A changepoint analysis of spatio-temporal point processes

Published Version:
A changepoint analysis of spatio-temporal point processes / Altieri, L; Scott, E. M; Cocchi, D; Illian, J. B.. - In: SPATIAL STATISTICS. - ISSN 2211-6753. - STAMPA. - 14:Part B(2015), pp. 197-207. [10.1016/j.spasta.2015.05.005]

Availability:
This version is available at: https://hdl.handle.net/11585/524155 since: 2015-12-13

Published:
DOI: http://doi.org/10.1016/j.spasta.2015.05.005

(Article begins on next page)

A changepoint analysis of spatio-temporal point processes

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

May 25, 2014

[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 spatially 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. In this paper, the context of the problem and the research questions are introduced, then the method is presented and discussed in detail. 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, Scotland.

Acknowledgements

The authors would like to thank . . .

1 Introduction

With this work, we aim at proposing a method for carrying out a changepoint analysis in the complex context of spatio-temporal point processes. Our study is motivated by questions on the monitoring and recovery of radioactive particles from Sandside beach, North of Scotland, due to the presence of a former nuclear reactor **?**; the distribution of the particles and their behaviour over time in the offshore and foreshore areas are of interest for cleaning purposes. Over the past 15 years, two major changes in the equipment used to detect the particles have taken place, representing known potential change points. In addition, 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, potentially generating multiple unknown change points in the intensity function. Questions on how to build a method able to detect changepoints in such a complex dataset are raised; the proposed method has to deal with the issues of spatial inhomogeneity, spatial dependence among points and temporal dependence of the process.

1.1 Theoretical issues

Changepoint analysis is a well-established area of statistical research, frequently applied in a temporal context, and less frequently over space. While some of the existing changepoint methods can potentially be extended to the general spatio-temporal context, for spatio-temporal point processes this branch of analysis 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 **?**. The other common assumption is that observations are *i.i.d.*. The aim of our work is to propose a method to find change points even when the mentioned assumptions do not hold. Modelling dependence within data segments in the context of unknown multiple change points is currently a challenge, and there is a need for fast approximate methods such as Integrated Nested Laplace Approximation (INLA), an alternative, computationally efficient approach to MCMC methods to obtain the posterior distribution of both the number of change points and their positions [ref]. The computational speed and flexibility of INLA has not been exploited for a spatio-temporal changepoint analysis yet.

Some substantial differences with regard to the standard changepoint analysis in time or in space have to be taken into account: firstly, an individual datum is not a single point but a pattern of points; secondly, the measured response variable is the point location. Frequently, point process data are collected over space, and it is not usual to have repeated measurements on the same observation window over time, in a number large enough to make a changepoint analysis sensible. Nevertheless, most of the studies on point processes aim at describing the behaviour of the intensity function, therefore its changes over time are certainly of interest, and the provision of tools for changepoint analysis on spatio-temporal processes would enlarge the number of questions that can be answered. Furthermore, in this context both the issues of spatial dependence among points and of temporal dependence within time segments have to be faced, which add further complexity to the analysis.

We do not have knowledge of a changepoint analysis carried over a spatio-temporal point process with recently developed techniques. For all the mentioned reasons, we believe a statistical analysis of change-point detection methods in the context of spatio-temporal point processes is a challenging and interesting study area.

1.2 Background

To our knowledge, the issue of dependence between data in a changepoint analysis has only been faced with Bayesian methods so far. Fearnhead (2006) proposes a method for simulating from the posterior distribution of multiple changepoints using a recursive technique that should theoretically extend to dependent data; when dependence is allowed, though, the segment marginal likelihood required by Fearnhead's method usually becomes intractable. Including any type of dependence increases the computational complexity of the problem, and fast methods providing an accurate and tractable approximation of the likelihood even in complex situations have to be developed. Recent work by Wyse, Friel and Rue **?** extends the method to allow for dependence within segments, using Integrated Nested Laplace Approximation **?** to face the well known difficulty in analytically obtaining the posterior distribution of the parameters. The authors combined recursive methods with INLA, to produce estimates for the segment marginal likelihoods, and approximations for the posterior of both the number of changepoints and their position.

Our work is set in the context of spatio-temporal log-Gaussian Cox point processes (LGCPs). Cox processes assume the point distribution over space (and potential aggregation) is due to stochastic environmental heterogeneity, modelled as a random intensity function $\Lambda(s)$ **?**; given $\Lambda(s)$, the distribution of points follows a inhomogeneous Poisson process. In LGCPs the logarithm of the intensity surface over an observation window W is assumed to be a Gaussian (latent) field $Z(s)$, i.e. $\Lambda(s) = \int_W \lambda(s)ds = \exp(Z(s),$ and conditional on $Z(s)$ the number of points $X \sim Poi(\Lambda(s))$. LGCPs constitute a very flexible class of models that can be extended to the spatio-temporal case and implemented using INLA **?**. The INLA approach has several fundamental advantages: above all, it is an effective computational tool for model implementation; this is fundamental in our context as the dataset is very complex (every single datum is a point pattern), therefore computations easily become very slow and demanding. Moreover, the efficiency of INLA allows an extension from the temporal to the spatio-temporal context; furthermore, the likelihood values resulting from different changepoint positions can be evaluated, and a posterior distribution can be approximated to choose the best change-point position a posteriori.

We address the analysis of temporal change points in a spatially inhomogeneous intensity function defining a point process observed over a window. An approximate likelihood based methodology is developed to detect change points and obtain estimates of the two-dimensional intensity function at each time point. We present a simulation study of this approach in the spatio-temporal point process context; unlike traditional changepoint detection algorithms (see **?**), with this method the 3 dimensions of the problem (two spatial and one temporal) are maintained. We propose two different Bayesian techniques allowing decisions on whether, how many and when temporal change points occur.

