Fiedler, B., Zöller, G., Holschneider, M., Hainzl, S. (2018): Multiple Change-Point Detection in Spatiotemporal Seismicity Data. - *Bulletin of the Seismological Society of America*, *108*, 3A, pp. 1147—1159.

DOI: http://doi.org/10.1785/0120170236

Multiple change-point detection in spatio-temporal seismicity data

August 28, 2018

Bernhard Fiedler, Institute of Mathematics, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany. Email bfiedler@uni-potsdam.de

Gert Zöller, Institute of Mathematics, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

Matthias Holschneider, Institute of Mathematics, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

Sebastian Hainzl, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany

Abstract

Earthquake rates are driven by tectonic stress buildup, earthquake- induced stress changes, and transient aseismic processes. Although the origin of the first two sources is known, transient aseismic processes are more difficult to detect. However, the knowledge of the associated changes of the earthquake activity is of great interest, because it might help identify natural aseismic deformation patterns such as slow-slip events, as well as the occurrence of induced seismicity related to human activities. For this goal, we develop a Bayesian approach to identify change- points in seismicity data automatically. Using the Bayes factor, we select a suitable model, estimate possible change-points, and we additionally use a likelihood ratio test to calculate the significance of the change of the intensity. The approach is extended to spatiotemporal data to detect the area in which the changes occur. The method is first applied to synthetic data showing its capability to detect real change-points. Finally, we apply this approach to observational data from Oklahoma and observe statistical significant changes of seismicity in space and time.

Introduction

Natural seismicity is a nonstationary process with vari- ous kinds of transient behavior on different spatiotemporal scales, for example, aftershocks, foreshocks, swarm activity, and quiescence lasting from hours to decades. Man-made earthquakes, for example, arising from fluid injection in geothermal areas or wastewater disposals (Ellsworth, 2013) have similar statistical features, but on smaller spatial scales with transient boundary conditions.

For example, the grow- ing amount of industrial projects related to the injection of fluids at depth has led to the question, to which degree the seismic hazard changes at an injection site. Figure **1** shows a clear increase of the earthquake number in Oklahoma at around the year 2010. Several authors including Keranen et al. (2013), Langenbruch and Zoback (2016), Walsh and Zoback (2015) and Weingarten et al. (2015) reported a correlation between the injection volume and the observed increase of the seismicity.

In our study, we propose a Bayesian approach to detect transients in seismicity. Using the Poisson assumption for the occurrence of earthquakes, we apply a method which was first introduced by Raftery and Akman (1986) and further applied to earthquake data by Gupta and Baker (2015), Montoya and Wang (2017) and Gupta and Baker (2017). We go beyond these works and present an algorithm that allows the identification of multiple change-points that occur in space and time. Moreover, we note that for observational data, signals for change-points might be weak and difficult to distinguish from random fluctuations. Therefore, we put special emphasis on the development of an appropriate significance test and apply the concept of the Bayes factor for model selection.

Our model approach is based on the assumption that the earthquake occurrence follows a piecewise homogeneous Poisson process (HPP) in time. In particular, the system is assumed to suddenly change from one Poisson rate into another. Such transitions are defined as change-points in time. This approach is then extended to space–time in a straightforward way by subdividing the area into smaller segments of a specific size. For every subarea, we obtain a time series contain- ing change-points or not. In both cases, we first address the question which model is statistically preferable, for example, a model with or without a change-point. If a specific change- point model is preferred, we then use an extended approach of Raftery and Akman (1986) to estimate the change-points and calculate associated Bayesian credibility intervals at a given significance level (e.g., 95%). This is described in the *Estimation of Change-Points* section. Additionally, we use a likelihood ratio test to calculate the significance (p-value) of the change of the Poisson intensity (the *Likelihood Ratio Test* section) and extend the approach to the space–time prob- lem (the *Spatiotemporal Change-Point Problem* section). By means of synthetic data, we demonstrate the performance of the method (the *Illustration for Synthetic Data* section) before applying it to the observed data from Oklahoma (the *Application to Seismicity in Oklahoma* section).

Method

Estimation of Change-Points

First we give a brief overview on the detection of temporal change-points according to Raftery and Akman (1986). In comparison to this work we extend the method to a general case with more than one change-point.

An observation period of $[a,\ b]$ is given with n events at times

$$a \leq t_1 < t_2 < \ldots < t_n \leq b. \tag{1}$$

We assume the existence of k change-points

$$\tau_1,\ \tau_2,\ \ldots,\ \tau_k \in [a, b] \tag{2}$$

with $k < n$. Moreover in $[a,\ \tau_1]$ we have $N(\tau_1)$ events which come from a Poisson process with rate λ_1, and $N(\tau_i) - N(\tau_{i-1})$ events in $(\tau_{i-1},\ \tau_i]$ with rate λ_i for $i = 2,\ \ldots,\ k$. Finally, in $(\tau_k,\ b]$ the number of events follows a Poisson process with rate λ_{k+1}.

Let $\underline{t} = \{t_1, \ldots, t_n\}$ and $\theta = \{\lambda_1,\ \ldots,\ \lambda_{k+1},\ \tau_1,\ \ldots,\ \tau_k\}$. It can easily be shown that the mutual likelihood function is given by

$$\begin{aligned} p(\underline{t} \mid \theta) &= \lambda_1^{N(\tau_1)} e^{-\lambda_1(\tau_1 - a)} \lambda_2^{N(\tau_2)-N(\tau_1)} e^{-\lambda_2(\tau_2-\tau_1)} \cdot \ldots \cdot \lambda_{k+1}^{N(b)-N(\tau_k)} e^{-\lambda_{k+1}(b-\tau_k)} \\ &= \lambda_1^{N(\tau_1)} e^{-\lambda_1(\tau_1 - a)} \lambda_{k+1}^{N(b)-N(\tau_k)} e^{-\lambda_{k+1}(b-\tau_k)} \prod_{i=2}^{k} \lambda_i^{N(\tau_i)-N(\tau_{i-1})} e^{-\lambda_i(\tau_i-\tau_{i-1})}. \end{aligned} \tag{3}$$

Using Bayes' Theorem

$$p(\theta \mid \underline{t}) = \frac{p(\underline{t} \mid \theta) p(\theta)}{\int_\Theta p(\underline{t} \mid \theta)\, p(\theta)\, d\theta} \propto p(\underline{t} \mid \theta) p(\theta) \tag{4}$$

we can calculate the posterior density $p(\theta \mid \underline{t})$ for the parameter θ given the data represented by $\underline{t} = \{t_1, \ldots, t_n\}$. Here $p(\underline{t} \mid \theta)$ denotes the likelihood function and $p(\theta)$ is the prior density of θ.

Let $p(\tau_1),\ \ldots,\ p(\tau_k)$ and $p(\lambda_1),\ \ldots,\ p(\lambda_{k+1})$ be the prior densities. Then the posterior density is given by

$$\begin{aligned} p(\theta \mid \underline{t}) \propto\ & p(\tau_1) p(\lambda_1) p(\lambda_{k+1}) \lambda_1^{N(\tau_1)} e^{-\lambda_1(\tau_1-a)} \lambda_{k+1}^{N(b)-N(\tau_k)} \\ & \times e^{-\lambda_{k+1}(b-\tau_k)} \prod_{i=2}^{k} p(\tau_i) p(\lambda_i) \lambda_i^{N(\tau_i)-N(\tau_{i-1})} e^{-\lambda_i(\tau_i-\tau_{i-1})}. \end{aligned} \tag{5}$$

Assuming now a flat prior, we calculate the marginal posterior density of $\underline{\tau} = \{\tau_1,\ \ldots,\ \tau_k\}$ by integrating with respect to $\lambda_1,\ \ldots,\ \lambda_{k+1}$ (see also the *Derivation of the Marginal Posterior Density* section in the Appendix).

$$
\begin{aligned}
p(\underline{\tau} \mid \underline{t}) &= c \int\limits_0^\infty \dots \int\limits_0^\infty \lambda_1^{N(\tau_1)} e^{-\lambda_1(\tau_1 - a)} \lambda_{k+1}^{N(b)-N(\tau_k)} e^{-\lambda_{k+1}(b-\tau_k)} \\
&\quad \times \prod_{i=2}^{k} \lambda_i^{N(\tau_i)-N(\tau_{i-1})} e^{-\lambda_i(\tau_i - \tau_{i-1})}\, d\lambda_1 \dots d\lambda_{k+1} \\
&= c(\tau_1 - a)^{-[N(\tau_1)+1]} \Gamma[N(\tau_1)+1](b-\tau_k)^{-[N(b)-N(\tau_k)+1]} \\
&\quad \times \Gamma[N(b)-N(\tau_k)+1] \prod_{i=2}^{k} (\tau_i - \tau_{i-1})^{-[N(\tau_i)-N(\tau_{i-1})+1]} \Gamma[N(\tau_i)-N(\tau_{i-1})+1]
\end{aligned}
\tag{6}
$$

We consider two special cases of Eq. (6).

Special case: one change-point

$$
p(\tau \mid \underline{t}) = c(\tau - a)^{-[N(\tau)+1]} \Gamma[N(\tau)+1](b-\tau)^{-[N(b)-N(\tau)+1]} \Gamma[N(b)-N(\tau)+1] \tag{7}
$$

Special case: two change-points

$$
\begin{aligned}
p(\tau_1,\ \tau_2 \mid \underline{t}) &= c(\tau_1 - a)^{-[N(\tau_1)+1]} \Gamma[N(\tau_1)+1](\tau_2 - \tau_1)^{-[N(\tau_2)-N(\tau_1)+1]} \\
&\quad \times \Gamma[N(\tau_2)-N(\tau_1)+1](b-\tau_2)^{-[N(b)-N(\tau_2)+1]} \Gamma[N(b)-N(\tau_2)+1]
\end{aligned}
\tag{8}
$$

We note that in Eq. (6), (7) and (8) c is a normalizing constant which ensures that the conditions for a probability density function is fulfilled. Alternatively to a flat prior density, a conjugated prior for the parameters $\lambda_1,\ \dots,\ \lambda_{k+1}$ (e.g. a gamma distribution) and uniformly distributed prior densities for $\tau_1,\ \dots,\ \tau_k$ can be used (see also Raftery and Akman (1986)). By maximizing $p(\underline{\tau} \mid \underline{t})$ in Eq. (6) with respect to $\underline{\tau} = \{\tau_1,\ \dots,\ \tau_k\}$ we obtain the estimation $\underline{\hat{\tau}} = \{\hat{\tau}_1,\ \dots,\ \hat{\tau}_k\}$ for the change-points.

