Applied Geography 44 (2013) 112–120

Contents lists available at ScienceDirect

Applied Geography

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

Spatial interpolation of temperature in the United States using residual kriging

Tingting Wu [a], Yingru Li [b,*]

[a] Department of Agronomy and Soils, Auburn University, Auburn, AL 36849-5412, USA
[b] Department of Geology and Geography, Auburn University, Auburn, AL 36849-5412, USA

Keywords:
ArcGIS
Residual kriging
Semivariogram
Spatial interpolation
Temperature

A B S T R A C T

Temperature is one of the most important factors influencing every aspect of life. In response to the increasing greenhouse effect in recent years, the demand for understanding the spatial variability of temperature in the U.S. has risen dramatically. To meet this need, we developed a statistical model for constructing a gridded temperature dataset over the mainland United States. Based on the data collected from 922 meteorological stations in the U.S., temperatures at over 5000 unknown locations were predicted in January and July, 2010. This study utilized variables of latitude and longitude (model 1), and latitude, longitude and elevation (model 2) as inputs in a residual kriging method to interpolate the average monthly temperature. We also estimated temperatures at the same locations with the kriging function of ArcGIS and compared the performances of our models with that of ArcGIS. We found that, by adding an elevation factor, our model (model 2) had a better predicting performance than that of ArcGIS kriging function in both January and July. However, only estimation in July was not different from the observation. This suggests that our kriging model is capable of capturing the spatial variability of temperature, but it is sensitive to season. The successful interpolation of July temperature indicates that the accuracy of interpolation can be improved by adding appropriate variables. Seasonal models developed in future research can be valuable tools for meteorological and climatological research.

Introduction

Temperature is one of the most important atmospheric variables which directly impact physical and biological processes (Li et al., 2013; Stahl et al., 2006). Temperature can be a basis for understanding many processes such as evapotranspiration and plant distribution (Dodson & Marks, 1997; Trisurat, Shrestha, & Kjelgren, 2011). With increasing concentration of greenhouse gases, climate change causes growing concerns (Boykoff & Boykoff, 2007). Since 1895, the average annual temperature of the contiguous United States increased by 0.07 °C per decade. In 2010, it was 12.1 °C, 0.6 °C above normal (NOAA National Climate Data Center, 2010). 2010 was among the two warmest years by 2011. Twelve eastern US states had their warmest summer in 2010 (Blunden, Arndt, & Baringer, 2011).

Knowledge of the spatial variability of temperature is required to meet the need of assessing the recent climate change and greenhouse effect in the U.S. However, this is limited by the current spatial coverage of climate datasets. Temperature data is always collected from irregularly arrayed and discretely distributed meteorological stations. Especially in mountainous regions, there are insufficient meteorological stations (Rolland, 2003). Therefore, there is a rising demand for creating quality gridded climatological datasets through interpolation methods. Previously, spatial interpolations of temperature in the U.S. were mostly at regional bases (Brown & Comrie, 2002; Hunter & Meentemeyer, 2005; Serbin & Kucharik, 2009). To our knowledge, only few studies developed interpolation models for the entire country. Willmott and Matsuura (1995) proposed annual interpolation techniques incorporating spatially high-resolution digital elevation information, an average environmental lapse rate, and high-resolution long-term average temperature for 1920–1987. Another study used a combination of daily observations and Parameter-Elevation Regression on Independent Slope Model (PRISM) maps as inputs to interpolate daily temperatures over the conterminous United States for 1960–2001 (Di Luzio et al., 2008). In another study, PRISM interpolation method was also used to develop datasets of monthly minimum and maximum temperature from 1971 to 2000 for the conterminous United States (Daly et al., 2008). Qi et al. (2012)

* Corresponding author. Tel.: +1 334 844 4069; fax: +1 334 844 4486.
E-mail addresses: tzw0010@tigermail.auburn.edu (T. Wu), yingru.li@auburn.edu (Y. Li).

0143-6228/$ – see front matter
http://dx.doi.org/10.1016/j.apgeog.2013.07.012

created county-level monthly average temperature using elevation as a covariate in the US in 2007. Spatial interpolation of mean temperature in short-term (monthly) always presents more challenges than annual or long-term mean temperature, and the implement of temperature interpolation in the recent year with limited interpolators presents even greater complexity.

A number of methods have been developed for interpolating climate data. Over the past several decades, kriging has become an essential tool in the field of geostatistics. Previous researches have demonstrated that kriging is a better interpolation method with high accuracy and low bias compared to other methods (Li, Cheng, & Lu, 2005; Mahdian, Bandarabady, Sokouti, & Banis, 2009; Yang, Wang, & August, 2004). The advantage of kriging over other non-geostatistical methods is that the spatial variation structure is estimated through variogram and takes spatial autocorrelation into consideration (Aalto, Pirinen, Heikkinen, & Venäläinen, 2013). The principle of kriging is straightforward. It is a linear combination of weights which are determined by the spatial variation structure (Hattis, Ogneva-Himmelberger, & Ratick, 2012). There are different forms of kriging: ordinary kriging, simple kriging, universal kriging, cokriging as well as residual kriging (Sluiter, 2009).

