A geographically-aware multilevel analysis on the association between atmospheric temperature and the “Emergency and transitional shelter population”

Carlos Siordia\textsuperscript{a*}, Curtis Smith\textsuperscript{b}, Ernesto Castañeda-Tinoco\textsuperscript{b}

\textsuperscript{a} University of Pittsburgh, USA \\
\textsuperscript{b} University of Texas at El Paso, USA

Understanding the geographical distribution and correlates of special segments of the population has the potential for offering insight into human behavior. Our study examines the Emergency and Transitional Shelter Population (ETSP)—which includes what are commonly referred to as “homeless” people. We use 2010 data from two sources: United States (US) Census Bureau county-level ETSP estimates; and North America Land Data Assimilation System Phase 2 (NLDAS-2). We investigate the ecological correlates of ETSP concentration by using a geographically-aware multilevel linear model. The specific aim is to investigate if an how atmospheric temperature is related with ETSP concentration by county—after accounting for population density and percent non-Hispanic-White. We use ArcGIS® 10.1 to create a spatial weight matrix of the ten most proximal counties and use SAS® 9.3 to create an algorithm that estimates County Cluster Dyadic Averages (CCDAs). By nesting the 31,090 CCDAs over the 3,109 counties in the continental US, we find a positive and statistically significant relationship between ETSP density and atmospheric temperature. Ecological studies should continue to explore the spatial heterogeneity of the ETSP.

Key Words: homeless, ETSP; multilevel, ecological, weather, spatial, mobile populations.

Article Info: Received: July 23, 2014; Accepted: October 24, 2014; Online: November 30, 2014.

Introduction

The growing number of “homelessness” in the United States (US) was recognized as a serious issue decades ago. Researchers have tried since then to use

* Corresponding author
Address: 130 North Bellefield Ave, Pittsburgh, PA 15213, USA.
Phone: 1-412-383-1708 | Email: cas271@pitt.edu

©2014 Human Geographies; The authors DOI:10.5719/hgeo.2014.82.5
scientifically acceptable methods for estimating the size, characteristics, and geographical distribution of their population (Rossi, et. al., 1987). In the US, the federal government makes extensive efforts to develop information on the demographic and geographic profile of the Emergency and Transitional Shelter Population (ETSP)—which includes what are commonly referred to as “homeless” individuals. The acquisition of data on the ETSP (which includes a portion of the population experiencing homelessness) provides valuable information to federal and local agencies charged with social welfare programs. For example, in fiscal year 2009, the US federal government appropriated $4.2 billion for homeless assistance programs (NAEH, 2009). In 2013, a proposed $4.7 billion budget would constitute a 17% increase on the 2012 fiscal year budget (US Interagency Council on Homelessness, 2012).

While it may be safe to assume that the migratory behavior of individuals in the ETSP may be influenced by various push and pull factors, no study has examined the ecological correlates of their density over the continental US. The US federal government has developed county-level estimates on the concentration of the ETSP. A “county” is a political subdivision within “states” with clear geographical boundaries. The specific aim of our ecological study is to investigate if the ETSP, at the county-level, is distributed over the US mainland as a function of: atmospheric temperature (i.e., weather); population density; and race-ethnic composition at the county-level.

Decades ago, social scientists sought to describe the condition of individuals who lost a stable physical residence (Anderson, 1961) and experienced a “winter of despair” (Snow & Anderson, 1993). Since then, the plea of the homeless has been given much research attention (Passaro, 1996; Jenck, 1995). The counting of the ETSP by the US Census Bureau is important because it directly affects the programs and funding for homeless services (Gabbard et al., 2007). Gathering demographic information and geographic distribution of the ETSP is complicated by the fact that defining the population has shifted over the decades (Voeten 2010; Donley & Wright, 2012). Debates around the use of technical labels to describe the ETSP are ongoing (Wright, 2009). Defining “homelessness” remains a core issue because the conceptual definitions influence data gathering procedures (Wiegand, 1986; Cordray & Pion, 1991; Wiliams, 2011).

In 1984, Housing and Urban Development (HUD 2014) made an extensive effort to estimate the homeless population (Gabbard et al., 2007). The lack of an adequate definition of homelessness was center stage in the 1990 decennial count when the US Census Bureau made efforts to collect information on the nation’s homeless population—which would later be used to produce the Shelter and Street Night (S-Night) dataset (Wright & Devine, 1992; Link et al., 1994; Hewitt, 1996; ). National statistics on the ETSP are gathered by the US Census Bureau and local statistics by HUD (NAEH, 2010). Compare to counting unsheltered/homeless individuals, numbering the population who actually uses shelters is more reliable. Hence, the US Census Bureau decided to enumerate the ETSP instead of the “homeless” population. Our study uses the most recent county-level data on the ETSP.

