A Point Process Analysis of Cloud-to-Ground Lightning Strikes in Urban and Rural Oklahoma Areas

I. Hernandez-Magallanes[a*] and M. Genton[b]

Summary: Lightning is a natural event that can cause severe human and financial losses. This work introduces a probability risk assessment of the occurrence of the cloud-to-ground (CG) lightning in urban and rural areas of Oklahoma. CG lightning which although not the most common type, is the most damaging. Previous studies have reported that urban areas experience an increase in the frequency of CG lightning events, during warm months. This increase poses serious threats to urban industries and electronic systems. Lightning strikes are point process in nature, although this quality has not been exploited in previous studies. We utilize a probability model for the spatio-temporal point process of CG lightning to estimate the risk of a CG lightning strike for a particular location and time. The data is discretized into small spatio-temporal cells (voxels), then we fit a generalized additive model with a complementary log-log link function using the location and the day of occurrence of the strike as explanatories. Based on this model, we compared the urban and rural monthly fitted rates of CG lightning strikes. We found that the rate in the rural area is smaller than the rate in the Tulsa metropolitan area during the warm months, however it is larger than the rate in the Oklahoma metropolitan area during May and June.

Keywords: Cloud to ground lightning; Probability risk assessment; Spatio-temporal; Point process; Complementary log-log model; Urban and rural Oklahoma

[a] *Department of Statistics, Texas A&M University and Instituto Tecnologico Autonomo de Mexico*
[b] *CEMSE Division, King Abdullah University of Science and Technology*
* Correspondence to: I. Hernandez-Magallanes, Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA. E-mail: irma@stat.tamu.edu

This paper has been submitted for consideration for publication in *Environmetrics*

Environmetrics   I. Hernandez-Magallanes and M. G. Genton

---

1. INTRODUCTION

1.1. Motivation

Lightning is a natural event that can cause severe human and financial losses. For example, in 2014, the Insurance Information Institute reported that just in homeowners insurance claims, lightning strikes cost nearly $440 millions in the US. The types of events associated with these lightning losses include (a) accidents in industrial facilities due to the release of hazardous substances (Necci *et al.*, 2013); (b) interruption of or damage to electrical or electronic systems exposed to lightning strikes (Cummins *et al.*, 1998); and (c) lightning-ignited fires in remote areas that can be difficult to detect and respond to in a timely manner (Flannigan *et al.*, 2000). Therefore predicting this type of phenomena is of interest for planning purposes.

This work focuses on cloud-to-ground (CG) lightning, which, although not the most common type of lightning, is the most damaging and hazardous to structures on the ground. In particular, we examine the frequency of CG lightning events in urban settings. Past research has found increases in CG lightning events in urban areas as compared to nearby rural areas (Westcott, 1995; Orville *et al.*, 2001; Steiger *et al.*, 2002; Rose *et al.*, 2008; Farias *et al.*, 2014). In Westcott (1995) and Orville *et al.* (2001), it is suggested that the increase in urban areas can be due to three mechanisms: (a) anthropogenic heating and moisture, (b) surface roughness due to the built environment, and (c) air pollution. Downwind augmentation of CG lightning flashes is also one of the most consistent findings in studies of urban lightning. Since CG lightning poses a serious threat to urban industries and electronic systems, it is important for decision-makers to estimate the risk of CG lightning strikes. Here, we present a probability risk assessment of the occurrence of CG lightning in urban and rural areas in Oklahoma.

To the best of our knowledge, most of the existing work on CG lightning strikes is based on

2

the aggregate number of observations per region. Existing studies include statistical analyses that use some of the following methodologies: exploratory data analysis (EDA) tools, like the five-number summary and smoothing (Tukey, 1993; Hidayat and Ishii, 1998); time series analysis (Villarini and Smith, 2013); and hypothesis tests between urban and rural regions (Westcott, 1995; Orville *et al.*, 2001). The point process nature of CG lightning strikes has not yet been extensively exploited, with the exception of the analysis of the temporal properties of lightning (De Miranda *et al.*, 2003; Telesca *et al.*, 2005, 2008) and their association with lightning-caused fires (Podur *et al.*, 2003; Hering *et al.*, 2009). Here, we utilized a probability model to estimate the intensity of the CG lightning spatio-temporal point process. The point process approach provides findings on the natural temporal and spatial variation in the intensity of CG lightning without the need for arbitrary aggregation of individual strikes into counts.

The data were obtained from the U.S. National Lightning Detection Network (NLDN) and covered the period from 2006-2010 for the two largest metropolitan areas in Oklahoma, Tulsa and Oklahoma City, as well as a rural region, as shown in Figure 1. The data are grouped into small spatial-temporal cells (voxels). A complementary log-log model was used under the assumption of the existence of a latent process with UTM coordinates and day of the year as explanatory variables. We employed a maximum likelihood fit using the conditions included in Schoenberg (2005), under which parametric estimates of the intensity of a spatial-temporal point process are consistent although the actual point process being estimated may not be Poisson. The model was assessed based on the fitted linear predictor. The analysis of the effects of the urban and rural conditions on the frequency of CG lightning strikes was based on a comparison of their monthly expected number of CG lightning strikes for some selected subregions.

[Figure 1 about here.]

