

Estimation of monthly near surface air temperature using geographically weighted regression in China

M. M. Wang^{1*}, G. J. He², Z. M. Zhang², Z. J. Zhang¹, X.G. Liu¹

¹ Faculty of Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China - (wangmm, zhangzj, liuxg)@cug.edu.cn

² Institute of Remote Sensing and Digital Earth, Chinese Academy of sciences, Beijing 10094, China – (hegj, zhangzm)@radi.ac.cn

Commission VI, WG III/8

KEY WORDS: Near Surface Air Temperature, Land Surface Temperature, Geographically Weighted Regression, MODIS, Remote Sensing, Geographically Information System.

ABSTRACT:

Near surface air temperature (NSAT) is a primary descriptor of terrestrial environment conditions. The availability of NSAT with high spatial resolution is deemed necessary for several applications such as hydrology, meteorology and ecology. In this study, a regression-based NSAT mapping method is proposed. This method is combined remote sensing variables with geographical variables, and uses geographically weighted regression to estimate NSAT. The altitude was selected as geographical variable; and the remote sensing variables include land surface temperature (LST) and Normalized Difference vegetation index (NDVI). The performance of the proposed method was assessed by predict monthly minimum, mean, and maximum NSAT from point station measurements in China, a domain with a large area, complex topography, and highly variable station density, and the NSAT maps were validated against the meteorology observations. Validation results with meteorological data show the proposed method achieved an accuracy of 1.58°C. It is concluded that the proposed method for mapping NSAT is very operational and has good precision.

1. INTRODUCTION

Near surface temperature air temperature (NSAT) is a primary descriptor of terrestrial environment conditions (Guan et al. 2013). NSAT is the most important component of global climate change and is sensitive to local anthropogenic disturbance (Hansen et al. 2006). Thus, the availability of NSAT with high spatial resolution is deemed necessary for several applications such as hydrology, meteorology and ecology (Zhu et al. 2013).

Considering the high spatial autocorrelation of NSAT, several spatial interpolation methods have been employed to generate spatially continuous NSAT from point station measurements, including inverse distance weighting (IDW), Spline, Kriging, and even more sophisticated methods, such as co-Kriging and elevation-de-trended Kriging techniques (Duhan et al. 2013). However the performance of interpolation methods is highly dependent on the spatial density and distribution of weather stations. Satellite remote sensing provides the ability to extract spatially continuous information of land surface characteristics such as land surface temperature (LST) and the vegetation index (VI), which are closely relative to NSAT. The regression analysis methods for estimating NSAT take advantage of the correlations between NSAT and other environmental variables. Multiple linear regression (MLR) analysis using both remote sensing and geographical variables, including LST, VI, latitude, altitude, and so on, as predictors was performed to model NSAT (Cristóbal et al. 2008). However, a global regression analysis may miss local details that can be significant if the relationship is spatially non-stationary. The geographically weighted regression (GWR) is a local regression model, in which the

contribution of an observation site to the point to be calculated is weighted using a distance decay function based on the assumption that the observations near to the point to be calculated would have more influence on the estimate than those further away (Fotheringham et al. 2003).

In this paper, a GWR based NSAT mapping method is proposed, in which remote sensing variables and geographical variables were considered. The adaptive bi-square function is selected as the kernel type for GWR model; and the golden section search is used to determine the optimal bandwidth. MODIS LST and Normalized Difference vegetation index (NDVI) data were employed to predict NSAT. The performance of the proposed method was assessed by mapping monthly minimum, mean and maximum NSAT in China for a period of 12 months of 2010, and the estimated NSAT were validated against the meteorology observations.

2. DATA

2.1 Satellite Data

The MOD13A3 is the monthly VI product at a 1 km spatial resolution produced by averaging one month of daily VI product. MOD11A2 is a tile of the eight-day LST product at a resolution of 1 km produced by averaging eight days of the daily LST product. In this study, NDVI from MOD13A3 and the daytime LST from MOD11A2 data were employed to predict the monthly NSAT. The MOD13A3 and MOD11A2 products covering China territory in 2010 were collected. The MODIS products were preprocessed, including projection, mosaicking, and clipping, using MRT software. In addition,

* Corresponding author

monthly LST data were generated by averaging four MOD11A2 data sets for each calendar month of 2010.

2.2 Station Data

Daily NSAT (i.e., minimum, maximum, and mean NSAT) data in 2010 were provided by the China Meteorological Data Service Center. These data were collected from 2132 meteorological stations in China. To predict monthly NSAT, the daily NSAT were aggregated to monthly NSAT. In this study, we have selected 80% of the meteorological stations (i.e., 1706 stations) for predicting NSAT and the remaining 20% (i.e., 426 stations) were used for validation.