From an ecological perspective, it may be argued that selecting which place to inhabit is affected by various factors (Garasky, 2002; Clark 2013). For example, amongst individuals who make the unsheltered outdoor their residence, the
weather may be an important consideration. It could be that areas with warm weather make for more comfortable sleeping during the night but harsher days during the summer. In contrast, areas with cold weather could make for harsh nights and softer days during the summer. The ETSP database shows that that homeless density is high in the warm-weather states of: California (13.2%); Florida (6.1%); and Texas (5.2%) (Smith et al., 2012). Some studies have shown that areas with higher temperatures can have higher rates of homelessness (Appelbaum et al., 1991; Grimes & Chressanthis, 1997; Quigley, 1990; Raphael, 2010). Another weather-related study observed homeless people did not make hospital visits more frequently during cold weather (Brown, Goodacre, & Croos 2010). However, a small study concluded good weather in Boston, did not act as a magnet for homeless males (Gray et al., 2011). The literature may be said to be inconclusive with regards to the ecological association between atmospheric temperature and the geographical distribution of the ETSP.

Despite the large literature on the homeless population, no study has investigated if the ETSP density varies as a function of outdoor weather conditions. Our study will help inform literature attempting to better understand the ecological correlates of the ETSP.

**METHODS**

*Data*

County level counts of the ETSP, total population, and number of non-Hispanic-Whites come from Summary File 1 (SF1) data in table form through the American Fact Finder. In 2010, individuals who experienced homelessness in overnight stay locations (e.g., shelters, missions, motels, etc.) constitute the ETSP. Their numbers were enumerated as part of the Service Based Enumeration (SBE) operation which included soup kitchens, targeted non-sheltered outdoor locations, and regularly scheduled mobile food vans.

The US Census Bureau designated March 29, 2010 as the single day to enumerate the ETSP population at emergency and transitional shelters. They designated March 3, 2010 as the single day to enumerate ETSP counts at soup kitchens and regular mobile food vans. They then used March 31, 2010 as the only day where they collected ETSP information from targeted non-sheltered outdoor locations. SBE service providers were given the option to be enumerated on any of the 3 days outlined above. The US Census Bureau has been clear that their ETSP count should not be misconstrued as a count of the entire homeless population. In 2010, the US Census Bureau counted 209,000 individuals in the ETSP.

*ETSP Concentration*

The homeless population is not distributed evenly across the US (Lee & Price, 2004). Our empirical models predict ETSP concentration per-10,000 people (ETSP10,000). We compute it as follows: \[(\text{ETSP count} \div \text{total population})\times10,000\]. In our analysis, ETSP10,000 is the dependent variable. The count
of ETSP per-10,000 is done at the county-level (a clearly defined and administrative geography in the US).

**Atmospheric Temperature**

Atmospheric temperature is our independent variable. Temperature refers to the average, in Fahrenheit degrees (°F), measured across the whole year in each county—where the daily measure is the average F° of readings through the day. Using yearly averages limits our models as the measures do not allow us to capture how within-year temperature variability may impact ETSP mobility. If, with this very crude measure, it turns out that ETSP density varies by atmospheric temperature, then the relationship may be much stronger than that depicted in our equation results.

County-level average atmospheric temperature was estimated using data gathered by the North America Land Data Assimilation System (NLDAS) Phase 2 from 2010. The NLDAS is a quantitative system driven by ‘atmospheric forcing’ that covers the continental US (Ek et al., 2011). The National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction/Environmental Modeling Center (NCEP/EMC) in collaboration with others developed NLDAS Phase 2 (Ek et al., 2011). NLDAS data calculates retrospective and real-time water (e.g., soil moisture) and energy (e.g., surface fluxes) cycles to help monitor important weather events (e.g., climate forecasting and drought monitoring) (Ek et al., 2011). The retrospective component of NLDAS Phase 2 (NLDAS-2) completed in 2010 is used in this study to measure the “Average Daily Minimum Air Temperature” in °F.