2 Methodology

2.1 Models

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

$$\begin{aligned} H_0: &\quad \log(\lambda_t) = \mu + \varepsilon_t \quad && \text{for } t = 1,\ldots,T \\ H_1: &\quad \log(\lambda_t) = \mu_1 + \varepsilon_t \quad && \text{for } t < \tau^* \\ &\quad \log(\lambda_t) = \mu_2 + \varepsilon_t \quad && \text{for } t \geq \tau^* \end{aligned} \tag{1}$$

where μ is the fixed effect and ε is an unstructured 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. Note that a single change point in location τ^* splits the dataset into two time segments with a different value for the intensity function (i.e. two equations under the alternative hypothesis), so 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 $M \geq 2$ change points, the equation under H_1 is split into $M+1$ segments, time intervals defined by the ordered changepoint locations $\tau_1, \tau_2, \ldots, \tau_M$. Note that each changepoint position τ_m, $m = 1,\ldots,M$, corresponds to the first point of a new segment.

The second model adds a temporal effect:

$$\begin{aligned} H_0: &\quad \log(\lambda_t) = \mu + \phi + \varepsilon_t \quad && \text{for } t = 1,\ldots,T \\ H_1: &\quad \log(\lambda_t) = \mu_1 + \phi_1 + \varepsilon_t \quad && \text{for } t < \tau^* \\ &\quad \log(\lambda_t) = \mu_2 + \phi_2 + \varepsilon_t \quad && \text{for } t \geq \tau^* \end{aligned} \tag{2}$$

and within each time segment ϕ is a random effect modelled as an AR(1), i.e. the logarithm of the intensity function at every time point is supposed to depend on its own value at the previous time. Hyperparameters are needed for the precision $\tau_\phi \sim Gamma(\alpha_\phi, \beta_\phi)$. Other time dependence structures can be easily modelled using INLA.

The first two models both consider a spatially homogeneous intensity function, therefore there is no space index s because at each time point the intensity takes a single value over the window. We now allow the intensity to vary over space as well as over time, and build a model with a spatial effect:

$$
\begin{aligned}
H_0: \quad & \log(\lambda_{ts}) = \delta + \psi_s + \varepsilon_{ts} && \text{for } t = 1, \ldots, T \text{ and } s = 1, \ldots, S \\
H_1: \quad & \log(\lambda_{ts}) = \delta + \psi_{1s} + \varepsilon_{ts} && \text{for } t < \tau^* \text{ and } s = 1, \ldots, S \\
& \log(\lambda_{ts}) = \delta + \psi_{2s} + \varepsilon_{ts} && \text{for } t \geq \tau^* \text{ and } s = 1, \ldots, S
\end{aligned}
\tag{3}
$$

where δ is a common intercept and ψ_s describes spatial dependence; it is indexed by s as we assume that the basic space unit is the grid cell, and that the intensity function is constant inside the cell. This approximation is needed for tractability reasons, but thanks to INLA we can build as fine a grid as we wish without encountering computational issues, so that the approximation error is very low and can be controlled. Under H_1, a single value defines the intensity for each cell over all the time segment, and after the change point the value for each cell changes. The spatial effect is modelled as an intrinsic CAR, i.e. as a Random Walk in two dimensions on a lattice; the model is easily specified with INLA, with a neighbourhood structure that gives non-zero (decreasing) weights to the first 12 neighbours in the lattice. This produces a very smooth spatial structure which is suitable for LGCPs, where the hypothesis is that there is a smooth underlying driver defining the intensity function. Again, the precision hyperparameter can be defined as $\tau_\psi \sim Gamma(\alpha_\psi, \beta_\psi)$.

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

$$
\begin{aligned}
H_0: \quad & \log(\lambda_{ts}) = \delta + \phi + \psi_s + \varepsilon_{ts} && \text{for } t = 1, \ldots, T \text{ and } s = 1, \ldots, S \\
H_1: \quad & \log(\lambda_{ts}) = \delta + \phi_1 + \psi_{1s} + \varepsilon_{ts} && \text{for } t < \tau^* \text{ and } s = 1, \ldots, S \\
& \log(\lambda_{ts}) = \delta + \phi_2 + \psi_{2s} + \varepsilon_{ts} && \text{for } t \geq \tau^* \text{ and } s = 1, \ldots, S
\end{aligned}
\quad . \tag{4}
$$

Please remember that in these models the temporal dependence is only assumed to be within, not across, segments. The precision parameter for both temporal and spatial effects has a *Gamma* prior that is by default set as non-informative but can be tuned according to a specific context.

When looking for a single change point, each model is run one time for every possible changepoint position, i.e. for every time point with a non-zero prior probability of being a change point. By fitting every model with INLA, a series of likelihood values is then produced, and normalised (in absence of prior knowledge) to obtain the posterior distribution of the change points: this gives, for every time point, the posterior probability of being a change point. Once the posterior is produced, methods for identifying significant change points are proposed in Section 2.2. Since each model is run many times assuming different changepoint positions, there is a need for efficient computational tools in order to obtain results in a reasonable time, and that is one of the reasons why we fit the models using INLA.

The approach for detecting multiple unknown change points is described in Section 2.3.