2.3 Elevation Data

The global digital elevation model (DEM) at the spatial resolution of 90 m that was produced by the NASA Shuttle Radar Topographic Mission (SRTM) was collected. In this study, the SRTM DEM data were resampled from 90 m to 1 km to render them consistent with the MODIS product

3. METHODOLOGY

3.1 Principle of Geographically Weighted Regression

The GWR is a regional regression method that can be used to investigate the non-stationary relationship between the dependent and explanatory variables (Foody 2003; Fotheringham et al. 2003). The GWR expands the standard multiple linear regression model for use with spatial data. With geographically weighted regression, the relationship between the dependent variable Y and explanatory variables X_i can be expressed as:

$$Y_j = \beta_0(u_j, v_j) + \sum_{i=1}^p \beta_i(u_j, v_j) X_{ij} + \varepsilon_j \quad (1)$$

where $\beta_0(u_j, v_j)$ and $\beta_i(u_j, v_j)$ are the intercept and the slope estimated at the j th point, respectively; ε_j is the regression residual at the j th point; and (u_j, v_j) are the coordinates of the j th point. Unlike a global regression method, the coefficients in Equation (2) are estimated by the observations around the j th point, and the contribution of an observation site to the coefficients estimate for the j th point is weighted using a distance decay function based on the assumption that the observations near to the j th point would have more influence on the estimate than those further away. Therefore, the coefficients can be obtained from:

$$\hat{\beta}(u_j, v_j) = (X^T W(u_j, v_j) X)^{-1} X^T W(u_j, v_j) Y \quad (2)$$

where $\hat{\beta}(u_j, v_j)$ represents the local coefficients to be estimated at location (u_j, v_j) ; X and Y are the vectors of the explanatory and the dependent variables, respectively; and $W(u_j, v_j)$ is the weight matrix. Gaussian and bi-square kernel functions are two common kernel types for the GWR model. The Gaussian kernel weights gradually decrease from the center of the kernel, but never reach zero. The bi-square kernel function has a clear-cut range where the weighting is non-zero (Chen et al. 2015). In this study, the adaptive bi-square function is used to derive the weight matrix:

$$w_{ij} = \begin{cases} \left[1 - \left(d_{ij}/b\right)^2\right]^2 & \text{when } d_{ij} \leq b \\ 0 & \text{when } d_{ij} > b \end{cases} \quad (3)$$

where d_{ij} is the Euclidean distance between the j th point and neighboring observation i and b is the kernel bandwidth. Golden section search is used to determine the optimal bandwidth, and is excluded in the GWR model.

3.2 Mapping Near Surface Air Temperature based on Geographically Weighted Regression

Because the GWR is a regional model, the effect of latitude on NSAT can be assumed to be constant. In this study, predictor variables used for GWR-based NSAT mapping include altitude, LST, and NDVI. The basic assumption of this method is that altitude, LST, and NDVI have a significant correlation with NSAT. However, the values of altitude and NDVI are usually constant over regions covered by snow and lakes, which contradicts this assumption, so the pixels of water body and snow are removed from further analysis. Figure 1 represents the flow chart of the GWR based mapping NSAT method proposed in this study.

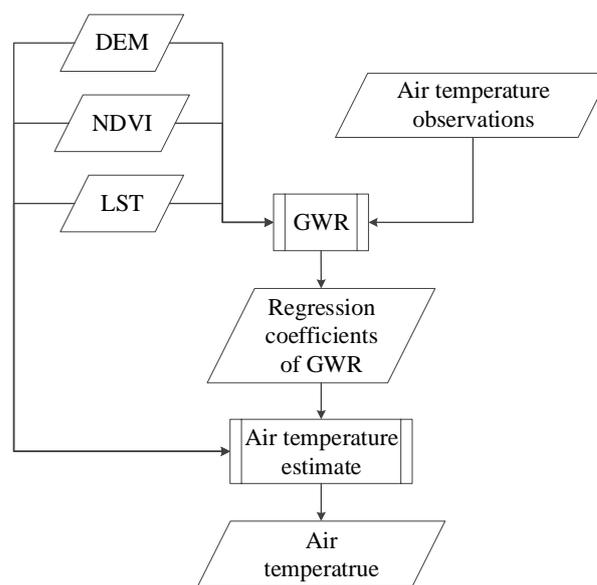


Figure 1 The flow chart of method for mapping near surface air temperature proposed in this study.

4. RESULT

4.1 Result

Figure 2 represents regression residual map and NSAT map derived using the GWR model in June 2010. The NSAT map shows some texture information, and includes some 'Nodata' due to the missing data (e.g., snow cover and water body). The estimated NSAT in China continent in June 2010 is between 5 °C and 33 °C. As shown in Figure 2 (b), the NSAT in most regions of China is greater than 20 °C, except for the high terrain regions, such as the Tibetan Plateau and Tianshan Mountains. The residuals derived using GWR model range from -5 °C to 5 °C, and most of them range from -2 °C to 2 °C.

