A Bayesian changepoint analysis of spatio-temporal point processes, with application to radioactive particle data

L. Altieri[1], E. M. Scott[2], D. Cocchi[1] and J. B. Illian[3]

[1] *University of Bologna, Department of Statistics; linda.altieri@unibo.it, daniela.cocchi@unibo.it*
[2] *University of Glasgow, School of Mathematics and Statistics; Marian.Scott@glasgow.ac.uk*
[3] *University of St Andrews, CREEM, School of Mathematics and Statistics; janine@mcs.st-and.ac.uk*

Abstract. *This work introduces a Bayesian approach for detecting multiple unknown change points over time in the inhomogeneous intensity of a spatio-temporal point process with spatial and temporal dependence within segments. We propose a new method for detecting changes by fitting a spatio-temporal log-Gaussian Cox process model using the computational efficiency and flexibility of INLA, and studying the posterior distribution of the potential changepoint positions. A simulation study assesses the validity and properties of the proposed method, before the approach is applied to examine potential unknown change points in the intensity of radioactive particles found on Sandside beach, Dounreay.*

Keywords. *Changepoint; Spatio-temporal point processes; LGCP; INLA*

1 Introduction

Changepoint analysis is a well-established area of statistical research, nevertheless for spatio-temporal point processes this appears to be as yet relatively unexplored. The basic assumption in a changepoint analysis is that data are ordered and split into segments, following the same model but under different parameter specifications [5]. The other common assumption is that observations are *i.i.d.*. Modelling dependence within data segments in the context of unknown multiple change points is currently a challenge: when dependence is allowed, the segments' marginal likelihood usually becomes intractable, and the computational complexity of the problem increases; there is a need for fast methods providing an accurate and tractable approximation of the likelihood. Recent work by Wyse et al. [5] uses Integrated Nested Laplace Approximation (INLA) to build a model allowing for dependence within segments. INLA is an alternative, computationally efficient approach to MCMC methods to obtain the posterior distribution of both the number of change points and their positions.

We address the analysis of temporal change points in a spatially inhomogeneous intensity function defining a point process observed over a window. The issues to face concern the complexity of such data and the inclusion of spatial dependence between points and temporal dependence within time segments, the latter being currently a challenge even in the changepoint analysis of simply temporal data series. We develop an approximate likelihood based methodology to detect change points and obtain estimates of the two-dimensional intensity function at each time point. Our work is set in the context of spatio-temporal log-Gaussian Cox point processes (LGCPs). Cox processes assume that the point distribution over space is due to stochastic underlying heterogeneity, modelled as a random intensity function $\Lambda(s)$; given $\Lambda(s)$, the distribution of points follows an inhomogeneous Poisson process. In LGCPs the logarithm of the intensity surface in an observation window W is assumed to be a (latent) Gaussian field $Z(s)$, *i.e.* $\Lambda(s) = \int_W \lambda(s)ds = \exp(Z(s))$. LGCPs constitute a very flexible class of models that can be adapted to the spatio-temporal case and fitted using INLA [3]. The INLA approach has several fundamental advantages: it is an effective computational tool for model implementation; its efficiency allows an extension from the temporal to the spatio-temporal context; the likelihood values resulting from different changepoint positions can be evaluated, to choose the best change-point position a posteriori. We present a simulation study of this approach in the spatio-temporal point process context; unlike traditional changepoint detection algorithms (see [2]), with this method the 3 dimensions of the problem (two spatial and one temporal) are maintained. We propose a Bayesian technique allowing decisions on whether, and how many, temporal change points are present.

