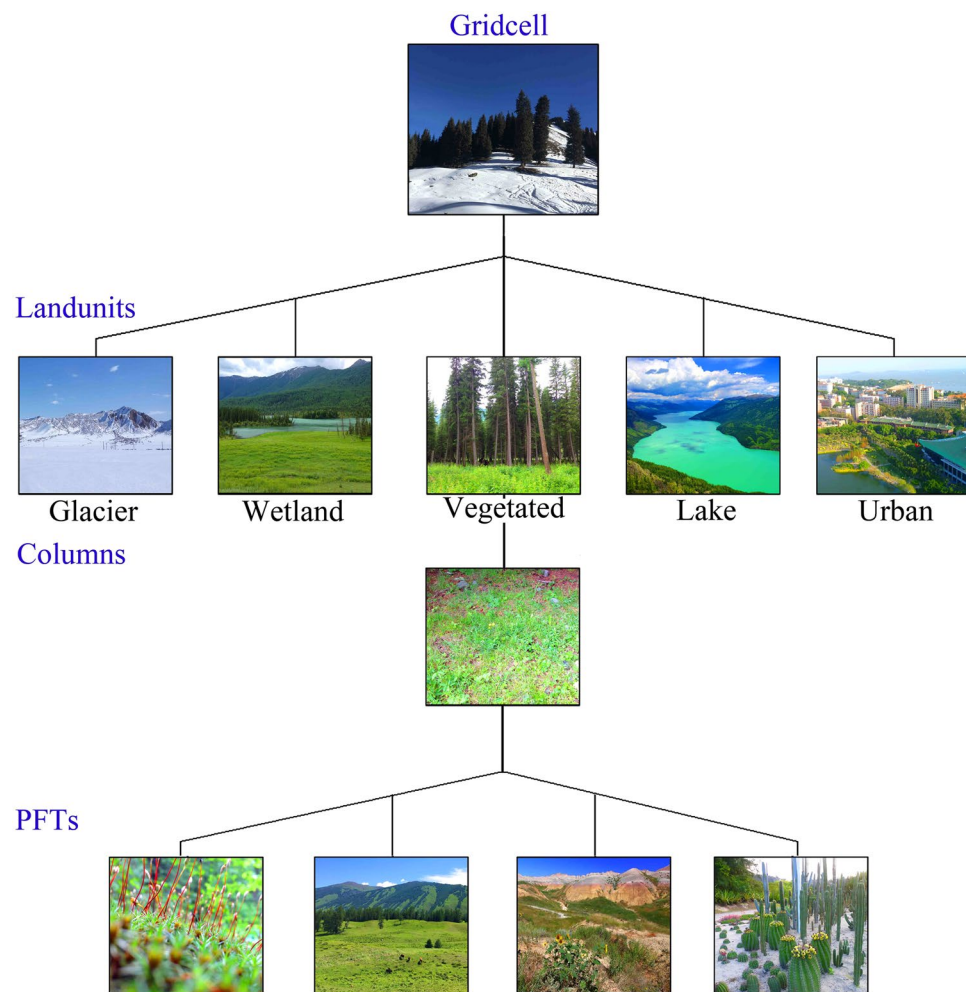
Five additional snow cover layers are added on top of the 10 layers of soil columns. In addition, boundary condition  $h$ , as heat flux, enters from the above atmosphere layer to the snow/soil layer on the surface. The bottom layer of the soil volume has zero heat flux. Initially, there is no change in the period when calculating the temperature profile, and then the adjustments are made during the process in which the period changes.

**Experiments.** In this study, the areas of lakes and wetlands are from the perennial freshwater lake and marshland data released by Cogley<sup>37</sup>. The soil colors are determined based on the results of the study by Zeng *et al.*<sup>38</sup> and Dickinson *et al.*<sup>39</sup>. We use the geosphere biosphere program (IGBP) data of soil sand and clay content with 4931 image elements for each soil layer<sup>40</sup> to generate a set of soil database that changes with the depth and soil texture<sup>41,42</sup>. PFTs and their contents along with the leaf area index are obtained based on the retrieval of the 1 km satellite data performed by Bonan *et al.*<sup>42</sup>. The glacier data are obtained from the IGBP data and the 1 km land surface cover database of the global information system (IGBP DISCover)<sup>43</sup>. The data of the stem area index and the heights of the upper and lower layers of each canopy are from Bonan *et al.*<sup>42</sup>.