The remainder of this paper is organized as follows: in Subsection 1.2 we include a

3

description of the physical properties of the CG lightning data. Section 2 includes a description of the data available and the definition of an urban and rural area. The theory and assumptions behind our model are included in Section 3, in this section we introduce the conditional intensity function of the spatio-temporal point process, as well as the 0-1 valued process used in practice to approximate the log-likelihood function. Section 4 includes the exploratory analysis of the temporal and spatial properties of the collected CG lightning data for the three regions. Section 5 includes the results of the fitted generalized additive model to the voxel based data with the complementary log-log link function and the respective model assessment. In Section 6, we compare the fitted monthly rates of CG lightning strikes between the urban and rural regions based on our probability model.

1.2. Cloud-to-ground lightning

The initial stage of all lightning events requires the separation of positive and negative charges into different regions of a cloud. Most of the time, the positive charges accumulate in the upper part of the cloud while negative charges accrue in the lower part. This separation of charges produces enormous electrical potential both within clouds, in the form of cloud-to-cloud (CC) lightning, and from the cloud to the surface, in the form of CG lightning. Eighty percent of all lightning is CC, while the remainder is CG. Lightning occurs exclusively in clouds above the freezing level, and it is restricted to precipitating clouds. The definitions included in this section and further details can be found in Aguado *et al.* (2007), Weidman (2013), and NASA (2013).

Most cloud-to-ground discharges begin with negatively charged air that moves downward in a rapid sequence of individual steps, called a stepped leader. A spark surges upward from the ground to connect to the leader, initiating the first in a sequence of illuminated strokes (return strokes). The combination of strokes forms a lightning flash. Most of the flashes consist of only two or three strokes, but some consist of as many as twenty individual

4

strokes. The number of strokes in a flash is called the flash multiplicity. Because each flash can be made up of several return strokes, the location used in this study corresponded to the first return stroke.

Around ninety-five percent of flashes originate near the lower-negative charge center, delivering negative charges to the ground. Only five percent carry positive charges to the ground, but they are more lethal and often occur during the dissipating stage of a thunderstorm. The polarity of CG lightning can be positive, $+$CG, or negative, $-$CG, based on whether the lightning carries a positive or negative charge. The peak current of the return stroke averages 30 kiloamperes (KA) for a typical $-$CG flash, while the peak charge for $+$CG can be ten times greater.

In this analysis, we studied two CG lightning characteristics that have been reported as being influenced by urban conditions: 1) the number of CG lightning flashes experiences a significant increase in urban regions compared with nearby rural areas (Westcott, 1995; Orville *et al.*, 2001; Steiger *et al.*, 2002); and 2) the percentage of positive flashes experiences a decrease in urban areas compared with nearby rural areas (Orville *et al.*, 2001; Farias *et al.*, 2009; Stallins *et al.*, 2013).

2. CLOUD-TO-GROUND LIGHTNING DATA

The CG lightning data analyzed in this paper were provided by NLDN that is owned and operated by Vaisala, Inc. The data cover the period from 2006 to 2010 for the Tulsa and Oklahoma City metropolitan areas and a selected rural region. The CG ligthning occurrences are acquired from 114 sensors that detect electromagnetic radiation from lightning return strokes (Orville, 2008; Cummins and Murphy, 2009; Vaisala, 2011).

Figure 1 shows the approximately one and a half million of CG lightning flashes recorded during 2006-2010, in the three regions of interest. For each CG lightning flash included we

5

analyzed the following features: (a) the location using the UTM coordinates, denoted by (x, y); (b) the day of occurrence denoted by t and (c) the polarity.

The analysis was performed on the data collected in Oklahoma because of the rich atmospherical and climatological information available for this state, including the Oklahoma Climatological Survey (2014) database. The two urban areas studied here were chosen based on the definition of a metropolitan area provided by the United States Census Bureau (2010). Both have a population of 50,000 or more and each metropolitan area comprises of one or more counties, including the county containing the core urban area as well as any adjacent counties that have a high degree of social and economic integration (as measured by commuting to work) with the urban core. In particular, we studied the two largest metropolitan regions in Oklahoma: Tulsa and Oklahoma City.

The Tulsa metropolitan area includes the following seven counties: Tulsa, Rogers, Wagoner, Creek, Osage, Okmulgee, and Pawnee. The Oklahoma City metropolitan area also includes seven counties: Canadian, Cleveland, Grady, Lincoln, Logan, McClain, and Oklahoma. The CG lightning data in these metropolitan regions were compared with the CG lightning flashes that occurred in the selected rural area, which includes five counties: Okfuskee, Seminole, Hughes, Coal, and Mcintosh. We choose the rural area shown in Figure 1 because it borders both Tulsa and Oklahoma City metropolitan areas.

