GRASPA 2015
BARI 15-16 JUNE

Environmental SmartCities: statistical mapping of environmental risk for natural and anthropic disasters in Chile

O. Nicolis

Instituto de Estadística, Universidad de Valparaíso, Av.Gran Bretaña 1111, 234000 Valparaíso, Chile; orietta.nicolis@uv.cl

Abstract. *The main aim of the work is to build statistical environmental risk maps for natural disasters in Chile, using spatio-temporal models in order to improve the assessment, prevention and mitigation of impacts. To this end, we analyze the spatial and temporal variability of the observed points, we study the dependence from the exogenous variables, and we create risk maps. Finally, we display the results in web platforms for mobile devices. Several environmental phenomena are considered such as earthquakes, wildfires, and air pollution. In all cases the methodology is based on the assumption that data can be modeled as a spatio-temporal process, although specific models are proposed for each category.*

Keywords. *Natural disasters; Spatio temporal modeling; Hazard map.*

1 Introduction

A natural risk can be defined as the probability that a natural phenomenon result in a natural disaster, called extraordinary event. Events that can potentially result in natural disasters can be classified as earthquakes, tsunamis, forest fires, avalanches, volcanic eruptions, etc.. The probability of their occurrence may not be homogeneous in space and time. Then, spatial variations can be displayed on a map, giving rise to a environmental hazard map. Predictive maps show the probability of occurrence taking account the time of occurrence of natural events. To construct this type of maps, it is required to model risk, which can be done by considering the event occurrence as a random point process. Such kind of processes can be defined as a random collection of points falling in a specific space. In most applications, each point represents the time and/or location of an event, such as a the epicenter of an earthquake or the centroid locations of forest fires. A spatio-temporal point process is defined as a random collection of points, where each point represents the time and location of an event ([10]). Typically, the spatial locations are recorded in three spatial coordinates, such as longitude, latitude and height or depth, though sometimes only one or two spatial coordinates are available or of interest. Catalogs of spatio-temporal data may also include explanatory variables, which could may be given by a spatial function $Z(x,y)$ defined at all spatial locations (x,y) (for example, as mentioned, altitude, temperature, wind speed and wind direction in the study of wildfires) or by another spatial pattern or line segment pattern (for example, the geological faults for evaluating the earthquake risk). Any analytical spatio-temporal point process is characterized uniquely by its associated conditional rate process or conditional intensity, which is usually indicated by

(t,x,y) ([7], [4]). The conditional intensity (t,x,y) may be thought as the frequency with which events are expected to occur around a particular location (t,x,y) in spatio-temporal, conditional on the prior history, H_t, of the point process up to time t. Formally, the conditional rate (t,x,y) may be defined as a limiting conditional expectation, provided the limit exists. The behavior of a spatio-temporal point process N is typically modeled by specifying a functional form for (t,x,y), which may be estimated non-parametrically or via a parametric model (see [5], [9], and [22]). In general, (t,x,y) depends not only on (t,x,y), but also on the times and locations of preceding events. Processes that display substantial spatial heterogeneity, such as earthquake epicenters, are sometimes modeled by stationary processes in time but not space. A commonly used form for such models is a spatial-temporal generalization of the Hawkes model, known as ETAS models proposed by [14]. The conditional intensity of ETAS models can be written as:

$$\lambda_\theta(t,s|\mathcal{H}_t) = \mu f(s) + \sum_{t_j<t} g(t-t_j|m_j)\ell(x-x_j, y-y_j|m_j) \tag{1}$$

where the sum is over all points (t_i,x_i,y_i) with $t_i < t$. The functions μ and g represent the deterministic background rate and clustering density (with magnitude $m > m_c$), respectively. Often μ is modeled as merely a function of the spatial coordinates (x,y), and may be estimated nonparametrically as in [14]. A variety of forms has been proposed for clustering the density g ([13]; [14]; [23]). Also, different estimation algorithms have been proposed for reducing the computational time ([15]; [20]; [1]). Sometimes, the conditional intensity λ is modeled as a product of marginal conditional intensities

$$\lambda(t,x,y) = \lambda_1(t)\lambda_2(x,y),$$

where forms embody the notion that the temporal behavior of the process is independent of the spatial behavior and, in the latter case, that furthermore the behavior along each of the spatial coordinates can be seen as independent. A wide range of models describe processes in which aggregation or repulsion between events is presented, known as "shot noise" [12]. In these models, the intensity function has the form

$$\lambda(x,y) = \lambda_1(x,y)\lambda_2(t)S(x,y,t)$$

, where $\lambda_1(x,y)$ is the intensity in the space, which can be modeled as a function of environmental and climatic variables; $\lambda_2(t)$ is the temporal intensity which depends on temporal variables, and $S(x,y,t)$ is the shot noise term, which allows us to model variability.