Table 1 lists the land-surface data needed for each land-surface grid, including the proportions of glaciers, lakes and cities in each grid (the rest of the grid is occupied by plants). In addition, Fig. 7 lists the data of four PFTs with the highest contents, including the cover rates of each PFT, the leaf area index and stem area index of each PFT and the heights of the upper and lower layers of the canopy of each PFT. The percentage of each PFT is based on the plant block in the grid. The sum of the percentages of all of the PFTs is 100%. The proportions of lakes, wetlands, glaciers and cities refer to their proportions in the grid.

Surface data	Resolution	Ensemble method
Glaciers(percentage)	0.5°	Area averaged values
Lakes(percentage)	1°	Area averaged values
Wetlands(percentage)	1°	Area averaged values
Mud and sand(percentage) Clay (percentage)	5 min	Distribution of soil in the extension range of the largest area in the grid cell
Soil colors	2.8°	Classification of soil colors in the extension range of the largest area in the grid
PFTs(percent of vegetated land)	0.5°	Area averaged values (a maximum of four full PFTs can be selected)
Stem and leaf area indexes (monthly)	0.5°	Area averaged values
Canopy heights (upper and lower layers)	0.5°	Area averaged values

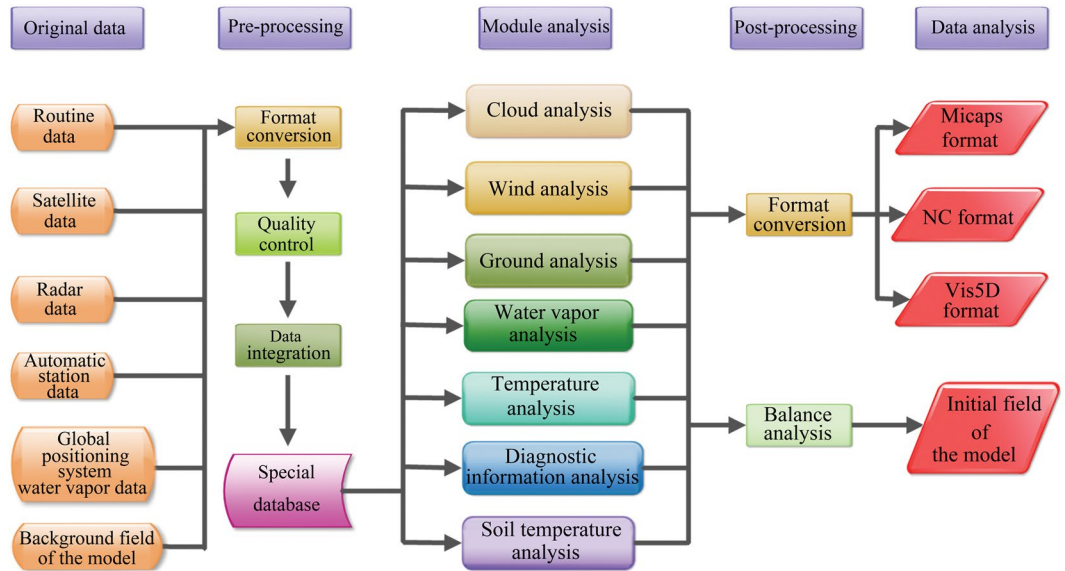
**Table 1.** Surface data needed for the CLM.



**Figure 7.** The sub grid distribution in CLM3.5 model.

**Data Preparation.** In addition to the description of the physical processes of the land surface, the land-surface model is sensitive to the atmospheric driving field. Therefore, the quality of the atmospheric driving field is an important factor to the models simulations<sup>44</sup>. In other words, the spatial-temporal resolution of atmospheric driving field and observation data determines the performance of the land-surface model. The atmospheric driving field of the CLDAS integrates data from various sources such as ground observation, satellite observation and numerical simulation products using integration and assimilation techniques. This product includes temperature, pressure, specific humidity, wind speed, precipitation and solar shortwave radiation. The atmospheric driving field of the CLDAS has a spatial resolution of  $1/16^\circ \times 1/16^\circ$  and a temporal resolution of 1 h, and this meets the study requirements for spatiotemporal precision of input data.

