The 2011 heat wave in Greater Houston: Effects of land use on temperature

Weihe Zhou, Shuang Ji, Tsun-Hsuan Chen, Yi Hou, Kai Zhang

Division of Biostatistics, University of Texas School of Public Health, Houston, TX, USA
Division of Epidemiology, Human Genetics and Environmental Sciences, University of Texas School of Public Health, Houston, TX, USA
CDM Smith, 8140 Walnut Hill Ln, Dallas, TX, USA

A R T I C L E   I N F O

Article history:
Received 31 May 2014
Received in revised form 30 July 2014
Accepted 11 August 2014

Keywords:
Heat wave
Land use regression
Quantile regression
Temperature

A B S T R A C T

Effects of land use on temperatures during severe heat waves have been rarely studied. This paper examines land use-temperature associations during the 2011 heat wave in Greater Houston. We obtained high resolution of satellite-derived land use data from the US National Land Cover Database, and temperature observations at 138 weather stations from Weather Underground, Inc (WU) during the August of 2011, which was the hottest month in Houston since 1889. Land use regression and quantile regression methods were applied to the monthly averages of daily maximum/mean/minimum temperatures and 114 land use-related predictors. Although selected variables vary with temperature metric, distance to the coastline consistently appears among all models. Other variables are generally related to high developed intensity, open water or wetlands. In addition, our quantile regression analysis shows that distance to the coastline and high developed intensity areas have larger impacts on daily average temperatures at higher quantiles, and open water area has greater impacts on daily minimum temperatures at lower quantiles. By utilizing both land use regression and quantile regression on a recent heat wave in one of the largest US metropolitan areas, this paper provides a new perspective on the impacts of land use on temperatures. Our models can provide estimates of heat exposures for epidemiological studies, and our findings can be combined with demographic variables, air conditioning and relevant diseases information to identify ‘hot spots’ of population vulnerability for public health interventions to reduce heat-related health effects during heat waves.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Heat waves have been reported to be associated with elevated risk of mortality and morbidity (Kovats and Hajat, 2008). Some people are more vulnerable to heat than others, including those with existing chronic disease, athletes, the elderly, children, the mentally ill, poor, among others (O’Neill et al., 2003, 2005). Because temperature is not homogeneous in space, higher exposures at certain areas related to types of land use such as vegetation, urbanicity and built environment (Harlan et al., 2006) may magnify the health risks of heat to vulnerable populations. Improved understanding of the associations between land use and temperature on hot days, particularly during heat waves, is therefore critical in identifying ‘hot spots’ and thus in targeting public health interventions to prevent heat-related illness and death during heat waves.

Effects of land use on air temperatures measurements during heat waves have been rarely studied. Although some studies have examined the impacts of land use on land surface temperature estimates derived from satellite images (Buscail et al., 2012; Kestens et al., 2011), few studies have explored the associations between air temperature observations and selected land use types and/or physical geographic information, e.g., imperviousness, distance to water bodies and distance to a city center (Oswald et al., 2012; Zhang et al., 2011); imperviousness, elevation and vegetation (Kloog et al., 2012); longitude, latitude, elevation, distances to coastline, and land use types of agriculture, forest, water bodies and build environment (Yunus et al., 2012). Additionally, limited studies have investigated how land use affects ground temperature measurements during severe heat waves when land use might have larger impacts on temperature than on the average temperature days, although some studies have examined the impacts of land use on land surface temperature during heat waves (Buscail et al., 2012). Moreover, previous studies have primarily used linear regression (Oswald et al., 2012), kriging (Zhang et al., 2011) and mixed models (Kloog et al., 2012), and have not addressed a research question: whether certain land use types have larger impacts on temperatures at ‘hot spots’ compared to non-hot spots.
This paper aims to examine effects of a variety of land use types on temperatures during the 2011 heat wave in Greater Houston using a land use regression approach, and then to conduct a quantile regression analysis to assess whether the impacts of selected land use on temperatures varied with a spectrum of quantiles. The 2011 heat wave in Houston led to the hottest August since records began in 1889 (National Weather Service, 2011) and the most consecutive days with issued heat wave alerts (August 2 to 30) by National Weather Service (Wood, L, personal communication, January 30, 2013). This heat wave lasted much longer than typical severe heat waves usually lasting for one week. It provides a unique research opportunity to explore land use-temperature associations. This analysis first takes advantage of the state-of-the-art technique developed in air pollution exposure assessment: land use regression, a linear regression built on a variety of variables related to land use, traffic and other variables derived from Geographic Information System (GIS) (Health Effects Institute, 2010; Zhang et al., 2014a). This analysis also applies a quantile regression method, which has been rarely used in heat-related studies.

