SDS
Statistica e
Data Science

Palermo, 11-12 April 2024

Proceedings of the Statistics and Data Science 2024 Conference

---

New perspectives on Statistics and Data Science

Edited by

Antonella Plaia – Leonardo Egidi
Antonino Abbruzzo

Proceedings of the SDS 2024 Conference
Palermo, 11-12 April 2024
Edited by: Antonella Plaia – Leonardo Egidi - Antonino Abbruzzo
-
Palermo: Università degli Studi di Palermo.

ISBN Ebook (Pdf)
978-88-5509-645-4

Spatio-temporal analysis of lightning point process data in severe storms

Nicoletta D'Angelo, Milind Sharma, Marco Tarantino and Giada Adelfio

Abstract This paper deals with the analysis of the 19 May 2013 tornadic supercell in central Oklahoma, exploiting point process theory to estimate the intensity of lightning flashes. We develop a computational strategy to fit spatio-temporal Poisson models, including external covariates, to understand the role of environmental spatio-temporal covariates on the occurrence of such lightning events.

Key words: Lightning data, Spatial Statistics, Point processes, Spatio-temporal analysis

1 Introduction

This paper analyses data from the 19 May 2013 tornadic supercell in central Oklahoma, USA. Per the U.S. National Weather Service (NWS) storm survey reports, the supercell produced multiple tornadoes throughout its lifetime. The first tornado had a damage rating of 0 on the enhanced Fujita (EF) scale (McDonald and Mehta, 2006) around 2122 UTC. Subsequently, two more EF1 tornadoes were reported at 2133 and 2153 UTC near Arcadia and Fallis, Oklahoma. The most intense tornado occurred around 2213 UTC, inflicting an EF3 damage as it passed over the city of Carney, Oklahoma. A detailed study of the electrical and polarimetric characteristics of this storm can be found in Sharma et al. (2021). A comparison of the polarimetric indicators of the updraft intensity with vertical velocity retrieved from dual-Doppler wind synthesis for the same storm can be found in Wienhoff et al. (2018).

Nicoletta D'Angelo, Marco Tarantino, Giada Adelfio
Department of Business, Economics and Statistics, University of Palermo
e-mail: nicoletta.dangelo@unipa.it; marco.tarantino01@unipa.it; giada.adelfio@unipa.it

Milind Sharma
Department of Atmospheric Sciences, Texas A&M University, USA
e-mail: milindsharma@tamu.edu

A lightning mapping array (LMA) can map the initiation and subsequent propagation of a lightning discharge in four dimensions (space-time; x, y, z, and t). A lightning flash propagates as a bidirectional leader, emitting very high frequency (VHF) signals from both negative and positive leaders. These signals, falling in the 60-66 MHz frequency range, can be detected by the LMA, which then triangulates the spatio-temporal location of each VHF source (Rison et al., 1999). Lightning data can be thought as spatio-temporal point patterns in different ways: (1) raw VHF source points (can be binned in fixed size time window; too many); (2) flash initiation centroids (one point per flash); (3) gridded flash products (e.g., flash extent or flash initiation density). Scenarios (2) and (3) can further be classified as marked point patterns since we have additional metadata for each flash or each grid point.

Point process methodology, rooted in statistical theory, enables the modelling of events occurring in space and time, making it an ideal tool for studying flash initiation patterns. In this paper, we exploit point process theory to estimate the intensity of lightning flash initiation. In particular, the aim of the paper is to understand the role of environmental spatio-temporal covariates on the occurrence of flash initiation events. To this end, we develop a computational strategy to fit spatio-temporal Poisson models, including external spatio-temporal covariates.

The structure of the paper is as follows. Section 2 introduces the data. In Section 3 we illustrate the proposed methodology to fit spatio-temporal Poisson processes depending on external covariates. Then, Section 4 contains the results of the application to the lighting data, and the paper ends with conclusions in Section 5.