2 Methodology

We define a change point under two increasingly complex point process models, and consider the case of multiple change points at unknown locations; we discretise the observation window into a fine grid, and assume $y_{ts} \sim Poi(|C|\lambda_{ts})$ is the number of points at time $t = 1,\dots,T$ in cell $s = 1,\dots,S$, where $|C|$ is the cell area. We initially consider a model with a fixed and a temporal effect which assumes a spatially homogeneous intensity λ_t, expressed under each hypothesis (for the alternative, the simple case is of a single change point) as:

$$\begin{array}{lll} H_0: & \log(\lambda_t) = \mu + \phi_1 + \varepsilon_t & \text{for } t = 1,\dots,T \\ H_1: & \log(\lambda_t) = \mu_1 + \phi_{1,1} + \varepsilon_t & \text{for } t < \tau^* \\ & \log(\lambda_t) = \mu_2 + \phi_{1,2} + \varepsilon_t & \text{for } t \geq \tau^* \end{array} \tag{1}$$

where μ is the fixed effect and within each time segment $\phi_{1,t}$ is a random effect modelled as an AR(1) and ε is an error term. Under H_0 all values over both space and time depend on a single value for μ, while under H_1 μ_t is constant over space but allowed to vary over time. For a single change point, we first have to detect where the change occurs, and then we estimate two values for μ; in the more general case of k change points, we split the equation under H_1 into $k+1$ segments defined by the ordered changepoint locations $\tau_1, \tau_2, \dots, \tau_k$. The same holds for the temporal effect.

For the most complicated model we consider an offset term, a temporal effect and a spatial effect, and allow for spatially inhomogeneous intensity:

$$\begin{array}{lll} H_0: & \log(\lambda_{ts}) = \alpha + \phi_1 + \phi_{2,s} + \varepsilon_{t,s} & \text{for } t = 1,\dots,T \text{ and } s = 1,\dots,S \\ H_1: & \log(\lambda_{ts}) = \alpha + \phi_{1,1} + \phi_{2,1s} + \varepsilon_{t,s} & \text{for } t < \tau^* \text{ and } s = 1,\dots,S \\ & \log(\lambda_{ts}) = \alpha + \phi_{1,2} + \phi_{2,2s} + \varepsilon_{t,s} & \text{for } t \geq \tau^* \text{ and } s = 1,\dots,S \end{array} \tag{2}$$

where α is an intercept and ϕ_2 describes spatial dependence and is modelled as an intrinsic CAR, i.e. as a Random Walk in two dimensions on a lattice. In both models the temporal dependence is only assumed to be within, not across, segments. The precision parameter for both effects has a *Gamma* prior. Here we allow for spatial inhomogeneity: we build the model assuming that the spatial structure is the same

over time up to a scale parameter, and the changepoint detection identifies the time point that describes the change of scale in our data.
A Bayesian technique is applied to detect change points in the process intensity, by looking at the posterior distribution of the potential changepoint positions: once the posterior distribution is produced, values above a certain threshold are considered to identify significant change points.

3 Simulation results

Our simulation study [1] consists in fitting model scenarios (1) and (2) over both *iid* and time dependent data, generated with single or multiple change points and with a homogeneous or inhomogeneous intensity function; 100 replicates are produced for each case, with a time series length $T = 50$, over a square observation window. For the multiple changepoint detection, 3 change points are fixed and a binary segmentation algorithm is implemented [2]. In order to reduce the arbitrariness of threshold choice, this has been fixed as the lowest possible value allowing the probability of committing a type I error to be kept below the usual limits (0.01, 0.05 and 0.1). The results, given the threshold choice, produce some grey zones (wrong conclusions in some replicates) but also sensible overall conclusions, which hold over both models and all types of data. If change points are detected in H_0 time dependent generated data and more than one data series is available, the change points due to the time dependence will be discarded as they take different positions from one replicate to the other, while the fixed change points are consistently identified in the replicates. All estimated changepoint locations computed with INLA are precise and accurate: in all scenarios and cases, the spatial structure of the intensity function and its smoothness have been perfectly captured; as for the estimated value (homogeneous process) or range of values (inhomogeneous process) for the intensity, this is extremely close to the true values in most cases, with possible departures from the initial values due to the lack of detection of a change point, or to the time dependence within a segment.

