Remote Sensing of Environment 154 (2014) 38–45

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Mapping maximum urban air temperature on hot summer days

Hung Chak Ho [a,*], Anders Knudby [a], Paul Sirovyak [a], Yongming Xu [b], Matus Hodul [a], Sarah B. Henderson [c,d]

[a] Department of Geography, Simon Fraser University, Burnaby, BC, Canada
[b] School of Remote Sensing, Nanjing University of Information Science & Technology, Nanjing, China
[c] School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
[d] BC Centre for Disease Control, Vancouver, BC, Canada

ARTICLE INFO

Article history:
Received 4 April 2014
Received in revised form 4 July 2014
Accepted 6 August 2014
Available online xxxx

Keywords:
Landsat
Air temperature
Urban
Spatial modeling
Remote sensing application
Random forest
Statistical model
Urban heat island

ABSTRACT

Air temperature is an essential component in microclimate and environmental health research, but difficult to map in urban environments because of strong temperature gradients. We introduce a spatial regression approach to map the peak daytime air temperature relative to a reference station on typical hot summer days using Vancouver, Canada as a case study. Three regression models, ordinary least squares regression, support vector machine, and random forest, were all calibrated using Landsat TM/ETM+ data and field observations from two sources: Environment Canada and the Weather Underground. Results based on cross-validation indicate that the random forest model produced the lowest prediction errors (RMSE = 2.31 °C). Some weather stations were consistently cooler/hotter than the reference station and were predicted well, while other stations, particularly those close to the ocean, showed greater temperature variability and were predicted with greater errors. A few stations, most of which were from the Weather Underground data set, were very poorly predicted and possibly unrepresentative of air temperature in the area. The random forest model generally produced a sensible map of temperature distribution in the area. The spatial regression approach appears useful for mapping intra-urban air temperature variability and can easily be applied to other cities.

1. Introduction

Near-surface air temperature, defined as the temperature 2 m above the land surface, is a key variable in studies of meteorology, climate, and environmental health (Garske, Ferguson, & Ghani, 2013; Harvell et al., 2002; Katsouyanni et al., 1993; Koken et al., 2003; Kuhn, Campbell-Lendrum, & Davies, 2002; Maria & Renganathan, 2008; Nichol, Fung, Lam, & Wong, 2009; Oke & Maxwell, 1975; Saaroni & Baruch, 2010). Previous studies have widely used air temperature to estimate the intensity of urban heat islands (Kolokotroni & Giridharan, 2008; Unger, Sümeghya, & Zobokib, 2001), to study the relationship air temperature and air pollution (Koken et al., 2003), and to predict risks of heat-related mortality (Laaidi et al., 2012). Air temperature is traditionally monitored by stationary meteorological instruments (weather stations), which provide point data with high temporal frequency, typically recorded on an hourly basis. However, such observations are often unable to adequately describe spatial heterogeneity over small geographic extents (Benali, Carvalho, Nunes, Carvalhais, & Santos, 2012). This is particularly important in thermally complex environments such as urban settings, where local microclimatic variability is influenced by factors such as land cover (Saaroni & Baruch, 2010), exposure to wind and sun, soil and vegetation moisture, and the thermal properties of upwind areas (Oke & Maxwell, 1975). Spatial patterns in these variables exist at very fine scales (10^1–10^2 m) compared with the sampling density typically provided by weather station networks (10^4–10^5 m), suggesting that spatial interpolation between station observations is not an optimal solution for mapping air temperature in the urban environment (Vogt, Viau, & Paquet, 1997). Remote sensing provides an additional source of data that can provide high-resolution spatially explicit information on many of the factors that influence air temperature and thus assist with mapping it in spatially heterogeneous environments. Three principal approaches have been used to map air temperature from remote sensing data: 1) the Temperature–Vegetation Index (TVX), 2) energy balance models, and 3) statistical analyses (Benali et al., 2012; Zakšek & Schroedter-Homscheidt, 2009).