2 Lightning data and spatio-temporal covariates

Since each lightning flash comprises multiple points, we usually use spatio-temporal clustering techniques like DB-SCAN (density-based clustering algorithm; Ester et al., 1996) to cluster points close to each other (in space-time) into lightning flashes (Bruning, 2015). We retain all lightning flashes comprising ten or more VHF sources when all sources occur within 3 km distance and 150 ms from the first identified source in that flash (MacGorman et al., 2008).

Figure 1 (a) shows the plot of the total lightning flash rate (sum of intra-cloud and cloud-to-ground flashes per minute) with three distinct phases in the lightning flash rate: (First) Monotonic increase in flash rate between 20:02:00 and 20:30:00 UTC; (Second) Minor fluctuations but stable flash rate of $\sim$150 flashes per minute between 20:30:01 and 21:05:00 UTC; (Third) Gradual decline in flash rate from 21:05:01 UTC onwards until the demise of the storm around 22:30:00 UTC. Figure 1 (b) depicts the multitype spatial point pattern, with the number of flashes in each time frame being 2278, 5687 and 3457, respectively.

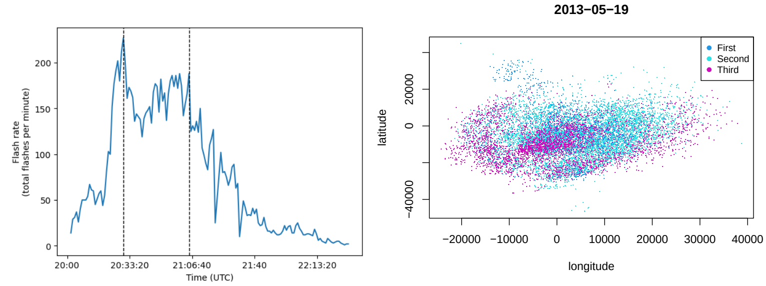

Fig. 1: (a) total count of flashes per minute; (b) multitype spatial point pattern.

We use the surface meteorological variables data from the Oklahoma mesonet as environmental covariates. For the purposes of this study, we use the 5-minute interval data for dry bulb temperature (TAIR), relative humidity (RELH), and equivalent potential temperature (THETA_E), displayed in Figure 2. All of them are available at 5-minute intervals within a 60 km $\times$ 60 km box around the storm centre throughout the analysis period and were measured at 1.5 m above the ground level.

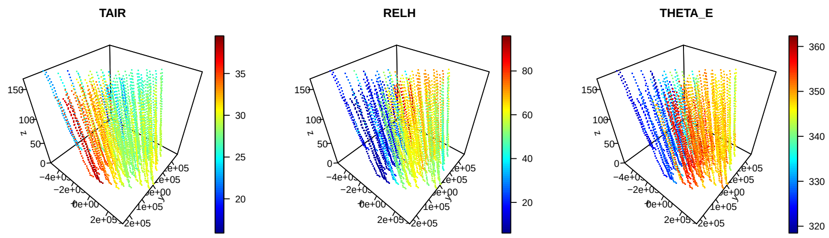

Fig. 2: Environmenal spatio-temporal covariates

3 Methodology

We assume that the template model is a Poisson process, with a parametric intensity or rate function $\lambda(u,t;\theta) = \exp(\theta^T Z(u,t))$ with spatial and temporal coordinates $u \in W, t \in T$, unknown prameters $\theta \in \Theta$, and some spatio-temporal covariates $Z(u,t)$ (D'Angelo and Adelfio, 2023). The log-likelihood is

$$\log L(\theta) = \sum_i \lambda(u_i,t_i;\theta) - \int_W \int_T \lambda(u,t;\theta) \mathrm{d}t \mathrm{d}u \quad (1)$$

up to an additive constant, where the sum is over all points u_i in the point pattern $\mathbf{x}$.