The integration of the temperature, pressure, humidity and wind speed data in the CLDAS is implemented using the LAPS/Space and Time Mesoscale Analysis System (STMAS). The LAPS is a comprehensive analysis system for multi-source data, including five major function modules—wind analysis, ground analysis, temperature



**Figure 8.** Flowchart of the Local Analysis and Prediction System (LAPS).

analysis, cloud analysis and water vapor analysis. The analyses must be carried out according to the order shown in Fig. 8, and the diagnostic analysis can be carried out on the analysis results of the five modules to obtain some diagnostic information about the weather. In addition, the numerical model can be accessed via balance analysis and thus, the model's hot start can be realized.

The STMAS<sup>45</sup> is a new-generation integration system that was developed within the framework of the LAPS. The variational method for the order of multiple grids is used as the algorithm of the STMAS, which is relatively different from the conventional LAPS. In terms of its function, the STMAS replaces the ground analysis of the LAPS with the ground analysis module and the wind analysis and temperature analysis of the LAPS with the STMAS3D module, respectively. The input, output, the cloud analysis, water vapor analysis and balance analysis of the STMAS continue to rely on the LAPS. The developers of the STMAS plan to gradually assimilate the diabatic initialization technology in the LAPS to create an independent system.

The multi-grid method was initially used to solve differential equations. A relatively coarse grid can allow a relatively low-frequency oscillation mode to converge rapidly. The multi-grid method was later introduced to data assimilation. The use of the objective function of a coarse grid in analyzing an error longwave can result in the rapid correction of the longwave of the analysis field. Afterwards, the use of the objective function of a fine grid in analyzing an error shortwave can reduce the confusion among different scales and improve the analysis results. The wavelength determines the correlation scale between lattice points. Therefore, the multi-grid method contains a built-in correlation. An analytical wave does not need a definite covariance. Thus, we can control the correlation scale of variation by controlling the number of lattice points. We can adjust the range of influence of the observed information by altering the coarseness of the grid to adjust the error covariance matrix of the background field. Under circumstances in which the accurate covariance is unknown, matrix B can be simplified into a diagonal matrix. The diagonal elements are the error variances of the background. Thus, the demand for computation resources is reduced. In the variational assimilation technique for the order of the multi-grid method, the objective function on each grid is expressed in the following form:

$$J^{(n)} = \frac{1}{2}X^{(n)T}X^{(n)} + \frac{1}{2}(H^{(n)}X^{(n)} - Y^{(n)})^T O^{(n-1)}(H^{(n)}X^{(n)} - Y^{(n)}) \quad (n = 1, 2, 3 \dots, N) \quad (5)$$

where  $Y = Y_{obs} - HX_b$  and  $X = X_a - X_b$ ,  $O$  represents the observation error covariance matrix,  $X_b$  represents the vector of the background field (forecast field),  $X_a$  represents the vector of the analysis field,  $Y_{obs}$  represents the constant vector of observation,  $H$  represents the bilinear interpolation operator from the model grid to the observation point,  $X$  represents the correction vector corresponding to the vector of the model field, which is calculated from the variational data assimilation system,  $Y$  represents the interpolation of the observation field and the model field,  $n$  represents the  $n^{th}$  grid, and  $N$  represents the number of grids.

$$\begin{cases} Y^{(1)} = Y^{obs} - H^{(1)}X^b & (n = 1) \\ Y^{(n)} = Y^{(n-1)} - H^{(n-1)}X^{(n-1)} & (n = 2, 3, \dots, N) \end{cases} \quad (6)$$

The STMAS first analyzes the coarsest grid among the multiple grids. When  $n = 1$ ,  $Y^{(1)}$  represents the difference between the projection of the background field of observation on the observation location and the projection of the background field of the model on the observation location, and  $X^{(n-1)}$  represents the solution or approximate solution of  $J^{(n-1)}$ .

Names of the statistical indicators	Equations
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^n (V_{oi} - V_{pi})^2}{n}}$
MAE	$MAE = \frac{1}{n} \cdot \sum_{i=1}^n ABS(V_{oi} - V_{pi})$
ME	$ME = \frac{1}{n} \cdot \sum_{i=1}^n (V_{oi} - V_{pi})$
COR	$COR = \text{corr}(V_{oi}, V_{pi}) = \frac{\text{cov}(V_{oi}, V_{pi})}{\sigma_{V_{oi}} \sigma_{V_{pi}}}$

**Table 2.** Four statistical indicators for examining the simulation precision of the model.  $V_{oi}$ : observed soil temperature value (K);  $V_{pi}$ : simulated soil temperature value (K);  $n$ : number of stations.

