Science of the Total Environment 615 (2018) 875–881

Contents lists available at ScienceDirect

Science of the Total Environment

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

Mapping relative humidity, average and extreme temperature in hot summer over China

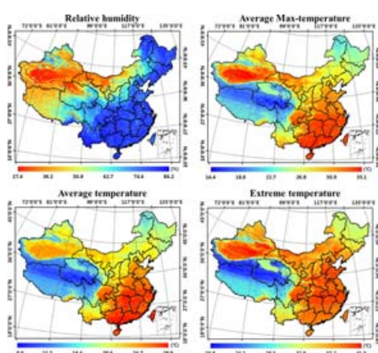
Long Li, Yong Zha *

Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Key Laboratory of Virtual Geographic Environment of Ministry of Education, College of Geographic Science, Nanjing Normal University, Nanjing 210023, China

HIGHLIGHTS

- Relative humidity (RH) and temperatures were estimated in hot summer over China.
- Finer RH and temperature gradients were produced.
- Social economic activity, RH and vegetation tend to affect extreme heat events.
- Geographical position significantly affects the spatial pattern of temperatures.
- Acceptable prediction errors for RH (RMSE = 7.4%) and temperatures (RMSE < 2.6 °C)

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 18 July 2017
Received in revised form 29 September 2017
Accepted 2 October 2017
Available online xxxx

Editor: D. Barcelo

Keywords:
Relative humidity
Air temperature
Remote sensing application
Urban heat island
Random forest
China

ABSTRACT

Air temperature and relative humidity are the key variables in environmental health research. Both of them are difficult to map especially at national scale because of spatial heterogeneity. This paper presents a methodology for mapping relative humidity, average and extreme temperature in hot summer (June to August) over China. Several data as explanatory variables were applied to random forest regression models to predict relative humidity and temperatures, including surface reflectance, land cover, digital elevation model (DEM), enhanced vegetation index (EVI), latitude, nighttime lights (NLs), as well as buffer zones of road, railroad, river system and administration center. Results based on cross-validation reflect acceptable prediction errors in estimating relative humidity (RMSE = 7.4%), average temperature (RMSE = 2.4 °C), average maximum temperature (RMSE = 2.5 °C), and extreme maximum temperature (RMSE = 2.6 °C). Despite the strong correlation between average and extreme temperatures, significant differences exist in their spatial distribution along the latitude direction, especially in the areas such as Hebei, Szechwan, Hubei, Henan, Shandong, and Inner Mongolia. Specifically, social economic activity, relative humidity and vegetation tend to affect extreme heat events, and both latitude and DEM (i.e., geographical position) determine the average level of temperature. Compared with interpolation technology and statistical methods, the proposed methodology demonstrates the ability to generate relative humidity and temperature maps with finer gradients in hot summer over China.

1. Introduction

China has experienced urbanization at an unprecedented pace over the past three decades (United Nations, 2010), causing anthropogenic

* Corresponding author.
E-mail address: yzha@njnu.edu.cn (Y. Zha).

https://doi.org/10.1016/j.scitotenv.2017.10.022
0048-9697/

climate change. Near-surface relative humidity and air temperature, defined as the observed values taken at the height 1.5 m above the ground monitored by stationary meteorological instruments, are the key variables in the studies of anthropogenic climate change. Both of them significantly affected the regional climate, microbial exposures, air quality, and human health (Sherwood et al., 2010; Grimm et al., 2008; Frankel et al., 2012; Fang et al., 2004; Piver et al., 2003; Harvell et al., 2002; Willett et al., 2007). Extreme temperature during the heat wave leads to excess mortality and morbidity related to vulnerable populations (Basu, 2009; Huang et al., 2011; Kovats and Hajat, 2008). The health risks associated with extremely hot weather was affected by the variability in heat exposure and social vulnerability (Ho et al., 2017; Hondula et al., 2012). Therefore, mapping relative humidity, average and extreme temperature in hot summer is of great significance to facilitate the public service and environmental policy.

Three main strands of research have evolved for estimating relative humidity and temperatures, including interpolation, simulation and regression techniques. As the simplest approach, interpolation generates relative humidity and temperature maps with high spatial and temporal resolution but leads to considerable uncertainties and errors (Vincent and Mekis, 2006; Yang, 2004). Four groups of simulation techniques are used to estimate air temperature, including micro-scale computational fluid dynamic (CFD) models, mesoscale numerical weather prediction (NWP) models, energy balance models, and coupled models (Taheri-Shahraiyni and Sodoudi, 2016). However, there are some limitations (e.g., small-scale, need many parameters, expensive time and computer load) so that it is not applicable for estimating relative humidity and temperatures with wide coverage. Satellite remote sensing provides high-resolution spatially contiguous information on the earth surface. It was used to analyze the thermal environment across 419 global big cities using MODIS land surface temperature (LST) products (Peng et al., 2012). Nevertheless, remote sensing technology is unable to derive air temperature, which is an important parameter that differ from LST for a wide range of applications such as vector borne, disease bionomics, hydrology and climate change studies (Benali et al., 2012; Jin and Dickinson, 2010). Regression approaches such as linear (Nichol et al., 2009) and complex non-linear models (Jang et al., 2004; Ho et al., 2016) thus were established to estimate surface air temperature (SAT) with satellite remote sensing. Several studies have attempted to estimate SAT using linear correlation between SAT and LST at regional and national scale (Benali et al., 2012; Williamson et al., 2013; Vogt et al., 1997; Gallo et al., 2011; Shen and Leptoukh, 2011; Zhu et al., 2013; Chen et al., 2015; Shen and Leptoukh, 2011; Xu and Liu, 2015). However, estimating SAT based on its linear correlation with LST was significantly affected by limited satellite time-series data, cloud contamination, and spatial heterogeneity (Shen et al., 2016; Jang et al., 2014; Ho et al., 2014).