The TVX method is based on the hypothesis that while an unvegetated surface can be substantially warmer than the surrounding air, the surface temperature of an infinitely thick vegetation canopy will approximate the air temperature because the canopy consists primarily of air, with branches and leaves volumetrically a minor component. On that basis, Prihodko and Goward (1997) used the observed negative correlation between the Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST), as well as an estimate of the NDVI value for an infinitely thick canopy, to estimate air temperature. To employ the TVX method, a sample window with varying vegetation cover is needed for the establishment of local NDVI–LST correlations. In addition to requiring local variations in vegetation cover, the use of a sample

* Corresponding author.
E-mail address: hohungh@sfu.ca (H.C. Ho).

http://dx.doi.org/10.1016/j.rse.2014.08.012
0034-4257/

window effectively reduces the spatial resolution of the predicted air temperatures and makes the TVX method best suited for large regions with gradual temperature changes (e.g. Sandholt, Rasmussen, & Andersen, 2002; Stisen, Sandholt, Norgaard, Fensholt, & Eklundh, 2007; Vancutsem, Ceccato, Dinku, & Connor, 2010; Zhu, Lu, & Jia, 2013) but unsuitable for urban areas.

The energy balance approach considers air temperature to be controlled by Earth system energy dynamics, including the radiation balance as well as soil, sensible, and latent heat fluxes (Meteotest, 2010; Oke, 1988; Sun et al., 2005). It is thus directly grounded in thermodynamics, but it relies on comprehensive parameterization for which spatially distributed data are rarely available, specifically at the resolution necessary for application to urban studies (Mostovoy, King, Reddy, Kakani, & Filippova, 2006).

Statistical analyses are mostly based on empirical regression modeling, which can take the form of linear (Nichol et al., 2009) or more complex statistical models like neural networks, genetic algorithms and regression trees (Emamifar, Rahimikhoob, & Noroozi, 2013; Jang, Viau, & Anctil, 2004; Singh, Joshi, & Kishtawal, 2006). Predictors of air temperature can be limited to land surface temperature (LST) (Mostovoy et al., 2006), or also include one or more additional environmental variables (Benali et al., 2012) such as NDVI, elevation, and land cover. Regression models can be suitable for areas with complex landscape characteristics, such as urban areas, although application will generally be limited to the environment for which they were developed.

Remote sensing-based air temperature mapping has typically focused on relatively large (>100,000 km^2) and homogeneous geographic regions (Benali et al., 2012; Mostovoy et al., 2006; Stisen et al., 2007; Vogt et al., 1997; Xu, Qin, & Shen, 2012). Few existing studies have attempted to map air temperature distributions at the city scale; the only notable example is provided by Nichol and To (2012) studying the distribution of air temperature in Kowloon, Hong Kong. The development and validation of approaches optimized to map air temperature distributions in urban environments is of particular importance in the context of extreme heat events and their impacts on human health, which are expected to increase in severity in the future. Specifically, development of a method to map peak daytime air temperature (Tmax) is of importance because this variable is commonly used to quantify the relationship between extreme heat and mortality (Kunst, Looman, & Mackenback, 1993; Medina-Ramon, Zanobetti, Cavanagh, & Schwartz, 2006), and recent studies indicate that maps of temperature during extreme heat events can help explain the spatial pattern of heat-related risk (Anderson & Bell, 2011; Buscail, Upegui, & Viel, 2012; Hondula et al., 2012; Laaidi et al., 2012; Tomlinson, Chapman, Thornes, & Baker, 2011). However, to be useful for heat emergency planning purposes, Tmax maps must be valid for typical (as opposed to specific) hot summer days, which precludes mapping of absolute temperature values that vary depending on the severity of the heat wave.