After solving  $J^{(n-1)}$ ,  $X^{(n-1)}$  is interpolated to the finer grid in the  $n^{\text{th}}$  layer.  $X^{(n)}$  can be solved through the minimization of  $J^{(n)}$ .

By altering the coarseness of the grid and analyzing the observation data in order from the longwave to the shortwave, we realized the extraction of the observation information of various scales. In this continuous multi-scale analysis process, the observation error covariance matrix of all of the scales ( $O^{(n)}$ ) is the same as that of the complete observation data because  $O^{(n)}$  reflects the observation error covariance and is unrelated to the multiple grid layers.

The final analysis result is the superposition of the analysis results of each grid:

$$X^a = X^b + X_L = X^b + \sum_{n=1}^N X^{(n)} \quad (7)$$

Therefore, unlike the three-dimensional variational (3D-Var) method, the multi-grid method does not confuse the longwave information with the shortwave information in the observation data. Using the 3D-Var method will bring about a relatively larger error in the analysis. This problem is more prominent when analyzing China's Xinjiang region, which has extremely unevenly distributed observation data. In contrast, the application of the multi-grid method to the preparation of the CLDAS data effectively assures the accuracy of this study's input data.

The main sources of the hourly CLDAS precipitation grid data include the integrated regional hourly precipitation produced by the National Meteorological Information Center of China, the hourly precipitation retrieved from the Fengyun-2E geostationary satellite by the National Satellite Meteorological Center of China<sup>46</sup> and the Climate Prediction Center Morphing Technique (CMORPH) integrated satellite precipitation produced by the Climate Prediction Center of the US National Oceanic and Atmospheric Administration. The ground-observed precipitation data used in this study is from the hourly precipitation observed at more than 30,000 automatic observation stations in China. Quality control (the climatological threshold value check, the regional threshold value check, the temporal consistency check and the spatial consistency check) is carried out on the precipitation data collected at more than 30,000 automatic observation stations (including both national and regional automatic observation stations) that have been built in China<sup>47</sup>. The CMORPH integrates the precipitation products retrieved from the microwave sensors of multiple satellites and uses infrared cold-cloud data for time extrapolation to obtain a global half-hourly precipitation product with a resolution of 8 km<sup>48</sup>. Shen *et al.*<sup>49</sup> evaluated this global half-hourly precipitation product, finding that the advantages of ground observation and the retrieval of precipitation from satellite data are effectively integrated in the CMORPH precipitation product. This CMORPH precipitation product is more reasonable in terms of the amount of precipitation and spatial distribution.

The discrete ordinate method proposed by Stamnes *et al.*<sup>50</sup> is used as the retrieval algorithm for the ground incident solar radiation of the CLDAS data for radiation transfer calculation. The discrete ordinate method can calculate the radiance in any arbitrary direction. Therefore, the discrete ordinate method considers the anisotropy of the solar radiation reflected by the top of the atmosphere. That is to say, the radiance of the solar radiation reflected by the top of the atmosphere in the satellite's observation direction is first calculated and then converted to the bidirectional visible albedo that is observed by the visible channel of the satellite. During the transmission process of the solar radiation incident from the top of the atmosphere to passing through the atmosphere to reaching the ground, there is a series of physical processes, including interactions with the atmosphere and ground. The following factors are primarily considered in the retrieval model: (1) absorption by the ozone layer, (2) multiple molecular Rayleigh scattering, (3) multiple scattering and absorption by cloud droplets, (4) absorption by water vapor, (5) multiple scattering and absorption by aerosols and (6) multiple reflection by the ground and atmosphere<sup>32,51</sup>.

This study uses the high spatiotemporal resolution land-surface atmospheric driving data of the CLDAS produced by the National Meteorological Information Center of China Meteorological Administration (temporal resolution: 1 h; spatial resolution:  $1/16^\circ \times 1/16^\circ$  and approximately 6.25 km; time scale: 2009–2012; elements: atmospheric temperature, pressure, specific humidity, wind speed, precipitation and shortwave solar radiation) to drive the CLM3.5 to conduct a numerical simulation of the land surface. The process is repeated 10 times to spin up the CLM model. A basically stable initial field of the model is generated to obtain a high temporal-spatial resolution soil temperature dataset. In addition, the bilinear interpolation method is used to interpolate the soil temperature grid data of the CLDAS to the observation stations (Fig. 6) to verify and analyze the hourly data generated by the CLDAS model and the matched sample data generated by observation. In this way, we are able



to understand the distribution of soil temperature in the Xinjiang region and the changes in the soil temperature of each layer.

