



Theme V – Models and Techniques for Analyzing Seismicity

Seismicity Declustering

Thomas van Stiphout¹ • Jiancang Zhuang² • David Marsan³

1. Swiss Seismological Service, ETH Zurich
2. Institute of Statistical Mathematics
3. Institut des Sciences de la Terre, CNRS, Universite de Savoie

How to cite this article:

van Stiphout, T., J. Zhuang, and D. Marsan (2012), Seismicity declustering, Community Online Resource for Statistical Seismicity Analysis, doi:[10.5078/corssa-52382934](https://doi.org/10.5078/corssa-52382934). Available at <http://www.corssa.org>.

Document Information:

Issue date: 02 February 2012 Version: 1.0

Contents

| | | |
|---|--------------------------------|----|
| 1 | Motivation | 3 |
| 2 | Starting Point | 4 |
| 3 | Ending Point | 4 |
| 4 | Theory | 5 |
| 5 | Available Algorithms | 9 |
| 6 | Final Remarks | 24 |
| 7 | Software | 24 |

Abstract Seismicity [declustering](#), the process of separating an [earthquake catalog](#) into [foreshocks](#), [mainshocks](#), and [aftershocks](#), is widely used in seismology, in particular for seismic hazard assessment and in earthquake prediction models. There are several declustering algorithms that have been proposed over the years. Up to now, most users have applied either the algorithm of [Gardner and Knopoff \(1974\)](#) or [Reasenberg \(1985\)](#), mainly because of the availability of the source codes and the simplicity of the algorithms. However, declustering algorithms are often applied blindly without scrutinizing parameter values or the result. In this article we present a broad range of algorithms, and we highlight the fundamentals of seismicity declustering and possible pitfalls. For most algorithms the source code or information regarding how to access the source code is available on the CORSSA website.

1 Motivation

Generally, scientists understand seismicity to consist of two parts: (1) earthquakes that are independent and (2) earthquakes that depend on each others like aftershocks, foreshocks, or multiplets. Independent earthquakes are assumed to be mostly caused by secular, tectonic loading or, in the case of seismic swarms, by [stress](#) transients that are not caused by previous earthquakes. The second part corresponds to earthquakes triggered by static or dynamic stress changes, seismically-activated fluid flows, after-slip, etc., hence by mechanical processes that are at least partly controlled by previous earthquakes. The process of separating earthquakes into these two classes is known as seismicity declustering.

There is a wide range of terminology for these two classes. Independent earthquakes are also known as [background earthquakes](#), mainshocks, or parent earthquakes, while dependent earthquakes are also called aftershocks, foreshocks, [triggered earthquakes](#), or offspring. The ultimate goal of declustering is therefore to isolate the class of background earthquakes, i.e. earthquakes that are independent of all preceding earthquakes. Alternatively, this corresponds to removing the dependent earthquakes that form seismicity clusters, hence the name 'declustering'. For large enough tectonic regions, the subset of independent earthquakes is expected to be homogeneous in time, i.e., a [stationary Poisson](#) process. Seismic swarms, typically caused by magma or fluid intrusions, are a special case. Although swarms, by definition, consist of independent earthquakes, they are more appropriately modeled as a strongly non-homogeneous Poisson process. Such a process is characterized by a time-varying rate (which models the intrusion) that is not conditioned on the earthquake history of the area: the intrusion generates earthquakes, but is itself not caused by earthquakes.

The identification of background earthquakes is important for many applications in seismology, including seismic hazard assessment, development of clustered seismicity models, earthquake prediction research, and [seismicity rate](#) change estimation. However, this is an ill-posed problem in that it does not have a unique solution. When studying large tectonic areas, one can construct many distinct earthquake subsets that are modeled as a stationary Poisson process. Indeed, for any such subset, any randomly-picked (thinned) sub-subset, e.g., by randomly keeping background earthquakes with a given fixed probability, will also by construction be a stationary Poisson process. The requirement that the selected earthquakes are independent of each other is therefore not sufficient by itself.

All declustering methods must therefore rely on a conceptual model of what is a mainshock. It is this underlying model that distinguishes declustering methods, and also what makes their comparison of interest to seismologists. Since all methods are model-dependent to some extent, there does not exist an a priori 'best' method. As will be detailed in section 5, there are many declustering algorithms. Until recently, most users have been applying variants of the methods proposed by [Gardner and Knopoff \(1974\)](#) or [Reasenber \(1985\)](#) - mostly because they are readily available and relatively simple to apply. The goal of this article is (1) to present an extensive list of available declustering algorithms, (2) to mention if there is a code available, and where to get it, and (3) to discuss pros and cons of each method.

2 Starting Point

The process of seismicity declustering starts with a seismicity catalog containing source parameters such as occurrence time, [hypocenter](#) or [epicenter](#) location, and [magnitude](#). It is important to understand [seismicity catalogs and their problems](#) (for example, [artifacts and quality control](#)) to avoid pitfalls. Due to [incompleteness](#), changes in the magnitude scales, and artifacts in seismicity catalogs can bias the process of declustering, preliminary quality control of the catalog is required. Since declustering methods generally have parameters related to spatio-temporal clustering, as well as epicenter and source depth distributions, reading the Introduction to Basic Feature of Seismicity is also beneficial for this chapter ([Link to Poisson distributions III 7](#)).

3 Ending Point

This article will provide you with practical and theoretical know-how about seismicity declustering. You will find the codes, or information where to get the codes or the original publication, example data as well as explanations to support you in

its applications. For the reasons already mentioned, we avoid making an absolute judgment on which method is best or worst.

4 Theory

The goal of seismicity declustering is to separate earthquakes in the seismicity catalog into independent and dependent earthquakes. Aftershocks, which are dependent earthquakes, cannot be distinguished by any particular, outstanding feature in their waveforms. They can thus only be selected on the basis of their spatio-temporal proximity to other, previous earthquakes, and / or by the fact that they occur at rates greater than the seismicity rate averaged over long durations. To relate an aftershock to a mainshock therefore requires defining a measure of the space-time distance between the two, and a criterion based on this measure that needs to be met. All declustering methods follow this general scheme. Before discussing in depth the available algorithms in detail in Section 5, it is of some interest to give a short summary of how research on declustering has developed over the years.