In Akman and Raftery (1986) it was shown that the estimator for a single change-point is consistent and asymptotically normal. Moreover in Ghosal et al. (1999) and other related papers (e.g. Ghosh et al. (1994) and Ghosal and Samanta (1995)) it was also demonstrated that in this case the posterior distribution asymptotically behaves like an exponential function on both sides of the detected change-point. The asymptotic behavior for the general case with more than one change-point is shown in Ghosal et al. (1999).

For model selection, we use the Bayes factor to get an idea which model should be preferred, that is, whether to prefer a change-point model ($\mathcal{M}_1$) or a model without a change-point (model $\mathcal{M}_0$). The Bayes factor is defined by the ratio of the marginal or integrated likelihood for both models, that is

$$B_{lm} = \frac{p(\underline{t} \mid \mathcal{M}_l)}{p(\underline{t} \mid \mathcal{M}_m)}. \tag{9}$$

Here $\mathcal{M}_l$ and $\mathcal{M}_m$ denote a model with l respectively with m change-points where $l, m = 0, 1, \ldots, k$. Apart from the goodness of fit, the complexity of the assumed model has to be taken into account in order to assess the most capable model describing the data and thus performing the estimation. As an example, if we test the hypothesis of no change-point ($\mathcal{H}_0$) against a change-point model, the value of the Bayes factor quantifies the evidence of the supported model, e.g. $B_{01} < 0.01$ can be interpreted as a decisive evidence against $\mathcal{H}_0$, compare Kass and Raftery (1995). Equation (9) strongly depends on the choice of the prior distributions. When an improper prior is used, the Bayes factor is, however, not well-defined and depends on an arbitrary ratio of constants. To handle this problem we use the idea of an imaginary training sample which involves the smallest possible sample size permitting a comparison of $\mathcal{M}_0$ and $\mathcal{M}_m$ and provides maximum possible support for $\mathcal{M}_0$. In this case the Bayes factor should be approximately one. This approach was introduced in Spiegelhalter and Smith (1982) and was adopted and discussed in several other works, e.g. Raftery and Akman (1986), Kass and Raftery (1995) or Berger et al. (2004). Using improper prior densities for the intensities of the shape $p(\lambda_i) \propto \lambda_i^{-\frac{1}{2}}$ and a uniform distributed prior for τ_i, i.e. $p(\tau_i) = \frac{1}{b-a}$ (Raftery and Akman, 1986) and taking into consideration the approach of Spiegelhalter and Smith (1982), the Bayes factor can be calculated by

$$B_{01} = \frac{4\sqrt{\pi}(b-a)^{-n}\Gamma(n+\frac{1}{2})}{\sum_{i=0}^{n}\Gamma(i+\frac{1}{2})\Gamma(n-i+\frac{1}{2})\int_{t_i}^{t_{i+1}}(\tau-a)^{-(i+\frac{1}{2})}(b-\tau)^{-(n-i+\frac{1}{2})}d\tau}. \tag{10}$$

This approach can be extended and the derivation for the general case B_{lm} is shown in the Appendix *Derivation of the Bayes factor*. For example, for the hypothesis of no change point ($\mathcal{H}_0$) against a model with two change-points we get

$$\begin{aligned} B_{02} = 4\pi^2(b-a)^{-n+\frac{1}{2}}\Gamma(n+\frac{1}{2}) &\left[\sum_{i=0}^{n}\sum_{j=i+1}^{n}\Gamma(i+\frac{1}{2})\Gamma(j-i+\frac{1}{2})\Gamma(n-j+\frac{1}{2})\right. \\ &\left.\times\int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}}(\tau_1-a)^{-(i+\frac{1}{2})}(\tau_2-\tau_1)^{-(j-i+\frac{1}{2})}(b-\tau_2)^{-(n-j+\frac{1}{2})}d\tau_1 d\tau_2\right]^{-1}. \end{aligned} \tag{11}$$

We note that the computation of Eq. (9) for large l and m is numerically very difficult to handle because of the high- dimensional integrals. We remark that the function evaluations grow exponentially as the number of dimensions increases. If the quadrature rules do not lead to a desirable result, Monte Carlo methods should be used instead. In our work, we apply a likelihood ratio test in addition to the established methods we considered before. As an advantage, we get the significance (p-value) of the change of the Poisson intensity. Needless to say that the Bayes factor is a powerful tool for the model selection, but although we know the preferred model, we cannot yet prove that the estimated change-points are significant. This problem can be solved with the aid of the likelihood ratio test.

Likelihood Ratio Test

We consider two Poisson processes with intensities λ_1 and λ_2 in the time intervals $[s_1, s_2]$ and $[s_3, s_4]$ with $s_3 > s_2$. In the first period we have n_1 events, and in the second period the number of events is n_2. We aim at testing whether or not the intensities are equal. In detail we test hypothesis $\mathcal{H}_0$ versus $\mathcal{H}_1$ with

$$\begin{aligned} &\mathcal{H}_0 : \lambda_1 = \lambda_2 = \lambda \\ &\mathcal{H}_1 : \lambda_1 \neq \lambda_2. \end{aligned} \tag{12}$$

The likelihood function is given by

$$p(\underline{t} \mid \lambda_1, \lambda_2) = \lambda_1^{n_1} \exp(-\lambda_1 \Delta_1) \lambda_2^{n_2} \exp(-\lambda_2 \Delta_2), \tag{13}$$

with $\Delta_1 = s_2 - s_1$ and $\Delta_2 = s_4 - s_3$.

For $\mathcal{H}_0$ we get

$$p(\underline{t} \mid \lambda) = \lambda^{n_1+n_2} \exp[-\lambda(\Delta_1 + \Delta_2)]. \tag{14}$$

As shown in the Appendix *Derivation of the Likelihood Ratio Test*, we can derive the statistic of this test by calculation of the maximum likelihood estimators for λ, λ_1 and λ_2 and by using a general result of Witting and Müller-Funk (1995). It follows that the test statistic of this likelihood ratio test is equal to

$$Z = 2\left[n_1 \log\left(\frac{n_1}{\Delta_1}\right) + n_2 \log\left(\frac{n_2}{\Delta_2}\right) - (n_1 + n_2)\log\left(\frac{n_1 + n_2}{\Delta_1 + \Delta_2}\right)\right]. \tag{15}$$

$\mathcal{H}_0$ is rejected, if $z > \chi^2_{1,\ 1-\alpha}$ or if the p-value $= P(Z \geq z) < \alpha$. Here $\alpha \in (0,1)$ is a given significance level and $\chi^2_{1,\ 1-\alpha}$ is the $(1-\alpha)$-quantile of the chi-squared distribution with one degree of freedom.

To investigate the properties of this test, we perform calculations with artificially generated data resulting in a reasonable resemblance to the error of the first kind α, as summarized in Table **1**. As an estimator for the error probability, we use the number of rejected hypotheses divided by the number of generated samples. Moreover, Figure **2** illustrates the behavior of the power for fixed values of λ and n. As expected, the simulations show that the test can distinguish between $\mathcal{H}_0$ and $\mathcal{H}_1$ in a suitable way.

Spatiotemporal Change-Point Problem

In this section, we extend our approach for time series in a straightforward way towards spatiotemporal change-point problems. For this aim, we scan an area $\mathcal{D}$ to find change-points in space and time. Figure **3** illustrates the algorithm. First, the investigated domain is subdivided into m subareas $\mathcal{A}_1, \ldots, \mathcal{A}_m$ with $\mathcal{D} = \bigcup_{i=1}^{m} \mathcal{A}_i$. For simplicity, we use equidistantly centered subareas with the same size in the following way: We consider a set of circles, where $\mathcal{A}_i$ has the radius r and the center $(x_i,\ y_i)$ for all $i = 1, \ldots, m$. However, any other subdivision is also possible, see Gupta and Baker (2017).

In the next step we investigate the time series of all events that occurred in $\mathcal{A}_i$ given by

$$\mathcal{S}_i = \{t_{i1}, t_{i2}, \ldots, t_{in_i}\}. \tag{16}$$

Hence the data is a set of triples

$$\bigcup_{i=1}^{m} (\mathcal{A}_i \cup \mathcal{S}_i) = \bigcup_{i=1}^{m} \{(x_{ij}, y_{ij}, t_{ij}) \mid j = 1, \ldots, n_i\}. \tag{17}$$

For $\mathcal{S}_i$ we use our method to detect and evaluate change-points as described before in the

Estimation of Change-Point and *Likelihood Ratio Test* sections.

In detail, for every time series $\mathcal{S}_i$ we use the Bayes factor (9) to decide which model fits best to the given data. If a specific change-point model is preferred, we maximize $p(\underline{\tau} \mid \underline{t})$ in Eq. (6) and receive a set of possible change-points. For every estimated change-point in this set we use the likelihood ratio test and define a change-point as significant, if the p-value is smaller than a given significance level α. The result is a set $\hat{\underline{\tau}}_i = \{\hat{\underline{\tau}}_{1i}, \ldots, \hat{\underline{\tau}}_{ki}\}$ of significant change-points in $\mathcal{S}_i$. Finally we provide the mathematical definition of a transition event within a global statistical model $\mathcal{M}_{trans}$. For this aim, we define a set of transition events as triples in the following way

$$\mathcal{M}_i = \begin{cases} (x_i, y_i, \hat{\underline{\tau}}_i), & \mathcal{S}_i \text{ has at least one change-point} \\ \emptyset, & \mathcal{S}_i \text{ has no change-point} \end{cases}. \tag{18}$$

$$\mathcal{M}_{trans} = \bigcup_{i=1}^{m} \mathcal{M}_i. \tag{19}$$

Evaluation and Application

The derived methodology from the *Method* section is for test and illustration purposes first applied on synthetic data and in the following part applied to real seismicity data recorded in Oklahoma, United States.