The CLM3.5 can simulate the temperatures of the soil of 10 layers. The depths of the node layers are 0.007 m, 0.0279 m, 0.0623 m, 0.1188 m, 0.212 m, 0.366 m, 0.619 m, 1.038 m, 1.727 m and 2.846 m, respectively. Comprehensively considering the observation depths of the national soil temperature stations of China and the typical temperatures of the soil at various depths, this study selects three soil layers, namely, layer 1 (at a depth of 5 cm), layer 2 (at a depth of 20 cm) and layer 3 (at a depth of 80 cm), which are observed at 105 soil temperature observation stations in Xinjiang (the stations' locations are shown in Fig. 6), for soil-temperature verification experimentation.

**Validation method for the simulation results.** The data are verified in this study to objectively analyze the accuracy of the spatial prediction method. First, the bilinear interpolation method is used to interpolate the planar raster data for the soil temperature simulated using the model to the 105 observation stations in Xinjiang to compare the model simulated temperatures with the observations. We select the root-mean-square error (RMSE), the mean absolute error (MAE), the mean error (ME) and the correlation coefficient (COR) for the aforementioned validation (Table 2).

## References

- Boone, A. A. *et al.* The interactions between soil–biosphere–atmosphere land surface model with a multi-energy balance (ISBA-MEB) option in SURFEXv8 – Part 1: Model description. *Geosci. Model. Dev.* **10**, 843–872 (2017).
- Diego, G. M., Adriaan, J. T., Chiel, C. & Jordi, V. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature Geosci.* **7**, 345–349 (2014).
- Jatin, K. *et al.* Impact of the representation of stomatal conductance on model projections of heatwave intensity. *Sci. Rep.* **6**, 23418 (2016).
- Meng, X. Y. *et al.* Hydrological Modeling in the Manas River Basin Using Soil and Water Assessment Tool Driven by CMADS. *Teh. Vjesn* **24**, 525–534 (2017).
- Morten, A. D. L., Jens, H. C., Martin, D., Michael, B. B. & Jens, C. R. Local control on precipitation in a fully coupled climate-hydrology model. *Sci. Rep.* **6**, 22927 (2016).
- Osuri, K. *et al.* Improved prediction of severe thunderstorms over the Indian Monsoon region using high-resolution soil moisture and temperature initialization. *Sci. Rep.* **7**, 41377 (2017).
- Ramarao, M. V. S., Krishnan, R., Sanjay, J. & Sabin, T. P. Understanding land surface response to changing South Asian monsoon in a warming climate. *Earth. Syst. Dynam.* **6**, 569–582 (2015).
- Ruiz-Barradas, A. & Nigam, S. Atmosphere–Land Surface Interactions over the Southern Great Plains: Characterization from Pentad Analysis of DOE ARM Field Observations and NARR. *J. Climate.* **26**, 875–886 (2013).
- Sulis, M., Paniconi, C., Rivard, C., Harvey, R. & Chaumont, D. Assessment of climate change impacts at the catchment scale with a detailed hydrological model of surface–subsurface interactions and comparison with a land surface model. *Water. Resour. Res.* **47**, 128–139 (2011).
- Trent, W. F. & Oliver, W. F. Surface–Atmosphere Moisture Interactions in the Frozen Ground Regions of Eurasia. *Sci. Rep.* **6**, 19163 (2016).
- Charney, J. G. Dynamics of deserts and drought in the Sahel. *Q. J. Roy. Meteor. Soc.* **101**, 193–202 (1975).
- Koster, R. D. *et al.* On the Nature of Soil Moisture in Land Surface Models. *J. Climate.* **22**, 4322–4335 (2009).