2. Methods

2.1. Study area

The study area is Houston-Sugar Land-Baytown Metropolitan Statistical Area (MSA) in Texas that was delineated in December, 2009, a 10-county area including Austin, Brazoria, Chambers, Fort Bend, Galveston, Harris, Liberty, Montgomery, San Jacinto and Waller counties in Texas (US Census Bureau, 2013a). It had a total area of 25,890 km² with a population of 5.95 million in 2010, and it was ranked the 6th largest US metropolitan statistical area (US Census Bureau, 2013b). According to the 2010 Census, the total Houston Metro population included a racial-ethnicity makeup of 62.7% white (39.7% non-Hispanic), 18% black or African American, 7.2% Asian, among others (US Census Bureau, 2013b).

2.2. Data source

Temperature observations were obtained from Weather Underground Inc. (WU). Specially, we downloaded daily average, minimum and maximum temperature measurements at all available weather stations from August 1 to August 31 in 2011 (Weather Underground, 2013). We then excluded observations with outliers or invalid values or a lack of internal consistency following the procedure suggested by Durre et al. (2010). There were a total of 138 weather stations in Greater Houston in the study period (Fig. 1).

Land cover and imperviousness data at the 30 m x 30 m resolution were obtained from the National Land Cover Database (U.S. Geological Survey, 2013a). These data were used to calculate the percentage of land cover types and imperviousness around Weather Underground stations within buffer distances varying from 50 m to 750 m using ArcGIS 10.0 (ESRI, Redlands, CA). The nearest distance to the coastline and the nearest distance to water bodies were derived by the shortest straight line distance between weather stations and the coastline of the Gulf of Mexico and the Galveston Bay or between weather stations and lake/pond or reservoirs using ArcGIS 10.0. Coastline and water body data were obtained from the National Hydrography Dataset (U.S. Geological Survey, 2013b). Table 1 illustrates all independent variables used in this study.

2.3. Statistical analysis

We utilized land use regression and quantile regression methods to estimate effects of land use and geographic variables (Table 1) on temperatures during the 2011 heat wave in Greater Houston. We examined three temperature measures as response variables: the monthly averages of daily minimum, mean, and maximum at each weather station in August, 2011.

Land use regression models have been receiving considerable attention recently due to its ability to predict long-term exposures to air pollutants, especially in the studies examining the health effects caused by traffic-related air pollution (Health Effects Institute, 2010; Zhang, 2014a). They are essentially built on multiple linear regressions using many variables generally derived using GIS, e.g., land use, traffic intensity, nearby emission sources, and distance to major roads. This study uses GIS-derived variables listed in Table 1.

Quantile regression is a statistical technique for estimating the associations between quantiles of a response variable and independent variables (Koenker, 2005). Quantile regression has two major advantages compared to linear regression. First, quantile regression can provide a more complete picture of the estimated associations by characterizing the effects of the covariates on a spectrum of quantiles of the response variable, while linear regression only quantifies the associations between the means of a response variable and independent variables (Koenker, 2005). This feature is particularly important when the covariates’ effects vary with quantiles. Second, quantile regression is more robust than linear regression (Koenker, 2005). Linear regression assumes that variables are independent and identically distributed and residuals follow a normal distribution. However, quantile regression can estimate the associations without distributional assumptions.