3. THE SPATIO-TEMPORAL POINT PROCESS METHODOLOGY

A spatio-temporal point process is a random collection of points, where each point represents the time and location of an event (Guttorp *et al.*, 2002). The class of spatio-temporal point processes described by Fishman and Snyder (1976) is uniquely characterized by its associated conditional intensity process, λ. Suppose that the space-time domain is broken into voxels $(x, x + dx] \times (y, y + dy] \times (t, t + dt]$. Consider the spatio-temporal point process, N, with

6

Point Process Analysis of Cloud-to-Ground Lightning **Environmetrics**

conditional intensity function for (x, y, t), where $(x, y) \in \mathbb{R}^2$ and $t = 1, \ldots, T$ is assumed to exist and is defined by:

$$\lambda(x, y, t) = \mathrm{P}\{dN(x, y, t) = 1 | H_t\}/dxdydt, \tag{1}$$

where $dN(x, y, t) = N(dx, dy, dt)$ counts the number of flashes in the voxel $(x, x + dx] \times (y, y + dy] \times (t, t + dt]$ and H_t is the prior history of the point process up to time t (Brillinger *et al.*, 2003). The conditional intensity function may be thought of as the frequency with which events are expected to occur around a particular location, conditional on the prior history of the point process up to a particular time.

In order to estimate the risk of a CG lightning, we model the conditional intensity of the spatio-temporal CG lightning point process for a particular location and time as a function of other covariates, such as the UTM coordinates and the day of the year. The parameters can be estimated by maximizing the following log-likelihood function (Fishman and Snyder, 1976):

$$\ell(\boldsymbol{\theta}) = \int_0^T \int_y \int_x \log\left[\lambda(x, y, t | \boldsymbol{\theta})\right] dN(x, y, t) - \int_0^T \int_y \int_x \lambda(x, y, t | \boldsymbol{\theta}) dxdydt, \tag{2}$$

where $\boldsymbol{\theta}$ is a vector of free parameters. Under general conditions, the maximum likelihood estimates (MLEs) have been shown to be consistent and asymptotically normal (Ogata, 1978; Rathbun and Cressie, 1994; Schoenberg, 1997).

In practice, the space-time domain can be broken up into voxels and the spatio-temporal point process, $N(dx, dy, dt)$, can be replaced by a 0-1 valued process, $Y_{x,y,t}$ on a lattice, with $Y_{x,y,t} = 1$ if there is a flash in the corresponding voxel and 0 otherwise for $(x, y) \in \mathbb{R}^2$ and $t = 1, \ldots, T$ (Brillinger and Segundo, 1979; Berman and Turner, 1992; Baddeley *et al.*, 2010, 2015). Suppose that:

$$\mathrm{P}\left(Y_{x,y,t} = 1 | H_t\right) = \lambda_{x,y,t},$$

7

Environmetrics   I. Hernandez-Magallanes and M. G. Genton

where H_t is the prior history of the point process up to time t.

In order to estimate the conditional intensity function we assume that $Y_{x,y,t}$ are independent given the covariates, therefore the Bernoulli approximation to the log-likelihood in Equation (2) is now:

$$\sum_{x,y,t} Y_{x,y,t} \log \left(\lambda_{x,y,t}\right) + \sum_{x,y,t}(1 - Y_{x,y,t}) \log \left(1 - \lambda_{x,y,t}\right). \tag{3}$$

Using the result in Schoenberg (1997, 2005), the estimates obtained by maximizing the log-likelihood function in Equation (3) will be consistent assymptotically even though the CG lightning process is not really a Poisson process.

To simplify the notation, the pixels can be indexed by $k = 1, \ldots, K$ rather than (x, y, t) and we express the log-likelihood function as follow:

$$\sum_{k=1}^{K} Y_k \log \left(\lambda_k\right) + \sum_{k=1}^{K}(1 - Y_k) \log \left(1 - \lambda_k\right). \tag{4}$$

To find the parameters, we maximize this log-likelihood function where the linear predictor (η) based on the explanatories, uses a complementary log-log link function and is expressed as $\eta = \log \left(-\log \left(1 - \lambda\right)\right)$. We used the complementary log-log link function since the theoretical and empirical results established in Baddeley *et al.* (2010), showed that the estimates with this link function are typically more accurate than the logit link function for a given grid size.

4. EXPLORATORY DATA ANALYSIS

4.1. Temporal analysis

In this section, we present the analysis of the purely temporal behavior of the resulting marginalized CG lightning point process. We analyzed the two characteristics of interest

8

Point Process Analysis of Cloud-to-Ground Lightning   **Environmetrics**

described in Section 2, the rate of CG strikes per km^2 and the proportion of +CG of the total count of CG lightning strikes (both positive and negative polarities) over the time period 2006-2010 for each of the three regions. The top of Figure 2 includes the time plots and the monthly boxplots of the square root of the rate of CG flashes per km^2 and at the bottom the respective plots for the logarithmic transformation of the proportion of +CG of the total count of CG lightning strikes. The four plots include a smoother and the respective confidence intervals. The smoother is based on a local polynomial regression (Cleveland, 1979; Chambers and Hastie, 1991), in this analysis the R function $loess()$ with a window size of 2/3 was used.

[Figure 2 about here.]

Plot a) in Figure 2 suggests that Tulsa's rate per km^2 has a trend that is characterized by a quasi-parabolic behaviour with a maximum in 2008. Compared to the other two regions, Tulsa's rate is higher from 2008 to 2010. The trends of the CG lightning flashes per km^2 in Oklahoma City and the rural region behave very similar and both experience a drop during 2008, the year where Tulsa Metropolitan Region reaches the highest level. The boxplots in plot b) show an annual seasonal effect, with the summer having the highest rate while the winter experiences the lowest rate of flashes; the outliers are associated to storms and unusual weather phenomena experienced in those months. For example, in September 2009, an outlier in the Tulsa metropolitan area is due to a large number of strikes that occurred on the afternoon of September 21, 2009 during a strong storm with hail.