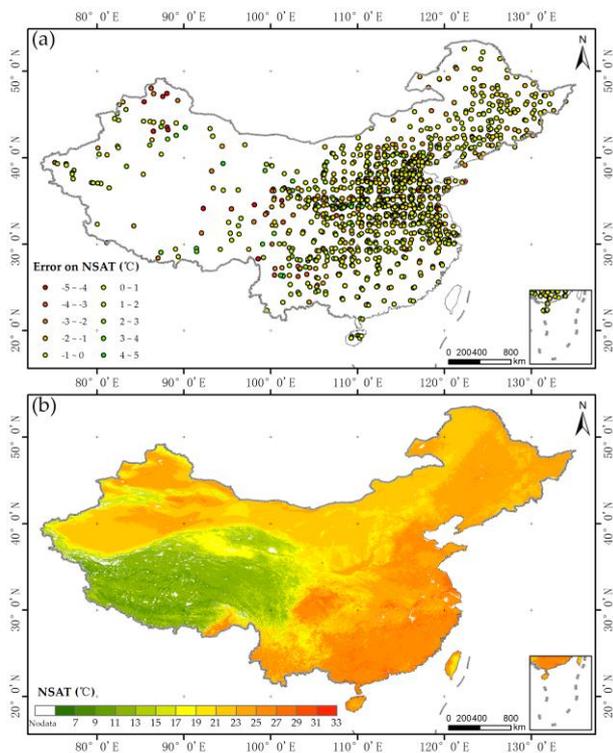


Figure 2 (a) The regression residual derived from GWR model in June 2010, (b) NSAT map derived using the GWR model in June 2010.

Figure 3 represents the RMSE and R^2 of the predicted monthly minimum, mean, and maximum NSAT using the GWR model in China in 12 months of 2010. In the colder months (i.e., from January to March and from October to December), the RMSEs of the predicted monthly mean NSAT using the GWR model are lower than those of the monthly maximum NSAT, and the RMSEs of the predicted monthly maximum NSAT are lower than those of the predicted monthly minimum NSAT. In the warmer months (i.e., from April to September), the RMSEs of the predicted monthly mean and minimum NSAT are similar, and both of them are lower than those of the predicted monthly maximum NSAT. The mean RMSEs for 12 months using the GWR model are 1.52 °C for monthly mean NSAT, 1.62 °C for monthly minimum NSAT, and 1.62 °C for monthly maximum NSAT, respectively. The Total RMSE for three NSAT variables is 1.58 °C. The R^2 for monthly minimum, mean, and maximum NSAT are similar in the colder months. The R^2 decrease in the order from monthly minimum, to mean, to maximum NSAT in the warmer months.

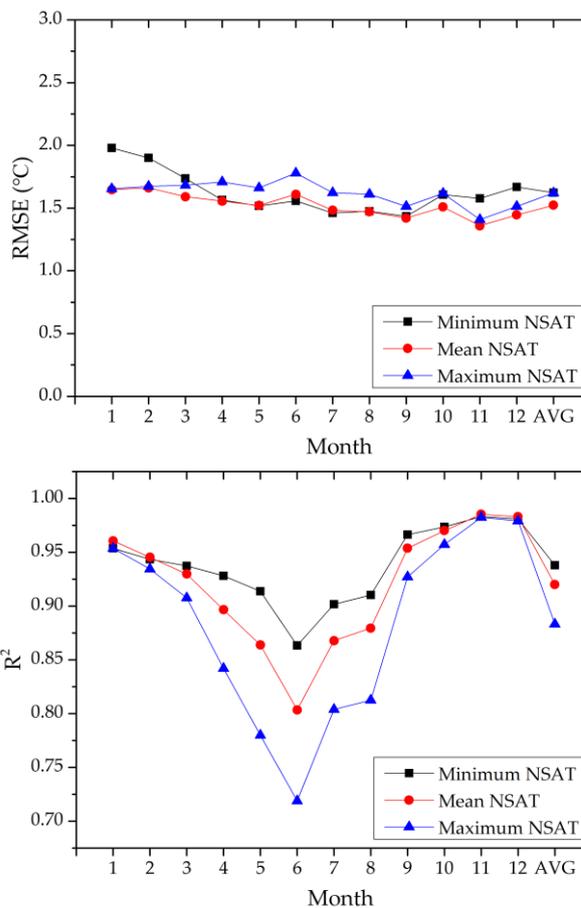


Figure 3 RMSE and R^2 of the predicted monthly minimum, mean, and maximum near surface air temperature using the geographically weighted regression model in China in 12 months of 2010.