- Godfrey, C. M. & Stensrud, D. J. Soil Temperature and Moisture Errors in Operational Eta Model Analyses. *J. Hydrometeorol. Sci. Rep.* **9**, 367–387 (2007).
- Fan, X. Impacts of Soil Heating Condition on Precipitation Simulations in the Weather Research and Forecasting Model. *Mon. Weather. Rev.* **137**, 2263–2285 (2009).
- Zhou, H. X. *et al.* Horizontal Heat Impact of Urban Structures on the Surface Soil Layer and Its Diurnal Patterns under Different Micrometeorological Conditions. *Sci. Rep.* **6**, 18790 (2016).
- Hu, Q. & Feng, S. How have soil temperatures been affected by the surface temperature and precipitation in the Eurasian continent? *Geophys. Res. Lett.* **32**, 57–76 (2005).
- Zhang, Y., Chen, W. J., Smith, S. L., Riseborough, D. W. & Cihlar, J. Soil temperature in Canada during the twentieth century: Complex responses to atmospheric climate change. *J. Geophys. Res. Atmospheres.* **110**, D03112 (2005).
- Wang, X., Guo, W., Zhong, Z. & Cui, X. Long term trends of soil moisture and temperature change in east China in relationship with climate background. *Adv. in. Earth. Sci.* **24**, 181–191 (2009).
- Bonan, G. B. & Levis, S. Evaluating Aspects of the Community Land and Atmosphere Models (CLM3 and CAM3) Using a Dynamic Global Vegetation Model. *J. Climate.* **19**, 2290–2301 (2006).
- Oleson, K. W. *et al.* Improvements to the Community Land Model and their impact on the hydrological cycle. *J. Geophys. Res.: Biogeosciences* **113**, G01021 (2008).
- Sakaguchi, K. & Zeng, X. B. Effects of soil wetness, plant litter, and undercanopy atmospheric stability on ground evaporation in the Community Land Model (CLM4.0). *J. Geophys. Res.* **114**, D01107 (2009).
- Oleson, K. W. *et al.* Technical description of version 4.5 of the Community Land Model (CLM). Ncar Tech. Note NCAR/TN-503 + STR. National Center for Atmospheric Research, Boulder, CO, 422 pp, doi:10.5065/D6RR1W7M, <http://opensky.ucar.edu/islandora/object/technotes:515> (2013).
- Whitfield, B., Jacobs, J. M. & Judge, J. Inter comparison Study of the Land Surface Process Model and the Common Land Model for a Prairie Wetland in Florida. *J. Hydrometeorol.* **7**, 1247–1258 (2006).
- Huang, W., Guo, Z. H. & Yu, R. Numerical simulation of CLM over Huaihe basin. *Acta. Meteorol. Sin.* **62**, 764–775 (2004).
- Niu, G. Y., Yang, Z. L., Dickinson, R. E. & Gulden, L. E. A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models. *J. Geophys. Res.* **110**, D21106 (2005).
- Chen, H., Xiong, M. & Sha, W. Simulation of land surface processes over China and its validation Part I: Soil temperature. *Sci. Meteor. Sin.* **30**, 621–630 (2010).
- Plauborg, F. Simple model for 10 cm soil temperature in different soils with short grass. *Eur J. of Agron.* **17**, 173–179 (2002).
- Tomoyoshi, H., John, W. P., Raoul, J. G. & Charles P. M. An extension of the force-restore method to estimating soil temperature at depth and evaluation for frozen soils under snow. *J. Geophys. Res.: Atmos.* **107**, ACL 11-1-ACL 11-10 (2002).
- Plummer, N., Allsopp, T. & Lopez, J. A. Guidelines on Climate Observation Networks and Systems. WMO/TD 1185, World Meteorological Organization. [https://library.wmo.int/pmb\\_ged/wmo-td\\_1185\\_en.pdf](https://library.wmo.int/pmb_ged/wmo-td_1185_en.pdf) (2003).
- Paimazumder, D. *et al.* Evaluation of Community Climate System Model soil temperatures using observations from Russia. *Theo. Appl. Climatol.* **94**, 187–213 (2008).

