Supplementary Materials for: *Analyzing an Incomplete Marked Point Pattern of Heat-Related 911 Calls*

August 29, 2014

1. WEIGHTING MEASURE

The weighting measure $w(\mathbf{z})$, intuitively, should reflect a “null hypothesis” of no change in 911 call volume from the population characteristics of Houston. For purposes of this study, the weighting measure is given by,

$$w(\mathbf{z}) = w(\mathbf{s}, y, d, h, a, g, e, C) = \frac{1}{|\mathcal{T}|} \frac{1}{|\mathcal{G}|} w_{\mathcal{S}}(\mathbf{s} \mid e) w_{\mathcal{C}}(C \mid d) w_{\mathcal{A}}(a) w_{\mathcal{E}}(e) \quad (1)$$

where $|\mathcal{T}|$ and $|\mathcal{G}|$ denote the cardinality of the sets $\mathcal{T}$ and $\mathcal{G}$. The measures $w_{\mathcal{S}}(\mathbf{s} \mid e)$, $w_{\mathcal{C}}(C \mid d)$, $w_{\mathcal{A}}(a)$, and $w_{\mathcal{E}}(e)$ define weights for the spatial location $\mathbf{s}$, temperature C , age a , and race/ethnicity e , respectively, and will be defined in further detail below. Uniform weights are assigned to all years (y), days of the year (d), hours of the day (h) and genders (g). The uniform weighting measure for gender adequately reflects the distribution of genders within Houston which, according to the 2010 U.S. Census was 50.2% male and 49.8% female. Equal weights to all years, days of the year and hours of the day were chosen to reflect our “null hypothesis” that 911 calls for dehydration can occur at any time. The weighting measure for each ethnicity $w_{\mathcal{S}}(\mathbf{s} \mid e)$ is shown in Figure 1 and denotes the number of people (in thousands) of ethnicity e located in the census block group in which $\mathbf{s}$ resides.

The weight measure $w_{\mathcal{C}}(C \mid d)$ is chosen to reflect the seasonal temperature changes in Houston. Obtaining data from the National Climatic Data Center (NCDC), $w_{\mathcal{C}}(C \mid d)$ is set be the

Figure 1. Weighting measure $w_S(s \mid e)$ for (a) Caucasian, (b) African American, (c) Hispanic and (d) Asian populations.

Figure 2. Weighting measure $w_C(C | d)$ across the days of the year.

height of a Gaussian density function with mean equal to the 30-year average, average temperature on day d and standard deviation equal to $1/2$ the 30-year average diurnal temperature range (i.e. $\mathcal{N}(C; \mu_d, \sigma_d^2)$). The temperature weighting measure is shown in Figure 2.

The weight measure for age and race/ethnicity are obtained from the 2010 U.S. Census. Specifically, $w_A(a)$ is taken to be the proportion of the population within 5-year age ranges up to age 85 (i.e. the proportion of the population aged less than 5 years, 5-9 years, 10-14 years, ..., 80-84 years, over 85 years). In this sense, $w_A(a)$ is a step function with 5-year step lengths. The race/ethnicity measure $w_E(e)$ is taken to be the proportion of the population with race/ethnicity $e \in \{\text{Caucasian, African American, Hispanic, Asian}\}$ and $w_E(e) = 0$ otherwise. We note that this excludes other ethnicities (e.g. American Indian) but because the 911 call data only included information on these 4 ethnicities, we can only make inferences for these groups. However, these ethnicities comprise approximately 99.3% of the population of Houston.

