METMA 9

Montpellier, 13-15 June 2018

Book of Extended Abstracts

Contents

A space-time branching process with covariates

Giada Adelfio[1,2,*], Marcello Chiodi[1,2]

[1] *Dipartimento di scienze Economiche Aziendali e Statistiche, University of Palermo;*
[2] *Istituto Nazionale di Geofisica e Vulcanologia (INGV), Palermo;*
giada.adelfio@unipa.it, marcello.chiodi@unipa.it
[*] *Corresponding author*

Abstract. *The paper proposes a stochastic process that improves the assessment of seismic events in space and time, considering a contagion model (branching process) within a regression-like framework. The proposed approach develops the Forward Likelihood for prediction (FLP) method including covariates in the epidemic component.*

Keywords. *Space-time Point Process; FLP; covariates; ETAS model*

1 Introduction

Contagious phenomena are well described in space and time by self-exciting point processes, where the conditional intensity function is obtained as the sum of the long-term variation component (called endemic) and the short-term variation one (named epidemic). This kind of models have been widely used in the literature: infectious disease [10], crime [11], quakes [12], [1]. To model earthquake activity in space and time accounting both for the endemic (background activity) and epidemic (aftershocks) effect, the Epidemic-Type Aftershock Sequences (ETAS) model is used. It describes events starting from their space-time coordinates (and magnitude as mark) and incorporates seismological laws in a mechanistic approach (e.g. the Omori law) as a natural one in the context of earthquake data. In this paper, we aim at providing an improved computational framework for further theoretical and empirical developments for studying and describing epidemic phenomena, where there is a contagious effect of the previous history, in space an time, and of specific covariates. In particular, we suggest the use of a branching-type model for earthquake description (the ETAS model) in a regression-oriented version modelling, accounting also for external covariates, for explaining some of the overall variability of the studied phenomenon and reducing the unpredictable variability. We provide developments of the Forward Likelihood for prediction (FLP) method [5] for estimating the ETAS model components, introducing covariates for the epidemic part, for a more realistic description of observed patterns.

2 Branching point processes and ETAS model with covariates

Branching processes are used to model reproduction phenomena. These models have been recently considered for the description of different applicative fields: biology [4], demography [9], epidemiology

[3]. Any analytic space-time point process defined in $[0,T] \times W \subset \mathbf{R}^2, T > 0$ is uniquely characterized by its associated *conditional intensity function* (CIF) [8]:

$$\lambda(t,\mathbf{s}|H_t) = \lim_{\Delta t, \Delta \mathbf{s} \to 0} \frac{\mathrm{E}\left[N([t,t+\Delta t],[\mathbf{s},\mathbf{s}+\Delta \mathbf{s}]|H_t)\right]}{\Delta t \Delta \mathbf{s}}$$

where H_t is the space-time occurrence history of the process up to time t, $\Delta t, \Delta \mathbf{s}$ are time and space increments, $\mathrm{E}\left[N([t,t+\Delta t],[\mathbf{s},\mathbf{s}+\Delta \mathbf{s}]|H_t)\right]$ is the history-dependent expected number of events occurring in the volume $\{[t,t+\Delta t) \times [\mathbf{s},\mathbf{s}+\Delta \mathbf{s}]\}$. Generally, intensities $\lambda(\cdot)$ depend on some unknown parameter $\boldsymbol{\theta}$, so that we have $\lambda(\cdot,\boldsymbol{\theta})$. The CIF represents the instantaneous rate or hazard for events at time t and location $\mathbf{s}$ given all the observations up to time t, conditioning on the random past history of the process. In general, the conditional intensity function of the branching model is defined as the sum of a term describing the large-time scale variation (spontaneous activity or background) and one relative to the small-time scale variation due to the interaction with the events in the past (induced activity or offsprings):

$$\lambda_{\boldsymbol{\theta}}(t,\mathbf{s}|H_t) = \mu f(\mathbf{s}) + \tau_{\boldsymbol{\phi}}(t,\mathbf{s}) \tag{1}$$

with H_t the past history of the process, $\boldsymbol{\theta} = (\boldsymbol{\phi},\mu)'$, the vector of parameters of the induced intensity ($\boldsymbol{\phi}$) together with the parameter of the background general intensity (μ), $f(\mathbf{s})$ the space density, and $\tau_{\boldsymbol{\phi}}(t,\mathbf{s})$ the induced intensity (or self-exciting component), given by:

$$\tau_{\boldsymbol{\phi}}(t,\mathbf{s}) = \sum_{t_j<t} \nu_{\boldsymbol{\phi}}(t-t_j,\mathbf{s}-\mathbf{s}_j).$$