Plots c) and d) in Figure 2 display the logarithmic transformation of the proportion of +CG lightning strikes. Plot c) exhibits that the urban's and rural's trends are characterized by a sigmoid behaviour. The boxplots in plot d) display an annual seasonality in the three regions, with the proportion of +CG reaching its lowest level during the summer and its highest peak during the winter. By comparing plots a) and c), the trends of the rate of CG flashes per km^2 and of the proportion of positive flashes appeared to be out of phase. This

9

out of phase finding, coincides with the behaviour previously reported (Orville *et al.*, 2001; Farias *et al.*, 2009; Stallins *et al.*, 2013) over urban areas: a decrease in the percentage of positive CG flashes during the season with the highest rate of CG lightning strikes (May to September) (Orville *et al.*, 2001; Farias *et al.*, 2009; Stallins *et al.*, 2013).

An interesting finding of our EDA is that the trend and seasonality of the analyzed variables (CG rate per km^2 and proportion of +CG lightning strikes) in the rural region are very similar to the ones of the Oklahoma City metropolitan area. Also, contrary to what we were expecting there is no evidence that non-metropolitan region experiences a lower rate per km^2 than the two urban regions, instead the non-metropolitan region's rate is higher than the Oklahoma City.

4.2. Spatial Analysis

In this section, we present the analysis of the purely spatial behavior of the resulting marginalized CG lightning point process. Due to the large number of CG lightning occurrences, we used a hexagonal binning of the observations per year (Carr *et al.*, 1987) to produce Figure 1. The maps in this figure depict the regions with the highest number of CG lightning flashes, for example, in 2007 these regions were located in the Oklahoma City metropolitan area, in 2008 they were found in the Tulsa metropolitan area, and finally in 2009 they included the area that extends from Tulsa to the south and east territory of the rural area. The same Figure 1 shows that the CG lightning flashes occurr in clusters, the weaker clustering effect revealed during the years 2006 and 2010, could be explained by the smaller number of CG lightning events that occurred during those years.

The monthly maps in Figure 3 display the smoothed CG lightning intensity function estimated with a biweight kernel function (Diggle, 1985); these maps also reveal a monthly spatial clustering effect. We can think of these clustered regions as locations with a higher risk of CG lightning flashes. We found that during the second quarter of the year, there is

10

Environmetrics

a higher risk of CG lightning flashes. For example, in May the Oklahoma City metropolitan area experiences the highest risk of CG lightning and this phenomenon affects the boundary with Tulsa and extends to the rural region. In June, the map suggests the highest risk region moves to the Tulsa metropolitan area while the rural territory keeps experiencing a high risk of CG lightning.

In order to confirm the suspected clustering effect, the summary statistic L-function has been frequently used (Hering *et al.*, 2009; Møller and Rasmussen, 2015). We are interested in identifying if there is a departure from complete spatial randomness (CSR) (Baddeley *et al.*, 2015); if the CSR hypothesis is rejected, then there must be a tendency toward either clustering (having flashes close to each other) or regularity (repulsion between flashes). The L-function defined in Besag (1977), is a variance-stabilizing transformation of the K-function:

$$L(r) = \sqrt{\frac{K(r)}{\pi}}. \tag{5}$$

The K-function (Baddeley *et al.*, 2015) is a second-order parameter of a point process. It is used to analyze a spatial point process and to summarize the spatial dependence over a wide range of scales (Cressie, 1993). In the stationary and isotropic case, Ripley's K-function is defined as $K(r) = \lambda^{-1}$ E(number of points within r of an arbitrary point), where $r \geq 0$ and λ is the conditional intensity.

We used the R function $Lest()$ of the spatstat package (Baddeley *et al.*, 2015) to estimate the L-function, given the observations of the process inside a known, bounded window. Due to the unobservability of points of the random pattern the estimation of the L-function is modified with an edge correction to reduce the estimate's bias (Ripley, 1991; Barndorff-Nielsen *et al.*, 1999).

[Figure 3 about here.]

11

For interpretation purposes, we employed the centered L-function (L-function minus the identity), this summary statistic under clustering, is expected to be above the zero-line, whereas under regularity tends to be less than the zero-line (see e.g. Moller and Waagepetersen (2003)). In order to better identify the clustering effect over the different annual seasons, we estimate the quarterly centered L-functions using the data for the period 2006 to 2010. Figure 4 includes the estimated centered L-function with a border edge correction and the corresponding simulated 95% envelopes. For visualization purposes, this figure depicts the estimated L-functions for distances (r) up to 200 meters; the values for larger distances are omitted as the trends remain the same after 200 meters.

As expected, the centered L-functions exhibit that the CG lightning flashes occur in clusters (Zajac and Rutledge, 2001; Podur *et al.*, 2003; Villarini and Smith, 2013), this finding was identified in the urban and rural regions. As CG lightning flashes accompany thunderstorms, the clustering effect that characterized the thunderstorms (Strauss *et al.*, 2013; Chronis *et al.*, 2015) could be one of the reason why the CG flashes occur in clusters. We also noticed a regularity effect within 5 to 10 meters during the second quarter (Q_2) of the year for the three regions. Tulsa experiences this regularity effect during the third quarter (Q_3) too.