Illustration for Synthetic Data

We first test our method by applying it to synthetic data under controlled conditions. For this aim we generate synthetic time series with $t \in [0, 1]$ with a single change-point at

$\tau_{real} = 0.5$ and investigate the goodness of the estimator. To test how the method works, we calculate the standard deviation of $\hat{\tau} - \tau_{real}$ and Bayes factors depending on the number of events and the ratio of the intensities, see Figure **4**. It can be seen e.g. that a change-point can be detected in sequences of 100 events with a high probability and precision, if the intensity ratio exceeds a value of 2.

For the spatiotemporal approach, we generate random realizations of a 3D HPP with a given intensity. From these data, we cut out cylinders and replace it by new cylinders with data from HPPs with different intensities, as illustrated in Figure **5**. Using our algorithm we calculate the transition events $\mathcal{M}_{trans}$. Therefore we scan the whole domain as explained in Figure **3**. The "training" sample is a 3D HPP with rate $\lambda \approx 0.8$ (per unit area) in a cylinder with center $(0, 0)$, radius $r_1 = 6$ and height $h_1 = 20$, corresponding to the time interval $t \in [0, 20]$. The replaced cylinders follows a HPP with rate $\lambda_{\text{cp}} \approx 8$ (per unit area). One cylinder has center $(1, 1)$, radius $r_2 = 1$ and height $h_2 = 10$. Here the related time interval is $t \in [5, 15]$ and the second replaced cylinder has the center $(-3, -3)$, radius $r_3 = 1$ and height $h_3 = 5$. Here the related time interval is $t \in [15, 20]$. In other words, the transitions are given by the sets

$$\mathcal{C}_1 = \{(x, y, t) \mid (x + 3)^2 + (y + 3)^2 \leq 1 \ \wedge \ t = 15\}. \tag{20}$$

and

$$\mathcal{C}_2 = \{(x, y, t) \mid (x - 1)^2 + (y - 1)^2 \leq 1 \ \wedge \ t \in \{5, \ 15\}\}. \tag{21}$$

The chosen sample size is $n = 2000$, and approximately 15% of the data is located within the replaced cylinders. For this test case, we set the selection radius to $r_0 = 0.3$. In general, our results presented in Figure **4** can guide the choice for this selection: To be able to detect a certain rate change, the event number within the selection radius must exceed a minimum number, e.g. 20 events for a ten-times increased intensity as in our example. For the change-point domain $\mathcal{C}_1$ the method yields an average value of $\overline{\hat{\tau}} = 15.173$ and for $\mathcal{C}_2$ we get average values of $\overline{\hat{\tau}_1} = 5.094$ and $\overline{\hat{\tau}_2} = 14.971$. The estimated areas are illustrated in Figure **6**. It is remarkable that apart from a small number of outliers the complete transition area was detected correctly by the method.

Additionally we investigate the sensitivity of the method depending on the selection radius. Therefore, we generated synthetic data from a HPP in the time interval $t \in [0, 20]$, where in the circular region with radius $r_0 = 2$ around the center occurs a change at time 10 to a five-times increased rate, particularly the change-point domain is given by the set

$$\mathcal{C} = \{(x, y, t) \mid x^2 + y^2 \leq 4 \ \wedge \ t \in [10, 20]\}. \tag{22}$$

The chosen sample size is $n = 2000$, where 50 events are within the change-point domain. The intensities are given by $\lambda \approx 0.08$ and $\lambda_{\mathrm{cp}} \approx 0.4$ (per unit area). For 100 simulations, we calculate the Bayes factors and the standard deviation of $\hat{\tau} - \tau_{\mathrm{real}}$ for increasing radii of the event selection around the center. The results are illustrated in Figure **7**. The test results show that the estimation uncertainty is lowest and the success rate is highest for the case that the selection radius equals the radius of the change-point region. A too small selection radius leads to time series with a non-significant number of events, while a too high value results in a systematical error and the precision of the method decreases. However, the results are found to be almost the same for a rather broad range of selection radii within $0.5r_0$ and $2r_0$. This indicates that the results should be rather robust, if the selection radius is chosen in a reasonable range.

Application to Seismicity in Oklahoma

We now apply the method to real earthquake data. Because of its drastic seismic activity changes, Oklahoma probably counts as one of the most interesting study areas for the application of the above estimations. Therefore, we consider an earthquake catalog from Oklahoma with 18,997 events from 1 January 1980 to 31 December 2015, obtained from the Oklahoma Geological Survey, compare *Data and Resources*. We declustered the catalog using the method of Reasenberg (1985) with standard parameters (van Stiphout et al. (2012), Tab. 3) and taking into account all events with magnitude $m \geq 3$. The declustered catalog contains 1,199 events. Using all $m \geq 3$ events, the Bayes factor from Eq. (9) leads to a model with two change-points (see detailed results in Table **2**). The estimated 95% credibility intervals for the (significant) change-points $\hat{\tau}_1$ and $\hat{\tau}_2$ are given by [12/01/2009; 28/03/2009] and [14/12/2013; 30/01/2014]. This result is illustrated in Figure **8**. We note that the application of the likelihood ratio test leads to p-values $\ll 1$ which means that both change-points are considered to be significant and the result strongly supports our model selection. As depicted in Table **2**, the calculation of the Bayes factor B_{01}, B_{02} and B_{03} always leads to the preference of a change-point model. For comparison, a model with one change-point leads to a 95% credibility interval [24/10/2013; 10/11/2013]. A model with three change-points would detect a further change-point in August 2014. If we use the non-declustered catalog a model with three change-points leads to the detection of the $M_W = 5.6$ earthquake at Prague in November 2011 with a subsequent aftershock sequence in addition to the induced seismic changes in 2009 and 2013 (see Figure **8**). Here we observe a natural change-point, caused by the aftershock sequence. In comparison to the works of Gupta and Baker (2017) and Montoya and Wang (2017) we note that they have found similar results for the change-points. The study of Montoya and Wang (2017) used another method for multiple change-point detection in time series and included four different areas in Oklahoma according to the towns Jones, Perry, Cherokee and Waynoka. In all of the four areas their method lead to the choice of a model with two change-points. In the Jones area they found two change-points in May, 2008 and August, 2011. For the other areas they calculated change-points in 2013 until 2015. The work of Gupta and Baker (2017) used the method of Raftery and Akman (1986) to detect single change-points in spatiotemporal data. They used a 25 km radius and found changes in seismicity rates between 2008 and the end of 2015.

By scanning the spatial domain shown in Figure **1** with a total area of approximately $260,000$ square kilometers, our method leads to the results shown in Figure **9** and Figure **10**. We used a radius of 5 km leading to $3,500$ evaluations of time series. This choice is a compromise between optimizing the spatial resolution and increasing the detectability which requires that the considered circles contains enough events to get robust results (see Figure **4**). In the Appendix *Case study Oklahoma: Evaluation with different choices of the radius*, we show the results for alternative values of $r = 2$ km and $r = 10$ km indicating that the main features are robust with regard to the choice of the selection radius. For a better illustration of the results, we only take into account the models $\mathcal{M}_0$, $\mathcal{M}_1$ and $\mathcal{M}_2$. Interestingly, the significant change-point locations show a spatial migration pattern from south to north in both figures and overlap with the injection wells. Moreover a correlation with the injection volume could be a reason for this result as illustrated in Figure **10**.

Furthermore we show the related times of the detected transition events. It is remarkable, that most of the corresponding times of the change-points occur after 2009. This result supports the hypothesis that the detected change-points are correlated with the onset dates of the wastewater injections. Here we have recorded an discernable increase of approval dates after 2010 for wells with an approved injection volume of at least $10,000$ barrels per day.

Conclusions

The main objective of this article is to present an algorithm for the automatic detection of change-points in seismicity data. We use a Bayesian algorithm to identify rate changes in time and space. Tests with synthetic earthquake data show a good agreement of detected change-points with real change-points in space and time. For the Oklahoma case study, the significant change-points show a correlation with the onset of injection wells and especially with the high-volume wells. The method leads to reasonable findings of significant change-points between 2008 and the end of 2015, which correspond to the results of Gupta and Baker (2017) and Montoya and Wang (2017). This makes us confident that our method is powerful for the automatic detection of change-points, even for cases with less drastic activity changes as in Oklahoma.

Nevertheless we only consider a fixed radius for the subdivision of the space. As we have shown the choice of the radius depends on the number of events, and the systematic error should be taken into account. Here the method could be extended for example by using a

Voronoi partition (Okabe et al., 2008) or by using the approach of Gupta and Baker (2017). Furthermore the likelihood ratio test assumes that we have two fixed intervals. Although our method leads to preferable results, an adaptive test could be useful. Another idea for such a test has been proposed in Csörgő and Horváth (1997). Another issue is the deviation from Poissonian behavior, e.g. due to aftershock sequences. In this respect, it is desirable to consider also cluster models like the Epidemic Type Aftershock Sequences (ETAS) model (Ogata, 1988; Zhuang et al., 2002). The ETAS approach to detect seismic changes within the framework of wastewater injections was presented by Wang et al. (2016). In our work we use the declustering approach of Reasenberg (1985) but also other methods could be used to fulfill the Poisson assumption for the considered catalogs (van Stiphout et al., 2012).

Data and Resources

The data used in this article are from the websites
http://www.ou.edu/ogs/research/earthquakes/catalogs.html, last accessed August 28, 2018 and
http://www.occeweb.com/og/ogdatafiles2.htm, last accessed August 28, 2018.

Figure 1, Figure 9 and Figure 10 were made using the Generic Mapping Tools version 4.2.1 (www.soest.hawaii.edu/gmt, last accessed March 2018; Wessel and Smith (1998)).

Simulations were made using the open source software packages R version 3.2 and Python version 2.7.12.

Acknowledgments

We are grateful to Hannelore Liero for fruitful discussions and comments. The manuscript benefitted from constructive comments of two anonymous reviewers. This work was supported by the DFG Research Training Group "Natural hazards and risks in a changing world"(NatRiskChange). GZ also acknowledges support from the DFG (SFB 1294).

References

Akman, V. E. and Raftery, A. E. (1986). Asymptotic inference for a change-point Poisson process, *The Annals of Statistics*, **14**(4), 1,583–1,590.

Berger, J. O., Pericchi, L. R. and others. (2004). Training samples in objective Bayesian model selection, *The Annals of Statistics*, **32**(3), 841–869.