The self-exciting component of the model essentially provides a description of the intensity at a space-time location $(t,\mathbf{s})$ caused by each previous event. In such models, we have to simultaneously estimate the different components of the intensity function (large-time scale and small-time scale). If the large-time scale component $\mu f(\mathbf{s})$ in (1) is known, the parameters $\boldsymbol{\phi}$ can be usually estimated by Maximum Likelihood method. In applications, the large-time scale component $\mu f(\mathbf{s})$ is usually estimated trough nonparametric techniques, like kernel estimators.

In seismological context, the branching process ETAS model has been introduced [12]. Starting from model (1), the ETAS conditional intensity function can be written as follows:

$$\lambda_{\boldsymbol{\theta}}(t,\mathbf{s}|H_t) = \mu f(\mathbf{s}) + \sum_{t_j<t} g(t-t_j|m_j)\ell(\mathbf{s}-\mathbf{s}_j|m_j) \tag{2}$$

with m_j the magnitude of the j-th event, $g(\cdot)$ the Omori law for occurrence density of aftershocks in time and $\ell(\cdot)$ the spatial distribution, conditioned to magnitude of the generating event. Since the criticality of the simultaneous estimation of the background intensity and the triggered intensity components of a Epidemic type model, the FLP approach was developed (see [5], [7]). It is a nonparametric estimation procedure, used for the large time scale component, based on the subsequent increments of log-likelihood obtained adding an observation one at a time, to account for the information of the observations until t_k on the next one. That provides a simultaneous estimation of the two parametric components of a branching-type model, alternating the standard likelihood method, to estimate the parameters, with the FLP approach, to estimate the nonparametric part. Given the lack of specific open-source tools, the package etasFLP [7] [6] provides tools to implement this mixed approach for a wide class of ETAS models for the description of seismic events, developed in the R environment.

In this paper, we propose an additive-multiplicative model for the conditional intensity function of a space-time point process, incorporating a forward predictive likelihood estimation approach for semi-parametric intensity function. Starting from the definition provided in eq. (1), we propose to modify the offspring component, accounting for a vector of covariates. As proposed by [10] in a context of infection occurrences, we incorporate the space-time phenomenological laws of the triggering part of the ETAS model with the effects of covariates. This triggering function is factorized into separate effects of external information, time and relative location, such that:

$$\lambda_{\theta}(t,\mathbf{s}|H_t) = \mu f(\mathbf{s}) + \sum_{t_j<t} exp(\eta_j)\tilde{g}(t-t_j|m_j)\tilde{\ell}(\mathbf{s}-\mathbf{s}_j|m_j) \tag{3}$$

where $(t_j, \boldsymbol{s}_j)$ is the time and location of individual occurrence j, $\eta_j = \boldsymbol{\beta}'\boldsymbol{z}_j$ is a linear predictor based on the vector of unpredictable variables $\boldsymbol{z}_j$ of each event, and $\tilde{g}$ and $\tilde{\ell}$ are defined as in eq (2), accordingly modified. In the seismic context, the proposed approach would provide a more general formalism for the earthquake occurrence in space and time. Indeed, the main idea is that the effect on the future activity does not depend only on the closeness of the previous events, but also on specific characteristics of the main event, like magnitude, as usual, and also further information, such as geological features.

3 Application to the Italian earthquakes and comments

We report some of the results of the proposed ETAS-FLP approach with covariates, starting from the Italian catalogue of the space-time Italian seismicity, from May 5th, 2012 to May 7th, 2016, with 2.5 as the threshold magnitude (i.e. the lower bound for which earthquakes with higher values of magnitude are surely recorded in the catalogue). The catalogue reports the usual hypocentral coordinates (longitude, latitude, depth, time) together with the magnitude of the events, and also some additional information, such as: the hypocentral uncertainty, the distance from the nearest station (for shallow earthquakes, this distance should be sufficiently small), a measure of the quality of the location (named rms), the number of stations that recorded the event (this number is heavily influenced by the magnitude of the event and strongly influences the accuracy of the location) and the distance from the nearest fault (i.e. the identified earthquakes sources in that area). Estimating the ETAS model as in eq. (2) by using the FLP approach and accounting for the epicentral coordinates (longitude, latitude and time) and the magnitude of the inducing event, results (not here reported) are not completely satisfying, suggesting, as usual, some lack of fitting mostly due to the triggered component. However, adding also the available covariates, that is considering the model in eq. (3), the best estimated one includes, together with the magnitude, also the depth, the distance from the nearest station and the distance from the nearest fault. In particular, the last two covariates have both a negative effect on the space-time reproducing activity. Diagnostic results suggest a satisfying fitting, as shown in the Figure 3 (see [2] for the residual description).