For estimation purposes, we use a finite quadrature approximation of the log-likelihood. Renaming the data points as $x_1,\dots,x_n$ with $(u_i,t_i) = x_i$ for $i = 1,\dots,n$, then generate m additional "dummy points" $(u_{n+1},t_{n+1})\dots,(u_{m+n},t_{m+n})$ to form a set of $n+m$ quadrature points (where $m > n$). Then we determine quadrature weights $a_1, \dots, a_m$ so that integrals in (1) can be approximated by a Riemann sum

$$\int_W \int_T \lambda(u,t;\theta) \mathrm{d}t \mathrm{d}u \approx \sum_{k=1}^{n+m} a_k \lambda(u_k, t_k; \theta)$$

where a_k are the quadrature weights such that $\sum_{k=1}^{n+m} a_k = l(W \times T)$ where l is the Lebesgue measure. Then, writing $y_k = e_k / a_k$, the log-likelihood (1) of the template model can be approximated by

$$\log L(\theta) \approx \sum_j a_k (y_k \log \lambda(u_k, t_k; \theta) - \lambda(u_k, t_k; \theta)) + \sum_k a_k.$$

Apart from the constant $\sum_k a_k$, this expression is formally equivalent to the weighted log-likelihood of a Poisson regression model with responses y_k and means $\lambda(u_k, t_k; \theta) = \exp(\theta Z(u_k, t_k))$. This means that the model can be maximised using the standard `glm` function, but also that covariate values must be known in every data and dummy point location. As this is typically unfeasible in practice, we first interpolate covariate values at a very fine regular grid, and then attribute to each data or dummy points the value of the closest point in three dimension. As the covariate location within the analysed region is just 205 sites, the interpolation is performed on a $12^3 = 1728$ point grid. Then, the interpolation at a (data or dummy) point location x_k is performed through the inverse-distance weighting smoothing procedure of the covariate values $Z(x_j)$ at their sampling locations $j = 1, \dots, J$. In such a case, the smoothed value at location x_k is

$$Z(x_k) = \frac{\sum_j w_j Z(x_j)}{\sum_j w_j},$$

where the weight w_j is the j-th element of the inverse pth powers of distance, $\mathbf{w} = \{w_j\}_{j=1}^J = \big\{\frac{1}{d(x_k - x_j)^p}\big\}_{j=1}^J$, with $d(x_k - x_j) = ||x_k - x_j||$ the Euclidean distance from x_k to x_j. To fit such a model, the spatio-temporal quadrature scheme is obtained by defining a spatio-temporal partition of $W \times T$ into cubes C_k of equal volume ν, assigning the weight $a_k = \nu / n_k$ to each quadrature point (dummy or data), where n_k is the number of points that lie in the same cube as the point u_k. The number of dummy points should be sufficient for an accurate estimate of the likelihood. We start with a number of dummy points $m \approx 4n$, increasing it until $\sum_k a_k = l(W \times T)$.

4 Results

We propose to model the flash lighting data by a Poisson model with a linear predictor including a non-parametric term for spatio-temporal coordinates and parametric expression for the spatio-temporal covariates, excluding the time phases, as

$$\lambda(u,t) = \exp(f(u,t) + \theta_1 Z_{TAIR}(u,t) + \theta_2 Z_{RELH}(u,t) + \theta_3 Z_{THETA\_E}(u,t))$$

where $f(\cdot)$ is a nonparametric function for $(u,t) \in W \times T$, estimated here through thin plate regression splines (Wood, 2003).

Coefficients	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-189.342	31.497	-6.011	0.000***
TAIR	-5.033	0.689	-7.302	0.000***
RELH	-0.893	0.109	-8.224	0.000***
THETA_E	1.039	0.161	6.469	0.000***

Table 1: Parametric Coefficients of the fitted Poisson model

The results in Table 1 indicate that even though the individual effects of dry air temperature and relative humidity tend to have a negative correlation with the total lightning flash rate, equivalent potential energy (which represents a combined effect of dry air temperature and humidity) has a mild positive effect. This makes physical sense since air parcels with higher potential energy being lifted upward by the storm updraft are more likely to reach cold temperatures and since lightning is a manifestation of interaction between graupel and ice crystals, both of which are active between 0 and -40 °C, more surface air parcels (with higher energy), reaching to higher altitudes can increase the lightning flash rate.