**Population Density**

Our regression models control for population density. We compute county-level population density as follows: (total population ÷ county square mile area). We content that as population density increases, resources (e.g., services, food, and money) will increase for ETSP residents. Percent Non-Hispanic-White

Our models also control for percent non-Hispanic-White. We calculated county-level percent non-Hispanic-White as follows: [(count of non-Hispanic-Whites ÷ total population*100)]. “Non-Hispanic-Whites” are the race-majority group in the US. In our study, we follow existing practices by using percent of non-Hispanic-White as a proxy measure for area-level SES (Siordi & Farias, 2013) and assume there will be decrease in the tolerance of ETSP as percent non-Hispanic-White increases (Iceland & Sharp, 2013; Siordia & Leyser-Whalen, 2014).

**Geographically-Aware Multilevel Modeling**

Presumably, the ETSP is not distributed over geographical space at random. If we assume that members of the ETSP choose (or are forced into) their area of residence as a function of how their compositional characteristic (i.e., person-level factors) interact with environmental attributes (i.e., place-level factors);
then this population must be analyzed with a geographically-aware conceptual prism. The zero-inflated distribution of ETSP (i.e., large number of counties with no ETSP individuals) prohibits the geospatial modeling of counties with more straightforward techniques such as geographically weighted regressions (Siordia, Saenz, Tom, 2012). Thus we employ a novel approach to modeling the ecological correlates of ETSP10,000 (i.e., ETSP density) in the US.

In order to account for spatial non-stationarity, the idea that statistical relationships can vary as a function of geographical location (Siordia, 2013), we undertake a novel procedure to develop what we call “county cluster dyadic averages” (CCDAs). We created CCDAs as follows: using ArcGIS 10.1®, we identified the “10 most proximal county neighbors” for each county; a spatial weight matrix output (in SWM format which is transformed to a 'DBF' file using ArcGIS 10.1® software) is imported into SAS 9.3®; where data was managed and an algorithm was created to compute the average score between the “anchor county” and each of its 10 most proximal county neighbors within the CCDA. Our approach produces a total of 31,090 CCDAs for the 3,109 single counties found in the continental US.

We illustrate the methodology with an example. Assume Harris County (the host county for Houston, Texas, US) is the “anchor county.” Because Harris County is the polygon on which the county cluster is identified, it is referred to as the anchor county. Our procedure identifies the 10 most geographically proximal counties around Harris County (using the polygon intersect procedure in ArcGIS 10.1®). The Harris County cluster would then contain Harris county and another 10 most geographically proximal counties. Our algorithm then averages the ETSP between Harris County (i.e., the anchor county) and “county #1” in the cluster. It does this for all of 10 counties in the cluster and for each of the measures in the models (i.e., ESTP10,000, temperature, population density, and percent non-Hispanic White). Thus, for the Harris County cluster, we end up with variable measures that combine the anchor county and one of its neighbors—hence the use of the “dyadic averages” labeling in CCDAs.

**Statistical Approach**

In classical multivariate linear modeling, there is an assumption that observations are independent of one another (i.e., no between-unit dependence). The county clusters we have created used dyadic averages where the anchor county is present in all dyads. Hence, there is between-unit dependence within county clusters. In order to properly model their statistical relationships, we must account for this “within county cluster dependence.” To do this, we use a hierarchical (interchangeable referred to as multilevel) model where we compute an error structure for within-county clusters and a separate error term for between-county clusters.

The dyadic averages in county clusters are used in a multilevel linear model that nests the 31,090 CCDAs by geographically linking them to the anchor county. Extending the example above, the empirical model uses the 10 dyadic averages around Harris County at “Level-1” (CCDA-level) and identifies them as belonging to Harris County at “Level-2” (county-level). The multilevel equation computes the variance explained and error term for Level-1 separately.
from that of Level-2—this accounts for the within-county cluster dependence that would be otherwise violated in a non-hierarchical model.

Our CCDA approach is capable of capturing potential spatial non-stationary processes that may be affecting ETSP10,000. The county cluster, we assume, has the potential to identify micro-regions where social processes diffuse (as if in a kernel function) in ways that may affect the person-level decisions within the ETSP as they decide where to make their residence. For example, an individual staying in a shelter may seek areas where multiple locations for food and night lodging are available over a wide geographical range. Our county clusters may have the ability to capture the geographical space where the services for ETSP are denser. By arbitrarily choosing the 10 neighbor threshold, we are imposing an empirically untested assumption that this geographically varying square mileage contains a space where social and physical factors affect the “pull” or “push” of the ESTP. We return to this limitation in closing.