2.2 Changepoint detection methods

We propose two different Bayesian techniques for assessing the presence of change points. The first option derives from the Bayes Factor, used in absence of prior knowledge to decide if there is a change point **?**. The Bayes Factor can be written as

$$\gamma = \frac{\sum_{\tau} \pi(\tau) Q_1(\tau) Q_2(\tau)}{L_0} \tag{5}$$

where $Q_1(\tau)$ and $Q_2(\tau)$ are the segment maximum likelihood values, i.e. the maximum likelihoods for the two segments resulting from a changepoint position in $\tau \in \{1, \ldots, T\}$; for every value of τ, a pair of values $Q_1(\tau)$ and $Q_2(\tau)$ is returned. Besides, $\pi(\tau)$ is the prior probability of the time point τ of being a change point, and L_0 is the likelihood value obtained by running the model once under H_0.
The Bayes Factor expresses the evidence showed by data in support of the alternative model with regard to the null model. Since independence across segments is assumed, for every changepoint position the maximum likelihood value under the alternative hypothesis is $L_1(\tau) = Q_1(\tau)Q_2(\tau)$. The formula (5) can be extended to the case of a non-vague prior distribution by taking the posterior ratio, i.e. the product of likelihoods and prior ratios.
The prior weight $\pi(\tau)$ in the nominator sum shrinks each alternative likelihood value, still every element in the sum will be positive, and the greater the nominator is, the more likely it is to reject H_0. We choose a more conservative condition, by substituting the sum in the numerator with a single term:

$$\gamma_{\tau^*} = \frac{\pi(\tau^*) Q_1(\tau^*) Q_2(\tau^*)}{L_0} = \frac{\pi(\tau^*) L_1(\tau^*)}{L_0} \tag{6}$$

where τ^* is the most likely changepoint position, i.e. the one returning the highest likelihood value under H_1, $\pi(\tau^*)$ is the prior distribution on its position, and $L_1(\tau^*)$ is the maximum likelihood under the alternative hypothesis: a value for the likelihood is obtained for every potential changepoint position τ, the highest one is chosen and the corresponding location is the τ^* to test. Equivalently, we can take the logarithm of (6)

$$\gamma'_{\tau^*} = \log(\pi(\tau^*)) + q_1(\tau^*) + q_2(\tau^*) - l_0 = \log(\pi(\tau^*)) + l_1(\tau^*) - l_0. \tag{7}$$

For the model with no change points, the maximum log-likelihood value under H_0 is greater than the maximum log-likelihood value under H_1, therefore the absolute threshold for this statistic, irrespective of the model used, is zero. Indeed, differently from the frequentist likelihood ratio, when using the Bayes Factor models with more parameters do not necessarily produce higher likelihood values, as Bayes factors naturally incorporate penalization for model complexity, so there is no need for an extra penalization term as in AIC or SIC. If $\gamma'_{\tau^*} > 0$, we reject the null model of no change points, and the change point is estimated to occur at τ^*.

An alternative option we propose is another typical Bayesian way of taking decisions, i.e. by looking at the posterior distribution and fixing a posterior probability threshold for significant values: once the resulting curve is plotted, a threshold value is fixed in order to take decisions on which time points are to be considered change points.
As for the threshold choice, it is to bear in mind that greater values (closer to 1) will lead to more conservative conclusions, and smaller values (closer to 0) will detect change points more easily. Hints for discussion on the choice of the threshold are given in Section 5. This 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 model fitting the data and to the analysis context. In the special case of a known changepoint position to test, the method does not change: a posterior probability curve will be estimated all the same, and the threshold will be only used to evaluate the significance of that specific changepoint position.

2.3 Binary segmentation algorithm

Both the models presented in Section 2.1 and the detection methods presented in Section 2.2 refer to a single changepoint search. The method can be extended to a multiple unknown number of change points, the most complicated type of changepoint analysis. The hypotheses become:
H_0: no change points
H_1: ≥ 1 change points.
As for the single changepoint detection, note that H_1 is not bound to a specific changepoint position; the alternative hypothesis is very complex because it considers the presence of change points first, but then the number and positions also have to be estimated. If H_0 is rejected, $\tau^* = (\tau_1^*, \ldots, \tau_M^*)$ is a $M \times 1$ vector containing the estimated changepoint positions, a subset of $(1, \ldots, T)$.
The simplest and most straightforward way of running a multiple changepoint analysis is to use a binary segmentation method. For a general introduction to these methods we refer to [killickeckley], and in particular for point processes to the work by [park2012]. An alternative option would be to perform a simultaneous changepoint search; this method is discarded as, with our techniques, it proved to perform poorly as it tends to underestimate the number of change points: different change points will refer to changes of different magnitudes in the intensity function; when the posterior probability curve is normalised, posterior peaks will tend to flatten, and changepoint positions corresponding to smaller, but not negligible, changes happen to be considered non-significant. A binary segmentation algorithm allows to find local maxima and has proved itself better performing in our analysis.
The idea of a binary segmentation procedure, and the key to its simplicity, is to split the multiple search into a series of subsequent single changepoint searches. It is an iterative procedure, which in general can be structured into steps:

1. Run a changepoint analysis on the whole data series Y, testing the simple hypotheses
   H_0: no change points
   H_1: one change point.

2. a) If no change points are found, stop the algorithm.
   b) If one change point is found, its position is defined as τ_0^*, and data are split in correspondence of τ_0^* into two segments, Y_A $([S \times (\tau_0^* - 1)] \times 1)$ and Y_B $([S \times (T - \tau_0^* + 1)] \times 1)$. For each of the two resulting segments, go back to step 1.

3. a) If no more change points are found, the dataset has a single change point in τ_0^*.
   b) If change points τ_A^* and/or τ_B^* are detected, go back to step 2b and repeat the procedure for each segment containing a change point.

4. Repeat until either no more change points are detected in any segment, or a pre-fixed number of change points is reached, or a minimum segment length is reached for all segments.