2. EXPLORATORY ANALYSIS OF THE 911 CALL DATA

In this Section, we justify our model specification for $\lambda^*(z)$ as,

$$\lambda^*(z) = \lambda^*(s, y, d, h, a, g, e, C) = \lambda_{\mathcal{HS}}(h, s \mid e) \lambda_{\mathcal{ADC}}(a, d, C \mid e) \lambda_{\mathcal{G}}(g \mid e) \lambda_{\mathcal{E}}(e) \quad (2)$$

based on an exploratory analysis. Recall, $\lambda_{\mathcal{HS}}(h, s \mid e)$ describes deviations of $\lambda(z)$ from the spatial location and hour of the day joint weighting measure $w_S(s \mid e)/|\mathcal{H}|$, $\lambda_{\mathcal{ADC}}(a, d, C \mid e)$ describes deviations from the age, day of the year, and temperature weighting measure $w_C(C \mid d)w_A(a)/|\mathcal{D}|$, $\lambda_{\mathcal{G}}(g \mid e)$ describes deviations from the gender weighting measure $1/|\mathcal{G}|$ and $\lambda_{\mathcal{E}}(e)$ describes deviations from the marginal ethnicity weight measure $w_{\mathcal{E}}(e)$. We assume that each of $\lambda_{\mathcal{HS}}$, $\lambda_{\mathcal{ADC}}$, $\lambda_{\mathcal{G}}$ and $\lambda_{\mathcal{E}}$ are densities with respect to their corresponding weight measure. For example, $\int_{\mathcal{G}} \lambda_{\mathcal{G}}(g \mid e) |\mathcal{G}|^{-1} dg = 1$ so as to maintain the interpretation of the δ parameter discussed above.

Over the period 2006-2010, there were 407, 384, 414, 419 and 430 calls in year 2006-2010, respectively. The relatively constant number of calls between years, coupled with the fact that year-to-year trends are not of primary interest, we assume $\lambda^*(z)$ in (2) is not a function of y . In other words, we set $\lambda^*(z) = \lambda^*(s, d, h, a, g, e, C)$. Under this assumption, the overall intensity measure $\lambda(z) = \delta \lambda^*(z) w(z)$ does not deviate from the year weight measure $w_{\mathcal{Y}}(y) = 1/|\mathcal{Y}|$ and is constant with respect to year.

Figure 3 displays a kernel density estimate of calls by ethnicity. Obviously, from Figure 3, ethnicity plays a major role in determining the location of the 911 calls; yet, the patterns seen in Figure 3 do not exactly mimic the patterns seen in the weighting measure in Figure 1. Similar plots (not shown) also display the difference in the distribution of marks (e.g. age) by ethnicity. Hence, we allow the intensity surface of every other mark to vary by ethnicity. We factor $\lambda^*(s, d, h, a, g, e, C)$ as the product of a conditional and marginal intensity such that,

$$\lambda^*(s, d, h, a, g, e, C) = \lambda_1^*(s, d, h, a, g, C \mid e) \lambda_{\mathcal{E}}(e). \quad (3)$$

Under (3), the intensity of 911 calls at different spatial locations, temporal locations, and attributes varies by ethnicity (note that λ_1^* is conditioned on e in (3)) and $\lambda_{\mathcal{E}}(e)$ controls the deviations from

Figure 3. Kernel density estimate of the locations of the 911 calls for (a) Caucasian (b) African American (c) Hispanic and (d) Asian ethnicities. Location of the 911 calls vary by ethnicity suggesting a different intensity surface for each race.

the ethnicity weighting measure $w_{\mathcal{E}}(e)$.

To make the modeling simpler and feasible, we also need to factor $\lambda_1^*(s, d, h, a, g, C \mid e)$ into smaller pieces. However, determining how to factor $\lambda_1^*(s, h, d, h, a, g, C \mid e)$ needs to be supported by the data. Conditional on ethnicity, the temperature at the time of the 911 call should not vary by spatial location. Indeed, a plot of the temperature by spatial location of the 911 call (not shown for confidentiality reasons) suggests that, conditional on ethnicity, temperatures are seemingly evenly distributed over the spatial region. Hence, the data support modeling s and C separately.