The first attempts to define whether an earthquake catalog is Poissonian or not were made by [Aki \(1956\)](#) and [Knopoff \(1964\)](#) who found that earthquake catalog do not generally fit a Poisson distribution. [Knopoff \(1964\)](#) probably introduced for the first time a kind of declustering algorithm by excluding the aftershocks from the analysis. They counted earthquakes in successive ten-day bins and found a histogram showing many feature of a Poisson distribution. Ten years later, [Gardner and Knopoff \(1974\)](#) introduced a procedure for identifying aftershocks within seismicity catalogs using inter-event distances in time and space. They also provided specific space-time distances as a function of the mainshock magnitude to identify aftershocks but encouraged readers to try out other values. This method is known as a window method and is one of the simplest forms of aftershock identification. They ignored secondary and higher order aftershocks (i.e., aftershocks of aftershocks): if an earthquake C falls in the triggering windows of the two potential mainshocks A and B, then only the largest shock A or B is kept as the actual mainshock of C, regardless of the possibility that C might be significantly closer in space and time to the other shock. They also did not consider fault extension for larger magnitude earthquakes by assuming circular spatial windows. [Reasenberg \(1985\)](#)'s algorithm allows to link up aftershock triggering within an earthquake cluster: if A is the mainshock of B, and B the mainshock of C, then all A, B and C are considered to belong to one common cluster. When defining a cluster, only the largest earthquake is finally kept to be the cluster's mainshock. Another crucial development in this method is that the space-time distance is based on [Omori's law](#) (for its temporal dependence): as the time from the mainshock increases, the time one must wait for the next aftershock also increases in proportion. Another cluster method was

introduced by [Molchan and Dmitrieva \(1992\)](#) by applying a game theory approach to formulate the problem allowing a whole class of optimal methods of aftershock identification.

So far, the algorithms mentioned here have been [deterministic](#), i.e., each earthquake is classified either as a mainshock or as an aftershock. Another class of seismicity declustering algorithm came with the [stochastic](#) model of [Zhuang et al. \(2002\)](#) which is based on a space-time branching process model (see, for example, [this CORSSA article](#)) to describe how each event generates offspring events. This approach generalizes and improves previous methods in two ways: (1) the choice of the space-time distance is optimized in order to best model the earthquake dataset, within the limits of the [ETAS](#) model. As such, there is no need to assume arbitrary values for the parameters that enter the space-time distance, although the parameterized form of the distance is imposed a priori. This comes with a cost: this optimization can sometimes be time-consuming and delicate to perform. (2) Instead of binary linking an aftershock to only one mainshock, this method gives, for each earthquake, the associated probabilities that it is an aftershock of each preceding earthquake (i.e., all preceding earthquakes are thus potential mainshocks). This makes for a much more sophisticated approach: if the space-time distance is roughly the same between A and C and between B and C, then instead of only keeping either A or B as the mainshock of C, this method keeps both earthquakes as mainshocks of C with roughly equal probability, reflecting the difficulty to make a clear decision in such a case. A limit to this method stems from the use of the ETAS model, as it imposes the parameterized form of the space-time distance. While this is appropriate for the temporal dependence, given the ubiquity of the [Omori-Utsu](#) law for describing the decaying influence of a mainshock, this is not the case anymore when considering the spatial dependence, or the space-time coupling (i.e., change with time of the spatial dependence) as no firm consensus exists on these yet. [Marsan and Lengline \(2008\)](#) went therefore a step further, by generalizing this triggering kernel without assuming a specific form. Moreover, the optimization of their model, based on an Expectation-Maximization algorithm, is easier to compute and more robust than traditional schemes used to invert the ETAS model, thanks to the fact that the solution does not depend on the initial choice of the parameters. Note that the solution however depend on the binning of the kernels, both spatial and temporal. This dependence is very mild as long as each bin is sufficiently populated.

Elaborating on the space-time distance proposed by [Baiesi and Paczuski \(2004\)](#), [Zaliapin et al. \(2008\)](#) showed that the background earthquakes can be identified by exploring a space-time graphical representation of this distance. [Hainzl et al. \(2006\)](#) uses the interevent-time distribution to reconstruct non-parametrically the background earthquake rate and could therefore provide an alternative to standard declustering algorithms. The seismicity declustering algorithms mentioned above are

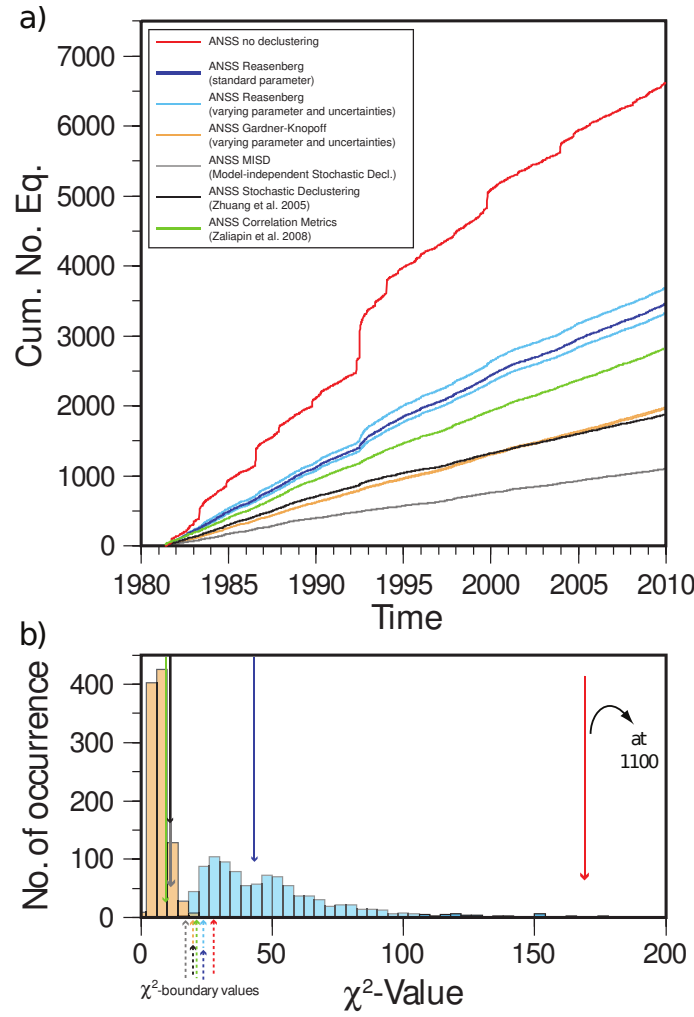


Fig. 1 Results of various declustering methods. (a) Cumulative number of $M \geq 3.5$ earthquakes after declustering the ANSS catalog in the California CSEP testing region ([Schorlemmer and Gerstenberger 2007](#)) between 1981 and 2010. We also report the 5%-and 95%-percentile of 1000 simulated declusterings produced by varying the parameter values of [Reasenberg \(1985\)](#) and [Gardner and Knopoff \(1974\)](#). (Note that the uncertainties of $\pm 10\%$ on the window parameter values of Gardner-Knopoff have only a small effect). (b) Histogram for these realizations with the χ^2 -values (color code is identical to a). The dashed arrows indicate the 5% significance level of being Poissonian (χ^2 -boundary values). For realizations with 1000 simulations histograms are shown, data based on a single run are indicated with a solid arrow. Thus, if the χ^2 -value is below the χ^2 -boundary value, the distribution follows a Poisson one.

described in section 5 together with information on where the code or algorithms is available. [Bottiglieri et al. \(2009\)](#) also uses interevent-time to distinguish Poissonian-like periods and clustered, aftershock activity. This is done by using a simple statistic (the coefficient of variation of the interevent-time) and iteratively finding a threshold for this statistic that allows for this separation in two classes.