We conducted the data analysis by following a procedure of screening, buffer size selection, variable selection, land use regression and quantile regression. First, given 114 geographic variables (Table 1), we excluded covariates with less variation when their 85th percentile values were the same with their 15th percentile values. Second, for the variables with varying buffer radii, we determined the ‘best’ buffer size where a variable had the highest Pearson correlation coefficient with a temperature variable of interest. Third, because some variables were dominated by zero values, we converted those with more than 50% zero values into binary variables. Forth, we built land use regression models by including all previous identified variables and then conducted the stepwise selection procedure to reduce the number of variables by setting p value < 0.15. Then we ran land use regression models based on those selected variables to keep variables with p value < 0.05. For those variables with marginal significance, we also used Bayesian information criterion (BIC) to compare the full model to the reduced model without the covariate of interest, and determined the model with smaller BIC values. Fifth, we ran quantile regression models based on the variables picked from previous stepwise selection procedure, and then made statistical inference on whether estimated effects of land use on temperature varied across different quantiles, following the procedures proposed by Peng and Huang (2008). We repeated the procedure for each of three response variables (monthly averages of daily average, maximum and minimum temperatures) separately. Statistical analysis was conducted using SAS (version 9.2, SAS Institute Inc., Cary, NC, USA) and R statistical software (R Development Core Team; http://R-project.org).

3. Results

3.1. Descriptive statistics

Monthly averages of daily maximum, mean and minimum at 138 weather stations in August, 2011, were 40.63, 32.38 and 25.23 °C, respectively. Daily maximum temperatures across stations have the largest variation (33.33–45.93 °C), followed by daily minimum temperatures (21.67–29.63 °C) and daily average temperatures (30.00–36.30 °C). The lowest temperatures occurred
Table 1
Variables considered for land use regression and quantile regression modeling.

<table>
<thead>
<tr>
<th>Predictor Variables</th>
<th>Variable name</th>
<th>Unit</th>
<th>Buffer zone radius (meters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to water bodiesa</td>
<td>Dist_Water</td>
<td>km</td>
<td>NA</td>
</tr>
<tr>
<td>Distance to the coastlineb</td>
<td>Dist_Coast</td>
<td>km</td>
<td>NA</td>
</tr>
<tr>
<td>Imperviousness</td>
<td>Imp_p &lt; radius&gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Open water</td>
<td>LU_water_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Developed open space</td>
<td>LU_dev_open_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Developed low intensity</td>
<td>LU_dev_lo_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Developed medium intensity</td>
<td>LU_dev_med_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Developed high intensity</td>
<td>LU_dev_hi_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Barren land (rock/sand/clay)</td>
<td>LU_barren_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Deciduous forest</td>
<td>LU_decid_forest_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Evergreen forest</td>
<td>LU_evergreen_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Mixed forest</td>
<td>LU_mix_forest_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Shrub/Scrub</td>
<td>LU_shrub_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Grassland/herbaceous</td>
<td>LU_grass_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Pasture/hay</td>
<td>LU_pasture_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Cultivated crops</td>
<td>LU_crop_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Woody wetlands</td>
<td>LU_woody_wetland_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
<tr>
<td>Emergent herbaceous wetlands</td>
<td>LU_herb_wetland_p &lt; radius &gt;</td>
<td>%</td>
<td>50, 100, 150, 300, 400, 500, 750</td>
</tr>
</tbody>
</table>

*a The nearest distance to lake/pond and reservoir.  
*b The nearest distance to the coastline.

Fig. 1. Weather Underground stations in Greater Houston, Texas, US (Weather stations in triangles with impervious surface image as the background; temperatures are monthly averages of daily maximum temperatures at each station).
generally along the coastline, and the highest temperatures were at inland stations (Fig. 1).

3.2. Land use regression modeling results

Model performance and selected land use variables vary with temperature metric (Table 2). Land use regression models have the best performance for the monthly averages of daily maximum temperature ($R^2=0.47$), followed by the models for minimum and mean temperatures ($R^2=0.42$ and 0.30 respectively). Distance to the coastline is the only variable appearing in all models. Lower monthly averaged daily maximum or mean temperatures are associated with shorter distance to the coast, more water and wetlands coverage. However, monthly averaged daily minimum temperatures is negatively associated with distance to the coast, but positively associated with variables representing open water, imperviousness and high developed intensity areas. Interestingly, daily maximum temperatures at a location could be lower 0.91 °C or 0.73 °C on average with one more percent of herbaceous wetlands within 750 m buffer of this location or open water area within 400 m buffer. Additionally, partial $R^2$ values (Table 2) illustrate that distance to the coastline is far more important than other predictors in predicting the monthly averages of daily highs and means; and high developed intensity area is the most important predictor for the monthly averages of daily lows, followed by distance to the coastline and others.