Many binary segmentation methods can be built, according to the criterion for detecting a change point and to the criterion for stopping the search; what they have in common is that at each step the algorithm runs a single changepoint search for every segment. Intuitively, the analysis can become computationally very demanding as T and M become large, and methods are available for reducing time and memory storage requirements [ref needed]. This is nevertheless the general idea we follow, and again the computational efficiency of INLA makes this algorithm feasible even for such complex spatio-temporal data.

3 Simulation study

3.1 Simulation design

In order to assess and compare the performance of the two methods proposed in Section 2.2, we carry out a simulation study covering different situations. We fix a time series of $T = 50$ time points, and a grid of $S = 20 \times 20 = 400$ cells. The observation window is a square of area 100 and the initial intensity value is $\lambda = 1$, generating 100 points on average in the window. We allow for spatial inhomogeneity: the value for λ gives the average number of points at each time point, but the spatial structure changes over the window. More precisely, we build a smooth spatial trend which is stronger in the top-right corner and then progressively decreases toward the bottom-left corner (see Fig.1 for an example before and after the change point). We build the series assuming that the spatial structure is the same over time up to a scale parameter, and the changepoint detection identifies the time point that corresponds to the change of scale in our data.

We generated both iid and AR(1) data series, under the hypotheses of no change point, one change point and three change points. For the single changepoint series, we tried two different change magnitudes: a big one, from $\lambda_1 = 1$ to $\lambda_2 = 2$, and a small one, from $\lambda_1 = 1$ to $\lambda_2 = 1.2$. As for the multiple changepoint series, we set two positive changes and a negative one: the segment intensity values are $\lambda_1 = 1$, $\lambda_2 = 1.4$, $\lambda_3 = 2.3$ and $\lambda_4 = 2$. The last change is extremely small, to further test the performance of the detection methods. Each one of these time series was replicated 100 times. A summary of the simulation design is in Fig. 2.

Both iid and time dependent data are generated as their behaviour is very different for what concern change points. Fig. 3 shows some time series made by counting the number of points for each time point. As it can be seen, iid data keep very close to the initial set value over the series, and the change points are easily recognizable. On the contrary, AR(1) data tend to drift far away from the initial value, and are far more variable. On one hand, this can result in the detection of spurious change points, i.e. change points that are due to the variability of the series and not to external factors; on the other hand changes set in the simulation may not be identifiable. It is therefore of interest to test the methods on both types of data.

On all the generated time series we fit the four models described in Section 2.1 and try to detect change points with both methods described in Section 2.2. All model fitting is done using INLA.

3.2 Simulation results

The two methods' performance was evaluated according to type I and type II errors, number and position of detected change points and values of the intensity estimates.

As for the errors, a summary of the performance is in Fig. 4. In general, the Bayes Factor method performs very well as regards the first two models: in most cases type I errors are very small (with the exception of one case with time dependent data, but we expect poorer performance on these data, for the reasons introduced in Section 3.1) and type II errors are negligible in all cases. When we fit more complicated models including spatial effect, though, the performance is very poor: the method is too conservative and does not detect change points, irrespective of their magnitude.

The posterior threshold method holds a better performance over all models; this is sensible, as the threshold value can be tuned according to the model. A few 'grey' zones are produced, but the overall conclusions are correct in most cases, and there is at least some ability to detect changes in all situations. A further summary of this performance can be found in Fig. 5: the first row in each table concerns data generated under H_0 and the second row concerns data generated under H_1, therefore numbers have to sum to 100 by row. It is very plain that the PT method has a better overall performance: as regards null data (first row), the behaviour of the two methods is very similar, but the PT method is 20 percentage points better in finding change points in H_1 data (second row).

As for the number of detected change points, results are linked, but not necessarily identical, to the previous results: committing or not a type II error only concerns the rejection of H_0 and tells nothing on the number and positions of change points found, which is of special interest in the multiple changepoint search. Fig. 6 shows a summary of the results. We can see that as far as H_0 data are concerned, results are correct in all cases: even in situation where some change points were found, as in AR(1) data, all the positions were different, and this indicates they are spurious change points and not 'true' ones. As regards the detection in H_1 data, the BF method suffers from the above mentioned issue: it is very precise in detecting the true change(s) in the first two models, but is too conservative when spatial dependence and inhomogeneity is introduced. The PT method performs much better: when change points are not detected in the majority of replicates, it is due to the small magnitude of the change, which means the method is not too sensible; despite the small size, a percentage of replicates still had a change detected. The only wrong conclusion concerns the multiple changepoint iid data series under the most complicated model; in all other cases, conclusions are very sensible and the detected position are correct or as close as makes no difference. It is interesting to note that spurious changes in the time dependent data do not affect the conclusions.

Lastly, a few comments about the intensity estimates, which again depend on the above presented results. A summary of the estimated values is given in Fig. 7. Note that the intensity is a inhomogeneous function which takes different values over space. In this table, for brevity reasons, only the mean value is reported, but the mean range (over the replicates) and credibility bands are also available. Given the detected change points, estimates are very accurate over all the simulated scenarios: when a change point was not detected, values are an average between the two segments' true values, and when a change point was only detected in part of the replicates (as it happens with very small changes), the true magnitude of the change is shrunk. In all cases the correct (increasing or decreasing) trend is captured. It is possible to see an example of the produced estimates in Fig. **??**: it represent a multiple changepoint data series, where the above panels show the true values for the intensity in the four segments, and the below panels show the three segments estimated by INLA, after detecting two change points (the last change has a very small magnitude and was not detected).

After assessing the performance of the methods, we applied both of them to the motivating dataset.