Csörgő, M. and L. Horváth (1997). Limit theorems in change-point analysis, *John Wiley and Sons Inc*, Vol. 18.

Ellsworth, William L. (2013). Injection-induced earthquakes, *Science*, **341**(6142), 1225942.

Ghosal, S., and Samanta, T. (1995). Asymptotic behaviour of Bayes estimates and posterior distributions in multiparameter nonregular cases, *Mathematical Methods of Statistics*, **4**(4), 361–388.

Ghosal, S., Ghosh, J. K. and T. Samanta (1999). Approximation of the posterior distribution in a change-point problem, *Annals of the Institute of Statistical Mathematics*, **51**(3), 479–497.

Ghosh, J. K., Ghosal, S., and T. Samanta (1994). Stability and convergence of the posterior in non-regular problems, *Statistical Decision Theory and Related Topics V*, Springer, 183–199.

Gupta, Abhineet, and Jack W. Baker (2015). A Bayesian change point model to detect changes in event occurrence rates, with application to induced seismicity. *In: 12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP12, Vancouver, Canada, 8 pp.*.

Gupta, Abhineet, and Jack W. Baker (2017). Estimating spatially varying event rates with a change point using Bayesian statistics: Application to induced seismicity, *Structural Safety*, **65**, 1–11.

Kass, Robert E. and Adrian E. Raftery (1995). Bayes factors, *Journal of the American Statistical Association*, **90**(430), 773–795.

Keranen, K. M., Savage, H. M., Abers, G. A. and Cochran, E. S. (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence, *Geology*, **41**(6), 699–702.

Langenbruch, C. and Zoback, M. D. (2016). How will induced seismicity in Oklahoma respond to decreased saltwater injection rates?, *Science advances*, **2**(11), e1601542.

Montoya-Noguera, S. and Wang, Y. (2017). Bayesian identification of multiple seismic change points and varying seismic rates caused by induced seismicity, *Geophysical Research Letters*, **44**(8), 3,509–3,516.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of the American Statistical Association*, **83**(401), 9–27.

Okabe, A., Boots, B., Sugihara, K., Chiu, S. N. and Kendall, D. G. (2008). Definitions and Basic Properties of Voronoi Diagrams. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, , *John Wiley and Sons, Inc.* , Vol. 2, 43–228.

Raftery, A. E. and Akman, V. E. (1986). Bayesian analysis of a Poisson process with a change-point, *Biometrika*, **73**(1), 85–89.

Reasenberg, P. (1985). Second-order moment of central California seismicity, 1969-1982, *Journal of Geophysical Research: Solid Earth*, **90**(B7), 5479–5495.

Spiegelhalter, D. J. and Smith, A. F. (1982). Bayes factors for linear and log-linear models with vague prior information, *Journal of the Royal Statistical Society. Series B (Methodological)*, **44**(3), 377–387.

van Stiphout, T., Zhuang, J. and Marsan, D. (2012). Seismicity declustering, *Community Online Resource for Statistical Seismicity Analysis*, **10**(1).

Walsh, F. R. and Zoback, M. D. (2015). Oklahoma's recent earthquakes and saltwater disposal, *Science advances*, **1**(5), e1500195.

Wang, P., Small, M. J., Harbert, W. and Pozzi, M. (2016). A Bayesian approach for assessing seismic transitions associated with wastewater injections, *Bulletin of the Seismological Society of America*, **106**(3), 832–845.

Weingarten, M., Ge, S., Godt, J. W., Bekins, B. A. and Rubinstein, J. L. (2015). High-rate injection is associated with the increase in US mid-continent seismicity, *Science*, **348**(6241), 1336–1340.

Wessel, P., and W. H. F. Smith (1998). New, improved version of generic mapping tools released, *EOS Trans. AGU*, **79**, 579.

Witting, H. and Müller-Funk, U. (1995). Mathematische Statistik: asymptotische Statistik: parametrische Modelle und nichtparametrische Funktionale, *Teubner*, Vol. 2, 215–235.

Zhuang, J., Ogata, Y. and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences, *Journal of the American Statistical Association*, **97**(458), 369–380.

Bernhard Fiedler, Institute of Mathematics, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany. Email bfiedler@uni-potsdam.de

Gert Zöller, Institute of Mathematics, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany.

Matthias Holschneider, Institute of Mathematics, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany.

Sebastian Hainzl, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany

Tables

Table 1

Table 1: Estimation of the α-error simulations

$\lambda_1 = \lambda_2$	theoretical α	estimated α	number of events
1	0.05	0.061	10
1	0.05	0.057	50
1	0.05	0.052	100
1	0.05	0.049	1000

Table 2

Table 2: Bayes factors for the declustered catalog of $M \geq 3$ earthquakes in Oklahoma. The results indicates that two change-points are preferable.

Bayes factor	Decision
$B_{01} = 3.73 \times 10^{-158}$	$\mathcal{M}_1$
$B_{02} = 1.67 \times 10^{-197}$	$\mathcal{M}_2$
$B_{03} = 1.16 \times 10^{-197}$	$\mathcal{M}_3$
$B_{12} = 4.47 \times 10^{-40}$	$\mathcal{M}_2$
$B_{13} = 3.12 \times 10^{-40}$	$\mathcal{M}_3$
$B_{23} = 0.69$	$\mathcal{M}_2$

Figure captions

Figure 1

(**A**) Magnitude-time plot for all earthquakes in Oklahoma from January 1, 1980 to December 31, 2015. (**B**) Cumulative number of earthquakes with $M \geq 3$ in Oklahoma from January 1, 1980 to December 31, 2015. Inset: Spatial map of all earthquakes with $M \geq 3$ (time color-coded).

Figure 2

Estimation of α error and power depending on λ and number of events n for a hypothesis test defined as $\mathcal{H}_0 : \ \lambda_1 = \lambda_2$ versus $\mathcal{H}_1 : \ \lambda_1 \neq \lambda_2$. Plots show the behavior of the empirical cumulative distribution function (ecdf) of the p-values generated under the null hypothesis $\mathcal{H}_0$ and its alternative $\mathcal{H}_1$. Here we have $n_1 + n_2 = 400$ events and 1000 random realizations were generated.

Figure 3

Schematic diagram presenting the steps for our scan algorithm. (**A**) A certain area is subdivided into m subareas $\mathcal{A}_1, \ldots, \mathcal{A}_m$. (**B**) Every subarea $\mathcal{A}_i$ is a disk with the same radius r. (**C**) Events within the subarea $\mathcal{A}_i$ occur at n_i times t_{ij}, so we can project it into a three-dimensional domain $\mathcal{A}_i \cup \mathcal{S}_i$. (**D**) The time series $\mathcal{S}_i$ is investigated with regard to (i) model selection with Bayes factors, (ii) estimation of change-points, (iii) significance of change-points, and (iv) credibility intervals.

Figure 4

Results based on 100 synthetic sequences for every evaluation: (**A**) Standard deviation of $\hat{\tau} - \tau_{real}$ and (**B**) percentage of change-point detections by means of the Bayes factor as a function of the number n of data points in the simulation and the ratio λ_1/λ_2 of intensities in the first and second half of the simulations.

Figure 5

Synthetic data: Generation of a 3D homogeneous Poisson process with different intensities. The sample size is 2000. (**A**) Poisson process with a rate $\lambda \approx 0.8$ (per unit area) and (**B**) Poisson processes with different rates within the replaced cylinders i.e. the intensity in the change-point domain is given by $\lambda_{\mathrm{cp}} \approx 8$ (per unit area).

Figure 6

Synthetic data: (**A**) Perspective view of the circle $\mathcal{C}_1$ and the change-point domain $\mathcal{C}_2$ with the estimated significant change-point locations. (**B**) Example for the marginal posterior $p(\tau \mid \underline{t})$ in the change-point domain $\mathcal{C}_1$. (**C**) Example for the marginal posterior $p(\tau_1,\ \tau_2 \mid \underline{t})$ in the change-point domain $\mathcal{C}_2$. The logarithmic values of the density are color coded.

Figure 7

Synthetic data of a Poisson process with an intensity of 0.08 (per unit area) in the time period [0, 20] in which a change-point domain is embedded (intensity $\lambda_{\mathrm{cp}} \approx 0.40$ within a cylinder with radius $r_0 = 2$, center $(0, 0)$ and $t \in [10, 20]$): (**A**) Standard deviation of $\hat{\tau} - \tau_{real}$ and (**B**) percentage of change-point detections by means of the Bayes factor as a function of the selection radius.

Figure 8

(**A**) Magnitude-time plot with the estimated change-points for the whole declustered time series. (**B**) Cumulative number of earthquakes with $M \geq 3$ for the declustered catalog with the estimated change-points (model with one change-point (green line) and two change-points (red lines). Inset: Cumulative number of earthquakes for the non-declustered catalog with the estimated change-points (model with three change-points), where the third change-point coincides with the occurrence time of the $M_W = 5.6$ mainshock.

Figure 9

Maps with transition events and the $M_W = 5.6$ earthquake for the case study Oklahoma. (**A**) and (**B**) Color-coded times of the first and second change-points at grid points where the algorithm prefers two change-points: (**A**) first change-point and (**B**) second change-points. (**C**) Illustration of all calculated transition times at grid points where the algorithm preferred a model with one change-point.

Figure 10

Locations and occurrence times of the first change-points (for models with one and with two change-points) in comparison to approval dates of injection wells from 1.1.2000 to

31.12.2015 for the Oklahoma case study. The high-volume injection wells (approved volume $> 10,000$ barrels per day) are illustrated in black. (**A**) Map view of the estimated change-points, (**B**) latitude-time plot, and (**C**) time-longitude plot with estimated transitions and injection wells.