4 Particle data analysis

Our study is motivated by questions concerning the detection and recovery of radioactive particles from beaches around Dounreay, North of Scotland [4]. Radioactive particles have been found on local beaches since the 1990s as a result of historic practices during nuclear fuel reprocessing at the Dounreay plant, which is currently being decommissioned (http://www.dounreay.com). The data set used gives the particles' locations on one of the local beaches, Sandside beach, during each of the years of monitoring. The underlying intensity and spatial structure are of interest, along with potential changes in its strength. There have been two major changes in the equipment used to detect the particles, representing known potential change points, as well, offshore particle retrieval campaigns are believed to have reduced the particle intensity for particles moved onshore with tides and currents with an unknown temporal lag, generating multiple unknown change points in the intensity. The dataset presents some difficulties when a changepoint analysis is carried out: the time series is not long ($T = 15$) and some patterns present very few points. Still, the questions are of interest, and the method performance has already been tested over simulated data.
Fig 1 shows a summary of the analysis: the posterior probability plot in the first panel shows that three change points have been detected over models (1) and (2); the second panel shows a simple estimate of the process intensity, without taking space into account, while the three following images are a two dimensional intensity estimate for each time segment. The first two detected change points correspond

to the periods of equipment changes and produce an increase in the point intensity; this supports that the changes in equipment has significantly improved the probability of detecting particles. The third change point is very close to the end of the series, therefore conclusions must be drawn carefully; it gives a hint of a decreasing intensity, which could be related to the offshore retrieval campaign, suggesting a reduction of the arrival of particles on Sandside beach.

5 Discussion

We have carried out a simulation study to assess the performance of a new method for detecting change points in the intensity of a spatio-temporal point process, and once evaluated the method has been applied to a small case study concerning radioactive particles found in the environment.
The posterior threshold method has the advantage of being visually immediate and easy to explain to non-statisticians; moreover, it is very flexible as the threshold choice can be adapted to the analysis context. It is interesting to observe the difference between data generated as *iid* replicates within time segments and data showing a very strong time dependence: in the latter, the series tend to drift away from the initial set values, therefore the variability of time dependent data is much higher, and spurious change points can be detected in the series; nevertheless, our method identifies those changes due to external conditions. The INLA method performs satisfactorily in terms of computational time, despite the dataset complexity and the need to fit every model multiple times, each one testing a different changepoint position.
In the case study, the detection of three change points is relevant given the context, reflecting expectations on both an increased ability to detect particles following equipment improvements, and a decreasing trend over the last few years due to an offshore retrieval campaign.

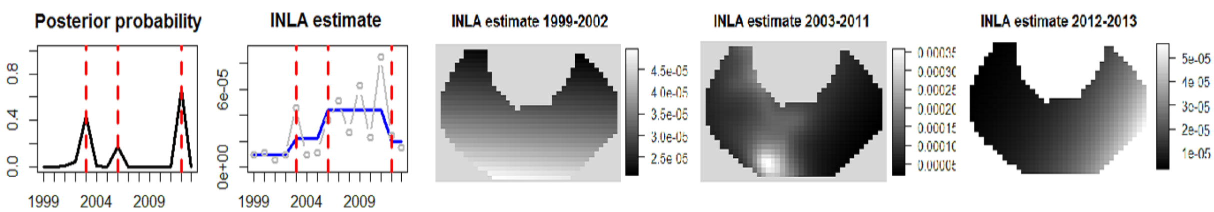

Figure 1: Changepoint analysis on particle data - results

References

[1] Altieri, L., Scott, M., Illian, J. (2014). A changepoint analysis on spatio-temporal point processes. Accepted for SIS scientific meeting 2014 conference proceedings

[2] Eckley, I.A., Fearnhead, P., Killick, R. (2011). Analysis of Changepoint Models. In: Barber, D., Cemgil, A.T., Chiappa, S. (eds.) *Bayesian Time Series Models*, Chapter 10. Cambridge University Press

[3] Illian, J., Sørbye, S., Rue, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Annals of Applied Statistics* **6**, 1499-1530

[4] Tyler, A.N., Scott, E.M., Dale, P., Elliott, A.T., Wilkins, B., Boddy, K., Toole, J., Cartwright, P. (2010). Reconstructing the abundance of Dounreay hot particles on an adjacent public beach in Northern Scotland. *Science of the Total Environment* **408**, 4495-4503

[5] Wyse, J., Friel, N., Rue, H. (2011). Approximate simulation-free Bayesian inference for multiple changepoint models with dependence within segments. *Bayesian Analysis* **6**, 501-528