[Figure 4 about here.]

5. THE SPATIO-TEMPORAL LIKELIHOOD APPROACH

5.1. Model fitting

The spatio-temporal likelihood approach is based on the 0-1-valued process, Y. The lightning data are grouped into small spatio-temporal cells (voxels) of size $1 \ km \ \times \ 1 \ km \ \times \ 1 \ day$. The parameters are estimated by maximizing the log-likelihood function defined in Equation

12

(4), where k refers to the k-th voxel and $Y_k = 1$ if there is a flash in the corresponding voxel and 0 otherwise. We assumed a smooth dependence of the risk probability of occurrence of a CG lightning flash on the spatial (x_k, y_k) and temporal (day_k) effects. Hence, we used the R function $gam()$ to fit the Bernoulli model with the following linear predictor:

$$\eta_k = \log\left(-log(1 - \lambda_k)\right) = \log\left(v_k\right) + g_1(x_k, y_k) + g_2(day_k). \tag{6}$$

with an offset of $\log\left(v_k\right)$ which is the volume of the k-th voxel. Here, (x_k, y_k) are the UTM coordinates of the k-th response and (day_k) the day of the year when it took place. The $g()$'s are (nonparametric) smooth functions to be estimated. Then the model is the following:

$$\lambda = P\left(Y = 1 | \eta\right) = 1 - \exp\left(-v \exp\left(\eta\right)\right) \tag{7}$$

as we used a regular grid for our analysis, the value of v_k is constant $(v_k = v)$ for any voxel. The log-likelihood function was maximized by applying penalized iteratively reweighted least squares (P-IRLS) (Wood, 2006a) and the variances were estimated by using the Bayesian approach (Wood, 2006b).

Figure 5 shows the contour and the perspective plots of the estimated spatial effect, $\hat{g}_1$, for the three regions of study. These plots indicate a higher risk of CG lightning occurrences in the Tulsa metropolitan area, in particular in Tulsa, Okmulgee and Pawnee counties. Tulsa County is the second most densely populated in the state and with the highest income. The higher risk identified in Tulsa could be explained by the large concentration of aviation and aerospace services provided by the Tulsa International Airport. For example, this airport is the global maintenance headquarters for American Airlines and is a large employer. The southern region of Tulsa is where most of airport employees live. The high-risk area in the northern part of the rural area might be due to the influence of the nearby Oklahoma City area.

The southeast area of the Oklahoma City is more prone to the occurrence of CG lightning.

As this area is located in a region that is seldom downwind of the larger urban region, we would expect little influence from the urban environment (Hand and Shepherd, 2009). Therefore, this finding doesn't support the downwind augmentation of CG lightning flashes, which is one of the most consistent findings in studies of urban lightning. Including other explanatory variables in our model might provide a better understanding of other conditions associated with the increase in lightning in this urban area.

[Figure 5 about here.]

The estimated temporal effect $\hat{g}_2$ with the corresponding 95% confidence interval is displayed in Figure 6. We identify an annual seasonal effect in the three regions, this finding coincides with the one described in the EDA included in Section 4. The plots suggest that the highest risk of CG lightning is experienced during the summer months, from May to September, with May being the month with the highest risk. Another finding is that the Tulsa metropolitan area and the rural area experience again a higher risk of CG lightning during August.

[Figure 6 about here.]

5.2. Model assessment

To verify the appropriateness of the model in Equation (7), we implemented an assessment method based on the fitted linear predictor liked the one used in Brillinger *et al.* (2003). Under the assumption of a complementary log-log link function, the fitted linear predictor values may be assigned to the bins of a histogram. For each bin, some of the corresponding Y's will be 0 and some will be 1. The "number of 1's" divided by "the number of 1's" plus "the number of 0's" provides a nonparametric estimate of the function $\lambda(\eta) = P(Y = 1|\eta = 1)$, with η representing the linear predictor of the employed generalized additive model. This estimate may be compared to $\lambda(\eta|\hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimate.

14

Figure 7 shows the assessment of the complementary log-log model in Equation (6). The points are the observed relative frequencies of CG lightning, after grouping the data into classes based on the fitted linear predictor $\hat{\eta}$. The solid red curve is the fitted complementary log-log curve. The dashed lines are approximate 95% confidence intervals, which are obtained via a binomial approximation. The fit of the model appears reasonable for the data in the three regions but there seems to be a departure from the complementary log-log model as some patterns of points are above the curve on the left of the plots. The departure from the complementary log-log model may disappear when we include other explanatory variables in the model. The confidence intervals are wider for the larger values of the fitted linear predictors.

[Figure 7 about here.]

