ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202006.00191

Estimation of Daily Maximum and Minimum Temperatures in Lanzhou City Based on MODIS and Random Forest: Postprint

Authors: Xing Liting, Li Jing, Jiao Wenhui, Li Jing

Date: 2020-06-21T00:00:00+00:00

Abstract

Near-surface air temperature is a crucial factor for measuring urban thermal environments and is considered an important variable for various urban issues; however, the severe shortage of meteorological stations in small areas limits the representation of spatially continuous air temperature distribution within heterogeneous cities. To obtain spatially continuous near-surface air temperature, this study employs remote sensing data combined with the random forest machine learning method to estimate urban near-surface air temperature. Taking Lanzhou City as the study area, we utilize land surface temperature data from the MODIS sensor at eight different time points from both the previous day and the current day, combined with a series of influencing factors, and employ random forest to estimate the city' s daily maximum and minimum air temperatures (Tmax/Tmin). Due to the varying correlation relationships between the eight temporal land surface temperature datasets and Tmax/Tmin, eight model schemes with different land surface temperature data inputs were designed based on these relationships. The results of different model schemes were validated using measured air temperature data to obtain daily maximum and minimum air temperatures estimated by the optimal scheme. The results indicate that using the random forest model combined with remote sensing data to estimate urban daily near-surface air temperature is feasible, and that land surface temperature from the previous day has a significant influence on air temperature, serving as a key parameter for temperature estimation.

Full Text

Preamble

Arid Zone Research (ChinaXiv Partner Journal)

Near-surface temperature is a critical factor for measuring the urban thermal environment. Urban temperature is considered to affect various urban problems. However, meteorological stations in small regions are insufficient to enable sufficient recording of temperature across heterogeneous cities. Remote sensing data combined with the random forest machine learning method was used to estimate the continuous near-surface temperature of Lanzhou. The random forest approach uses the previous day' s surface temperature data from a MODIS sensor for eight different time points in the day, combined with a series of influencing factors, to estimate the daily maximum and minimum temperature (Tmax/Tmin). Because the eight time points of surface temperature data have different correlations with Tmax/Tmin, these different relationships are used to create eight different model schemes with different input surface temperature data, and the results of the different models are verified using the measured temperature to obtain the best estimates of daily maximum and minimum temperatures. We found that it is feasible to use the random forest model combined with remote sensing data to estimate the daily and near-surface temperature of Lanzhou city, and that the previous day' s surface temperature has a large effect on the present temperature, making it a key parameter for estimating temperature.

Authors: t%u (born 1996-), v, wxwx, specializing in remote sensing applications. E-mail: 1453374906@qq.com. EFAB: z{. E-mail: li_{jinger}@163.com.

Funding: Supported by the National Natural Science Foundation of China (NWNULKQN-14-4).

1 Introduction

Near-surface temperature is an important factor for measuring the urban thermal environment. Urban temperature is considered to affect various urban problems. However, meteorological stations in small regions are insufficient to enable sufficient recording of temperature across heterogeneous cities. Remote sensing data combined with the random forest machine learning method was used to estimate the continuous near-surface temperature of Lanzhou. The random forest approach uses the previous day' s surface temperature data from a MODIS sensor for eight different time points in the day, combined with a series of influencing factors, to estimate the daily maximum and minimum temperature (Tmax/Tmin). Because the eight time points of surface temperature data have different correlations with Tmax/Tmin, these different relationships are used to create eight different model schemes with different input surface temperature data, and the results of the different models are verified using the measured temperature to obtain the best estimates of daily maximum and minimum temperatures. We found that it is feasible to use the random forest model combined with remote sensing data to estimate the daily and near-surface temperature of Lanzhou city, and that the previous day' s surface temperature

has a large effect on the present temperature, making it a key parameter for estimating temperature.

Previous studies have demonstrated that remote sensing data can be used to estimate near-surface temperature [1-4]. Cresswell et al. [15] showed that MODIS land surface temperature (LST) data can be used to estimate air temperature. Other researchers have applied NDVI, elevation, latitude, and longitude as auxiliary variables for temperature estimation [5-8]. Stisen et al. [18] used TVX methods to estimate air temperature from MODIS data. The heating effect of the Tibetan Plateau can cause temperature increases of 2.55-2.99°C, which significantly affects regional climate [9-12]. Additionally, four key factors have been identified: (1) surface temperature from remote sensing, (2) elevation, (3) latitude and longitude, and (4) temporal variation. Papé et al. [19] found that previous day' s temperature is an important predictor for current temperature estimation.