Our inclusion of atmospheric temperature, population density, and percent non-Hispanic-White is informed by the literature. Our approach assumes the three county-level measures will be statistically related with the concentration of the ETSP. Although we show interesting ecological correlations, we recommend these assumptions be investigated with person-level data. Our findings may not be used to infer that ecological relationships explain micro-level events—e.g., that atmospheric temperature, population density, and percent non-Hispanic-White jointly influence how each member of the ETSP determines his or her location over the geography.

RESULTS

From Table 1, descriptive statistics, we see the average ETSP10,000 is 3.17—over all the county dyads, the average number of ETSP individuals per every 10,000 people in the dyad-county-population was 3.17. This number may be skewed because of a few (n=34) counties where ETSP10,000 > 20. The most extreme case is Bronx County, New York, US (ETSP10,000=65) and Roanoke City, Virginia, US (ETSP10,000=63). From Table 1 we also see that the average daily minimum air temperature is about 46.72, population density is 261, and the average concentration of non-Hispanic Whites is about 83%.

Before moving on to the multilevel equation results, we highlight the geographical distribution of atmospheric temperature in Figure 1. Colder

<table>
<thead>
<tr>
<th>Table 1. Descriptive statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Units (n=31,090)</td>
</tr>
<tr>
<td>M (^1)</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Emergency and Transitional Shelter Population density</td>
</tr>
<tr>
<td>Average daily minimum air temperature F°</td>
</tr>
<tr>
<td>Population density per-square mile</td>
</tr>
<tr>
<td>Percent Non-Hispanic-White</td>
</tr>
</tbody>
</table>

\(^1\) Mean; \(^2\) Standard deviation; \(^3\) Minimum; \(^4\) Maximum
temperatures are on average more present in the north and diffuse as one moves southbound towards Texas and Florida. We also provide a map showing the amount of ETSP individuals for every 10,000 people. The ETSP10,000 is shows in Figure 2, where we see high concentrations at the county level are located in mid- to northern-areas of the US. A spatial clustering analysis (available upon request from corresponding author) did not reveal any stable patterns of ETSP concentrations.

Before we conducted the full-model, we first executed a simple two-level random intercept model with ETSP10,000 as the dependent variable. At Level-1 (where CCDAs are located), our equation is:

$$ETSP_{ij} = \beta_{0j} + r_{ij},$$

and at Level-2 (where counties are located), our equation is:

$$\beta_{0j} = \gamma_{00} + u_{0}.$$

From this intercept-only model, we use the outputs to calculate between county intra-class correlation (ICC) as follows: ICC = $\tau_{00} / (\tau_{00} + \sigma^2)$. From our HLM® 6.04 (Raudenbush et al., 2004) outputs, $\tau_{00} = 14.28$ (P-Value = 0.000) and a $\sigma^2 = 11.28$, so that: 14.28 / (14.28 + 11.28) = 0.55. This means 55% of the variance in ETSP10,000 can be explained from between-county cluster factors (Hox, 1995). Since ICC measures the proportions of variance in ETSP10,000 between-
Figure 2. The amount of ETSP individuals for every 10,000 people

county clusters, we could infer that about half of the variance in ETSP10,000 can be explained at the county cluster level. Since our \( \tau_0 \) is statistically significant and the ICC is greater than zero, we conclude a multilevel model is useful (Raudenbush & Bryk, 2002). After justifying the use of the multilevel linear model (with alpha level at 0.05), we specified the following “full equation” at Level-1:

\[
ETSP_{ij} = \beta_0 + \beta_1(F^*)_{ij} + \beta_2(\text{PopulationDensity})_{ij} + \beta_3(\text{PercentNHW})_{ij} + r_{ij}
\]

Where ETSP\(_{ij}\) represents the predicted ETSP10,000; \( i \) and \( j \) refer to the ith dyadic average for jth anchor county; \( \beta_0 \) is the intercept for jth anchor county; \( \beta_1 \) through \( \beta_3 \) are the three average slopes for the CCDAs, for jth anchor county; and \( r_{ij} \) is the error term for the ith dyadic average for jth anchor county. At Level-2, our equation is:

\[
\begin{align*}
\beta_0 &= \gamma_0 + u_0 \\
\beta_1 &= \gamma_1 \\
\beta_2 &= \gamma_2 \\
\beta_3 &= \gamma_3
\end{align*}
\]

where \( \gamma_0 \) is the main intercept of the full model; and \( u_0 \) is the error measurement for all 3,109 intercepts. No other \( \tau \)s (i.e., \( u_0 \)) are included because we assumed cross-level interactions to be functioning similarly across all county clusters.
Table 2. Multilevel linear model predicting average concentration of Emergency and Transitional Shelter Population