439 # Figures

440 ## Figure 1

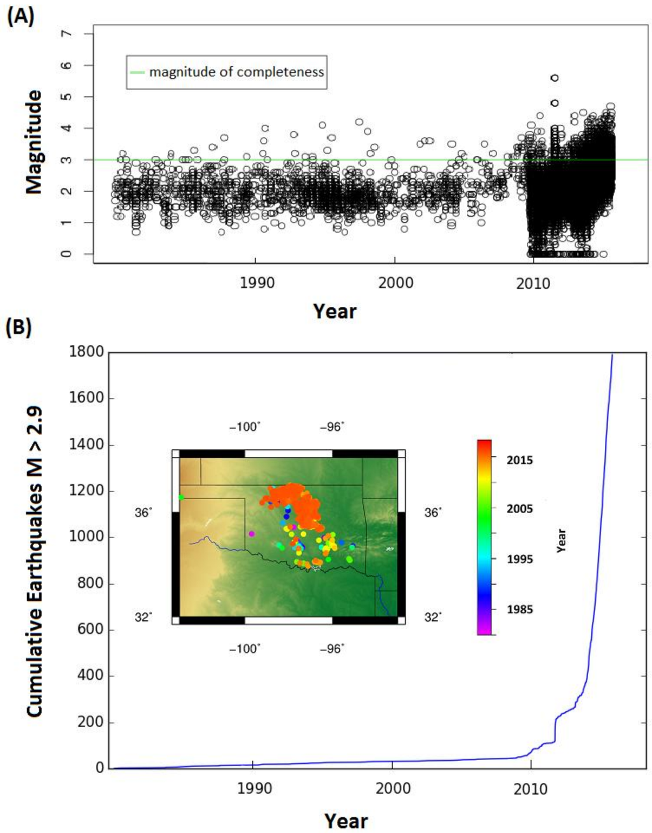

Figure 1: (**A**) Magnitude-time plot for all earthquakes in Oklahoma from January 1, 1980 to December 31, 2015. (**B**) Cumulative number of earthquakes with $M \geq 3$ in Oklahoma from January 1, 1980 to December 31, 2015. Inset: Spatial map of all earthquakes with $M \geq 3$ (time color-coded).

Figure 2

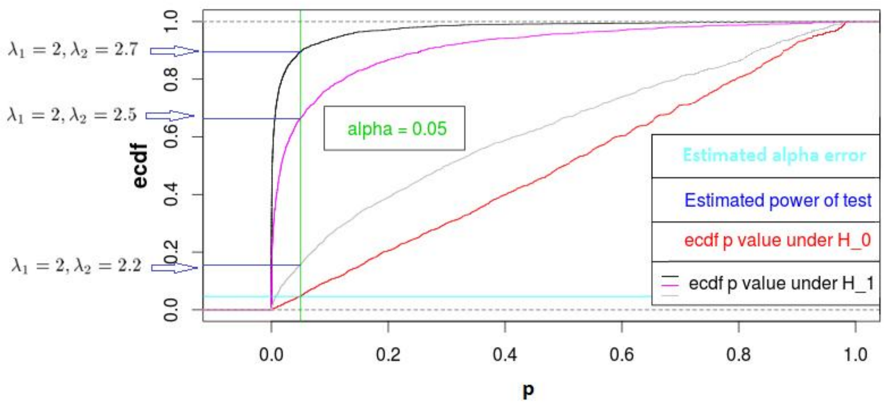

Figure 2: Estimation of α error and power depending on λ and number of events n for a hypothesis test defined as $\mathcal{H}_0:\ \lambda_1 = \lambda_2$ versus $\mathcal{H}_1:\ \lambda_1 \neq \lambda_2$. Plots show the behavior of the empirical cumulative distribution function (ecdf) of the p-values generated under the null hypothesis $\mathcal{H}_0$ and its alternative $\mathcal{H}_1$. Here we have $n_1 + n_2 = 400$ events and 1000 random realizations were generated.

Figure 3

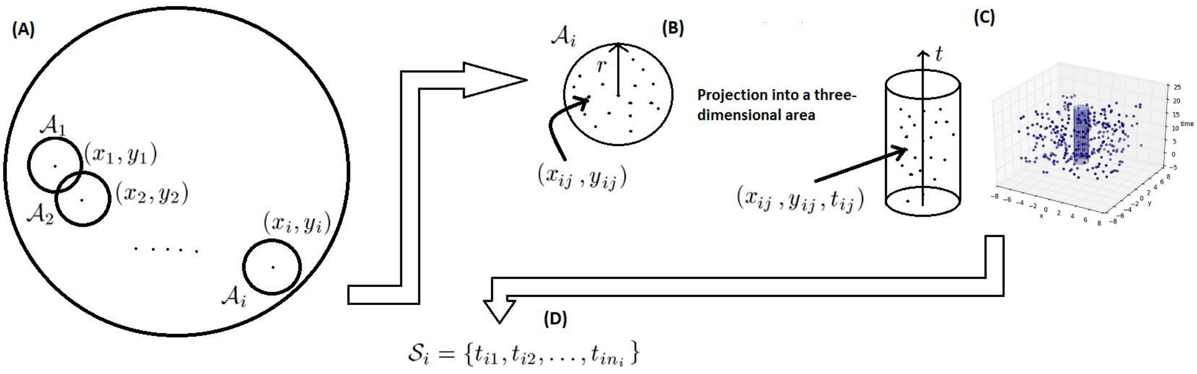

Figure 3: Schematic diagram presenting the steps for our scan algorithm. (**A**) A certain area is subdivided into m subareas $\mathcal{A}_1, \ldots, \mathcal{A}_m$. (**B**) Every subarea $\mathcal{A}_i$ is a disk with the same radius r. (**C**) Events within the subarea $\mathcal{A}_i$ occur at n_i times t_{ij}, so we can project it into a three-dimensional domain $\mathcal{A}_i \cup \mathcal{S}_i$. (**D**) The time series $\mathcal{S}_i$ is investigated with regard to (i) model selection with Bayes factors, (ii) estimation of change-points, (iii) significance of change-points, and (iv) credibility intervals.

Figure 4

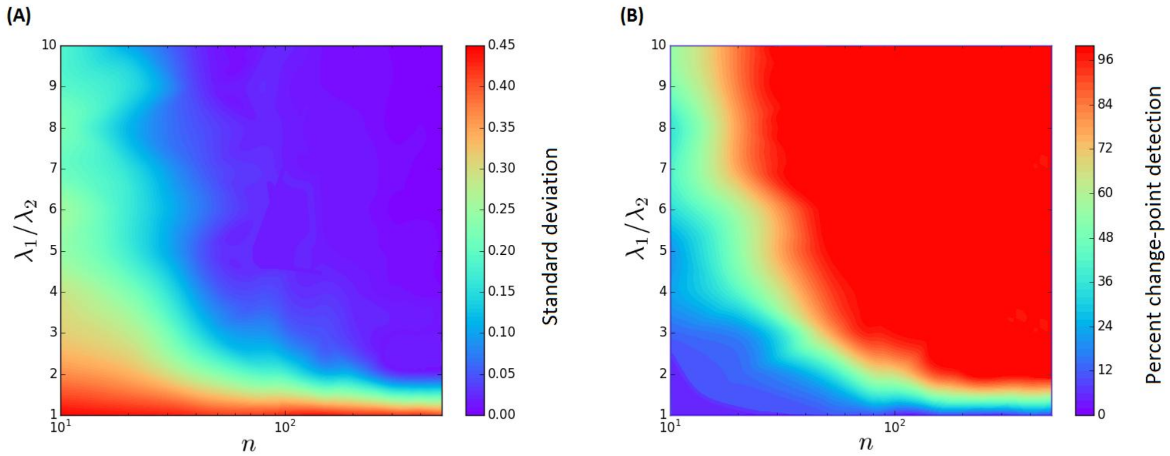

Figure 4: Results based on 100 synthetic sequences for every evaluation: (**A**) Standard deviation of $\hat{\tau} - \tau_{real}$ and (**B**) percentage of change-point detections by means of the Bayes factor as a function of the number n of data points in the simulation and the ratio λ_1/λ_2 of intensities in the first and second half of the simulations.

Figure 5

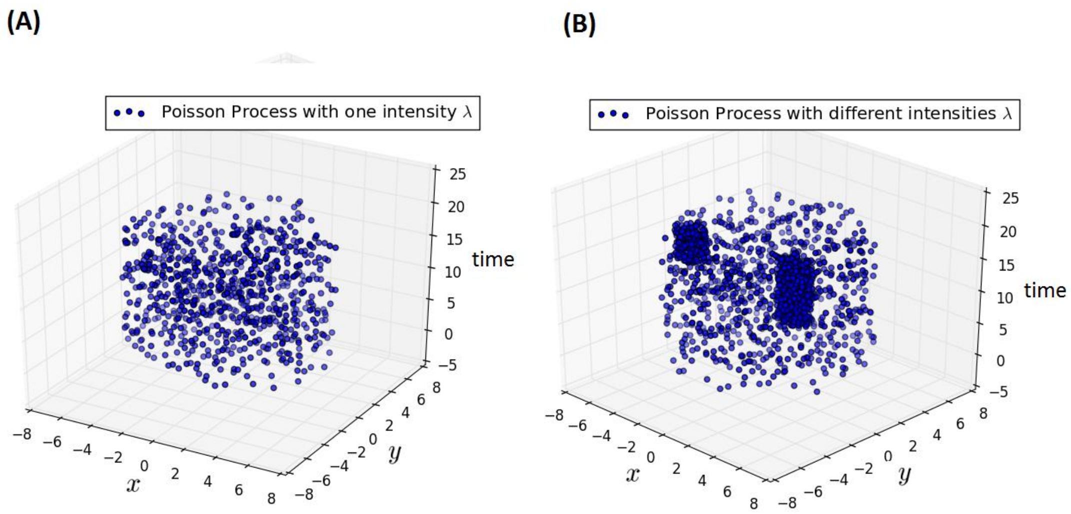

Figure 5: Synthetic data: Generation of a 3D homogeneous Poisson process with different intensities. The sample size is 2000. (**A**) Poisson process with a rate $\lambda \approx 0.8$ (per unit area) and (**B**) Poisson processes with different rates within the replaced cylinders i.e. the intensity in the change-point domain is given by $\lambda_{\rm cp} \approx 8$ (per unit area).