Some environmental disasters are caused by human activities that alter normal environment. Atmospheric pollution is an example. In this case, the risk is that the concentration of a contaminant is greater than a threshold, considered dangerous to human health. Mapping the concentration of a pollutant, it is possible to identify areas most at risk than others and estimate the human exposure. The data of air pollution are usually collected by a spatial monitoring network at regular intervals (say, every hour or day or week). Thus, the data analysis has to take account temporal correlations as well as spatial correlations. Geostatistical approaches to spatio-temporal prediction in environmental science rely on appropriate correlation/covariance models ([3]). Let Z be a spatial-temporal process (i.e. concentrations of PM2.5) observed at the spatial locations $s_1,\ldots,s_n \in D$, where $s_i = (x_i,y_i)$, for $i = 1,\ldots,n$, and times $t_1,\ldots,t_m \in T$, a simple spatio-temporal model can be defined as

$$Z(s,t) = \mu(s,t) + \varepsilon(s,t) \tag{2}$$

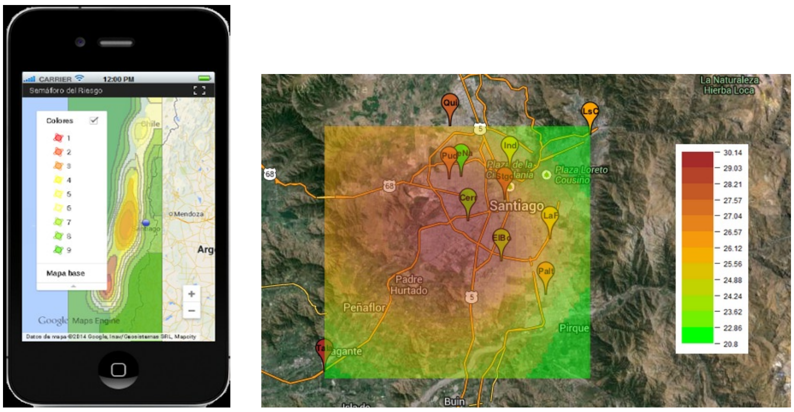
where $\mu(s,t) = X(s,t)\beta$ is a deterministic trend component depending on the exogenous variables $X(s,t)$ (that is, temperature, relative humidity, wind speed, wind direction, land use, elevation, etc.) and $\varepsilon(s,t)$ is a zero-mean intrinsically stationary spatio-temporal stochastic process which covariance structure is normally specified by an isotropic parametric function (that is, exponential, Gaussian, Matérn). Many extensions of the model 2 have been proposed in the literature ([6], [18], [2]). The following model has been recently proposed by [16] for modelling the trend of air pollutant concentrations

$$\mu(s,t) = \sum_{l=1}^{L} \beta_l X_l(s,t) + \sum_{k=1}^{K} \gamma_k(s)\psi_k(t). \tag{3}$$

where $X_l(s,t)$ are spatio-temporal covariates; β_l are the coefficients for the spatio-temporal covariates; the $\{\psi_k(t)\}_{k=1}^{K}$ is a set of (smooth) temporal basis functions with $\psi_1(t) = 1$ estimated by the modified singular value decomposition (see,[8] and [21]), and the $\gamma_k(s)$ are spatially varying coefficients for the temporal functions.

2 Visualizing on web and mobil devices

The outputs of the spatio temporal models described in Section 1, can be used for visualizing the environmental hazard through web platforms for mobil devices. As an example, we show the results of the ETAS model described by Eq. 1 for visualizing the seismic risk in Chile. The map of Fig. 1 (a) represents the estimated seismicity of Chile using the earthquake catalogue from the January 2000 to May 2014. The results of the ETAS model have been classified into nine categories of colors representing the different seismic hazard rate. The GPS system of the mobile device allows to show if the user is in a high level risk position. A similar result can be obtained for assessing the daily level of air pollution as represented in Fig. 1 (b).