3.3 Extension: changes in the spatial structure

All the simulated data series are generated taking a constant spatial structure for the intensity function and allowing for a change in scale, i.e. a change point corresponds to a greater or smaller number of points in the window, which follow the same spatial distribution. We are interested in relaxing the assumptions and allowing the intensity function to change in space as well, as it happens in many real situations. This might lead to two different types of change: a change in structure, when the overall number of points remain approximately the same but the spatial distribution changes, and a change in both scale and structure.

We believe our methods hold over this general situation as well: when looking for a change with the proposed algorithms, we never specify that we are looking for a different number of points. We try and split the data at all different time points and we look for the single equation (no change point) or $M+1$ equations (M change points) that describe the dataset best, irrespective of the type of change that occurred. Therefore, if we use a model that includes a spatial effect, we expect our methodology to be able to identify change points in both space and time.

For studying this situation, we only worked on inhomogeneous data generated under the alternative hypothesis of one change point. The spatially homogeneous case is of no interest here, and if the method works for a single changepoint search it is straightforward to extend it to multiple changes. We used the same values for T, S and W and the change point is again set in the centre of the time series. We cover both cases of only spatial change and spatial plus scale change. An example of generated data can be seen in Fig 8. As expected, results are very good and show that the methods are able to detect all types of change. A summary of the performance of the methods in terms of power is displayed in Tab 9. As expected, the first two models do not perform very well in detecting a spatial change (yellow area in the table), as the spatial effect is not included and they assume the intensity function is constant over time. There are no substantial differences in the performance of the BF and PT method. It is nevertheless interesting to point out that, in the minority of cases where the change point is detected, it is in the correct location.

The spatial and scale change is correctly detected in all replicates even in the homogeneous models, as a change in the number of points is recognized as change point over all models.

The most interesting result is that, as soon as the spatial effect is included (model 3 and 4), conclusions are perfectly correct. The BF method performs even better than in the only-scale change situation, as it does not suffer from too much conservationism.

The INLA estimates, again, reproduce very accurately both the scale and the spatial structure of the time segment intensity function in all cases.

4 Particle data

Since the 1950s, Dounreay has been the site of several nuclear research establishments, because of its isolation for safety reasons. In 1994, the last reactor ceased operation and the area is currently being decommissioned (http://www.dounreay.com). Radioactive particles have been found on local beaches in the North of Scotland since the 1990s as a result of historic practices during nuclear fuel reprocessing at the Dounreay plant. The data set used gives the particles' locations on one of the local beaches, Sandside beach, during each of the years of monitoring. The temporal data series is made of yearly point pattern realizations, and additional information about the retrieval and radioactivity level also labels each particle once it has been collected and examined. The underlying intensity and its spatial structure are of interest, along with potential changes in its strength. The dataset presents some difficulties when a changepoint analysis is carried out: the time series is not long ($T = 15$) and some yearly patterns present very few points. Still, the questions are of interest, and the method performance has already been tested over simulated data.

An exploratory analysis shows that Cox processes fit data very well; in particular, the flexible class of log-Gaussian Cox processes is realistically suitable for the problem as the distribution of particles could be due to an underlying driver (tides and winds). Moreover, it is very straightforward to complicate these models by adding fixed, random of smoothed effects to the structured predictor; the estimation with INLA is very fast (and precise) even for complex models and this allows to try many different models without high computational effort.

4.1 Results on particle data

Table 4.1 and 4.1 display a summary of the number and positions of detected change points in the data series for both a single and a multiple search.

Model	Change point (BF)	Change point (PT)
Fixed	2006	2006
Temporal	2003	2003
Spatial	—	2006
Sp-temp	—	—

Model	Change point (BF)	Change point (PT)
Fixed	2003, 2006, 2012	2003, 2006, 2012
Temporal	2003	2003, 2006, 2012
Spatial	—	2006, 2012
Sp-temp	—	—

Results must be interpreted carefully since the time series is very short, but they are sensible given the context. 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 with a special care; 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.
An example of the analysis output is given in Fig. 10: this shows the result for a multiple changepoint detection using the model including spatial dependence and the Posterior Threshold method. The posterior probability plot in the first panel shows that two significant change points have been detected, in 2006 and 2012; the right hand side part of the figure shows a comparison between non-parametric kernel estimates and INLA estimates. This dataset is similar enough to the case covered by our simulation study: the spatial structure of the intensity function is inhomogeneous, but can be considered constant over time up to a scale parameter, with a low density value in most of the window and a hot spot in the bottom-right area. The reported scale of values shows that there is a significant increase in the intensity after 2006, and then a significant decrease in the last two years of the series.

5 Discussion

In this work, we presented a new method which is able to find unknown multiple change points in the intensity of a spatio-temporal point process. The novelty of our method lies in the ability of modelling both spatial and temporal dependence on such a complex point dataset.
A few considerations can be done on the methodology we follow and the results we obtain.
All the models presented in Section 2.1 are very simple, but they contain the key elements for the analysis, i.e. spatial inhomogeneity, spatial dependence and temporal dependence. Once we find a method that allows to detect change points in these situations, it is straightforward to complicate the models by adding fixed effects, such as covariates, and random effects (in a limited number), up to very complex models able to give a good description of many real situations.
As for the fitting with INLA, a log-likelihood value is returned for every fitting. What we are interested in, as in all Bayesian inference, is the posterior distribution. In our work, this is simply obtained by normalising the likelihood values, as we set non-informative priors on both number of change points and their positions. According to the specific context, different prior distributions can be set, and the posterior distribution is found by following the general Bayes rule of multiplying prior and likelihood and then rescaling in order to have a proper distribution.
As for the threshold choice in the PT method, it is to bear in mind that greater values (closer to 1) will lead to more conservative conclusions, and smaller values (closer to 0) will detect change points more easily. The choice of the threshold can therefore be knowledge-driven, if information is available on the diffusion of change points in the data series. Note that useful knowledge can also be incorporated in the posterior through the prior distribution. Another important notion is that the height of peaks in the