3.3. Quantile regression modeling results

The averages of covariates' coefficients ranging from the 10th quantile to the 90th quantile called mean effect are generally similar to the estimated coefficients from land use regression models (Table 2). Figs. 2–4 show coefficients varying with quantiles for daily maximum, mean and minimum temperatures. According to formally integral tests, distance to the coastline and high developed area have larger impacts on daily mean temperatures at higher quantiles, and open water area has greater impacts on daily maximum temperatures at lower quantiles. Interestingly, the highest monthly averages of daily highs are not observed at those weather stations in the downtown area as expected, although an increasing trend can be seen from the coastline to the inland shown in Fig. 1. This may be explained by some microclimate factors which are not captured by land use variables used in this study, e.g., nearby housing types (Harlan et al., 2006) and distance between a weather station to the nearest driveway or vegetation area or water bodies (Zhang et al., 2011).

Partial $R^2$ values classify distance to the coastline as the most important variable for daily maximum and mean temperatures, but the highly developed area for daily minimum temperatures. This finding suggests that the local-specific feature (distance to the coast) has the strongest impacts on daily highs and means while the microclimate factor (high built-up density) has the largest impacts on daily lows. Daily minimum temperatures have been suggested by a few studies to be more related to heat-related health effects than daily mean and maximum temperatures (Kloog et al., 2012).

The associations between temperature metric and land use variables can be explained by the physical properties of those land use types. Built environment absorbs more heat than vegetation

### Table 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Land use regression</th>
<th>Quantile regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum temperature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to the coastline</td>
<td>0.04 (0.03, 0.05)</td>
<td>0.394</td>
</tr>
<tr>
<td>Open water area within 400 m buffer</td>
<td>−0.73 (−1.36, −0.10)</td>
<td>0.055</td>
</tr>
<tr>
<td>Herbaceous wetlands within 750 m buffer</td>
<td>−0.91 (−1.49, −0.33)</td>
<td>0.021</td>
</tr>
<tr>
<td>Average temperature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to the coastline</td>
<td>0.01 (0.01, 0.02)</td>
<td>0.166</td>
</tr>
<tr>
<td>Developed high intensity within 750 m buffer</td>
<td>0.04 (0.02, 0.07)</td>
<td>0.073</td>
</tr>
<tr>
<td>Woody wetlands within 400 m buffer</td>
<td>−0.32 (−0.60, −0.03)</td>
<td>0.065</td>
</tr>
<tr>
<td>Minimum temperature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to the coastline</td>
<td>−0.01 (−0.02, −0.01)</td>
<td>0.089</td>
</tr>
<tr>
<td>Open water area within 400 m buffer</td>
<td>0.52 (0.15, 0.89)</td>
<td>0.027</td>
</tr>
<tr>
<td>Developed high intensity within 750 m buffer</td>
<td>0.07 (0.04, 0.11)</td>
<td>0.285</td>
</tr>
<tr>
<td>Impervious surface within 750 m buffer</td>
<td>0.01 (0.001, 0.02)</td>
<td>0.022</td>
</tr>
</tbody>
</table>

* Estimated mean effect refers to the averages of estimated covariate effect across all quantiles (0.1—0.9).
and water bodies during the day, and releases more heat at night, resulting in elevated temperatures at high developed areas and decreased temperature at those areas with more vegetation (Zhang et al., 2011). The sea breeze from the coastline cools those locations close to the coast in the daytime, and warms nearby inland areas at night. Thus, shorter distances to the coastline and more water bodies have negative impacts on daily highs and positive impacts on daily lows.

Larger impacts of high developed area on daily minimum temperature at higher quantiles suggest that built-up environment intensifies the temperatures at the areas with elevated temperatures during the 2011 heat wave. This finding has significant implication for local health departments to implement public health interventions more effectively to reduce heat-related health effects. When a heat wave comes, ‘hot spots’ become much hotter than other areas, which may exacerbate health effects of a heat wave at those ‘hot spots’. Therefore, considerable attention may be paid to these areas, particularly on vulnerable populations living in those areas.