All these published declustering algorithms were designed with a specific research focus such as establishing Poissonian background seismicity or analyzing aftershock sequences and their properties. Figure 1 shows the cumulative number of earthquakes of the seismicity background established by several declustering algorithms and, as a quality estimate for declustering, the χ^2 goodness of fit test to determine how well they are fit by a Poisson distribution. For the χ^2 goodness-of-fit test, our null hypothesis is that earthquakes obey a Poisson distribution in time. We test the time distribution of the events in the declustered catalogs and reject the null hypothesis at the 5% significance level. In practice, if the computed χ^2 -statistics of a declustered catalog is smaller than the χ^2 -statistics of a theoretical Poisson distribution, the null hypothesis is accepted and we conclude that the temporal distribution of earthquakes in the declustered catalog follows a Poisson distribution. These results demonstrate the ambiguity of seismicity declustering and how difficult it can be to estimate the quality of the algorithm. According to this test, the seismicity background derived by the methods of *Zhuang et al. (2002)*, *Marsan and Lengline (2008)*, and *Gardner and Knopoff (1974)* follow a Poisson distribution while the absolute numbers varies almost by a factor of two. The resulting background seismicity based on declustering using the method of Reasenberg (1985) with standard parameter values does not follow a Poisson distribution, nor does it for most other parameter values.

There can exist quiet periods during which no large shock occurs and the earthquake rate remains roughly constant at a low level. For traditional methods like those of *Gardner and Knopoff (1974)* and *Reasenberg (1985)*, such periods are devoid of earthquake clusters, and the earthquake rate is then equal to the declustered, hence background rate. On the contrary, for the methods proposed by *Zhuang et al. (2002)* and *Marsan and Lengline (2008)*, even during such quiet periods there is triggering at work. This effect might be more pronounced depending on the earthquake catalog that is used and the magnitude of completeness, i.e. influence of linking earthquakes, and therefore might change the resulting background seismicity (see Figure 4 in *Woessner et al. (2010)*). However, this effect has not been systematically analyzed yet.

The large variation between declustered catalogs, derived using different methods and parameter values (Figure 1), indicates the non-unique and broad view of what earthquake activity is made of, in particular concerning the existence of secondary aftershocks. We therefore emphasize that, if declustering results are used for further studies, the effect of the choice of the algorithm on the results should in principle be tested, for example by varying the space-time distance parameters, or even better by using several distinct methods.

Ideally, seismicity declustering should be applied to a homogeneously recorded and complete seismicity catalog. The data should be free of artifacts such as those discussed in the CORSSA article on [catalog artifacts and quality control](#). Moreover, users should be aware of the possible effect of censored data on the result; censored data are earthquakes located outside the region of interest and occurring before the time of interest. The magnitude threshold (i.e., completeness) can also affect declustering. The censoring of earthquakes implies that triggering chains are severed, and can therefore result in defining too many clusters or can result in too many earthquakes not being identified as part of any cluster that should be.

5 Available Algorithms

5.1 Window Methods

Windowing techniques are a simple way of identifying mainshocks and aftershocks. For each earthquake in the catalog with magnitude M , the subsequent shocks are identified as aftershocks if they occur within a specified time interval $T(M)$, and within a distance interval $L(M)$. Foreshocks are treated in the same way as aftershocks, i.e., if the largest earthquake occurs later in the sequence, the foreshock is treated as an aftershock. Consequently, the time-space windows are reset according to the magnitude of the largest shock in a sequence. Usually, these algorithms do not distinguish between direct and indirect aftershocks, i.e., 1st-generation aftershocks and aftershocks of aftershocks. The aftershock identification windows can vary substantially from one study to the other (see Figure 2 or cf. [Molchan and Dmitrieva \(1992\)](#) Figure 1), and usually do not result from an optimization procedure. We give in Tables 1 and 2 the lengths and durations of these windows, according to ([Knopoff and Gardner 1972](#)) and ([Gardner and Knopoff 1974](#)). An approximation of the windows sizes according to [Gardner and Knopoff \(1974\)](#) is shown in equation 1. Additionally, we present in equation 2 and 3 alternative window parameter settings proposed by Gruenthal through personal communication to the authors of the MATLAB code (see CORSSA Website to this article) and by [Uhrhammer \(1986\)](#). This algorithm is straightforward and easy to implement. The [online supplement to this article](#) provides codes written in Java as well as MATLAB.

$$d = 10^{0.1238*M+0.983} [km] \quad t = \begin{cases} 10^{0.032*M+2.7389}, & \text{if } M \geq 6.5 \\ 10^{0.5409*M-0.547}, & \text{else} \end{cases} [days] \quad (1)$$

$$d = e^{1.77+(0.037+1.02*M)^2} [km] \quad t = \begin{cases} |e^{-3.95+(0.62+17.32*M)^2}|, & \text{if } M \geq 6.5 \\ 10^{2.8+0.024*M}, & \text{else} \end{cases} [days] \quad (2)$$

$$d = e^{-1.024+0.804*M} [km] \quad t = e^{-2.87+1.235*M} [days] \quad (3)$$

| M | L(km) | T(days) |
|-------------|-------|---------|
| ≤ 4.99 | 20 | 100 |
| 5.0–5.49 | 40 | 150 |
| 5.5–5.99 | 70 | 200 |
| 6.0–6.49 | 100 | 280 |
| 6.5–6.99 | 180 | 400 |
| 7.0–7.49 | 300 | 650 |
| 7.5–7.99 | 400 | 1000 |
| 8.0–8.49 | 700 | 1000 |
| 8.5–8.99 | 900 | 1000 |

Table 1 Aftershock identification windows (*Knopoff and Gardner 1972*)

| M | L(km) | T(days) |
|-----|-------|---------|
| 2.5 | 19.5 | 6 |
| 3.0 | 22.5 | 11.5 |
| 3.5 | 26 | 22 |
| 4.0 | 30 | 42 |
| 4.5 | 35 | 83 |
| 5.0 | 40 | 155 |
| 5.5 | 47 | 290 |
| 6.0 | 54 | 510 |
| 6.5 | 61 | 790 |
| 7.0 | 70 | 915 |
| 7.5 | 81 | 960 |
| 8.0 | 94.0 | 985 |

Table 2 Aftershock identification windows (*Gardner and Knopoff 1974*)

5.2 Cluster Method - Reasenber

Reasenber (1985) introduced a method for identifying aftershocks by linking earthquakes to clusters according to spatial and temporal interaction zones. Earthquake clusters thus typically grow in size when processing more and more earthquakes. This method is based on the previous work of *Savage* (1972). The spatial extent of the interaction zone is chosen according to stress distribution near the mainshock,