posterior distribution depends on the length of the time series: since the curve must integrate to 1, longer Ts will flatten its peaks. For example, Park et al. (2012 but CHECK REF!) use a threshold of 0.1 for a data series of $T = 1000$; the same value would certainly lead to the acceptance of too many change points in a shorter series. In order to find a sensible and not too arbitrary threshold h, it is possible to use simulated data under the null hypothesis for assessing the significance level α based on different values of h. Once we find a value for h such that the significance level does not exceed a certain limit (usually $\alpha \leq \{0.01, 0.05, 0.1\}$), we use that threshold on data generated under the alternative hypothesis in order to evaluate its power level, the ability to detect the correct change points and the accuracy of the estimates produced. This is the idea we follow in our simulation study.

As for what concerns the results, there are multiple aspects we can focus on. In some changepoint analysis, the interest only lies on where the change point(s) occur(s), and not on the parameter estimates. In many other cases, it is of interest to understand if the change is positive (an increase in the estimated values) or negative (a decrease in the estimated values). For all these cases, the accuracy of the estimates is not the main goal, and a good performing detection method is all that is needed. Nonetheless, we want to focus on the most general case, where the estimate (in our case, the intensity estimate) is of interest and an accurate estimation method is also required, once the change points are detected.

The performance of INLA is very satisfactory as regards both computational time and produced estimates. Please note that the ability of detecting change points does not depend directly on the INLA approach, but depends on the choice of the detection method: we have seen that the Bayes Factor method and the Posterior Threshold method have different performances, even if they are used on the same model, i.e. they are used on the same set of likelihood values produced by INLA. Given the detection of change points, INLA performs very well in reproducing both spatial trend and scale of values of the intensity function over all the simulation study. Note that the accuracy of INLA is high when the hypothesis underlying the use of INLA work: the random field has to be well approximated by a Gaussian field, with a smooth but limited spatial structure (i.e. a sparse covariance matrix for the parameters). As for the computational time, in the simulation study running the models took a few minutes for every replicate, on real data all results were obtained in less than 30 minutes in total. Should computational time issues be encountered, for example if working on an extremely long time series and an extremely dense dataset, there is an alternative to the grid approach, the Stochastic Partial Differential Equation approach (SPDE, see [ref]).

This work is a first step toward spatio-temporal changepoint analysis. Many interesting extensions are possible. A further step would be to generalise the intensity function in order to allow its spatial structure to change over time and look for changes in structure as well as scale changes. Looking for improved version of the detection methods would be of interest; Wyse, Friel and Rue (2011) propose a combination of INLA and recursive techniques to look for multiple change points: an extension of this methodology to the spatio temporal case may lead to better result with regard to the Bayes Factor method.

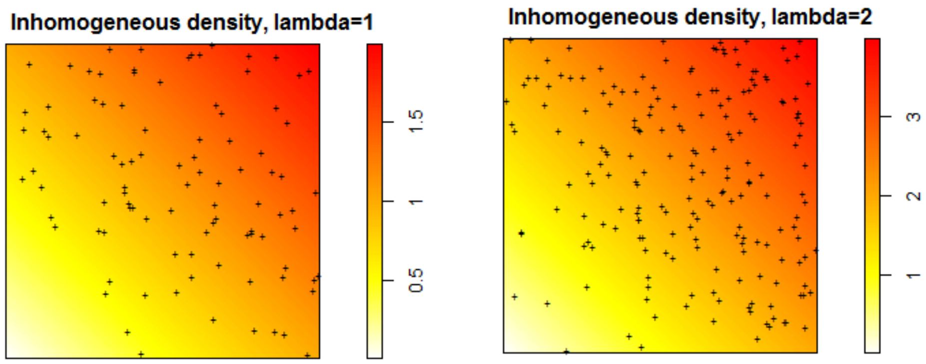

Figura 1: Simulated data - example

			MODEL 1		MODEL 2		MODEL 3		MODEL 4	
			BF	PT	BF	PT	BF	PT	BF	PT
IID	H_0	λ=1	100repl	100repl	100repl	100repl	100repl	100repl	100repl	100repl
	H_1	λ_2=2	100repl	100repl	100repl	100repl	100repl	100repl	100repl	100repl
		λ_2=1.2	100repl	100repl	100repl	100repl	100repl	100repl	100repl	100repl
		multiple	100repl	100repl	100repl	100repl	100repl	100repl	100repl	100repl
AR(1)	H_0	λ=1	100repl	100repl	100repl	100repl	100repl	100repl	100repl	100repl
	H_1	λ_2=2	100repl	100repl	100repl	100repl	100repl	100repl	100repl	100repl
		λ_2=1.2	100repl	100repl	100repl	100repl	100repl	100repl	100repl	100repl
		multiple	100repl	100repl	100repl	100repl	100repl	100repl	100repl	100repl