Our models can provide estimates of heat exposures at fine resolution for epidemiological studies and vulnerability mapping. Some studies have examined the associations between land

Fig. 2. Quantile regression estimates and 95% confidence intervals with the monthly averages of daily maximum temperature as the response variable (solid lines stand for estimated coefficients, dashed lines stand for 95% confidence intervals, and dotted lines stand for coefficients equal to 0 (A) Distance to the coastline, (B) Open water area within 400 m buffer, (C) Herbaceous wetlands within 750 m buffer.

Fig. 3. Quantile regression estimates and 95% confidence intervals with the monthly averages of daily mean temperature as the response variable (A) Distance to the coastline, (B) Woody wetlands within 400 m buffer, (C) Developed high intensity within 750 m buffer. Otherwise as Fig. 2.
surface temperature and heat-related deaths, e.g., Johnson et al. (2009) demonstrated that remote sensing estimates of land surface temperature were associated with increased risk of heat-related deaths during the 1993 Philadelphia heat wave. Our developed models based on ground-based air temperature measurements have a few advantages over land surface temperature estimates derived from satellites images. First, the interpretation of land surface temperature has some difficulties because the relationships between ground-based air temperature measurements and land surface temperature are complex and affected by many factors, e.g., cloud cover, wind, time of a day, urban geometry, land use/cover characteristics (Johnson et al., 2009; White-Newsome and Brines, 2013). Second, satellites only provide snapshots of thermal data during a day in an area (White-Newsome and Brines, 2013), and spatial resolution of land surface temperature can affect results (Johnson et al., 2009). In addition, temperature predictions generated from our models can serve as a surrogate of physical heat exposures, which can be combined with social-demographic variables, air conditioning and diseases data to identify where people vulnerable to heat live (Reid et al., 2011).

Our findings provide important insights for weather forecasters into their decisions on issuing heat alerts. A few metrics have been used to trigger heat wave and health warning systems and release heat alerts, however, there is no consensus on the 'best' trigger metric (Zhang et al., 2012; Zhang et al., 2014b). In addition, local weather forecasters typically make a decision on issuing heat alerts based not only weather forecasts, but also their local knowledge and other factors (Zhang et al., 2014c). Forecasts from numerical models usually have coarse spatial resolution ranging from tens of kilometers to hundreds of kilometers, depending on numerical model (Zhang et al., 2014c). The findings presented in this paper can add the knowledge of temperatures at fine spatial resolution into forecasters’ local knowledge.

This study has a few strengths. First, this study focuses on a recent 2011 heat wave in Greater Houston, one of the largest Metropolitan areas in US with around 6 million people (US Census Bureau, 2013b). Second, this study explores all available land use/land cover variables and local-specific geographic variable with a range of buffer radii while previous studies usually examined a few preselected land use variables (Kloog et al., 2012; Oswald et al., 2012; Zhang et al., 2011). Third, this study applies two new techniques in investigating effects of land use on temperature: land use regression and quantile regression. In particular, quantile regression allows us to address potential changing effects of land use on temperatures.

This study has several limitations. First, this study only uses available land use types from National Land Cover Database, and does not include detailed environment information around weather stations, e.g., urban morphology and nearby buildings (Zhang et al., 2011). Second, this study does not account for spatial correlations among weather stations explicitly. However, spatially-related land use variables can capture spatial correlation to some degree (Zhang et al., 2014a). Also, how to include spatial correlations in quantile regression is still an open research topic.

5. Conclusion

Advanced knowledge of the effects of land use on temperatures during heat waves is critical to understand heat exposures at ‘hot spots’ and thus better characterize heat stress of people particularly vulnerable populations in urban areas. A novel analysis was conducted to investigate effects of land use on temperatures during the 2011 heat wave in Greater Houston using land use regression and quantile regression based on satellite-derived fine-scale land use information. Our investigation shows statistically significant effects of distance to the coastline, open water, built environment and vegetation on temperature observations. Our quantile regression analysis demonstrates larger impacts of distance to the coastline and highly developed intensity area at ‘hot spots’ compared to non-hot spots. The methodology and findings presented in this paper are relevant to the research related to urban microclimate, heat-related exposure and epidemiology studies.
Conflict of interest

The authors declare they have no competing financial interests.

Acknowledgments

The research described in this paper was funded through the start-up funds from the University of Texas School of Public Health (UTSPH). This paper does not necessarily reflect the views of UTSPH.

References