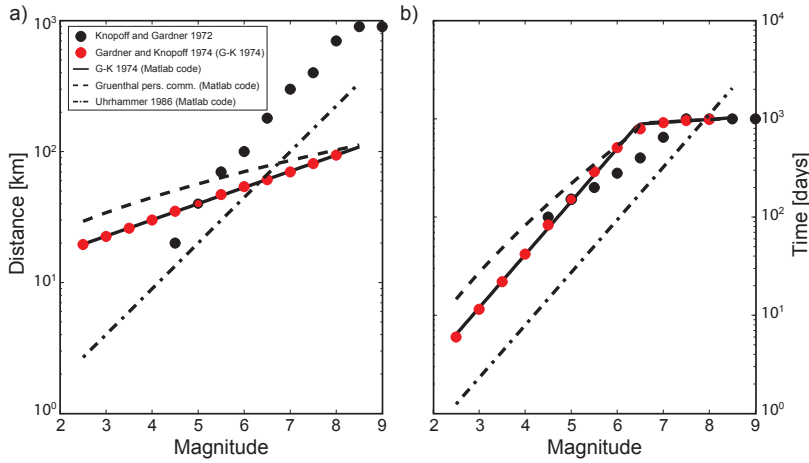


Fig. 2 Aftershock identification windows in space (a) and time (b) domain are shown as a function of the mainshock magnitude. The circles indicate the original parameter values according to [Knopoff and Gardner \(1972\)](#) (Table 1) and [Gardner and Knopoff \(1974\)](#) (Table 2). The function to approximate the values of [Gardner and Knopoff \(1974\)](#) is shown besides two alternative window parameter settings from Gruenthal and Uhrhammer.

and incorporates after-slip, although in a rudimentary way. [Reasenberg \(1985\)](#)'s spatial interaction relationship is defined by the threshold $\log d(km) = 0.4M_0 - 1.943 + k$ ([Molchan and Dmitrieva 1992](#)), where k is 1 for the distance to the largest earthquake and 0 for the distance to the last one. The temporal extension of the interaction zone is based on Omori's law. All linked events define a cluster, for which the largest earthquake is considered the mainshock and smaller earthquakes are divided into fore- and aftershocks. The details of the method can be found in the original paper of [Reasenberg \(1985\)](#), while [Molchan and Dmitrieva \(1992\)](#) provide a condensed summary of this original paper.

The original research focus of [Reasenberg \(1985\)](#) was the detection of fore- and aftershocks in central California between 1969 and 1982 for events with $M \geq 4$. Since then, this algorithm has been a very popular one among the seismological community. It has become common practice to use the standard parameter values of Table 3. However, a specific set of parameter values seems to be an arbitrary choice. In Table 3 we also provide parameter ranges that have been used in the RELM testing center ([Schorlemmer and Gerstenberger 2007](#)). It is recommended to analyze the effect of varying the parameter values.

The parameter τ_{\min} and τ_{\max} denote the minimum and maximum look-ahead time of observing the next earthquake at a certain probability, p_1 . These three parameters are associated according to Equation 4 ([Reasenberg 1985](#), eq. 13), assuming the

Omori rate decay exponent to be 1 and $\Delta M = M_{mainshock} - x_{meff}$, where x_{meff} denotes the minimum magnitude cutoff for the earthquake catalog. During clusters the effective cutoff magnitude, x_{meff} , is raised by a factor x_k of the largest earthquake in the cluster ($x_k, M_{mainshock}$). The parameter r_{fact} denotes the number of crack radii (see [Kanamori and Anderson 1975](#)) surrounding each earthquake within which to consider linking a new event into cluster.

$$\tau = -\ln(1 - p_1)t/10^{2(\Delta M - 1)/3} \quad (4)$$

The latest version of this declustering algorithm has been named **CLUSTER2000**, and can be downloaded from the USGS webpage. The [online supplement to this article](#) also includes a code of the Reasenber algorithm written in MATLAB.

| Parameter | Standard | Min | Max |
|---------------------|----------|-----|------|
| τ_{min} [days] | 1 | 0.5 | 2.5 |
| τ_{max} [days] | 10 | 3 | 15 |
| p_1 | 0.95 | 0.9 | 0.99 |
| x_k | 0.5 | 0 | 1 |
| x_{meff} | 1.5 | 1.6 | 1.8 |
| r_{fact} | 10 | 5 | 20 |

Table 3 Input parameters for declustering algorithm by [Reasenber \(1985\)](#), where τ_{min} is the minimum value of the look-ahead time for building clusters when the first event is not clustered, τ_{max} is the maximum value of the look-ahead time for building clusters, p_1 is the probability of detecting the next clustered event used to compute the look-ahead time, τ , x_k is the increase of the lower cut-off magnitude during clusters: $x_{meff} = x_{meff} + x_k M$, where M is the magnitude of the largest event in the cluster, x_{meff} is the effective lower magnitude cutoff for catalog, r_{fact} is the number of crack radii surrounding each earthquake within new events considered to be part of the cluster. The standard parameter derived for northern California are given in the first column, the second and third column show the ranges for the parameters used for the simulations in the χ^2 goodness of fit test to determine how well they fit a Poisson distribution (Figure 1).

5.3 Stochastic Declustering

The windowing and link-based declustering algorithms discussed in this article involve subjectively chosen parameter values for the window sizes or the link distance. Different choices of parameter values result in different declustered catalogs and different estimates of the background seismicity. The choices of these parameters are usually based on the experience of the researchers for specific data; sometimes they are chosen by a trial and error process driven by the declustering outcome, i.e. based on the temporal smoothness of the declustered catalog.

Alternative to deterministic declustering methods, ideas of probabilistic separation of the background component and clustering component first vaguely appeared in [Kagan and Jackson \(1991\)](#). [Zhuang et al. \(2002\)](#) suggested the stochastic declustering method to bring such a probabilistic treatment into practice based on clustering models, such as the ETAS model. The core of the stochastic declustering method is the estimated background intensity, assumed to be a function of space but not of time, and the parameters associated with clustering structures. Making use of the thinning operation for point processes, one can obtain the probabilities that each event is a background event or a triggered event.

The ETAS model, on which we base the stochastic declustering, can be represented by a [conditional intensity](#) (please see Article 14) in the form of

$$\lambda(t, x, y) = \mu(x, y) + \sum_{\{k: t_k < t\}} \kappa(m_k) g(t - t_k) f(x - x_k, y - y_k | m_k), \quad (5)$$

where $\mu(x, y)$ is the background intensity function and is assumed to be independent of time, and the functions $g(t)$ and $f(x, y | m_k)$ are respectively the normalized response functions (i.e., p.d.f.s) of the occurrence time and the location, the magnitude of an offspring from an ancestor of magnitude m_k . From the fact that the k th event excites a non-stationary Poisson process with intensity function $\kappa(m_k) g(t - t_k) f(x - x_k, y - y_k | m_k)$, it can be seen that $\kappa(m_k)$ represents the expected number of offspring from an ancestor of size m_k .