Figure 6

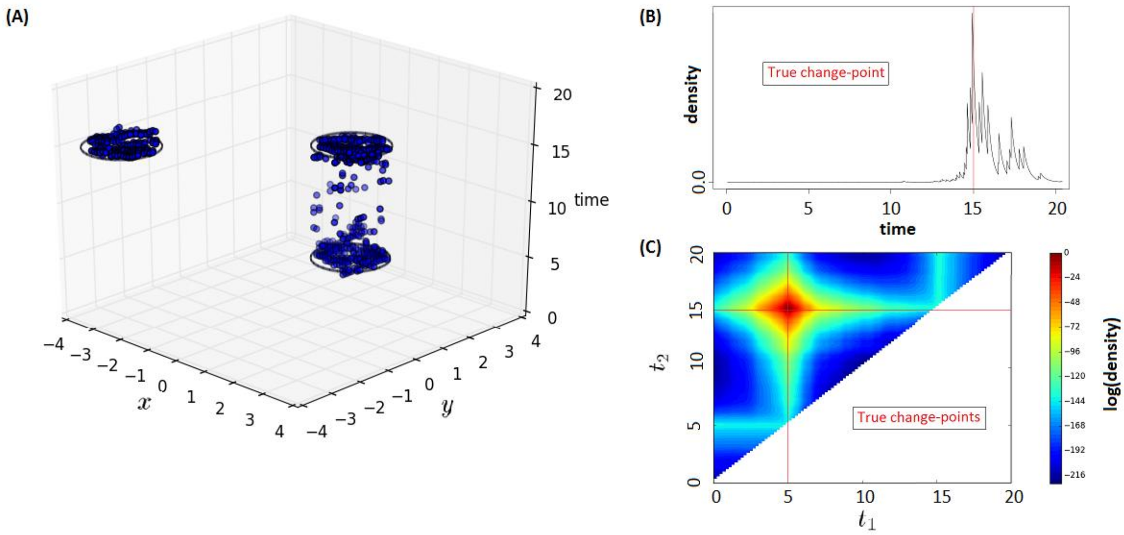

Figure 6: Synthetic data: (**A**) Perspective view of the circle $\mathcal{C}_1$ and the change-point domain $\mathcal{C}_2$ with the estimated significant change-point locations. (**B**) Example for the marginal posterior $p(\tau \mid \underline{t})$ in the change-point domain $\mathcal{C}_1$. (**C**) Example for the marginal posterior $p(\tau_1,\ \tau_2 \mid \underline{t})$ in the change-point domain $\mathcal{C}_2$. The logarithmic values of the density are color coded.

446 **Figure 7**

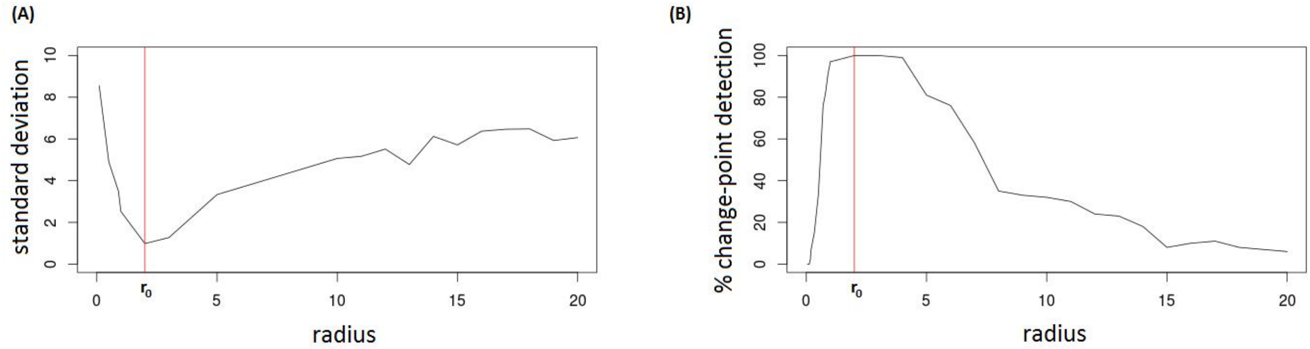

Figure 7: Synthetic data of a Poisson process with an intensity of 0.08 (per unit area) in the time period [0, 20] in which a change-point domain is embedded (intensity $\lambda_{\rm cp} \approx 0.40$ within a cylinder with radius $r_0 = 2$, center $(0, 0)$ and $t \in [10, 20]$): (**A**) Standard deviation of $\hat{\tau} - \tau_{real}$ and (**B**) percentage of change-point detections by means of the Bayes factor as a function of the selection radius.

Figure 8

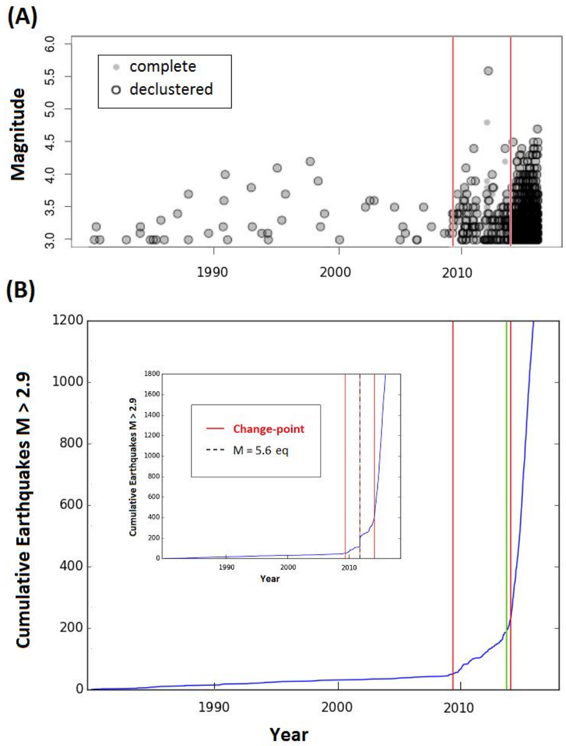

Figure 8: (**A**) Magnitude-time plot with the estimated change-points for the whole declustered time series. (**B**) Cumulative number of earthquakes with $M \geq 3$ for the declustered catalog with the estimated change-points (model with one change-point (green line) and two change-points (red lines). Inset: Cumulative number of earthquakes for the non-declustered catalog with the estimated change-points (model with three change-points), where the third change-point coincides with the occurrence time of the $M_W = 5.6$ mainshock.

Figure 9

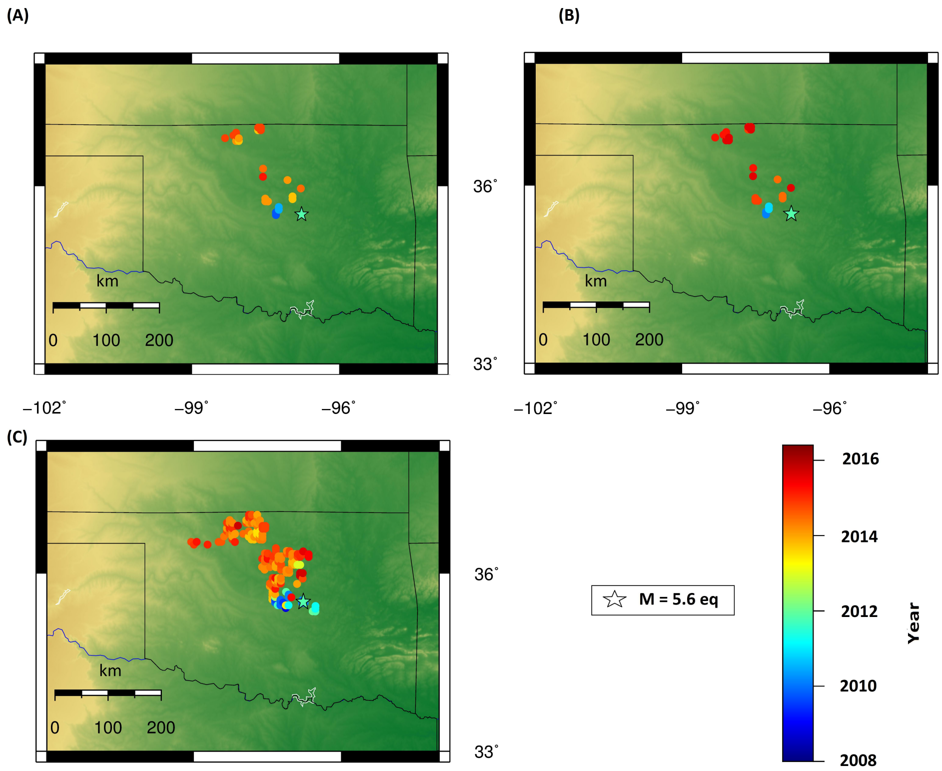

Figure 9: Maps with transition events and the $M_W = 5.6$ earthquake for the case study Oklahoma. (**A**) and (**B**) Color-coded times of the first and second change-points at grid points where the algorithm prefers two change-points: (**A**) first change-point and (**B**) second change-points. (**C**) Illustration of all calculated transition times at grid points where the algorithm preferred a model with one change-point.

449 **Figure 10**

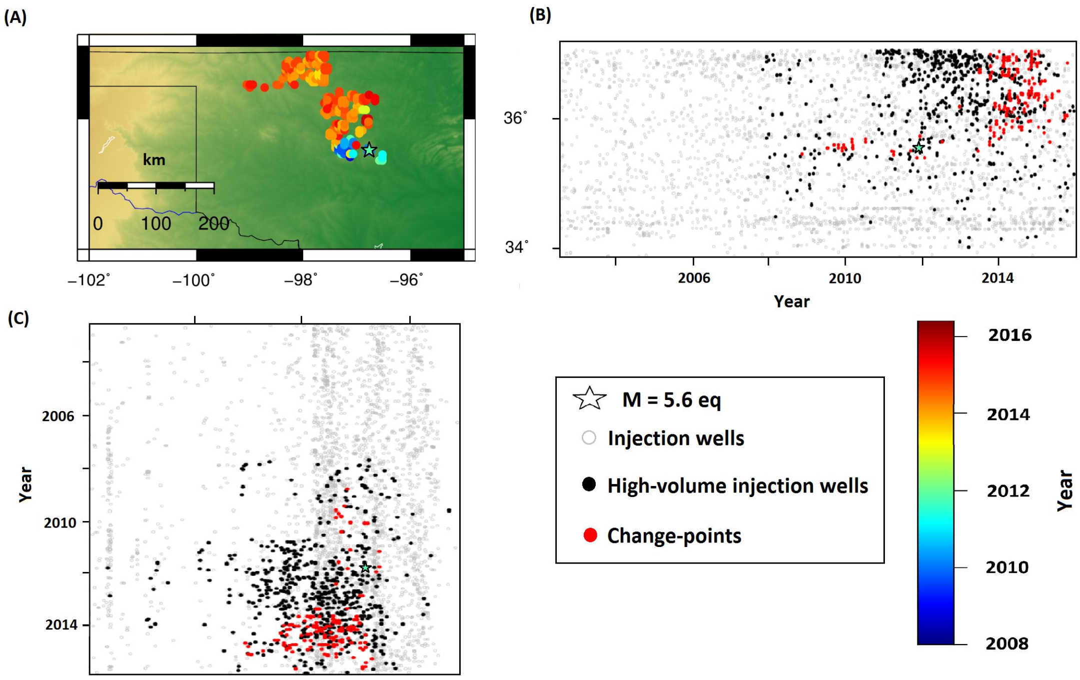

Figure 10: Locations and occurrence times of the first change-points (for models with one and with two change-points) in comparison to approval dates of injection wells from 1.1.2000 to 31.12.2015 for the Oklahoma case study. The high-volume injection wells (approved volume $> 10,000$ barrels per day) are illustrated in black. (**A**) Map view of the estimated change-points, (**B**) latitude-time plot, and (**C**) time-longitude plot with estimated transitions and injection wells.