<table>
<thead>
<tr>
<th></th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.174</td>
</tr>
<tr>
<td>Average daily minimum air temperature (^\circ)C</td>
<td>0.134</td>
</tr>
<tr>
<td>Population density per-square mile</td>
<td>0.001</td>
</tr>
<tr>
<td>Percent Non-Hispanic-White</td>
<td>-0.133</td>
</tr>
</tbody>
</table>

**Note:** *** < 0.0001; \(^\bullet\)*Coefficient

Our multilevel linear models execute a “global regression” using the 31,090 CCDAs as the level-1 units and the 3,109 counties as the level-2 units. Because a portion of the County Cluster Dyadic Averages (CCDAs) may be cross-nested (i.e., be linked to more than one county), the Gauss-Markov independence assumption may be violated. In closing, alternate approaches are discussed for future work.

Table 2 shows the results of our multilevel linear model predicting ETSP10,000. We see, as we expected, that as atmospheric temperature increases, ETSP10,000 increases (\(\beta=0.134\); \(\alpha < 0.000\)). We also see that as population density increases, ETSP10,000 increases (\(\beta=0.001\); \(\alpha < 0.000\)) and that a percent non-Hispanic White increases, ETSP10,000 decreases (\(\beta=-0.133\); \(\alpha < 0.000\)). In simpler terms, we find that atmospheric temperature is related with ETSP concentrations—after accounting for county-level population density and percent non-Hispanic-White.

CONCLUSIONS

Although our ecological study clearly indicates a statistical association between atmospheric temperature and the ETSP, it does have some limitations. Foremost is the fact that although our hierarchical model using CCDAs does incorporate a geographically-aware approach, it does not do so in a way that weights regression parameters as a function of geographical location—as is the case with geographically weighted regressions (Fotheringham et al., 1996; 2007). Future work should explore how a basic zero inflated spatial regression model (Rathbun, 2006) could be estimated.

Secondly, the 10 neighbor threshold was made arbitrarily and primarily driven by the need to provide our multilevel equation with enough Level-1 units so as to produce reliable parameter estimations. Because the Modifiable Areal Unit Problem (MAUP: Fotheringham & Wong, 1991; Openshaw, 1983) remains an unresolved issue, future work should explore alternate neighbor thresholds (e.g., 5 or 15 CCDAs). Because a portion of CCDAs may be linked to more than one county (are cross-classified in multiple counties), future work should explore the same data sources using a cross-classified multi-level model approach to account for the dependence between some of the level-2 clusters (see Arcaya & Subramanian, 2014; Goldstein, 2011). Another important issue to study in future research is how other components of weather (e.g., rain, snow)
are correlated with the geographical distribution of the ETSP in the US mainland and in other nations. Notwithstanding the limitations, our project clearly shows evidence that atmospheric temperature is associated with ETSP10,000 at the county-level in the continental US. Because estimating the correlates of the ETSP density may help inform policy seeking to aid an underserved population, studies should continue to explore the geo-spatial heterogeneity and ecological correlates of the ETSP. The paper helps fill a gap in the literature on the geography of homelessness by being the first to combine a geographically-aware multi-level model that investigates how atmospheric temperature is ecologically correlated with the concentration of the ESTP in the continental US—while accounting for population density and percent non-Hispanic-White.

**Funding**

CS is supported by the National Institutes of Health (NIH Grant: T32 AG000181 to Dr. Newman).

**References**


Fotheringham, AS & Wong, DW 1991, The modifiable areal unit problem in multivariate


Goldstein, H 2011, Multilevel statistical models, 4th edn, Wiley.


NAEH: The National Alliance to End Homelessness 2009, Federal Funding for Homeless Assistance Programs, Including for Youth, National Alliance to End Homelessness, Washington, DC.

NAEH: The National Alliance to End Homelessness 2010, What is a Point-in-Time Count?, The National Alliance to End Homelessness, Washington, DC.


Siordia, C 2013, Sociospatial Inequality: Combining Multilevel and Spatial Analysis, *Journal of Maps*, vol. 9, no. 1, pp. 50-54.