Suppose that events are numbered in chronological order from 1 to N . In equation 5, the probability of an event j being triggered by the i th event can be estimated naturally as the relative contribution from the i th event from the occurrence rate at the occurrence time and spatial location of the j th event, i.e.,

$$\rho_{ij} = \frac{\kappa(m_i) g(t_j - t_i) f(x_j - x_i, y_j - y_i | m_i)}{\lambda(t_j, x_j, y_j)}. \quad (6)$$

i.e., the relative contribution of the i th event to the occurrence rate at the time and the location of event j . Similarly, the probability that Event j is a background event or a triggered event are, respectively,

$$\varphi_j = \frac{\mu(x_j, y_j)}{\lambda(t_j, x_j, y_j)} \quad (7)$$

and

$$\rho_j = 1 - \varphi_j = \sum_{i=1}^{j-1} \rho_{ij} \quad (8)$$

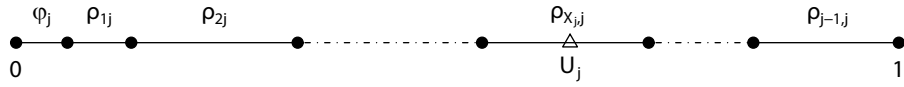


Fig. 3 An illustration of the stochastic declustering algorithm.

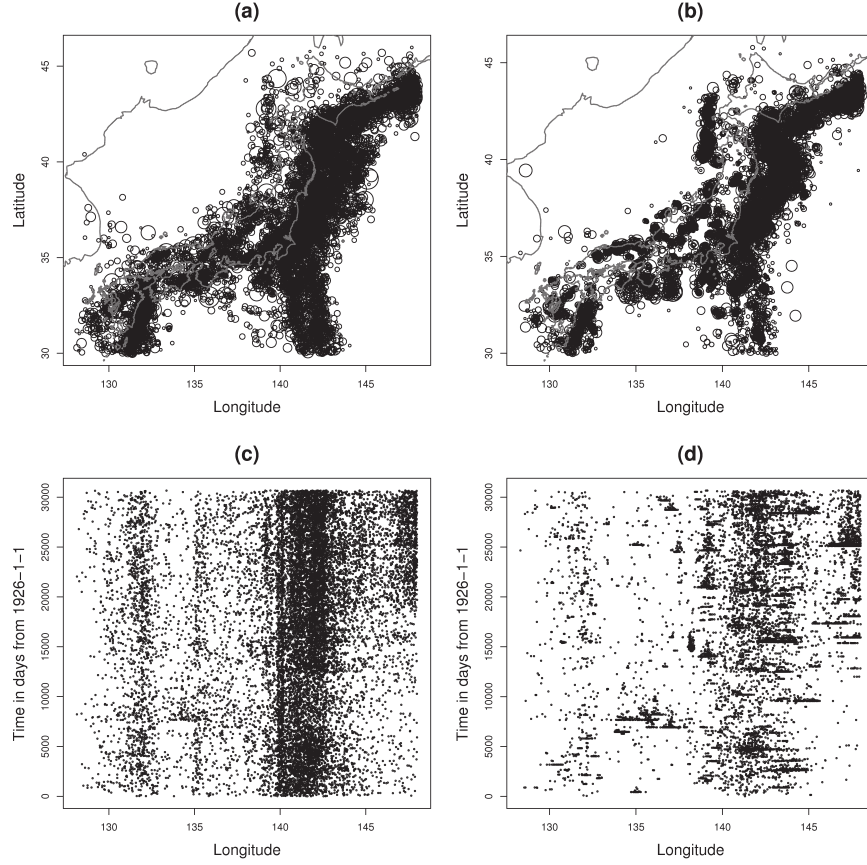


Fig. 4 A realization of stochastically declustered JMA catalog ($M_J \geq 4.0$). (a) and (b) are the location maps for background events and triggered events, respectively. (c) and (d) are space-time plots for background seismicity and triggered seismicity, respectively. Please see the definitions of φ_j and ρ_{ij} in equations 6 and 7.

That is to say, selecting each event j with probability ρ_{ij} , φ_j or ρ_j , we can realize the subprocess triggered by event i , the background subprocess or the clustering subprocess, respectively. Given the estimated model parameters (see Article 14) or [Zhuang et al. \(2002\)](#) for the iterative algorithm for simultaneously estimating the background rate and the model parameters), we can apply the following algorithm to separate the whole catalog into different family trees.

Algorithm: Stochastic classification of earthquake clusters (illustrated as Figure 3)

1. Calculate φ_j and ρ_{ij} by using equations 6 and 7, where $j = 1, 2, \dots, N$ and $i = 1, 2, \dots, j-1$, being the total number of events.
2. For each event j , $j = 1, 2, \dots, N$, generate a random variable U_j , uniformly distributed on $[0, 1]$.
3. For each j , let

$$I_j = \min\{k - 1 : \varphi_j + \sum_{i=1}^k \rho_{ij} \geq U_j \text{ and } 0 \leq k < j\}.$$

If $I_j = 0$, then select j as a background or initial event; else, set the j th event to be a direct offspring of the I_j th event.

Once the catalog is divided into different family trees, we can keep the initiating events in each family as representative of the background seismicity. The mainshock, which is the biggest event in its family, may not be selected as background in this way. However, if preferred, we can use the biggest events in each family instead of the initiating events to create the background catalog. Since the output of stochastic declustering is not unique, we usually generate many copies of the declustered catalogs and use them to test an hypothesis associated with background seismicity or earthquake clustering. Alternatively, we can also work directly on the probabilities φ_j and ρ_{ij} to test such hypotheses. This method is also called stochastic reconstruction, introduced by [Zhuang et al. \(2004\)](#) and [Zhuang \(2006\)](#). Figure 4 shows a realization of stochastically declustered JMA catalog and illustrates the stochastic declustering algorithm.

We provide with the [online supplement to this article](#) the contact (email) of the author of this method.

5.4 Model-Independent Stochastic Declustering

The stochastic declustering of [Zhuang et al. \(2002\)](#), as described above, can be extended to other classes of models (other than ETAS). Indeed, the generalization by [Marsan and Lengline \(2008\)](#) has no specific underlying model, and can therefore accept any (additive) seismicity model, hence the name *Model-Independent Stochastic Declustering* or MISD. Namely, seismicity is described as the following: an earth-

quake A of magnitude m_a in the magnitude interval $[m_i, m_{i+1}]$ and occurring at time t_a triggers aftershocks at location x and time $t > t_a$ with conditional intensity

$$\lambda_a(x, t) = \sum_j \sum_k \lambda_{ijk} \theta(t_j \leq t - t_a < t_{j+1}) \theta(r_k \leq r_a(x) < r_{k+1}) \quad (9)$$

where λ_{ijk} are the unknowns (the triple indices denote (i) magnitude (j) time (k) distance), $\theta(P) = 1$ if proposition P is true, 0 otherwise, $[t_j, t_{j+1}]$ and $[r_k, r_{k+1}]$ are the discretization intervals in time and distance, and $r_a(x)$ is the (2D or 3D) distance between the triggering earthquake and the location of interest x . Compared to ETAS, this triggering kernel also depends on time, distance and magnitude, but with no specific form imposed a priori. Indeed, this formulation is equivalent to a simple piecewise constant triggering kernel. On top of this triggering, which gives the clustered part of the seismicity, background earthquakes occur with constant and spatially uniform rate density μ . MISD first requires to define the discretization intervals in magnitude, time and distance, and then amounts to finding the best λ_{ijk} given the data.