6. THE URBAN AND RURAL EFFECT

Several authors (Westcott, 1995; Orville *et al.*, 2001; Steiger *et al.*, 2002; Rose *et al.*, 2008; Hand and Shepherd, 2009) have found that some urban areas, like: Dallas-Fort Worth, Milwaukee and Kansas City, experience an increase in the occurrence of CG lightning flashes mainly in the summer compared to nearby rural regions. To identify whether this same phenomenon occurs in Oklahoma, we compared the expected number of CG lightning strikes between two metropolitan areas with a contiguous rural region. The expected number of CG lightning flashes were calculated by month for equal-area circular regions, each region has a radius of approximately 20 kms and are shown in Figure 1. The centers of the circles are located in Tulsa and Oklahoma City, while in the rural area, the center is located in a zone less densely populated.

We used Hodges and Le Cam's Poisson Binomial distribution (Hodges and Le Cam, 1960) as an approximation of the distribution of the total number of CG lightning flashes in the

15

l-th region, named $N(R_l)$. Hodges and Le Cam's result shows that $N(R_l)$ has in the limit a Poisson distribution with parameter $\lambda = \sum_{(x,y) \in R_l} \lambda(x, y)$, where $\lambda(x, y)$ is the probability of occurrence of a CG lightning flash at the location (x, y). In this study, the number of voxels is sufficiently large and $\alpha = \max_{(x,y) \in R_l} [\lambda(x, y)]$ is small, satisfying the assumption of a Poisson Binomial distribution.

Recalling the binary-valued variable described in Section 3, Y_k, defined on a lattice, with $Y_k = 1$ if there is a flash in the corresponding voxel and 0 otherwise and considering a general distribution for Y_k, the monthly expected number of CG lightning events in the l-th region is:

$$E\left[N^{(m)}(R_l)\right] = \sum_{k \in R_l} \lambda_k^{(m)}, \tag{8}$$

where $N^{(m)}(R_l)$ stands for the total number of CG lightning flashes in the l-th region for the m-th month and $\lambda_k^{(m)}$ is the probability of occurrence of a CG lightning flash at the k-th voxel in R_l for the m-th month.

The empirical and fitted CG lightning rates by month are included in Figure 8; the solid lines are the fitted CG lightning rates by month based on the model in Equation (7), with location and day of the year as explanatory variables; the gray shaped region provides their approximate 95% marginal confidence limits. The empirical rates by months and their approximated 95% confidence intervals based on a Poisson approximation are included as bar plots.

[Figure 8 about here.]

The fitted rate by month was higher during the warm months in both urban and rural areas. Overall, the fitted CG lightning rate by month in the Tulsa metropolitan region seemed to be larger than that in the other two regions. The claim that urban areas experience a higher rate

16

of CG lightning flashes during the warm months than nearby rural regions, is not completely accurate based on the fitted rates shown Figure 8: the rate in the rural area is smaller than the rate in the Tulsa metropolitan area during the warm months, however it is larger than the rate in the Oklahoma metropolitan area during May and June. The plots also suggest that the Tulsa metropolitan area experiences an increase in the CG frequency rate in August while the rural region experiences this increase in September. The latter growths have been associated with large lightning events, as described by Westcott (1995), Orville *et al.* (2001), and Steiger *et al.* (2002). Large lightning events refer to days when a large number of flashes is recorded.

7. SUMMARY AND DISCUSSION

This study developed a risk probability model that uses the point process nature of the CG lightning flashes and jointly modeled the temporal and spatial domains. Our approach, compared with previous works, provided an additional step in the modeling of CG lightning, as our proposed model estimates the probability of the occurrence of a CG lightning flash for a particular time and location. Therefore, we were able to estimate the expected number of CG lightning strikes that can be used for planning and resource allocation. A complementary log-log model was used under the assumption of the existence of a latent process with UTM coordinates and the day of the year as explanatory variables. The proposed model is also flexible enough to include other lightning properties (polarity, peak current, and multiplicity) as marks of the process.

The estimated temporal effect revealed an increase in the frequency of CG lightning events during the warm seasons. The estimated spatial effect showed that the areas with a higher risk of CG lightning flashes were located in the Tulsa metropolitan area, in particular in Tulsa, Okmulgee and Pawnee counties; along the border between the Tulsa metropolitan

17

Environmetrics

I. Hernandez-Magallanes and M. G. Genton

region and the rural area; and in the southeast area of the Oklahoma City.

One of the hypothesized explanations for the high rate of CG lightning flashes in the Tulsa metropolitan region was air pollution. The urban areas experience elevated surface ozone concentrations during the summer season (Hu *et al.*, 2013). Tulsa's propensity for lightning might be partially explained by the activities associated with the Tulsa International Airport, not only because airports are among the largest sources of air pollution in the United States (Westerdahl *et al.*, 2008), but also due to the large number of aviation and aerospace services provided there.

Based on the predicted probabilities and their associated confidence intervals, the estimated number of CG lightning strikes in the Tulsa metropolitan area appears to be larger than the other two regions. Because of the results of previous works, we were expecting the estimated number of CG lightning strikes in the rural area to be smaller than the estimated number in the urban regions during the warm months. On the contrary, the rate in the rural area is larger than the rate in the Oklahoma metropolitan area during May and June.

In conclusion, we found that the Tulsa metropolitan area experiences conditions that promote the increased frequency of CG lightning compared to the Oklahoma City metropolitan area and the nearby rural region. A suggested extension of this work would be to include other explanatory variables, such as air pollution, population density, and topography, which might explain the CG flash increments. Also, it will be of interest to identify the reasons why urban regions like the Oklahoma City metropolitan area, do not experience a higher rate of CG lightning than the nearby rural regions.