The reported results, tough partial and provisional, confirm our intuition reported in previous studies (e.g. [2]). Indeed, the need of a more flexible model for the space-time triggered component of the ETAS model is often revealed, although the background seismicity is well described by the FLP estimated intensity. In our opinion, considering external information (such as geological information related to faults distribution) for the description of spatio-temporal earthquakes is a innovative and promising perspective of study, even relevant in different fields of research.

Acknowledgments. This paper has been supported by the national grant of the Italian Ministry of Education University and Research (MIUR) for the PRIN-2015 program “Prot. 20157PRZC4. PI: G.Adelfio".

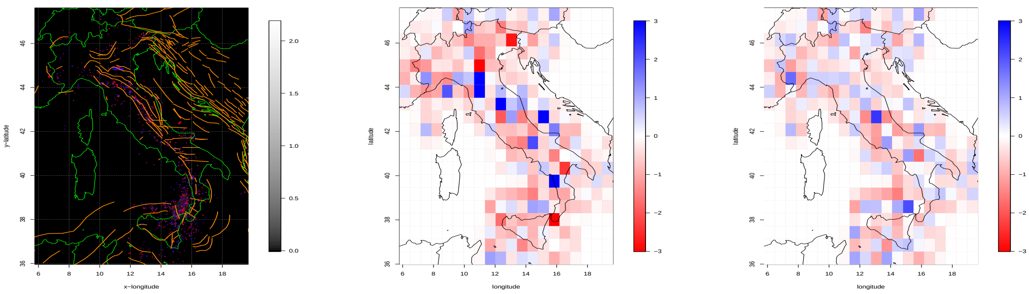

Figure 1: Output for the estimated ETAS-FLP model with covariates: estimated total intensity together with the observed points (old in blue and recent in red) and available line-faults (on the left); space residuals for the model (in the middle), space residuals for the background (on the right).

References

[1] G. Adelfio and M Chiodi, *Alternated estimation in semi-parametric space-time branching-type point processes with application to seismic catalogs*, Stochastic Environmental Research and Risk Assessment **29** (2015), 443–450.

[2] G. Adelfio and M. Chiodi, *Flp estimation of semi-parametric models for space-time point processes and diagnostic tools*, Spatial Statistics **14** (2015), 119–132.

[3] E. Balderama, F.P. Schoenberg, E. Murray, and P.W. Rundel, *Application of branching point process models to the study of invasive red banana plants in costa rica*, JASA **107** (2012), 467–476.

[4] G. Caron-Lormier, J.P. Masson, N. Menard, and J.S. Pierre, *A branching process, its application in biology: Influence of demographic parameters on the social structure in mammal groups*, Journal of Theoretical Biology **238** (2006), 564–574.

[5] M. Chiodi and G. Adelfio, *Forward likelihood-based predictive approach for space-time processes*, Environmetrics **22** (2011), 749–757.

[6] M. Chiodi and G. Adelfio, *etasflp: Estimation of an etas model. mixed flp (forward likelihood predictive) and ml estimation of non-parametric and parametric components of the etas model for earthquake description*, R package version 1.4.1. (2017).

[7] M. Chiodi and G. Adelfio, *Mixed non-parametric and parametric estimation techniques in R package etas-FLP for earthquakes' description*, Journal of Statistical Software **76** (2017), no. 3, 1–28.

[8] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes*, second ed., New York: Springer-Verlag, 2003.

[9] R. A. Johnson and J. R. Taylor, *Preservation of some life length classes for age distributions associated with age-dependent branching processes*, Statistics and Probability Letters **78** (2008), 2981–2987.

[10] S. Meyer, L. Held, and M. Hohle, *Spatio-temporal analysis of epidemic phenomena using the r package surveillance*, Journal of Statistical Software **77** (2012), no. 11.

[11] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, *Self-exciting point process modeling of crime*, Journal of the American Statistical Association **106** (2011), no. 493, 100–108.

[12] Y. Ogata, *Space-time point-process models for earthquake occurrences*, Annals of the Institute of Statistical Mathematics **50** (1998), no. 2, 379–402.