In order to find these unknowns, an Expectation-Maximization algorithm can be used. It is based on the iterative computation of the probabilities ω_{ab} that earthquake a triggered earthquake b , and ω_{0b} that b is a background earthquake, as already introduced in Section 5.3. Two steps are required:

Expectation: given a priori intensities λ_{ijk} and μ , then, for all earthquakes b , compute the probabilities ω_{ab} and ω_{0b} (please note that the nomenclature here follows the one in the original publication; ω_{ab} and ω_{0b} correspond to ρ_{ij} and ϕ_j in the previous section on stochastic declustering), defined as

$$\omega_{ab} = \frac{\lambda_a(x_b, t_b)}{\mu + \sum_{c < b} \lambda_c(x_b, t_b)} \quad (10)$$

and

$$\omega_{0b} = \frac{\mu}{\mu + \sum_{c < b} \lambda_c(x_b, t_b)} \quad (11)$$

where the intensities are defined in Equation 9.

Maximization: Knowing the probabilities ω , we now compute the Maximum Likelihood Estimates of λ_{ijk} and μ . It can be shown that these MLE are

$$\lambda_{ijk} = \frac{n_{ijk}}{n_i (t_{j+1} - t_j) \delta V_k} \quad (12)$$

and

$$\mu = \frac{n_0}{T V} \quad (13)$$

where n_i is the number of earthquakes with magnitude in the interval $[m_i, m_{i+1}]$, δV_k is the volume of the shell $r_k < r < r_{k+1}$, T is the total duration of the dataset, and V its total volume. The 'number' n_{ijk} is the number of earthquake pairs (a, b) such that a has magnitude in the interval $[m_i, m_{i+1}]$, and are separated by $t_b - t_a \in [t_j, t_{j+1}]$ and $r_{ab} \in [r_k, r_{k+1}]$, weighted by the probability ω_{ab} . Similarly, n_0 is the 'number' of background earthquakes: $n_0 = \sum_b \omega_{0b}$.

These two steps are iterated until convergence of λ_{ijk} and μ is obtained. Initial conditions for λ_{ijk} and μ must be provided; the strength of this algorithm is that the final solution does not depend on this initial choice, as long as it is not cumbersome, i.e., one must avoid zero values. Finally, the declustered catalog can be obtained from ω_{0b} : earthquake b is kept as a background earthquake if a random realization of a uniform random number between 0 and 1 is less than ω_{0b} . Figure 5 shows a comparison of the declustered catalogs obtained by using this method, and those obtained using other methods.

[Marsan and Lengline \(2010\)](#) further discuss how the choice of the distance between earthquakes has implications in the resulting declustering. In particular, the best choice is to define r_{ab} as the distance from the fault of earthquake a to the hypocenter of earthquake b .

We provide with the [online supplement to this article](#) the contact (email) of the author of this method.

5.5 Single-link cluster analysis

[Frohlich and Davis \(1990\)](#) proposed a space-time distance between two earthquakes i and j as:

$$d_{ij} = \sqrt{r_{ij}^2 + C^2(t_j - t_i)^2} \quad (14)$$

A scaling constant $C = 1 \text{ km.day}^{-1}$ was found to give satisfactory results. An earthquake j is then the aftershock of i^* if the distance d_{ij} is minimized when $i = i^*$, and if $d_{i^*j} < D$ where the threshold D depends on the background activity of the analyzed region. [Davis and Frohlich \(1991\)](#) further investigated how D can be optimized. They found that $D = 9.4 \text{ km}^{1/2} \cdot \sqrt{S_1} - 25.2 \text{ km}$ gives good results, where S_1 is the [median](#) of all d_{i^*j} distances.

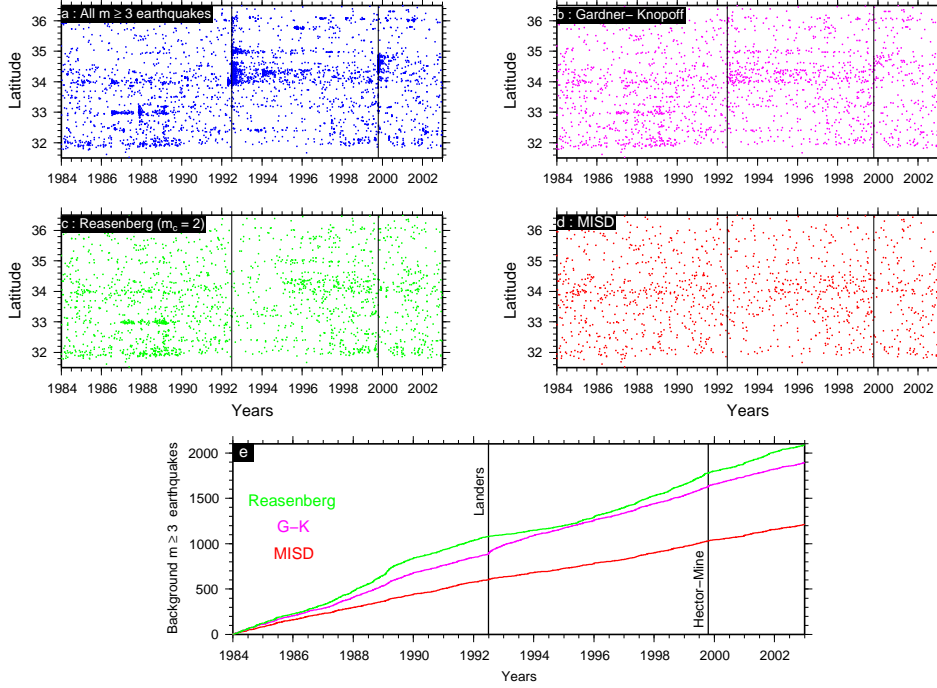


Fig. 5 Comparison of declustered catalog for southern California, 1984-2003. Graphs (a) to (d): latitude vs time of occurrence of earthquakes. (a) All earthquakes in the catalogue. (b) Declustered catalogue using the method by [Gardner and Knopoff \(1974\)](#). (c) Declustered catalogue using the algorithm by [Reasenberg \(1985\)](#). (d) Declustered catalogue using MISC. (e) Cumulative time series of the three declustered catalogues. The two vertical lines indicate the time of occurrence of the 1992 Landers and 1999 Hector Mine earthquakes. Taken from [Marsan and Lengline \(2008\)](#).