Figure 4 displays the observed number of calls for African Americans by day of the year and temperature minus the “expected” number of calls according to the associated weight measure $w_C(C \mid d)$ shown in Figure 2. Note that, while the weighting measure captures the overall seasonal pattern, temperature and day of the year are not separable as there is a clear (non-linear)

Figure 4. Observed number of calls for African Americans by day of the year and temperature minus the “expected” number of calls according to the associated weight measure $w_C(C \mid d)$. Temperature is clearly not separable with respect to day of the year.

relationship between temperature at the time of the 911 call and day of the year. Hence, C and d need to be modeled jointly. Figure 5 displays a scatterplot of age of the patient by temperature at the time of the 911 call with fitted regression line overlaid. The slope was significant at the 0.05 level suggesting that (a, d, C) should be non-separable (modeled jointly). This finding agrees with several previous studies that have found that heat affects the elderly more than younger individuals (see, e.g. Luber and McGeekin 2008, Wilhelmi and Hayden 2010).

Figure 6 displays the observed minus expected (according to the weighting measure w) male and female callers by temperature. Because both male and female patterns are similar across temperature, we can assume that gender is separable of temperature, conditional on ethnicity.

Considering that people change locations depending on the time of the day, we also assume that the spatial location of 911 calls will vary by hour of the day. For example, people may be at work during the day yet in their residence at night suggesting that 911 calls during daylight hours will tend to occur downtown whereas 911 calls will tend to occur in residential areas at night. This

Figure 5. Scatterplot of age of the patient by temperature with fitted smooth loess curve.

Figure 6. Observed minus expected counts for the number of male (solid line) and female (dashed line) callers by temperature.

Figure 7. Kernel density estimates for (a) location of the 911 calls between 12-1am and (b) the location of the 911 calls between 12-1pm. Solid circle denote observed location. Because the density estimates vary by space, we treat spatial location and hour of the day as non-separable.

change in spatial location by hour of the day was found in the 911 calls and is displayed in Figure 7.

Assuming separability of (h, s) and (a, d, C) , conditional on ethnicity, was made to prevent over-fitting. For example, if d (day of the year) and h (hour of the day) were assumed to be non-separable, then the intensity surface would be different for each hour of the year. Given only 2054 calls were observed over the 5-year time span and there are 8760 hours in a year, assuming a non-separable structure for d and h would likely over-fit the data. Admittedly, assuming a separable form for C (temperature) and h (hour of the day) is an over-simplification because of the diurnal cycle of temperature. However, because the variability of C is largely explained by d over days of the year, we made the simplifying assumption of separability between C and h .

We summarize the main results of this exploratory analysis as follows. First, each mark seemed

to vary by ethnicity. And, second, conditional on ethnicity $(h, \mathbf{s})$, (a, d, C) and g can be modeled separately. Hence, we arrive at the form for the intensity surface given in (2).

3. DISCRETE CONVOLUTION MODELS FOR $\lambda_{\mathcal{HS}}$ AND $\lambda_{\mathcal{ADC}}$

We desire a computationally tractable, yet flexible modeling strategy for each of $\lambda_{\mathcal{HS}}$ and $\lambda_{\mathcal{ADC}}$. Here, we chose to rely on the discrete process convolution approach of Higdon (2002) to construct a Gaussian process model for $\lambda_{\mathcal{HS}}$ and $\lambda_{\mathcal{ADC}}$. Let,

$$\log(\lambda_{\mathcal{HS}}(\mathbf{s}, h \mid e)) = \sum_{k_s=1}^{K_S} G_2(\mathbf{s} \mid \mathbf{s}_{k_s}^*, \sigma_S^2 \mathbf{I}) \phi_{k_s h e} \quad (4)$$