5 Conclusions

The analysed dataset represents just one supercell storm (single sample point), with limited spatiotemporal covariate data. Thereby, the preliminary results presented here should not be generalized. We plan to expand this study by applying the methodology developed for this case to a larger dataset representing multiple convective modes in different ambient environments. We would also like to compare the effect of environmental covariates for thunderstorms that predominantly lower either positive or negative charges to the ground. Other future paths include exploring more complex models, like the log-Gaussian Cox processes, multitype Poisson models, and local ones (D'Angelo et al., 2023). Indeed, multitype Poisson models could be useful to characterize the different phases of the storm by assessing the significance of the covariates' random effects and running a test of segregation.

6 Acknowledgments

The lightning data used in this study were obtained from the Oklahoma lightning mapping array data (OKLMA), provided by the U.S. NOAA National Severe Storms Laboratory. Environmental covariate data were obtained from the Oklahoma Mesonet (Brock et al., 1995; McPherson et al., 2007) stations can be found at https://doi.org/10.15763/dbs.mesonet.standard. This work was supported by the Mobilità e Formazione Internazionali - Miur INT project "Sviluppo di metodologie per processi di punto spazio-temporali marcati funzionali per la previsione probabilistica dei terremoti" and the European Union - NextGenerationEU, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP C93C22005270001).

References

Brock, F. V., Crawford, K. C., Elliott, R. L., Cuperus, G. W., Stadler, S. J., Johnson, H. L., and Eilts, M. D. (1995). The oklahoma mesonet: a technical overview. *Journal of Atmospheric and Oceanic Technology*, 12(1):5–19.

Bruning, E. (2015). lmatools: lmatools-v0.5z-stable.

D'Angelo, N. and Adelfio, G. (2023). *stopp: Spatio-Temporal Point Pattern Methods, Model Fitting, Diagnostics, Simulation, Local Tests*. R package version 0.1.0.

D'Angelo, N., Adelfio, G., and Mateu, J. (2023). Locally weighted minimum contrast estimation for spatio-temporal log-gaussian cox processes. *Computational Statistics & Data Analysis*, 180:107679.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231.

MacGorman, D. R., Rust, W. D., Schuur, T. J., Biggerstaff, M. I., Straka, J. M., Ziegler, C. L., Mansell, E. R., Bruning, E. C., Kuhlman, K. M., Lund, N. R., et al. (2008). Telex the thunderstorm electrification and lightning experiment. *Bulletin of the American Meteorological Society*, 89(7):997–1014.

McDonald, J. R. and Mehta, K. C. (2006). *A recommendation for an Enhanced Fujita scale (EF-Scale)*. Wind Science and Engineering Center, Texas Tech University.

McPherson, R. A., Fiebrich, C. A., Crawford, K. C., Kilby, J. R., Grimsley, D. L., Martinez, J. E., Basara, J. B., Illston, B. G., Morris, D. A., Kloesel, K. A., et al. (2007). Statewide monitoring of the mesoscale environment: A technical update on the oklahoma mesonet. *Journal of Atmospheric and Oceanic Technology*, 24(3):301–321.

Rison, W., Thomas, R. J., Krehbiel, P. R., Hamlin, T., and Harlin, J. (1999). A gps-based three-dimensional lightning mapping system: Initial observations in central new mexico. *Geophysical research letters*, 26(23):3573–3576.

Sharma, M., Tanamachi, R. L., Bruning, E. C., and Calhoun, K. M. (2021). Polarimetric and Electrical Structure of the 19 May 2013 Edmond–Carney, Oklahoma, Tornadic Supercell. *Mon. Wea. Rev.*, 149(7):2049–2078.

Wienhoff, Z. B., Bluestein, H. B., Wicker, L. J., Snyder, J. C., Shapiro, A., Potvin, C. K., Houser, J. B., and Reif, D. W. (2018). Applications of a Spatially Variable Advection Correction Technique for Temporal Correction of Dual-Doppler Analyses of Tornadic Supercells. *Mon. Wea. Rev.*, 146(9):2949–2971.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(1):95–114.