5.6 Estimating background rate based on Interevent-Time Distribution

According to the work of [Hainzl et al. \(2006\)](#) the interevent-time distribution of earthquakes correlates with the level of background activity. Analyses on real seismicity, as well as on synthetic seismicity with Poissonian background activity and triggered Omori-type aftershock sequences, indicate that interevent-times τ can be approximated by the gamma distribution:

$$p(\tau) = C \cdot \tau^{\gamma-1} \cdot e^{-\mu\tau} \quad (15)$$

where p is the probability density function of τ , $\mu = \text{var}(\tau)/\bar{\tau}$ is the background rate, $\gamma = \bar{\tau}^2/\text{var}(\tau)$ is the fraction of mainshocks among all earthquakes, and the normalizing constant is $C = \frac{\mu^\gamma}{\Gamma(\gamma)}$. This approach can be used as an independent,

nonparametric estimation of the background rate.

Although not described in [Hainzl et al. \(2006\)](#), this method can be extended to decluster earthquake catalogs by means of a thinning procedure. This is done by comparing $p(\tau)$ to the pdf of the background earthquakes alone, i.e., $p_0(\tau) = \mu e^{-\mu\tau}$, multiplied by the background fraction. We then obtain for each inter-event time τ a probability $P = \frac{\gamma p_0(\tau)}{p(\tau)}$ that τ is a 'normal' value for two consecutive earthquakes. One can draw random numbers x uniformly distributed between 0 and 1 and checking whether $x < P$ or not: that is, whether the second earthquake is a background earthquake or not. An illustration of this method is given in Figure 6 for southern California (1984-2002). A feature of this method is that the largest shocks are not more likely to be kept as background earthquakes than any other shock, since the thinning procedure is only based on inter-event times and is not conditioned on the magnitudes.

As a final remark on this method, we point out the fact that the estimate μ can be obtained with a very simple computation, but it is not the Maximum Likelihood Estimate given the gamma model of Equation 15. Indeed, for N inter-event times τ_1, \dots, τ_N characterizing a catalog of total duration T , the log-likelihood is given by

$$\ell = -\log \text{Likelihood} = N\gamma \left[1 - \log\left(\frac{N}{T}\gamma\right) \right] + N \log \Gamma(\gamma) - \gamma \sum_{i=1}^N \log \tau_i \quad (16)$$

where we only kept the terms that depend on parameter gamma. The minimum of ℓ can be found numerically. A MATLAB code that finds this MLE background rate is:

```
function mu=gamma_law_MLE(t);
% finds the Maximum Likelihood Estimate of the background rate mu for
% the earthquake time series t, based on a gamma distribution of the
% inter-event times.

dt=diff(t); I=find(dt>0); dt=dt(I);
T=sum(dt); N=length(dt); S=sum(log(dt));

dg=10^(-4); gam=dg:dg:1-dg;
ell=N*gam.*(1-log(N)+log(T)-log(gam))+N*log(gamma(gam))-gam*S;
[ell,i]=min(ell); gam=gam(i);
mu=N/T*gam;
```

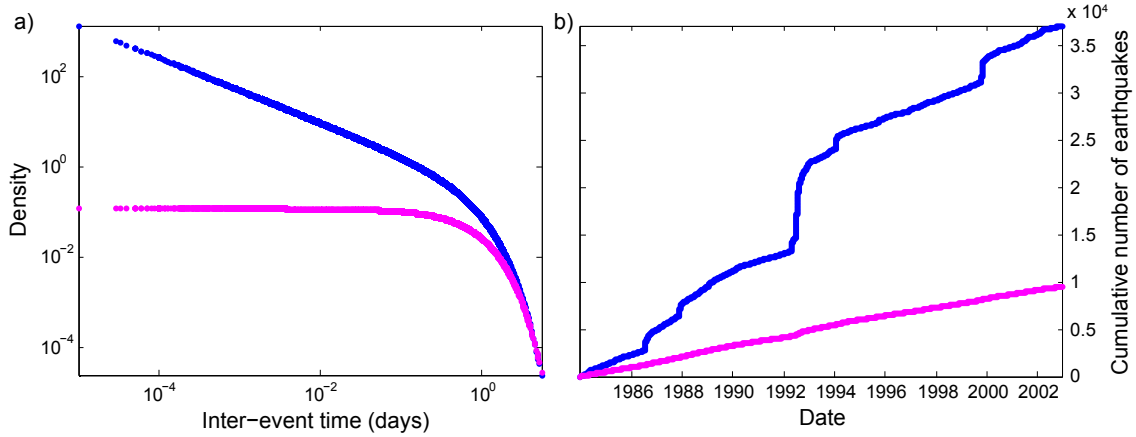


Fig. 6 Declustering based on the background estimates by [Hainzl et al. \(2006\)](#) for $m \geq 2.3$ earthquakes in southern California, 1984-2002. (a) Probability density function $p(\tau)$ of the inter-event times τ , in blue, compared to the pdf $p_0(\tau)$ of a homogeneous Poisson process with rate equal to background rate, times the fraction of background earthquakes γ , in pink. The ratio of the pink to blue curve gives the probability that the inter-event time can be kept when declustering. (b) Cumulative time series in blue, and for a realization of a declustered catalog, in pink.

5.7 Declustering based on the coefficient of variation of inter-event times

Also based on inter-event times τ , the method by [Bottiglieri et al. \(2009\)](#) uses the coefficient of variation of τ . Given a time series of earthquakes, this coefficient is defined as $COV = \frac{\sigma(\tau)}{\bar{\tau}}$ where σ is the [standard deviation](#). The rationale here is that the background rate is not easy to estimate (see however the method by [Hainzl et al. \(2006\)](#), for example), so that we cannot be certain when aftershock sequences start and end. However, the coefficient of variation must be close to 1 for a homogeneous Poisson process. The method thus searches for periods during which COV is less than 1, such periods being characterized by background activity only, intertwined between earthquake clusters that can be reduced to their mainshock for the purpose of declustering. An iterative approach that allows precise identification of the starting time and the duration of a cluster sequence is proposed by [Bottiglieri et al. \(2009\)](#). This is a slight sophistication to this method to provide a reliable way for identifying clusters, and can therefore be used for declustering purposes as well.

5.8 Ratios method of [Frohlich and Davis \(1985\)](#)

This method also exploits the inter-event times, but without examining their distribution. Consider a sequence of earthquakes such that N_a earthquakes occur exactly

in a time T_{N_a} following a given earthquake, and that N_b earthquakes occurred exactly in T_{N_b} before it. Under the null hypothesis of a homogeneous Poisson process, the distribution of the ratio $r = \frac{T_{N_a}}{T_{N_b}}$ is known - it is derived in [Frohlich and Davis \(1985\)](#). Anomalous small r values indicate that T_{N_a} is too short and cannot be explained by a homogeneous Poisson process, hence one or more of the N_a earthquakes are aftershocks. Interestingly, the r -distribution is independent of the seismicity rate that characterizes the Poisson process, which is a particularly appealing feature of this method.