where $\{\phi_{k_s h e}\}$ are a set of unknown coefficients and $\{G_2(\mathbf{s} \mid \mathbf{s}_{k_s}^*, \sigma_S^2 \mathbf{I}) \propto \mathcal{N}_2(\mathbf{s} \mid \mathbf{s}_{k_s}^*, \sigma_S^2 \mathbf{I}_2) : k_s = 1, \dots, K_S\}$ are a set of bivariate Gaussian kernel weights centered at spatial knot locations $\mathbf{s}_1^*, \dots, \mathbf{s}_{K_S}^*$ and normalized such that $\sum_{k_s} G_2(\mathbf{s} \mid \mathbf{s}_{k_s}^*, \sigma_S^2 \mathbf{I}) = 1$. By Equation (4), $\lambda_{\mathcal{HS}}$ is smooth with respect to spatial location ($\mathbf{s}$) but we also hypothesize that $\lambda_{\mathcal{HS}}$ is smooth with respect to hour of the day. To enforce smoothness with respect to h , let,

$$\phi_{k_s h e} = \sum_{k_h=1}^{K_{\mathcal{H}}} V_{24}(h \mid h_{k_h}^*, \sigma_{\mathcal{H}}^2) \eta_{k_s k_h e} \quad (5)$$

for $k_s = 1, \dots, K_S$ where

$$V_{24}(h \mid h_{k_h}^*, \sigma_{\mathcal{H}}^2) \propto \exp \left\{ \frac{1}{\sigma_{\mathcal{H}}^2} \cos \left(\frac{2\pi}{24} [h - h_{k_h}^*] \right) \right\}, \quad (6)$$

is the von-Mises kernel that, internally, maps $h \in [0, 23]$ to a corresponding angle on the unit circle, $h_1^*, \dots, h_{K_{\mathcal{H}}}^*$ are a set of hour knot locations and, again, $\sum_{k_h} V_{24}(h \mid h_{k_h}^*, \sigma_{\mathcal{H}}^2) = 1$. We emphasize that by using the von-Mises kernel, we appropriately account for the circular nature of hour of the day. We chose to use $K_S = 150$ and $K_{\mathcal{H}} = 10$ leading to 1500 coefficients which completely determine $\lambda_{\mathcal{HS}}(\mathbf{s}, h \mid e)$. The spatial and hourly knot locations were chosen according to a space-filling design over the spatial domain of Houston and the set $\{0, \dots, 23\}$, respectively, using the `cover.design` function from the `fields` package in R (Furrer et al. 2011).

As with hour of the day, a model for $\lambda_{\mathcal{ADC}}$ needs to account for the circular nature of day of the

year. As such, let

$$\log(\lambda_{\mathcal{ADC}}(a, d, C \mid e)) = \sum_{k_{dC}=1}^{K_{\mathcal{DC}}} G_1(C \mid C_{k_{dC}}^*, \sigma_C^2) V_{365}(d \mid d_{k_{dC}}^*, \sigma_D^2) \psi_{k_{dC}ae} \quad (7)$$

where $\{\psi_{k_{dC}ae}\}$ are a set of $K_{\mathcal{DC}}$ unknown coefficients at knot locations $(d^*, C^*)_{k_{dC}} \in \mathcal{D} \times \mathcal{C}$, $G_1(C \mid C_{k_{dC}}^*, \sigma_C^2)$ and $V_{365}(d \mid d_{k_{dC}}^*, \sigma_D^2)$ are the Gaussian and von-Mises kernel weights defined above but with $\sum_{k_{dC}} G_1(C \mid C_{k_{dC}}^*, \sigma_C^2) V_{365}(d \mid d_{k_{dC}}^*, \sigma_D^2) = 1$. The kernel weights in (7) are intentionally anisotropic with σ_D^2 and σ_C^2 controlling the smoothness in the day of the year and temperature direction, respectively. To smooth over age, let

$$\psi_{k_{dC}ae} = \sum_{k_a=1}^{K_{\mathcal{A}}} G_1(a \mid a_{k_a}^*, \sigma_A^2) \gamma_{k_{dC}k_ae} \quad (8)$$

for $k_{dC} = 1, \dots, K_{\mathcal{DC}}$ and G_1 is defined as above at the age knot locations $a_1^*, \dots, a_{K_{\mathcal{A}}}^*$. From (7) and (8), notice that $\lambda_{\mathcal{ADC}}$ is completely determined by the set of $K_{\mathcal{DC}} \times K_{\mathcal{A}}$ coefficients $\{\gamma_{k_{dC}k_ae}\}$; thus, the dimension of $\lambda_{\mathcal{ADC}}(a, d, C \mid e)$ is reduced to $K_{\mathcal{DC}} \times K_{\mathcal{A}}$. For this analysis, we use $K_{\mathcal{DC}} = 100$ and $K_{\mathcal{A}} = 15$ knot locations chosen according to the `cover.design` space-filling design over the respective discrete domains.