1.1 Data and Methods

1.1.1 MODIS Land Surface Temperature Data The study utilized MODIS LST products from both Terra and Aqua satellites. Terra MODIS provides daytime (10:30) and nighttime (22:30) LST, while Aqua MODIS provides daytime (13:30) and nighttime (01:30) LST, resulting in eight daily observations. The study area included 53 meteorological stations, with 30 stations reserved for validation. Out-of-bag (OOB) error was used for model validation. The data period covered 2014, with specific focus on days 97, 197, and 297.

The eight time points were designated as S1 through S8, representing different LST observations: S1 and S2 correspond to Terra and Aqua daytime LST, while other time points represent various combinations of daytime and nighttime observations from both satellites. The maximum temperature (Tmax) and minimum temperature (Tmin) were estimated using different combinations of these LST observations.

1.1.2 Ancillary Data In addition to LST data, the model incorporated several influencing factors including NDVI, elevation, latitude, and longitude. The spatial resolution of MODIS LST and NDVI data was 1 km, while DEM data had a resolution of 30 m. The DEM data were resampled to match the 1 km resolution of the MODIS products. AVHRR data at 0.25° resolution were also considered for comparison. All data were processed using Python 2.7.15 and ArcGIS 10.4 software platforms.

1.2.2 Random Forest Model Development

Eight different model schemes (S1-S8) were designed based on different combinations of LST inputs. The random forest algorithm was trained using surface temperature data from the eight time points combined with ancillary variables.

Each scheme tested different input combinations to determine the optimal configuration for Tmax and Tmin estimation. The models were validated using meteorological station observations through calculation of R^2, MAE, and RMSE metrics.

[Figure 2: see original paper] shows the correlation coefficients between Tmax/Tmin and surface temperature at different time points.

2 Results

2.1 Model Performance Evaluation

The correlation analysis revealed strong relationships between LST observations and air temperature. Using 53 stations for training and 40 for validation (with 13 stations providing high-quality data), the models achieved R^2 values exceeding 0.8. The OOB error estimation showed R^2 values above 0.9 for optimal schemes, with MAE as low as 1.344°C.

For Tmax estimation, scheme S6 demonstrated the best performance with $R^2 = 0.921$, while scheme S2 achieved $R^2 = 0.916$. For Tmin estimation, scheme S6 showed $R^2 = 0.733$, with MAE of 1.344°C and RMSE of 1.501°C. The results indicate that incorporating multiple LST time points significantly improves estimation accuracy compared to using single observations.

Specific design scheme of model

Scheme	Input Variables for Tmax	Input Variables for Tmin
S1	LST_{TBN} + α	LST_{AN} + α
S2	LST_{TBN} + LST_{ABN} + α	LST_{AN} + LST_{TN} + α
S3	LST_{TBN} + LST_{ABN} + LST_{ABD} + α	LST_{AN} + LST_{TN} + LST_{ABN} + α
S4	LST_{TBN} + LST_{ABN} + LST_{ABD} + LST_{TBD} + α	LST_{AN} + LST_{TN} + LST_{ABN} + LST_{TBN} + α
S5	LST_{TBN} + LST_{ABN} + LST_{ABD} + LST_{TBD} + LST_{TN} + α	LST_{AN} + LST_{TN} + LST_{ABN} + LST_{TBN} + LST_{TBD} + α
S6	LST_{TBN} + LST_{ABN} + LST_{ABD} + LST_{TBD} + LST_{TN} + LST_{AN} + α	LST_{AN} + LST_{TN} + LST_{ABN} + LST_{TBN} + LST_{TBD} + LST_{ABD} + α
S7	LST_{TBN} + LST_{ABN} + LST_{ABDD} + LST_{TBD} + LST_{TN} + LST_{AN} + LST_{AD} + α	LST_{AN} + LST_{TN} + LST_{ABN} + LST_{TBN} + LST_{TBD} + LST_{ABD} + LST_{TD} + α