The risk assesment of the CG lightning events included in this work can be employed to help prevent such things as electric power interruptions, forest fires, and human injuries and deaths caused by lightning strikes.

18

ACKNOWLEDGEMENTS

REFERENCES

Aguado E, Burt JE, Rohli RV, Schmidlin TW, 2007. *Understanding Weather and Climate*. Pearson Prentice Hall.

Baddeley A, Berman M, Fisher N, Hardegen A, Milne R, Schuhmacher D, Shah R, Turner R, 2010. Spatial logistic regression and change of support in Poisson point processes. *Electronic Journal of Statistics* **4**: 1151–1201.

Baddeley A, Rubak E, Turner R, 2015. *Spatial Point Patterns: Methodology and Applications with R*. CRC Press.

Barndorff-Nielsen OE, Kendall WS, Van Lieshout M, 1999. *Stochastic Geometry: Likelihood and Computation*, volume 80. CRC Press.

Berman M, Turner TR, 1992. Approximating Point Process Likelihoods with GLIM. *Applied Statistics* : 31–38.

Besag J, 1977. Contribution to The Discussion of Dr. Ripley's Paper. *Journal of the Royal Statistical Society. Series B (Methodological)* **39**(2): 193–195.

Brillinger DR, Preisler HK, Benoit JW, 2003. Risk Assessment: A Forest Fire Example. *The Institute of Mathematical Statistics Lecture Notes Monograph Series. Statistics and Science: A Festschrift for Terry Speed* **40**: 177–196.

Brillinger DR, Segundo JP, 1979. Empirical Examination of The Threshold Model of Neuron Firing. *Biological Cybernetics* (35): 213–220.

Carr DB, Littlefield RJ, Nicholson W, Littlefield J, 1987. Scatterplot matrix techniques for large n. *Journal of the American Statistical Association* **82**(398): 424–436.

19

Chambers JM, Hastie TJ, 1991. *Statistical Models in S.* CRC Press, Inc.

Chronis T, Carey LD, Schultz CJ, Schultz EV, Calhoun KM, Goodman SJ, 2015. Exploring lightning jump characteristics. *Weather and Forecasting* **30**(1): 23–37.

Cleveland WS, 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* **74**(368).

Cressie NA, 1993. *Statistics for Spatial Data, Revised Edition.* Wiley New York.

Cummins KL, Krider EP, Malone MD, 1998. The US National Lightning Detection Network and Applications of Cloud-to-Ground Lightning Data by Electric Power Utilities. *IEEE Transactions on Electromagnetic Compatibility* **40**(4): 465–480.

Cummins KL, Murphy MJ, 2009. An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the us nldn. *Electromagnetic Compatibility, IEEE Transactions on* **51**(3): 499–518.

De Miranda F, Pinto Jr O, Saba M, 2003. A Study of The Time Interval Between Return Strokes and K-changes of Negative Cloud-to-Ground Lightning Flashes in Brazil. *Journal of Atmospheric and Solar-Terrestrial Physics* **65**(3): 293–297.

Diggle P, 1985. A Kernel Method for Smoothing Point Process Data. *Applied Statistics* : 138–147.

Farias W, Pinto Jr O, Naccarato K, Pinto I, 2009. Anomalous Lightning Activity over The Metropolitan Region of Sao Paulo Due to Urban Effects. *Atmospheric Research* **91**(2): 485–490.

Farias W, Pinto Jr O, Pinto I, Naccarato K, 2014. The Influence of Urban Effect on Lightning Activity: Evidence of Weekly Cycle. *Atmospheric Research* **135**: 370–373.

Fishman PM, Snyder DL, 1976. The Statistical Analysis of Space-Time Point Processes. *IEEE Transactions on Information Theory* **22**(3): 257–274.

Flannigan M, Stocks BJ, Wotton B, 2000. Climate Change and Forest Fires. *Science of the Total Environment* **262**(3): 221–229.

Guttorp P, Brillinger DR, Schoenberg FP, 2002. Point Processes, Spatial. *Encyclopedia of Environmetrics* **3**: 1571–1573.

Hand LM, Shepherd JM, 2009. An Investigation of Warm-Season Spatial Rainfall Variability in Oklahoma City: Possible Linkages to Urbanization and Prevailing Wind. *Journal of Applied Meteorology and Climatology* **48**(2).

20

Hering AS, Bell CL, Genton MG, 2009. Modeling Spatio-Temporal Wildfire Ignition Point Patterns. *Environmental and Ecological Statistics, Special Issue on Statistics for Wildfire Processes* **16**(2): 225–250.

Hidayat S, Ishii M, 1998. Spatial and Temporal Distribution of Lightning Activity around Java. *Journal of Geophysical Research* **103**(D12): 14001–14009.

Hodges JL, Le Cam L, 1960. The Poisson Approximation to the Poisson Binomial Distribution. *The Annals of Mathematical Statistics* : 737–740.