4. MODEL COMPARISON

As described in Section 4.1 of the manuscript, the full model was compared to three alternative model specifications. The null model simply assumes that $\lambda(\mathbf{z}) = 1$ for all $\mathbf{z}$ so that δ is the only parameter and $w(\mathbf{z})$ completely explains the spatial, temporal and marked pattern of 911 calls. The separable model assumes,

$$\lambda(\mathbf{z}) = \lambda_{\mathcal{H}}(h) \lambda_{\mathcal{S}}(\mathbf{s}) \lambda_{\mathcal{A}}(a) \lambda_{\mathcal{D}}(d) \lambda_{\mathcal{C}}(C) \lambda_{\mathcal{G}}(g) \lambda_{\mathcal{E}}(e), \quad (9)$$

so that the deviation surface is separable across all marks. Note that under the separable model, each mark is independent. Finally, the separable-in-ethnicity model assumes:

$$\lambda(\mathbf{z}) = \lambda_{\mathcal{HS}}(\mathbf{s}, h) \lambda_{\mathcal{ADC}}(a, d, C) \lambda_{\mathcal{G}}(g) \lambda_{\mathcal{E}}(e) \quad (10)$$

Table 1. Model comparison using DIC

Model	p_D	DIC
Null	0.97	122200
Separable	85.37	39361
Separable-in-ethnicity	1039.84	37971
Full	4147.56	37235

which resembles the model described in Section 2 of the manuscript with the exception that the deviation surfaces $\lambda_{\mathcal{HS}}(\mathbf{s}, h)$, $\lambda_{\mathcal{ADC}}(a, d, C)$, $\lambda_{\mathcal{G}}(g)$ are common across each race/ethnicity category.

Similar basis functions as those described above were used for the separable and separable-in-ethnicity models. For example, the deviation surface $\lambda_{\mathcal{H}}(h)$ in (9) was modeled using $K_{\mathcal{H}} = 10$ basis functions given by Equation (6). The separable-in-ethnicity model used the same set of basis functions with the exception that the associated ψ and γ coefficients were constant across race/ethnicity.

The three models posited above represent increasing levels of model complexity yet maintain a similar set of basis functions. Hence, a comparison of these models suggests the degree of complexity necessary to model the heat-related 911 call data. Table 1 herein displays the model comparison results. Clearly, the null model is inadequate suggesting the need to include a deviation surface in the model specification. While the full model is preferred in terms of DIC, the decrease from the separable-in-ethnicity to the full model is small suggesting that including an additive race/ethnicity effect for the log-deviation surface may be an appropriate simplifying assumption.

REFERENCES

- Furrer, R., Nychka, D., and Sain, S. (2011). *fields: Tools for spatial data*. R package version 6.5.2.
- Higdon, D. (2002). Space and space-time modeling using process convolutions. In Anderson, C., Barnett, V., Chatwin, P. C., and El-Shaarawi, A. H., editors, *Quantitative methods for current environmental issues*, pages 37–56. Springer-Verlag.
- Luber, G. and McGeehin, M. (2008). Climate Change and Extreme Heat Events. *American Journal of Preventative Medicine*, 35(5):429–435.

Wilhelmi, O. V. and Hayden, M. H. (2010). Connecting People and Place: A New Framework for Reducing Urban Vulnerability to Extreme Heat. *Environmental Research Letters*, 5(1):014021.