Appendix

Derivation of the Marginal Posterior Density

With the notation of the *Estimation of Change-Points* section, we derive the formula for the Bayesian posterior density Eq. (6). Here we use Fubini's theorem and the definition of the gamma function

$$\Gamma(x) = \int_0^\infty z^{x-1} e^{-z}\, dz \tag{A1}$$

or more precisely the following calculation:

$$\begin{aligned}\int_0^\infty \lambda_i^{(N(\tau_i)-N(\tau_{i-1})} e^{-\lambda_i(\tau_i-\tau_{i-1})}\, d\lambda_i &= \int_0^\infty \left(\frac{z}{\tau_i-\tau_{i-1}}\right)^{N(\tau_i)-N(\tau_{i-1})} e^{-z} \frac{dz}{\tau_i-\tau_{i-1}} \\ &= (\tau_i-\tau_{i-1})^{-(N(\tau_i)-N(\tau_{i-1})+1)} \int_0^\infty z^{N(\tau_i)-N(\tau_{i-1})+1-1} e^{-z}\, dz \\ &= (\tau_i-\tau_{i-1})^{-(N(\tau_i)-N(\tau_{i-1})+1)} \Gamma(N(\tau_i)-N(\tau_{i-1})+1)\end{aligned} \tag{A2}$$

Derivation of the Likelihood Ratio Test

Based on the test problem

$$\mathcal{H}_0 : \lambda_1 = \lambda_2 \text{ versus } \mathcal{H}_1 : \lambda_1 \neq \lambda_2. \tag{A3}$$

the likelihood function for two different rates is given by

$$p(\underline{t} \mid \lambda_1, \lambda_2) = \lambda_1^{n_1} \exp(-\lambda_1 \Delta_1) \lambda_2^{n_2} \exp(-\lambda_2 \Delta_2), \tag{A4}$$

where $\Delta_1 = s_2 - s_1$ and $\Delta_2 = s_4 - s_3$.

The log-likelihood function is given by

$$l\left(\lambda_1, \lambda_2 \mid \underline{t}\right) = n_1 \log \lambda_1 - \lambda_1 \Delta_1 + n_2 \log \lambda_2 - \lambda_2 \Delta_2. \tag{A5}$$

Under $\mathcal{H}_1$ we have to calculate the maximum likelihood estimator (MLE) for λ_1 and λ_2.

From

$$\frac{\partial l\left(\lambda_1, \lambda_2 \mid \underline{t}\right)}{\partial \lambda_1} = \frac{n_1}{\lambda_1} - \Delta_1 \stackrel{!}{=} 0 \tag{A6}$$

we get

$$\hat{\lambda}_1 = \frac{n_1}{\Delta_1}. \tag{A7}$$

Furthermore

$$\frac{\partial^2 l\left(\lambda_1, \lambda_2 \mid \underline{t}\right)}{\partial \lambda_1^2} = -\frac{n_1}{\lambda_1^2} < 0 \text{ for all } \lambda_1 \in \mathbb{R}^+. \tag{A8}$$

So $\hat{\lambda}_1$ is the MLE for λ_1. In the same way we can show that $\hat{\lambda}_2 = \dfrac{n_2}{\Delta_2}$ is the MLE for λ_2.

Under $\mathcal{H}_0$ is $\lambda = \lambda_1 = \lambda_2$, so we get the likelihood

$$l\left(\underline{t} \mid \lambda\right) = \lambda^{n_1 + n_2} \exp[-\lambda(\Delta_1 + \Delta_2)]. \tag{A9}$$

The log-likelihood function is given by

$$l(\lambda \mid \underline{t}) = (n_1 + n_2)\log\lambda - \lambda(\Delta_1 + \Delta_2). \tag{A10}$$

Thus

$$\frac{\partial l(\lambda \mid \underline{t})}{\partial\lambda} = \frac{n_1 + n_2}{\lambda} - (\Delta_1 + \Delta_2) \stackrel{!}{=} 0, \tag{A11}$$

which leads to

$$\hat{\lambda} = \frac{n_1 + n_2}{\Delta_1 + \Delta_2}. \tag{A12}$$

Furthermore

$$\frac{\partial^2 l(\lambda \mid \underline{t})}{\partial\lambda^2} = -\frac{n_1 + n_2}{\lambda^2} < 0 \text{ for all } \lambda \in \mathbb{R}^+. \tag{A13}$$

So $\hat{\lambda}$ is the MLE for λ.

In general the test statistic is given by

$$Z = 2\ln\left[\frac{p(\underline{t} \mid \mathcal{H}_1)}{p(\underline{t} \mid \mathcal{H}_0)}\right]. \tag{A14}$$

Hence

$$Z = 2\left[l\left(\hat{\lambda}_1, \hat{\lambda}_2 \mid \underline{t}\right) - l\left(\hat{\lambda} \mid \underline{t}\right)\right] \tag{A15}$$

leads to

$$\begin{aligned} Z = 2\Bigg[&n_1\log\left(\frac{n_1}{\Delta_1}\right) - \frac{n_1}{\Delta_1}(\Delta_1) + n_2\log\left(\frac{n_2}{\Delta_2}\right) - \frac{n_2}{\Delta_2}(\Delta_2) \\ &- \left((n_1 + n_2)\log\left(\frac{n_1 + n_2}{\Delta_1 + \Delta_2}\right) - \frac{n_1 + n_2}{\Delta_1 + \Delta_2}(\Delta_1 + \Delta_2)\right)\Bigg] \end{aligned}$$

and finally to

$$Z = 2\left[n_1\log\left(\frac{n_1}{\Delta_1}\right) + n_2\log\left(\frac{n_2}{\Delta_2}\right) - (n_1 + n_2)\log\left(\frac{n_1 + n_2}{\Delta_1 + \Delta_2}\right)\right]. \tag{A16}$$

Derivation of the Bayes Factors

The Bayes factor is defined by the ratio of the marginal or integrated likelihood for the two considered models $\mathcal{M}_l$ (model with l change-points) and $\mathcal{M}_m$ (model with m change-points), i.e.

$$B_{lm} = \frac{p(\underline{t} \mid \mathcal{M}_l)}{p(\underline{t} \mid \mathcal{M}_m)}, \tag{A17}$$

with $l, m = 0, \ldots, k$ and $l \neq m$. For $\mathcal{M}_0$ and $\mathcal{M}_1$ we get

$$p(\underline{t} \mid \mathcal{M}_0) = \int_0^\infty p(\lambda)\lambda^n e^{-\lambda(b-a)}\, d\lambda \tag{A18}$$

and

$$p(\underline{t} \mid \mathcal{M}_1) = \int_a^b \int_0^\infty \int_0^\infty p(\tau)p(\lambda_1)p(\lambda_2)\lambda_1^{N(\tau)} e^{-\lambda_1(\tau-a)} \lambda_2^{N(b)-N(\tau)} e^{-\lambda_2(b-\tau)}\, d\lambda_1\, d\lambda_2\, d\tau. \tag{A19}$$

For $l \geq 2$ we obtain

$$\begin{aligned} p(\underline{t} \mid \mathcal{M}_l) = \int_\Lambda \int_T & p(\tau_1)p(\lambda_1)p(\lambda_{l+1})\lambda_1^{N(\tau_1)} e^{-\lambda_1(\tau_1-a)} \lambda_{l+1}^{N(b)-N(\tau_l)} e^{-\lambda_{l+1}(b-\tau_l)} \\ & \times \prod_{i=2}^{l} p(\tau_i)p(\lambda_i)\lambda_i^{N(\tau_i)-N(\tau_{i-1})} e^{-\lambda_i(\tau_i-\tau_{i-1})}\, d\lambda_1 \ldots d\lambda_{l+1}\, d\tau_1 \ldots d\tau_l. \end{aligned} \tag{A20}$$

Here is $\Lambda = (0, \infty)^{l+1}$ and $T = (a, b)^l$.