The real spatial processes always show surface trend or drift which is the variation of the attribute depending on other variables. Latitude and longitude are always the underlying effects relative to temperature (Zhao, Nan, & Cheng, 2005). Elevation was also found to be highly associated with temperature (Samanta, Pal, Lohar, & Pal, 2012; Stahl et al., 2006). In addition, other factors including water bodies and local terrain also contribute to the temperature variation (Daly et al., 2008; Minder, Mote, & Lundquist, 2010). The distance to large water body greatly influences the temperature pattern. The difference of summer maximum temperature in 5 km and 20 km distance to the water is above 20 °C (Daly et al., 2002). Terrain-related factors such as slope orientation and slope gradient affect the amount of sun radiation on the surface of the land and then temperature (Zhao et al., 2005). Residual kriging can address the spatial data with broad regional trend. It is a two-step algorithm. The first step is to identify and remove the systematic trend in the spatial process. The second step is that the remaining part, the stationary residuals, are kriged to obtain estimations. The final product is the sum of the trend and the kriging residuals (Holdaway, 1996). Residual kriging is extensively used in meteorological data analysis (Sluiter, 2009). This method of decomposing the observation into a surface trend plus a residual value is better than interpolating directly on the observed data, especially when it applies to a large spatial area (Holdaway, 1996; Liao & Li, 2004). Dryas and Ustrnul (2007) compared different interpolation methods and found the residual kriging method as the best method of all to interpolate the monthly and seasonal average temperatures in Poland. This method also performed as the best method to predict temperatures in Spain and Slovenia (Sluiter, 2009). So far, few have been known to use residual kriging in interpolating temperatures for the entire U.S. Given the diversity of spatial variability of monthly temperature in such a large region, we sought to develop a simplified interpolation model by using detrending to account for the latitude–longitude and the elevation effects.

Geographic Information System (GIS) plays an important role in geospatial analysis (Silberman & Rees, 2010) and produces reliable climatological datasets. Meanwhile, with the development and proliferation of computer technology, statistical software also becomes a popular tool in geostatistical analysis such as SAS software (Statistical Analysis System). However, comparison of the accuracy of kriging interpolations in ArcGIS and in statistical software has been rarely studied. This study also aims at identifying the performance of kriging model fitting in SAS as compared with ArcGIS.

Based on the above literature review, this study aims to develop a statistical model to interpolate the average monthly temperatures in the U.S. in January and July 2010, to test the accuracy and precision of the proposed model by comparing it with ArcGIS kriging function, and to identify the variability of average temperatures in January and July in 2010. The section that follows introduces data and data sources as well as model construction and validation. A discussion of results is presented in the next section regarding the model performance and spatial variation of air temperature in the U.S. Finally, conclusions based on the paper's major findings are summarized in the last section.

Methodology

Study area and data collection

The study area for this analysis mainly includes the mainland of the United States located at 24–50°N and 68 to 125°W (Fig. 1). The study region is characterized by a variety of climate patterns. It ranges from humid continental in the north to humid subtropical in the south. It also includes many other types such as the tropical climate in southern Florida, the semi-arid climate in the west of the Great Plains, the arid climate in the Great Basin as well as the alpine, desert, and oceanic climates.

The dataset was obtained from the global historical climatology network (GHCN-Monthly) database, version 3, in oceanic and atmospheric administration national climate data center for the year 2010. The version 3 dataset is subjected to quality control and homogeneity adjustment (NCDC, 2010a). The data file includes the daily-updated monthly average, maximum, and minimum temperatures. Likewise, it also includes basic geographical station information such as latitude, longitude, elevation, type of topography, surrounding vegetation and population class, etc. Analysis was restricted to stations with no missing observations. A total of 922 historical climate network stations were included in this study. The latitude, longitude and elevation were used as predictors to interpolate the monthly average temperature. Elevation of the selected stations ranges from −59.1 m to 2763 m. The categories of vegetation vary across warm grassland, warm crop and warm field woods in the south to cool grassland, cool crop and cool conifer in the north. 60% of the stations are in rural areas with a population of less than 10,000 persons, 24% are in suburban areas (10,000–50,000 persons), and 16% are in urban areas (>50,000 persons). January and July were selected as representatives of winter and summer months, respectively.

Interpolation

The trend estimation

The first multiple regression model (model 1) for estimating the underlying trend was developed on the geographic coordinates latitude and longitude. It was formulated as $T = a + bx + cy + dx^2 + ey^2 + fxy$. The dependent variable T is the monthly average temperature in °C, the independent variables are x (latitude in degree), y (longitude in degree), the second-order polynomials (x^2, y^2) and the interaction term (xy).

The second multiple regression model (model 2) was developed on latitude, longitude and elevation, and was formulated as $T = a + bx + cy + dx^2 + ey^2 + fxy + gz + hzx + izy$. Predictors of this model are the same as model 1 except for the addition of elevation (z), the interaction between elevation and latitude (zx) and the interaction between elevation and longitude (zy).

We ran the two models for the January and July datasets, respectively. A step-wise method was used to select the significant predictors. The first predictor entered into the model is considered