Hu XM, Klein PM, Xue M, Shapiro A, Nallapareddy A, 2013. Enhanced Vertical Mixing Associated with a Nocturnal Cold Front Passage and its Impact on Near-Surface Temperature and Ozone Concentration. *Journal of Geophysical Research: Atmospheres* **118**(7): 2714–2728.

Møller J, Rasmussen JG, 2015. Spatial cluster point processes related to poisson–voronoi tessellations. *Stochastic environmental research and risk assessment* **29**(2): 431–441.

Moller J, Waagepetersen RP, 2003. *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC.

NASA, 2013. Lightning & Atmospheric Electricity Research. *Global Hydrology Resource Center* .

Necci A, Antonioni G, Cozzani V, Krausmann E, Borghetti A, Alberto Nucci C, 2013. A Model for Process Equipment Damage Probability Assessment Due to Lightning. *Reliability Engineering & System Safety* **115**: 91–99.

Ogata Y, 1978. The Asymptotic Behaviour of Maximum Likelihood Estimators for Stationary Point Processes. *Annals of the Institute of Statistical Mathematics* **30**(1): 243–261.

Oklahoma Climatological Survey, 2014. The Oklahoma Mesonet. http://www.mesonet.org/index.phpl.

Orville RE, 2008. Development of the National Lightning Detection Network. *Bulletin of the American Meteorological Society* **89**(2): 180–190.

Orville RE, Huffines G, Nielsen-Gammon J, Zhang R, Ely B, Steiger S, Phillips S, Allen S, Read W, 2001. Enhancement of Cloud-to-Ground Lightning over Houston, Texas. *Geophysical Research Letters* **28**(13): 2597–2600.

Podur J, Martell DL, Csillag F, 2003. Spatial Patterns of Lightning-Caused Forest Fires in Ontario, 1976–1998. *Ecological Modelling* **164**(1): 1–20.

Rathbun SL, Cressie N, 1994. Asymptotic Properties of Estimators for the Parameters of Spatial Inhomogeneous Poisson Point Processes. *Advances in Applied Probability* : 122–154.

Ripley BD, 1991. *Statistical Inference for Spatial Processes*. Cambridge University Press.

21

Rose L, Stallins J, Bentley M, 2008. Concurrent Cloud-to-Ground Lightning and Precipitation Enhancement in the Atlanta, Georgia (United States), Urban Region. *Earth Interactions* **12**(11): 1–30.

Schoenberg FP, 1997. Assessment of Point Process Models. *Ph.D. Thesis* .

Schoenberg FP, 2005. Consistent parametric estimation of the intensity of a spatial-temporal point process. *Journal of Statistical Planning and Inference* (128): 79–93.

Stallins JA, Carpenter J, Bentley ML, Ashley WS, Mulholland JA, 2013. Weekend–Weekday Aerosols and Geographic Variability in Cloud-to-Ground Lightning for the Urban Region of Atlanta, Georgia, USA. *Regional Environmental Change* **13**(1): 137–151.

Steiger SM, Orville RE, Huffines G, 2002. Cloud-to-Ground Lightning Characteristics over Houston, Texas: 1989–2000. *Journal of Geophysical Research* **107**(D11): ACL–2.

Strauss C, Rosa MB, Stephany S, 2013. Spatio-temporal clustering and density estimation of lightning data for the tracking of convective events. *Atmospheric research* **134**: 87–99.

Telesca L, Bernardi M, Rovelli C, 2005. Intra-Cluster and Inter-Cluster Time Correlations in Lightning Sequences. *Physica A: Statistical Mechanics and its Applications* **356**(2): 655–661.

Telesca L, Bernardi M, Rovelli C, 2008. Time-Scaling Analysis of Lightning in Italy. *Communications in Nonlinear Science and Numerical Simulation* **13**(7): 1384–1396.

Tukey JW, 1993. Exploratory Data Analysis: Past, Present and Future. Technical report, DTIC Document.

United States Census Bureau, 2010. Geographic Terms and Concepts: Core Based Statistical Areas and Related Statistical Areas. http://www.census.gov/geo/reference/gtc/gtccbsa.html.

Vaisala, 2011. Vaisala's NLDN U.S. National Lightning Detection Network. NLDN Brochure Ref. B210412EN-F.

Villarini G, Smith JA, 2013. Spatial and temporal variability of cloud-to-ground lightning over the continental U.S. during the period 1995 to 2010. *Atmospheric Research* (124): 137–148.

Weidman C, 2013. Introduction to Weather and Climate. University of Arizona, Lecture Notes Atmo 170A1.

Westcott NE, 1995. Summertime Cloud-to-Ground Lightning Activity around Major Midwestern Urban Areas. *Journal of Applied Meteorology and Climatology* **34**(7): 1633–1642.

Westerdahl D, Fruin SA, Fine PL, Sioutas C, 2008. The Los Angeles International Airport as a Source of Ultrafine Particles and Other Pollutants to Nearby Communities. *Atmospheric Environment* **42**(13): 3143–3155.

Wood S, 2006a. *Generalized Additive Models: An Introduction with R*. CRC press.

22

Wood SN, 2006b. On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics* **48**(4): 445–464.

Zajac BA, Rutledge SA, 2001. Cloud-to-ground lightning activity in the contiguous united states from 1995 to 1999. *Monthly Weather Review* **129**(5): 999–1019.