(a) (b)

Figure 1: (a) Estimating seismic hazard using earthquake events in the period 2000-2007 on *Google Earth* platform for mobile device. The blu circle indicate the position of the user. (b) Visualization of the estimated average PM2.5 concentrations for the month June, 2011 on *Google Map* platform.

Acknowledgments. The present work has been partially supported by Fondecyt grant 1131147.

References

[1] Chiodi, M., Adelfio, G. (2011). Forward Likelihood-based predictive approach for spatio-temporal processes. *Environmetrics*, 22, 749–757.

[2] . Clark. J.S., Gelfand A.E. (eds.) (2006). *Hierarchical Modelling for the Environmental Sciences*. Oxford University Press, Oxford, England.

[3] Cressie, N., Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, New York.

[4] Daley, D., Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*, Springer-Verlag, Berlin.

[5] Diggle, P. (1985). A kernel method for smoothing point process data, Applied Statistics 34, 138–147.

[6] Fassò A., Cameletti M., Nicolis O. (2007). Air quality monitoring using heterogeneous networks, *Environmetrics*, **18**, 245-264.

[7] Fishman, P.M., Snyder, D.L. (1976). The statistical analysis of space-time point processes, *IEEE Transactions on Information Theory*, IT-**22**, 257–274.

[8] Fuentes M., Guttorp P., Sampson P.D. (2006). Using transforms to analyze space-time processes. In B Finkenstadt, L Held, V Isham (eds.), *Statistical Methods for Spatio-Temporal Systems*, 77–150. CRC/Chapman and Hall.

[9] Guttorp, P., Thompson, M. (1990). Nonparametric estimation of intensities for sampled counting processes, *Journal of the Royal Statistical Society, Series B* **52**, 157–173.

[10] Guttorp, P.M., Brillinger, D.R., Schoenberg, F.P. (2002). Point processes, spatial. in *Encyclopedia of Environmetrics*, El-Shaarawi. A., Piegorsch, W. (eds). Wiley, New York, **3**, 1571–1573.

[11] Karr, A. (1991). *Point Processes and Their Statistical Inference*, 2nd Edition, Marcel Dekker, New York.

[12] Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability*, **35**, 614–640.

[13] Musmeci, F., Vere-Jones, D. (1992). A spatio-temporal clustering model for historical earthquakes, *The Annals of the Institute of Statistical Mathematics*, **44**, 1–11.

[14] Ogata, Y. (1998). Spatio-temporal point process models for earthquake occurrences, *The Annals of the Institute of Statistical Mathematics*, **50**, 379–402.

[15] Ogata, Y., Zhuang, J. (2006). Spatio-temporal ETAS models and an improved extension, *Tectonophysics*, **413**, 13-23.

[16] Olives, C., Kaufman J.D., Sheppard L., Szpiro A.A., Lindström, J., Sampson P.D. (2014). Reduced-rank spatio-temporal modeling of air pollution concentrations in the multi-ethnic study of atherosclerosis and air pollution. *Annals of Applied Statistics*. **8**(4):2509–2537.

[17] Peng, R.D., Schoenberg, F.P., Woods, J. (2005). A spatio-temporal conditional intensity model for evaluating a wildfire hazard index. *Journal of the American Statistical Association*, **100**, 26-35.

[18] Sahu, S.K., Nicolis, O. (2008). An evaluation of European air pollution regulations for particulate matter monitored from a heterogeneous network. *Environmetrics*, **20**, 943–961.

[19] Schoenberg, F.P., Brillinger, D.R., Guttorp, P.M. (2002) Point processes, spatial-temporal. in *Encyclopedia of Environmetrics*, El-Shaarawi. A., Piegorsch, W. (eds). Wiley, New York, **3**, 1573–1577.

[20] Schoenberg, F.P. (2013) Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America*, **103**, 601-605.

[21] Szpiro, A. A., Sampson, P. D., Sheppard, L., Lumley, T., Adar, S. D. and Kaufman, J. D. (2010). Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics*, **21**, 606–631.

[22] Vere-Jones, D. (1992) Statistical methods for the description and display of earthquake catalogs, in S*tatistics in the Environmental and Earth Sciences*, A. Walden, Guttorp, P. (eds). Arnold, London, 220–246.

[23] Zhuang, J., Ogata, Y., Vere-Jones, D. (2002). Stochastic declustering of spatio-temporal earthquake occurrences. *Journal of the American Statistical Association*, **97**, 369–379.