Figura 2: Table 1 - Simulation design

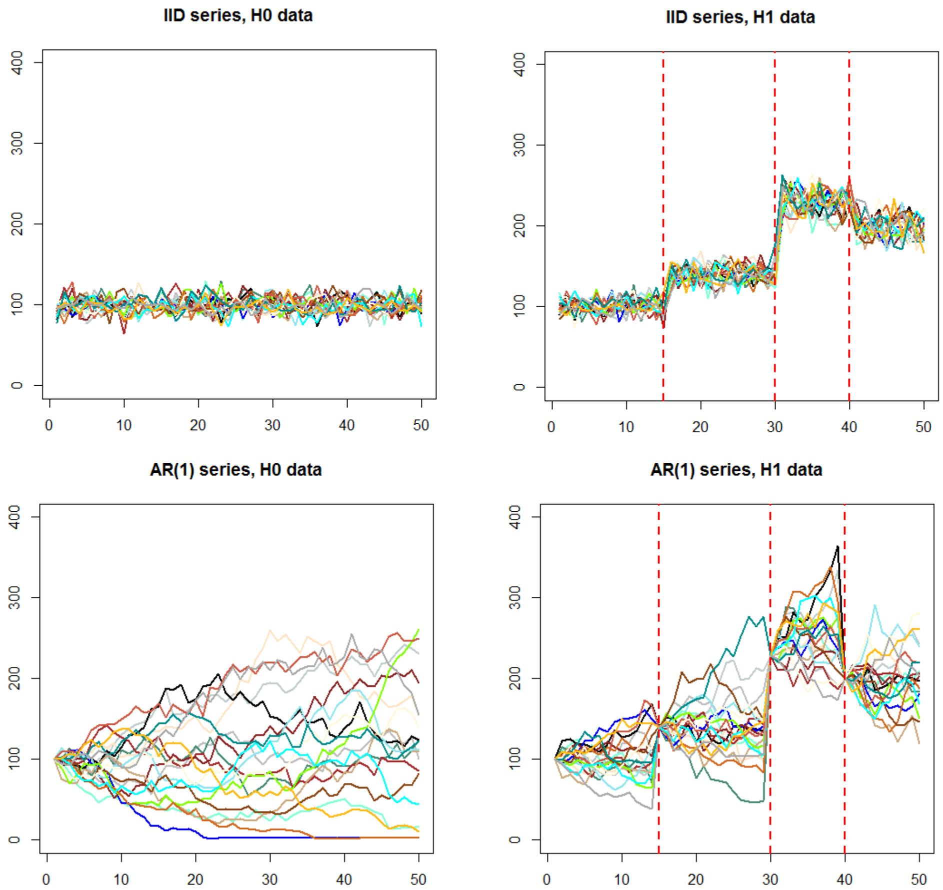

Figura 3: Simulated time series - iid vs AR(1) data

			MODEL 1		MODEL 2		MODEL 3		MODEL 4	
			BF	PT	BF	PT	BF	PT	BF	PT
IID	H_0	λ=1	α=0	$\alpha\leq$0.05	α=0	$\alpha\leq$0.05	α=0	$\alpha\leq$0.1	α=0	$\alpha\leq$0.1
	H_1	λ_2=2	β=1	β=1	β=1	β=1	β=1	β=1	β=0	β=1
		λ_2=1.2	β=1	β=1	β=1	β=0.98	β=0	β=0.34	β=0	β=0.3
		multiple	β=1	β=1	β=0.99	β=1	β=0	β=0.93	β=0	β=0.26
AR(1)	H_0	λ=1	α=0.96	α=0.66	α=0.38	α=0.24	α=0.15	α=0.26	α=0	α=0.18
	H_1	λ_2=2	β=1	β=0.97	β=0.81	β=0.75	β=0.53	β=0.81	β=0	β=0.52
		λ_2=1.2	β=1	β=0.73	β=0.43	β=0.36	β=0.19	β=0.54	β=0	β=0.37
		multiple	----	β=0.98	----	β=0.91	----	β=0.84	----	β=0.67

Figura 4: Simulation results - type I and II errors

H_0 correct result	False positive
False negative	H_1 correct result

BF PERFORMANCE (%)	
81.38	18.63
43.79	56.21

PT PERFORMANCE (%)	
79.75	20.25
23.92	76.08

Figura 5: Simulation results - type I and II errors - summary

			MODEL 1		MODEL 2		MODEL 3		MODEL 4	
			BF	PT	BF	PT	BF	PT	BF	PT
IID	H_0	λ=1	0	0	0	0	0	0	0	0
	H_1	λ_2=2	1	1	1	1	1	1	0	1
		λ_2=1.2	1	1	1	1	0	0	0	0
		multiple	3	3	3	3	0	2	0	0
AR(1)	H_0	λ=1	0	0	0	0	0	0	0	0
	H_1	λ_2=2	1	1	1	1	1	1	0	1
		λ_2=1.2	1	1	0	0	0	1	0	0
		multiple	----	3	----	3	----	3	----	3

Figura 6: Simulation results - Number and position of detected change points

			MODEL 1		MODEL 2		MODEL 3		MODEL 4	
			BF	PT	BF	PT	BF	PT	BF	PT
IID	H_0	λ=1	1	1.01	1	1.01	0.99	1	1	1
	H_1	λ_1=1, λ_2=2	0.99 - 2.00	1.04 - 2.00	1.00 - 2.00	1.04 - 2	0.99 - 2	1 - 2	1.5	1 - 2
		λ_1=1, λ_2=1.2	1.00 - 1.20	1.01 - 1.20	1.00 - 1.20	1.01 - 1.2	1.09	1.1	1.1	1.1
		λ_1=1, λ_2=1.4 λ_3=2.3, λ_4=2	1.00 - 1.40 2.17 - 2.07	1.00 - 1.38 2.22 - 2.05	1.01 - 1.40 2.12 - 2.10	1.00 - 1.38 2.15 - 2.08	1.58	1.20 - 1.50 2.20	1.51	1.60
AR(1)	H_0	λ=1	0.99	1.01	0.96	0.99	0.92	0.98	0.87	0.98
	H_1	λ_1=1, λ_2=2	1.05 - 2.04	1.05 - 2.03	1.09 - 2.10	1.11 - 2.05	1.18 - 1.90	1.08 - 1.99	1.41	1.20 - 1.78
		λ_1=1, λ_2=1.2	1.01 - 1.14	1.01 - 1.14	1.06	1.06	1.1	1.01 - 1.18	1.07	1.09
		λ_1=1, λ_2=1.4 λ_3=2.3, λ_4=2	----	1.02 - 1.40 2.22 - 2.07	----	1.07 - 1.38 2.14 - 2.04	----	1.10 - 1.42 2.18 - 2.08	----	1.15 - 1.41 2.03 - 1.97

