Environmental Research 146 (2016) 51–58

Contents lists available at ScienceDirect

Environmental Research

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

Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: A statistical modeling study

Liuhua Shi [a,*,1], Pengfei Liu [b,1], Itai Kloog [c], Mihye Lee [a], Anna Kosheleva [a], Joel Schwartz [a]

[a] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA 02215, USA
[b] John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
[c] Department of Geography and Environmental Development, Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva, Israel

ARTICLE INFO

Article history:
Received 6 October 2015
Received in revised form
2 December 2015
Accepted 4 December 2015

Keywords:
Air temperature
Surface temperature
MODIS
Reanalysis
Exposure error

ABSTRACT

Accurate estimates of spatio-temporal resolved near-surface air temperature (T_a) are crucial for environmental epidemiological studies. However, values of T_a are conventionally obtained from weather stations, which have limited spatial coverage. Satellite surface temperature (T_s) measurements offer the possibility of local exposure estimates across large domains. The Southeastern United States has different climatic conditions, more small water bodies and wetlands, and greater humidity in contrast to other regions, which add to the challenge of modeling air temperature. In this study, we incorporated satellite T_s to estimate high resolution (1 km × 1 km) daily T_a across the southeastern USA for 2000–2014. We calibrated T_s–T_a measurements using mixed linear models, land use, and separate slopes for each day. A high out-of-sample cross-validated R^2 of 0.952 indicated excellent model performance. When satellite T_s were unavailable, linear regression on nearby monitors and spatio-temporal smoothing was used to estimate T_a. The daily T_a estimations were compared to the NASA's Modern-Era Retrospective Analysis for Research and Applications (MERRA) model. A good agreement with an R^2 of 0.969 and a mean squared prediction error (RMSPE) of 1.376 °C was achieved. Our results demonstrate that T_a can be reliably predicted using this T_s-based prediction model, even in a large geographical area with topography and weather patterns varying considerably.

1. Introduction

Global warming adds urgency to better understand the impact of temperature on health, particularly in the warm areas such as the southeastern USA. Growing evidence has linked near-surface air temperature (T_a), an important environmental stressor, with morbidity and mortality (Gosling et al., 2009a, 2009b; Zanobetti and Schwartz, 2008; Medina-Ramón and Schwartz, 2007; Basu, 2009; Laaidi et al., 2006; Xu et al., 2012). Previous studies concerning human health and T_a exposure are primarily limited by the spatial and temporal availability of T_a measurements, leaving large areas uncovered. Temperature can vary greatly both in space and time, therefore these collected point-samples are insufficient to adequately capture the spatial and temporal variability within a large area (Guo et al., 2013). In addition, owing to the urban heat-island effect, higher temperatures are often observed in urban areas versus surrounding areas. For example, temperatures measured in a station near an airport may underestimate the true near-surface air temperatures in the urban area.

However, these small geographic changes in air temperature can have important health effects. Shi et al. (2015) found that both geographical contrasts and annual anomalies of local air temperature contribute to excess public health burden of climate change. They used daily local air temperature estimates at fine geographic scale (1 km × 1 km) in New England, to capture the exposure variability in space and time which may be driving the adverse health effects. Lower exposure measurement errors, and the inclusion of the entire region, largely alleviate the downward bias in health effect estimates.

In recent years, several methods were developed to address the lack of high-resolution exposure data. Geospatial statistical methods, such as land use regression and kriging, are most commonly used approaches. They allow characterizing the spatial heterogeneity of exposure by using time invariant geographical variables to expand the ground monitored measurements to large areas (Vicente Serrano et al., 2003; Hudson and Wackernagel, 1994). However these methods do not generally capture temporal variability in exposure, in that they are commonly based on a year

* Corresponding author.
E-mail address: lis678@mail.harvard.edu (L. Shi).
[1] L. Shi and P. Liu contributed equally to this work.