Scheme	Input Variables for Tmax	Input Variables for Tmin
S8	LST_{TBN} + LST_{ABN} + LST_{ABD} + LST_{TBD} + LST_{TN} + LST_{AN} + LST_{AD} + LST_{TD} + α	LST_{AN} + LST_{TN} + LST_{ABN} + LST_{TBN} + LST_{TBD} + LST_{ABD} + LST_{TD} + LST_{AD} + α

Verification of the model

Scheme	Tmax R^2	Tmax R^2_{test}	Tmax MAE (°C)	Tmax RMSE (°C)	Tmin R^2	Tmin R^2_{test}	Tmin MAE (°C)	Tmin RMSE (°C)
S1	0.689	0.816	2.332	2.598	0.876	0.784	1.261	1.637
S2	0.691	0.820	2.375	2.650	0.916	0.816	1.218	1.596
S3	0.697	0.892	1.663	1.801	0.886	0.630	1.305	2.155
S4	0.713	0.912	1.534	1.648	0.882	0.666	1.276	2.049
S5	0.729	0.918	1.513	1.632	0.864	0.686	1.297	2.036
S6	0.733	0.921	1.344	1.501	0.870	0.699	1.366	2.056
S7	0.733	0.912	1.352	1.516	0.855	0.701	1.319	2.014
S8	0.706	0.913	1.416	1.538	0.867	0.678	1.361	2.126

The validation results demonstrate that scheme S6 provides the optimal balance of accuracy and robustness for both Tmax and Tmin estimation. The scatter plot [Figure 3: see original paper] shows the relationship between estimated and observed temperatures for days 97, 197, and 297 of 2014. The spatial distribution of temperature estimates is presented in [Figure 4: see original paper].

2.2 Spatial Application

The optimal model (S6) was applied to generate continuous temperature surfaces for Lanzhou City. Using MODIS LST data from 2014 (days 97, 197, and 297), the model produced daily Tmax and Tmin maps at 1 km resolution. The results show clear spatial patterns related to elevation, land cover, and urban heat island effects. The model successfully captures the temperature gradient from urban centers to surrounding rural areas, with the previous day' s LST serving as a critical predictor variable.

[Figure 3: see original paper] Scatter plot of estimated temperature and site temperature

[Figure 4: see original paper] Maximum/minimum temperature simulation for 97, 197, 297 days, 2014

References

[1-4] Previous studies on remote sensing temperature estimation [5-8] Studies using NDVI, elevation, and geographic coordinates [9-12] Research on Tibetan Plateau heating effects [13] Key factors in temperature modeling [15] Cresswell et al. MODIS-based air temperature estimation [16-17] NDVI applications in temperature studies [18] Stisen et al. TVX methods [19] Papé et al. Temporal temperature modeling [20] Random forest methodology [21] Sentinel-2 data applications [22] NASA MODIS products [23] Landsat/TM comparisons [24-25] MODIS temperature retrieval methods [26] Breiman L. Random forests [27] Zhou Yi et al. Random forest algorithm design [28] Liu Jian et al. Solar radiation prediction model [29] Breiman L. Bagging predictors [30] Fang Kuangnan et al. Random forest technology review

---

Abstract: Near-surface temperature is an important factor for measuring the urban thermal environment. Urban temperature is considered to affect various urban problems. However, meteorological stations in small regions are insufficient to enable sufficient recording of temperature across heterogeneous cities. Remote sensing data combined with the random forest machine learning method was used to estimate the continuous near-surface temperature of Lanzhou. The random forest approach uses the previous day' s surface temperature data from a MODIS sensor for eight different time points in the day, combined with a series of influencing factors, to estimate the daily maximum and minimum temperature (Tmax/Tmin). Because the eight time points of surface temperature data have different correlations with Tmax/Tmin, these different relationships are used to create eight different model schemes with different input surface temperature data, and the results of the different models are verified using the measured temperature to obtain the best estimates of daily maximum and minimum temperatures. We found that it is feasible to use the random forest model combined with remote sensing data to estimate the daily and near-surface temperature of Lanzhou city, and that the previous day' s surface temperature has a large effect on the present temperature, making it a key parameter for estimating temperature.

Keywords: land surface temperature; MODIS; random forest; near-surface temperature; Lanzhou City

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.