In this study, we assess the ability of three remote sensing-based regression models to map Tmax for the Greater Vancouver region of British Columbia, Canada, using Landsat data and point observations from weather stations in the area. The methods are ordinary least squares regression, support vector machine, and random forest. We quantify Tmax relative to Vancouver International Airport (YVR), as forecasted and observed temperatures at this weather station form the basis for heat health emergency definitions for the area.

2 . Study Area

Our study area is Greater Vancouver, British Columbia, Canada (Fig. 1), a coastal metropolitan area with >2 million people (Statistics Canada, 2007). Greater Vancouver is bordered to the north by fold mountain ridges, to the west by the Pacific Ocean, and to the east by the semi-arid Fraser Valley, a geographic context that generates a complex microclimate in the area (Oke, 1976; Oke & Hay, 1994; Runnalls & Oke, 2000). During the summer, ocean breezes and winds from the mountain ridges can cool down the coastal regions, while the Fraser Valley can trap air masses and create a relatively hot zone (Oke & Hay, 1994). Temperature in the urban area is heavily influenced by cloud cover in the summer period, while evaporative cooling is of little influence due to limited vegetation cover. On a hot summer day Greater Vancouver is typically cloudless with light winds from the Fraser Valley, a weather situation that can generate a strong urban heat island effect and substantially higher temperatures in the urban areas compared with their surroundings (Oke & Hay, 1994).

3 . Data and methods

3.1 . Satellite data

The satellite data used in this study consist of all ($n = 6$) cloud-free Landsat 5 TM and Landsat 7 ETM+ images available from 2001 to 2010 for hot summer days in the study area, here defined as days with Tmax > 25 °C at YVR (Table 1). 25 m Canadian Digital Elevation Data (CDED, http://www.geobase.ca) were used to provide elevation information for the study area. Landsat 5 TM images and the DEM were resampled to 60 m to match the spatial resolution of the ETM+ thermal band, and all data were projected to UTM zone 10 N.

3.2 . Satellite-derived predictors

Several spatial data layers were derived from the Landsat and elevation data for use as predictors in regression models to map Tmax: LST, Normalized Difference Water Index (NDWI), elevation, skyview factor (SVF), and solar radiation. All layers except elevation were derived separately for each Landsat image.

LST was estimated from Landsat band 6. Top of atmosphere radiance values were atmospherically corrected using NASA's Atmospheric Correction Parameter Calculator to obtain at-surface radiance (Barsi, Barker, & Schott, 2003), and kinetic surface temperature was then derived by inversion of Planck's Law, applying emissivity values from the North American ASTER Land Surface Emissivity Database (Hulley & Hook, 2009).

$$LST = K_2 / \ln(\varepsilon K_1 L_\lambda + 1)$$

where K_1 and K_2 are the coefficients, ε is the emissivity and L_λ is the radiance.

NDWI is an index designed to quantify vegetation water content (Gao, 1996), which strongly influences surface cooling through evapotranspiration. NDWI is defined as:

$$NDWI = (\rho_{NIR} - \rho_{MIR}) / (\rho_{NIR} + \rho_{MIR})$$

In this study, Landsat bands 4 and 5 were used to calculate NDWI for land areas. NDWI values are not meaningful over water and, as both ρ_{NIR} and ρ_{MIR} are very small over water, tend to be noisy. To allow the NDWI layer to function as a proxy for cooling by evapotranspiration, we found the maximum NDWI value on land and applied it to all water surfaces, replacing the results from the original NDWI calculation for water.

SVF can be defined as the portion of unobscured sky, which is related to the radiation received or emitted in an area (Chen et al., 2012; Su, Brauer, & Buzzelli, 2008) and is influenced by both topography and building structure. SVF was mapped for each pixel using an empirically calibrated relationship with shadow proportion, which was derived using partial spectral unmixing. Full details of the method used to derive the SVF data layer will be forthcoming in a separate paper; validation using independent lidar data for Vancouver shows the method to perform well (SVF root mean square error = 0.056).