In this study, remote sensing-based random forest (RF) regression model is implemented to estimate relative humidity and temperatures during summer (June to August) over China, using meteorological data, basic geographic information and satellite remote sensing imagery. Meteorological parameters such as relative humidity, average temperature, and average maximum temperature and extreme temperature are mapped. We analyze the spatial pattern of relative humidity and temperatures, and determine the importance of explanatory variables in the regression model.

2. Materials and methods

2.1. Data preparation

Meteorological data from meteorological monitoring stations records relative humidity and air temperature. It was collected primarily from China meteorological data service center (CMDC) (http://data.cma.cn/site/index.html). Basic geographic information such as road, railroad, and river system (third-, fourth-, fifth-order stream) and administration center (34 provincial cities, 331 prefecture cities, and 2089 county-level cities) was used for buffer analysis and reclassification (Table 1), and then was considered as the prediction variables. Satellite remote sensing products were prepared, including red, green and blue bands from MODIS surface reflectance products (MOD09A1), MODIS EVI (MOD13A3, month composites) products, and nighttime lights products. Surface reflectance products download from NASA (https://modis.gsfc.nasa.gov/) covers China with 500 m spatial resolution. MODIS EVI over the period from 1 June to 31 August 2009 was obtained from the website of NASA with 1180 m spatial resolution (https://ladsweb.nascom.nasa.gov/). Version 4 of DMSP-OLS nighttime lights was obtained from national oceanic and atmospheric administration (NOAA) with 1 km spatial resolution (https://www.ngdc.noaa.gov/). Digital elevation model (DEM) was resampled to 1 km spatial resolution. It was obtained from the CGIAR consortium for spatial information with 90 m spatial resolution (http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp). Land cover map during 2009 was collected (http://due.esrin.esa.int/page_globcover.php).

2.2. Regression modeling

RF regression model, using predictors derived from satellite and elevation data, was used to estimate urban SAT on hot summer days, and produced the lower error in comparison with ordinary least squares regression and support vector machine (Ho et al., 2014). It does not overfit and can optimize importing data with the advantage of fast convergence and strong generalization ability (Breiman, 2001). RF regression model was proved to have finer spatial and temporal resolution to quantify UHI using meteorological data and satellite remote sensing imagery (Li et al., 2017). In this study, RF regression model was utilized to estimate relative humidity, average temperature, and average maximum temperature and extreme temperature in hot summer (June to August) over China. First, observed relative humidity as response variable and the buffer zones of basic geographic information, and satellite observation data as predictive variables were input into RF regression model to estimate relative humidity. Next, relative humidity derived from RF regression model, integrating sixteen predicted variables (red band, green band, blue band from MODIS surface reflectance products, land cover, DEM, EVI, latitude, nighttime lights (NLs), as well as buffer zones of road, railroad, river system and administration center) was used to estimate average and extreme temperature using RF regression model. The importance of predicted variables was evaluated based on the percentage increase in mean squared error (%IncMSE) from RF regression model by 100 iterations. The mean absolute error (MAE) and root mean square error (RMSE) was used to evaluate the prediction accuracy. The formula is as follows:

$$MAE = \frac{\sum_{i=1}^{n} \left| Y_{obs,i} - Y_{model,i} \right|}{n} \quad (1)$$

Table 1
Buffer analysis and reclassification.

Prediction variables	Number of subclasses	Classes (1 to 3 from left to right)
Proximity to third-order stream (PtTS)	3	<1500, 1500–2000, >2000
Proximity to fourth-order stream (PtFoS)	3	<1000, 1000–1500, >1500
Proximity to fifth-order stream (PtFiS)	3	<500, 500–1000, >1000
Proximity to main railroad (PtMR)	3	<1000, 1000–2000, >2000
Proximity to main highway (PtMH)	3	<1000, 1000–2000, >2000
Proximity to provincial city (PtProC)	3	<20,000, 20,000–40,000, >40,000
Proximity to prefecture city (PtPreC)	3	<10,000, 10,000–20,000, >20,000
Proximity to county city (PtCC)	3	<5000, 5000–10,000, >10,000