To evaluate Eq. (A18), Eq. (A19) and Eq. (A20) we use improper prior densities for the intensities so that $p(\lambda) = c_0\lambda^{-\frac{1}{2}}$ and $p(\underline{\lambda}) = c_k\lambda_1^{-\frac{1}{2}} \ldots \lambda_{k+1}^{-\frac{1}{2}}$, where c_i is a not further specified constant. Moreover we formulate uniform distributed priors for τ_i, i.e. $p(\tau_i) = \frac{1}{b-a}$ (compare Raftery and Akman (1986)). For this approach Eq. (A18) becomes

$$\begin{aligned} p(\underline{t} \mid \mathcal{M}_0) &= \int_0^\infty c_0\lambda^{-\frac{1}{2}}\lambda_1^n e^{-\lambda(b-a)}\, d\lambda \\ &= c_0(b-a)^{-(n+\frac{1}{2})}\Gamma(n+\frac{1}{2}). \end{aligned} \tag{A21}$$

Further Eq. (A19) becomes

$$\begin{aligned} p(\underline{t} \mid \mathcal{M}_1) &= \int_a^b \int_0^\infty \int_0^\infty \frac{c_1}{b-a} \lambda_1^{N(\tau)-\frac{1}{2}} e^{-\lambda_1(\tau-a)} \lambda_2^{N(b)-N(\tau)-\frac{1}{2}} e^{-\lambda_2(b-\tau)}\, d\lambda_1\, d\lambda_2\, d\tau \\ &= \frac{c_1}{b-a} \sum_{i=0}^{n} \Gamma(i+\frac{1}{2}) \Gamma(n-i+\frac{1}{2}) \int_{t_i}^{t_{i+1}} (\tau-a)^{-(i+\frac{1}{2})} (b-\tau)^{-(n-i+\frac{1}{2})} d\tau, \end{aligned} \tag{A22}$$

with $t_0 = a$ and $t_{n+1} = b$. The resulting Bayes factor B_{01} contains an unspecified constant c_0/c_1, which can determined by using the boundary condition $B_{01} \approx 1$, if we consider an observation period of $[a, b]$ consisting only a single event $t_1 = (a + b)/2$. So Eq. (A22) becomes

$$\begin{aligned} p(\underline{t} \mid \mathcal{M}_1) &= \frac{c_1}{b-a} \sum_{i=0}^{1} \Gamma(i+\frac{1}{2}) \Gamma(n-i+\frac{1}{2}) \int_{t_i}^{t_{i+1}} (\tau-a)^{-(i+\frac{1}{2})} (b-\tau)^{-(1-i+\frac{1}{2})} d\tau \\ &= \frac{c_1}{b-a} \Gamma(0.5)\Gamma(1.5) \left[\int_a^{(a+b)/2} (\tau-a)^{-\frac{1}{2}} (b-\tau)^{-\frac{3}{2}} d\tau \right. \\ &\quad \left. + \int_{(a+b)/2}^{b} (\tau-a)^{-\frac{3}{2}} (b-\tau)^{-\frac{1}{2}} d\tau \right] \\ &= \frac{c_1}{(b-a)^2} 4\sqrt{\pi}\Gamma(1.5). \end{aligned} \tag{A23}$$

If $c_0/c_1 =: c_{01}(a, b)$, we receive by solving $B_{01} \stackrel{!}{=} 1$ that $c_{01}(a, b) = 4\sqrt{\pi}(b-a)^{-\frac{1}{2}}$. Finally we get Eq. (10). In the same way we can evaluate Eq. (A20). Here we have to consider

$$\begin{aligned} p(\underline{t} \mid \mathcal{M}_2) &= c_2 \int_a^b \int_a^b \int_0^\infty \int_0^\infty \int_0^\infty \frac{1}{(b-a)^2} \lambda_1^{N(\tau_1)-\frac{1}{2}} e^{-\lambda_1(\tau_1-a)} \lambda_2^{N(\tau_2)-N(\tau_1)-\frac{1}{2}} \\ &\quad \times e^{-\lambda_2(\tau_2-\tau_1)} \lambda_3^{N(b)-N(\tau_2)-\frac{1}{2}} e^{-\lambda_3(b-\tau_2)}\, d\lambda_1\, d\lambda_2\, d\lambda_3\, d\tau_1\, d\tau_2 \\ &= c_2 \frac{1}{(b-a)^2} \sum_{i=0}^{n} \sum_{j=i+1}^{n} \Gamma(i+\frac{1}{2}) \Gamma(j-i+\frac{1}{2}) \Gamma(n-j+\frac{1}{2}) \\ &\quad \times \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} (\tau_1-a)^{-(i+\frac{1}{2})} (\tau_2-\tau_1)^{-(j-i+\frac{1}{2})} (b-\tau_2)^{-(n-j+\frac{1}{2})} d\tau_1 d\tau_2. \end{aligned} \tag{A24}$$

By using the training sample method we obtain

$$
\begin{aligned}
p(\underline{t} \mid \mathcal{M}_2) &= \frac{c_2}{(b-a)^2} \sum_{i=0}^{1} \sum_{j=i+1}^{1} \Gamma(i+\frac{1}{2})\Gamma(j-i+\frac{1}{2})\Gamma(1-j+\frac{1}{2}) \\
&\quad \times \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} (\tau_1 - a)^{-(i+\frac{1}{2})}(\tau_2-\tau_1)^{-(j-i+\frac{1}{2})}(b-\tau_2)^{-(1-j+\frac{1}{2})} d\tau_1 d\tau_2 \\
&= c_2 \frac{[\Gamma(0.5)]^2\Gamma(1.5)}{(b-a)^2} \int_a^{(a+b)/2} \int_{(a+b)/2}^{b} (\tau_1-a)^{-\frac{1}{2}}(\tau_2-\tau_1)^{-\frac{3}{2}}(b-\tau_2)^{-\frac{1}{2}} d\tau_1 d\tau_2 \\
&= \frac{c_2}{(b-a)^{\frac{5}{2}}} 2\pi^2\Gamma(1.5).
\end{aligned}
\tag{A25}
$$

Without loss of generality we assume that $\tau_2 > \tau_1$, so that we have to multiply the resulting constant with the factor 2. This finally leads to $c_{02}(a,b) = 4\pi^2(b-a)^{-1}$ and Eq. (11). To compare $\mathcal{M}_1$ and $\mathcal{M}_2$ we use

$$B_{12} = \frac{B_{02}}{B_{01}}. \tag{A26}$$

For the general case B_{lm}, we first calculate the Bayes factors B_{0l} and B_{0m} by using the training sample method to get the occurring constants as shown in Eq. (A23) or in Eq. (A25) and then straightforward

$$B_{lm} = \frac{B_{0m}}{B_{0l}}. \tag{A27}$$

Using the priors as explained before, evaluation of Eq. (A20) leads to

$$
\begin{aligned}
p(\underline{t} \mid \mathcal{M}_l) = c_l \int_\Lambda \int_T & p(\tau_1)\lambda_1^{N(\tau_1)-\frac{1}{2}} e^{-\lambda_1(\tau_1-a)} p(\tau_l)\lambda_{l+1}^{N(b)-N(\tau_l)-\frac{1}{2}} e^{-\lambda_{l+1}(b-\tau_l)} \\
& \times \prod_{i=2}^{l} p(\tau_i)\lambda_i^{N(\tau_i)-N(\tau_{i-1})-\frac{1}{2}} e^{-\lambda_i(\tau_i-\tau_{i-1})}\, d\lambda_1 \ldots d\lambda_{l+1}\, d\tau_1 \ldots d\tau_l \\
& = \frac{c_l}{(b-a)^l} \sum_{i_1=0}^{n} \cdots \sum_{i_l=i_{l-1}+1}^{n} \Gamma(i_1+\frac{1}{2})\Gamma(n-i_l+\frac{1}{2}) \prod_{j=2}^{l} \Gamma(i_j - i_{j-1} + \frac{1}{2}) \\
& \times \int_{t_{i_1}}^{t_{i_1+1}} \cdots \int_{t_{i_l}}^{t_{i_l+1}} (\tau_1-a)^{-(i_1+\frac{1}{2})}(b-\tau_l)^{-(n-i_l+\frac{1}{2})} \prod_{j=2}^{l} (\tau_j-\tau_{j-1})^{-(i_j-i_{j-1}+\frac{1}{2})} \\
& \times\, d\tau_1 \ldots d\tau_l.
\end{aligned}
\tag{A28}
$$

With the help of the training sample method, the occurring constants can be calculated. As a further example for B_{03} we get $c_{03}(a,b) = 4\sqrt{2}\pi^{\frac{5}{2}}(b-a)^{-\frac{3}{2}}$.

For model selection we use the following algorithm:

i) Define the maximum number k of possible change-points in the investigated data.

ii) Set $m = 0$.

iii) Calculate the Bayes factors B_{ml} with $l = m+1, \ldots, k$.

iv) Calculate $l_{\text{new}} = \underset{l \in \{m+1,\ldots,k\}}{\arg\min} \{B_{ml} < 0.3\}$.

v) If l_{new} exists, set $m = l_{\text{new}}$ and go to step iii). Otherwise, select a model where the number of change-points is equal to m.

Case Study Oklahoma: Evaluation with Different Choices of the Radius

In comparison to the results illustrated in Fig. **9** where we used a radius $r = 5$ km, Fig. **A1** shows the transition events for the radii $r = 2$ and $r = 10$ km.

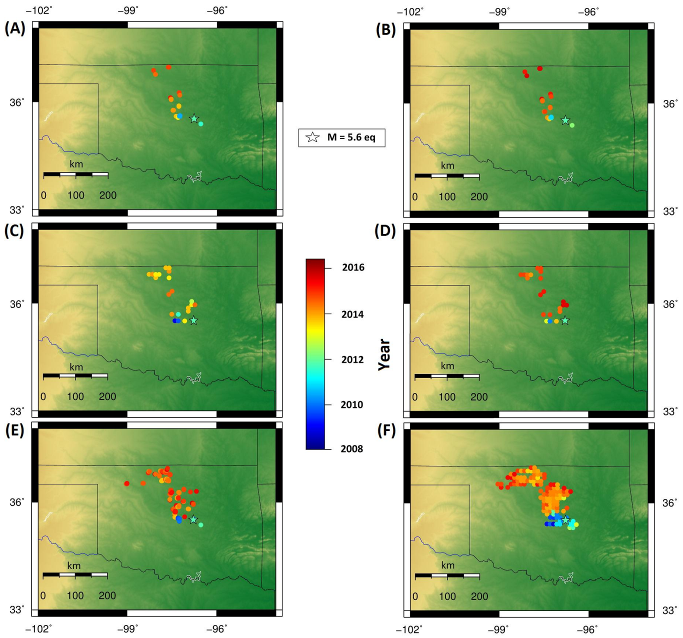

Figure A1: Maps with transition events and the $M_W = 5.6$ earthquake for the case study Oklahoma. (**A**) and (**B**) Illustration of all calculated change-point locations where the algorithm prefers two change-points by using a radius of 2 km. (**C**) and (**D**) Illustration of all calculated change-point locations where the algorithm prefers two change-points by using a radius of 10 km. (**E**) and (**F**) show all calculated transition events where the algorithm prefers a model with one change-point, e.g. (**E**) r = 2 km and (**F**) r = 10 km.