31. Blanchet, J. & Davison, A. C. Statistical modelling of ground temperature in mountain permafrost. *Proceedings: Mathematical, Physical and Engineering Sciences*. **468**, 1472–1495 (2012).
32. Liu, S. *et al.* Chapter 7: Baseline and projected future carbon storage, carbon sequestration, and greenhouse-gas fluxes in terrestrial ecosystems of the eastern United States. U.S. Geological Survey, Reston, Virginia. [http://pubs.usgs.gov/pp/1804/pdf/pp1804\\_chapter7.pdf](http://pubs.usgs.gov/pp/1804/pdf/pp1804_chapter7.pdf) (2014).
33. Oleson, K. *et al.* Technical description of the Community Land Model (CLM). NCAR Technical Note NCAR/TN-461 + STR, National Center for Atmospheric Research, Boulder. (2004).
34. Lawrence, D. M., Thornton, P. E., Oleson, K. W. & Bonan, G. B. The Partitioning of Evapotranspiration into Transpiration, Soil Evaporation, and Canopy Evaporation in a GCM: Impacts on Land–Atmosphere Interaction. *J. Hydrometeorol.* **8**, 862–880 (2007).
35. Niu, G. Y. & Yang, Z. L. Effects of Frozen Soil on Snowmelt Runoff and Soil Water Storage at a Continental Scale. *J. Hydrometeorol.* **7**, 937–952 (2006).
36. Niu, G. Y., Yang, Z. L., Dickinson, R. E., Gulden, L. E. & Su, H. Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. *J. Geophys. Res.* **112**, 227–287 (2007).
37. Cogley, J. G. GGHYDRO-Global hydrographic data, Release 2.3.1., Trent Climate Note 2003-1. <http://people.trentu.ca/~gcogley/glaciology/gghrls231.pdf> (2003).
38. Zeng, X., Shaikh, M., Dai, Y., Dickinson, R. E. & Myneni, R. Coupling of the Common Land Model to the NCAR Community Climate Model. *J. Climate*. **15**, 1832–1854 (2002).
39. Dickinson, R., Henderson-Sellers, A. & Kennedy, P. Biosphere-Atmosphere Transfer Scheme (BATS) Version 1. As coupled to the NCAR Community Climate Model, NCAR Tech Note 77 <http://opensky.ucar.edu/islandora/object/technotes:154> (1993).
40. Global Soil Data Task. Global soil data products CD-ROM (IGBP-DIS). International Geosphere-Biosphere Programme-Data and Information Available Services <http://www.daac.ornl.gov> (2000).
41. Robock, A. *et al.* The Global Soil Moisture Data Bank. *B. Am. Meteorol. Soc.* **81**, 1281–1299 (2000).
42. Bonan, G. B., Levis, S., Kergoat, L. & Oleson, K. W. Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochem Cy.* **16**, 5-1–5-23 (2002).
43. Loveland, T. R. *et al.* Development of a global land cover characteristics database and IGBP DIS Cover from 1 km AVHRR data. *Int. J. Remote Sens.* **21**, 1303–1330 (2010).
44. Dickinson, R. E. *et al.* The Community Land Model and Its Climate Statistics as a Component of the Community Climate System Model. *J. Climate*. **19**, 2302–2324 (2006).
45. Xie, Y. *et al.* A Space–Time Multiscale Analysis System: A Sequential Variational Analysis Approach. **139**, 1224–1240 (2011).
46. Ren, Z. H. *et al.* Quality control procedures for hourly precipitation data from automatic weather stations in China. *Meteor. Mon.* **36**, 123–132 (2010).
47. Shi, C., Xie, Z. H., Qian, H., Liang, M. L. & Yang, X. C. China land soil moisture EnKF data assimilation based on satellite remote sensing data. *Sci. China Earth Sci.* **54**, 1430–1440 (2011).
48. Joyce, R. J., Janowiak, J. E., Arkin, P. A. & Xie, P. CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *J. Hydrometeorol.* **5**, 487–503 (2004).
49. Shen, Y., feng, M., Zhang, H. & Gao, F. Interpolation methods of China daily precipitation data. *J. Appl. Meteorol. Sci.* **21**, 279–286 (2010).
50. Stammes, K., Tsay, S. C., Wiscombe, W. & Jayaweera, K. Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Appl. Optics*. **27**, 2502–2509 (1988).
51. Jia, B. H., Xie, Z. H., Dai, A. G., Shi, C. X. & Chen, F. Evaluation of satellite and reanalysis products of downward surface solar radiation over East Asia spatial and seasonal variations. *J. Geophys. Res. Atmospheres*. **118**, 3431–3446 (2013).

## Acknowledgements

This research was financially supported by the National Science Foundation of China (51479209, 41701076, 51609260), the National Key Technology R&D Program of China (2016YFA0601602, 2017YFC0404301, 2017YFC0404305, 2017YFB0203104), the Chinese Academy of Engineering Program (2014-2D-4-2, Z0202192014), the China Postdoctoral Science Foundation (2017M610950), and National One-Thousand Youth Talent Program of China (122990901606).

## Author Contributions

X.M. organized and wrote the manuscript. X.M. and Y.W. collected and analyzed the data. H.W., J.W. and A.L. assisted in analyzed the data. X.M., C.S., X.J. and Y.W. contributed to discussing the results. All authors reviewed the manuscript.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017