Figura 7: Simulation results - Time segment intensity estimates

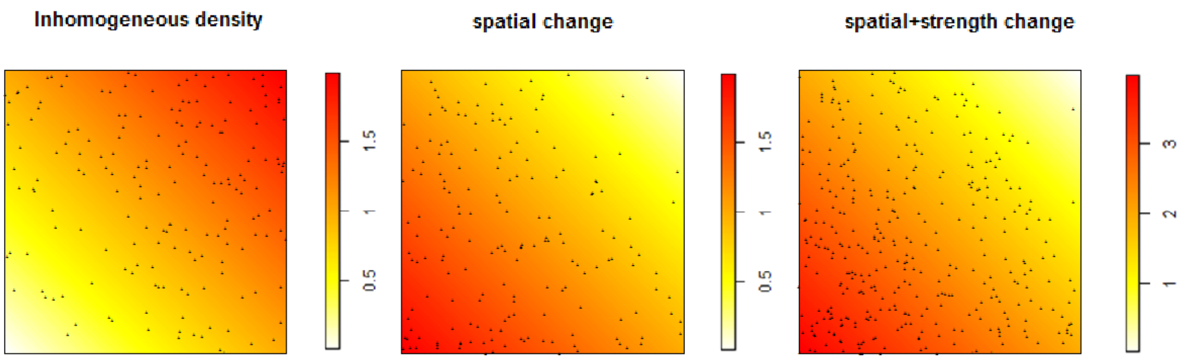

Figura 8: Spatial change - examples of simulated data

		MODEL 1		MODEL 2		MODEL 3		MODEL 4	
		BF	PT	BF	PT	BF	PT	BF	PT
SPATIAL CHANGE	Type II error	β=0.38	β=0.44	β=0.42	β=0.26	β=1	β=1	β=1	β=1
	Changepoint location	No chp	No chp	No chp	No chp	24	24	24	24
	Intensity estimate	0.95	0.95	0.97	0.97	1.05 - 1.00	1.03 - 1.01	1.06 - 1.02	1.04 - 1.00
SPATIAL AND SCALE CHANGE	Type II error	β=1	β=1	β=1	β=1	β=1	β=1	β=1	β=1
	Changepoint location	24	24	24	24	24	24	24	24
	Intensity estimate	1.00 - 1.95	1.00 - 1.95	0.99 - 1.96	0.99 - 1.97	1.06 - 2.00	1.01 - 2.00	1.03 - 2.01	1.02 - 2.00

Figura 9: Spatial change - summary of the results

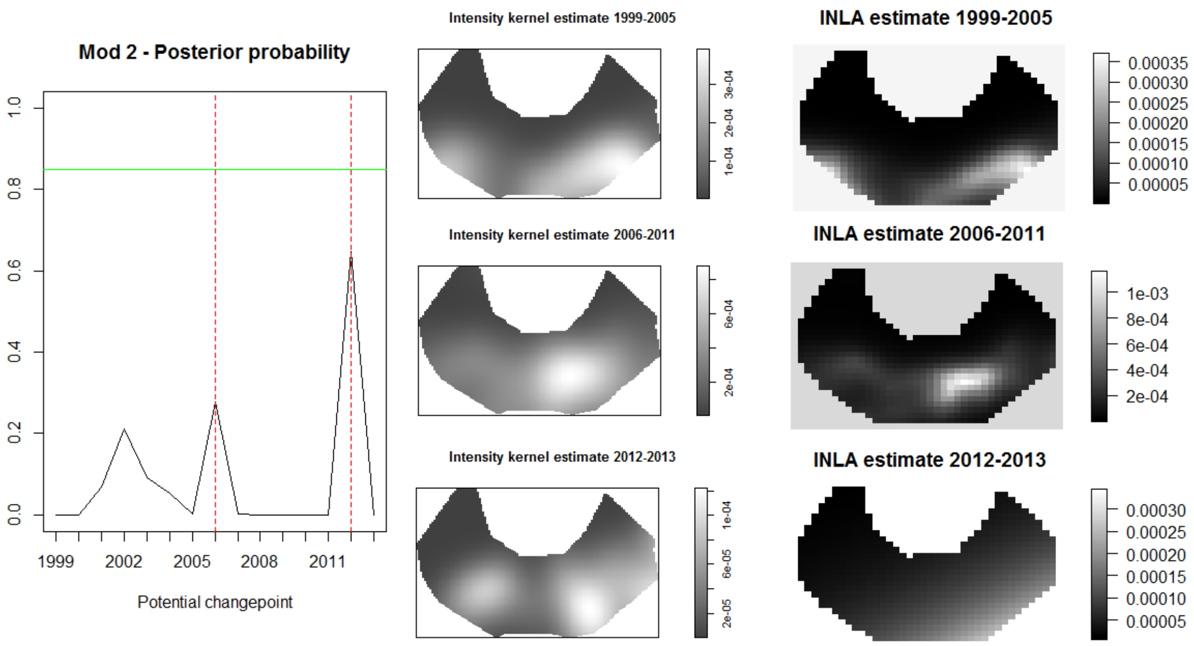

Figura 10: Spatial model and PT method - results