Figure 4 represents the RMSE and R^2 of the predicted monthly mean NSAT using GWR model for varied terrain types in China in 12 months of 2010. As shown in Figure 4, with month change, the RMSEs using the GWR model for plateaus, hills, and plains are stable, while the RMSEs for basins are variable. The mean RMSEs of 12 months using the GWR model are 2.09 °C for plateaus, 1.41 °C for basins, 0.48 °C for plains, and 1.13 °C for hills, respectively. The mean RMSE for plateaus and basins is higher than that of hills and plains. One possible reason for this is that the weather station density of hills and plains is greater than that of plateaus and basins. The R^2 values of GWR model decrease first and then increase as the month progresses for all terrain types. The mean R^2 values of 12 months using the GWR model are 0.85 for plateaus, 0.87 for basins, 0.94 for plains, and 0.94 for hills, respectively.

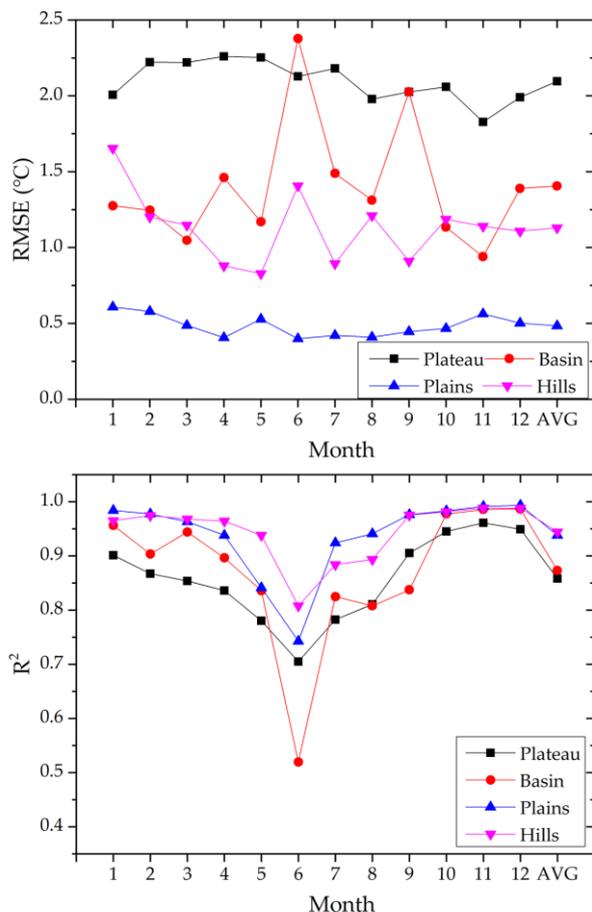


Figure 4 RMSE and R^2 of the predicted monthly mean near surface air temperature using the geographically weighted regression model for varied terrain types in China in 12 months of 2010.

5. CONCLUSION

In this study, a regression-based NSAT mapping method is proposed. This method is combined remote sensing variables with geographical variables, and uses geographically weighted regression to obtain continuous surface of NSAT. The meteorology observations were used to validate the NSAT retrieved using the proposed method. NSAT variable type, season and terrain type have impact on predicting NSAT using GWR model. Validation results with meteorological data show the proposed method achieved an total accuracy of 1.58°C.

ACKNOWLEDGEMENTS

This research has been supported by National Natural Science Foundation of China (61401461).

REFERENCES

Chen, F., Liu, Y., Liu, Q., & Qin, F., 2015. A statistical method based on remote sensing for the estimation of air temperature in China. *International Journal of Climatology*, 35, 2131-2143

Cristóbal, J., Ninyerola, M., & Pons, X., 2008. Modeling air temperature through a combination of remote sensing and GIS data. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 113

Duhan, D., Pandey, A., Gahalaut, K.P.S., & Pandey, R.P., 2013. Spatial and temporal variability in maximum, minimum and mean air temperatures at Madhya Pradesh in central India. *Comptes Rendus Geoscience*, 345, 3-21

Foody, G.M., 2003. Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI-rainfall relationship. *Remote Sensing of environment*, 88, 283-293

Fotheringham, A.S., Brunson, C., & Charlton, M., 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons

Guan, H., Zhang, X., Makhnin, O., & Sun, Z., 2013. Mapping mean monthly temperatures over a coastal hilly area incorporating terrain aspect effects. *Journal of Hydrometeorology*, 14, 233-250

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D.W., & Medina-Elizade, M., 2006. Global temperature change. *Proceedings of the National Academy of Sciences*, 103, 14288-14293

Zhu, W., Lü, A., & Jia, S., 2013. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sensing of environment*, 130, 62-73

APPENDIX (OPTIONAL)

Any additional supporting data may be appended, provided the paper does not exceed the limits given above.

Revised March 2018