http://dx.doi.org/10.1016/j.envres.2015.12.006
0013-9351/

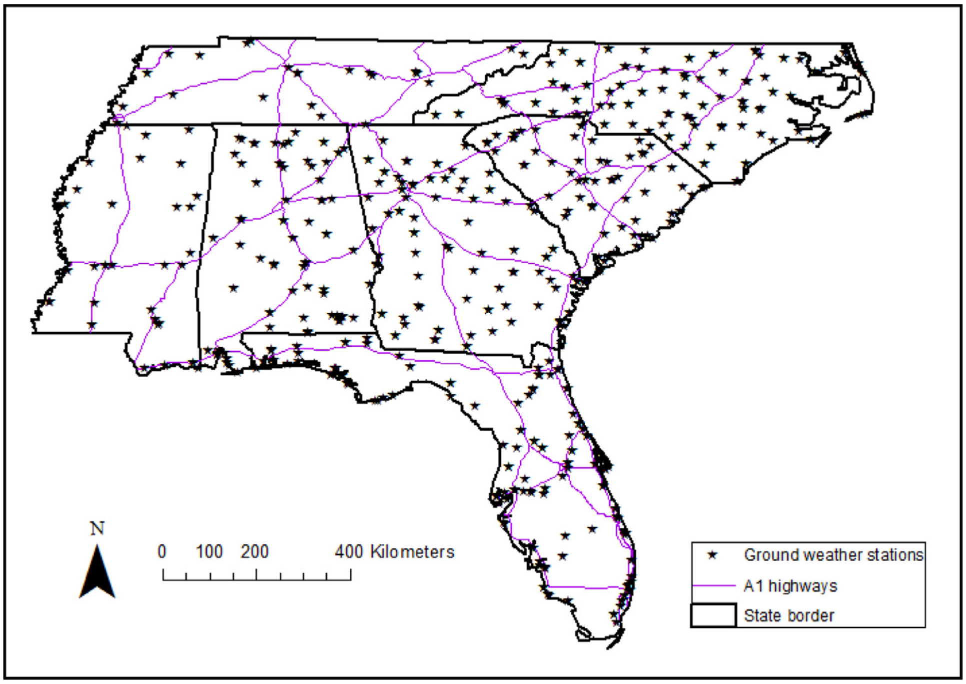

Fig. 1. Map of the study area showing all available NCDC air temperature monitor stations across southeastern USA for 2000–2014.

of intensive monitoring, and miss changes over years in the spatial pattern of T_a. Therefore they are primarily used to assess chronic health effects.

Satellite-based remote sensing can provide additional information at high spatial and temporal resolution. Satellite instruments, such as Moderate Resolution Imaging Spectroradiometer (MODIS), can provide global daily estimates of 1 km surface skin temperature (T_s), the temperature at the air-soil interface (Vancutsem et al., 2010; Zhu et al., 2013; Benali et al., 2012). T_s is derived from the thermal infrared signal received by the satellite sensor. T_s as an indicator of the net surface energy balance, depends on the presence of vegetation or plant cover, atmospheric conditions, and thermal properties of the underlying surfaces. It is different from T_a, which is measured at meteorological stations at the screen height of 1–2 m above the land surface.

Several studies have shown that T_s and T_a are correlated (Stoll and Brazel, 1992; Voogt and Oke, 1997). Even so, there are many factors that can influence the complex and geographically heterogeneous relationship between T_s and T_a, such as humidity, the type of underlying surfaces, the elevation, and other surface parameters. Additionally, Dousset (1989) stated that T_a has superior correlation with T_s at night because the solar radiation does not affect the thermal infrared signal. During nighttime T_s is close to T_a and during daytime T_s is generally higher than T_a (Vancutsem et al., 2010).

The value of T_a cannot be predicted by T_s using a simple linear relationship with reasonable accuracy. Fu et al. (2011) explored predicting T_a using satellite T_s and found an $R^2 > 0.55$. Recently, Kloog et al. (2014) presented a novel model and assessed daily mean T_a in the northeastern USA using MODIS-derived T_s measurements (Kloog et al., 2012, 2014). Better predictive performance was reported. For days with available T_s data, mean out-of-sample R^2 was 0.947. Even for days without T_s values, the model accuracy was also excellent. Although T_a estimation in the northeastern USA was excellent, it is uncertain that how well the satellite approach would perform in areas with different geographic features and weather patterns.

Predicting T_a for the southeastern USA is of top priority for epidemiological studies. Many researchers are of great interest in investigating the health effects of temperature using the Medicare cohort (aged 65+), because this elderly population is potentially most vulnerable to climate change. Due to its warm weather, the southeastern USA contains a very large Medicare population (13 millions). Thus, it is particularly urgent to provide a comprehensive temperature dataset for this particular region.

The aim of this paper was to estimate 15-year daily local air temperature in the southeastern USA, by extending and validating the previous hybrid-model approach to account for the unique geographical and climatological characteristics of the study area. Specifically, by incorporating satellite derived T_s, T_a measured at monitors, meteorological variables and land use terms, we employed a 3-stage statistical modeling approach to obtain daily air temperature predictions at 1 km × 1 km resolution across the southeastern USA for the years 2000–2014. The retrieved air temperature was cross validated against the measurements from weather stations. As an independent validation, the results were also compared with NASA's reanalysis products, Modern-Era Retrospective Analysis for Research and Applications (MERRA).

2. Materials and methods

2.1. Study domain

The spatial domain of our study includes the southeastern part of the USA, comprising the states of Georgia, Alabama, Mississippi, Tennessee, North Carolina, South Carolina and Florida (Fig. 1). The southeastern states include some populous metropolitan areas (Charlotte, Memphis, Raleigh, Atlanta and Miami), rural towns, large forested regions, mountains, water bodies, and the Atlantic sea shoreline. The study region covers an area of 916,904 km^2 with a population of 56,742,948 according to the 2010 census, and encompasses 1,013,408 discrete 1 km × 1 km satellite grid cells.

2.2. Surface temperature

Daily T_s data from the MODIS sensors located on polar orbiting