[Frohlich and Davis \(1985\)](#) therefore proposed a simple way to select aftershocks based on this ratio r : for two successive earthquakes i and $j = i + 1$ with an inter-event time τ , compute the ratio $r = \frac{\tau}{T_{N_b}}$ with $N_b = 5$ (or less if there are less than 5 earthquakes in the catalog before i), and check whether r is less than r_c or not, with r_c being the r -value that is obtained less than 1% of times in the case of a homogeneous Poisson process. If $r < r_c$, then one can be 99% confident that j is an aftershock of i . [Frohlich and Davis \(1985\)](#) provides a table giving the value of r_c for various N_b . In the case of $N_b = 5$, the 99% confidence level is given by $r_c = 0.0020$. A MATLAB program that computes r_c for any N_a , N_b and confidence level follows:

```
function r=frohlich_davis(Na,Nb,yc);
% compute the r-value such that there is only yc chance of
% obtaining less than this r-value by chance (confidence level
% equal to 1-yc). Cf Table 1 of Frohlich C. and Davis S. (1985),
% Geophys. Res. Lett., 12, 713-716.

r1=10^(-3); r2=Na/Nb+1;
y1=cdf(r1,Na,Nb); while(y1>yc) r1=r1/2; y1=cdf(r1,Na,Nb); end
y2=cdf(r2,Na,Nb); while(y2<yc) r2=r2*2; y2=cdf(r2,Na,Nb); end

while(y2-y1>yc*10^(-3))
r=(r1+r2)/2;
y=cdf(r,Na,Nb);
if(y<yc) r1=r; y1=y; else r2=r; y2=y; end
end

function y=cdf(r,Na,Nb);
y=1-sum(exp(gammaln(Nb+(0:Na-1))-gammaln(Nb)-gammaln(1:Na)...
+(0:Na-1)*log(r)-((0:Na-1)+Nb)*log(1+r))));
```

As emphasized in [Frohlich and Davis \(1985\)](#), this algorithm will result in 1% of the earthquakes being marked as aftershocks, even if the time series is actually a

homogeneous Poisson process.

Originally, this method was developed for studying deep earthquakes, which are known to generally generate relatively few aftershocks; it is therefore well designed for analyzing small datasets. While simple to implement, it must then be further developed to link mainshock - aftershock pairs together when analyzing larger datasets. As with the other methods using inter-event times, its simplicity must be balanced by the fact that it does not exploit important information that can be very discriminative when searching for aftershocks - namely: distance to mainshock, and mainshock magnitude.

5.9 Declustering methods based on correlation metric

Baiesi and Paczuski (2004) proposed a simple space-time metric to correlate earthquakes with each other. This distance between shocks i and $j > i$ is defined as

$$n_{ij} = (t_j - t_i) r_{ij}^{d_f} 10^{-b m_i} \quad (17)$$

where d_f is the [fractal](#) dimension characterizing the distribution of epi- or hypocenters, and b the parameter of the [Gutenberg-Richter](#) law. This metric decreases as the two earthquakes get closer in time and space, and as the first shock i is bigger. This is equivalent to the ETAS metric, defined as the inverse of the triggering kernel λ_{ij} , but here assuming a pure Omori's law (p -value set to 1) with no cut-off (c -value set to 0), and a distance dependence that does not account for rupture length.

For any earthquake j , the closest distance n_{ij} is sought by examining all preceding earthquakes $i < j$. We denote by i^* the index i giving this minimum distance, and by n_j^* the minimum distance. Linking j to i^* is equivalent, for *Baiesi and Paczuski (2004)*, as linking an aftershock to a mainshock. However, the distribution of n_j^* is very wide, with very small values indicating very strong links, and larger values that are a signature of a weak link. Severing the weak links above a threshold value $n_j^* > n_c$ separate the earthquakes into distinct clusters. No clear procedure is given as how to choose n_c . Declustering is then simply accomplished by reducing every cluster to its largest shock.

Elaborating on this, *Zaliapin et al. (2008)* further defined the rescaled distance and time as

$$T_j = (t_j - t_{i^*}) 10^{-b m_{i^*}/2} \quad (18)$$

$$R_j = r_{i^*j}^{d_f} 10^{-b m_{i^*}/2} \quad (19)$$

A plot of R vs. T for all earthquakes j then makes it possible to identify two distinct populations. These two populations correspond to the space-homogeneous, time-stationary Poissonian background seismicity and the clusters characterized by much smaller time and space inter-event distances. Namely, [Zaliapin et al. \(2008\)](#) analyzed such plots and found clearly separated populations for a global earthquake catalog (Figure 7), and for a simulated catalog using the ETAS model. This development can be seen as a way to avoid the use of the threshold n_c of [Baiesi and Paczuski \(2004\)](#), which is not strongly constrained.

We provide with the [online supplement to this article](#) the contact (email and website) of the author of this method.

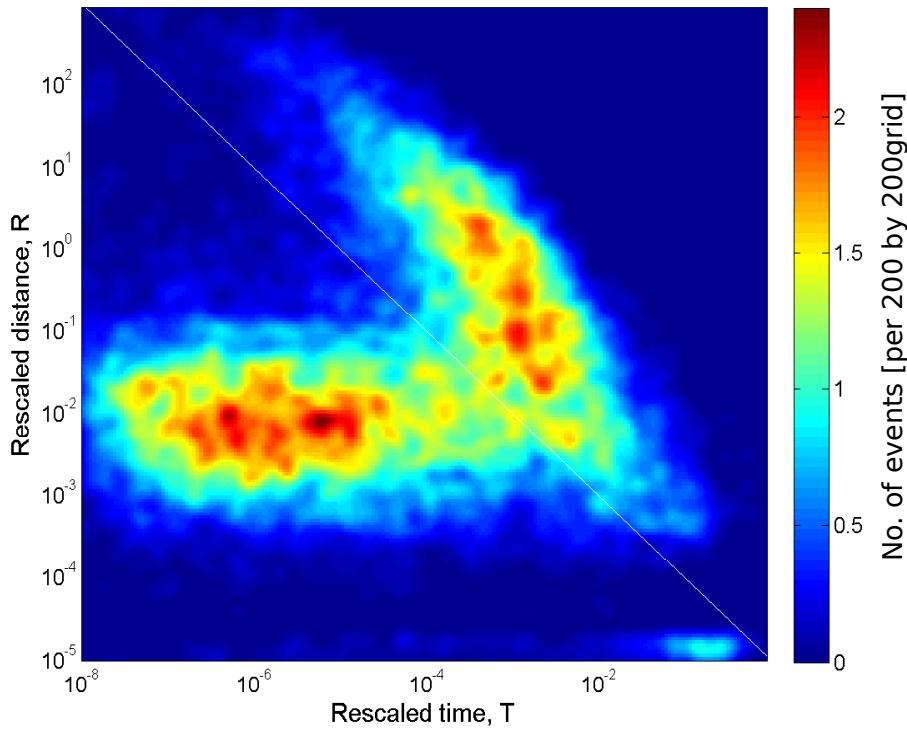


Fig. 7 Bimodal distribution of for ANSS seismicity catalog between 1981 and 2009 for the CSEP testing region (Courtesy of I. Zaliapin)

6 Final Remarks

This article provides an extended overview and practical and theoretical know-how on seismicity declustering. Focusing on the most popular seismicity declustering algorithms, we are aware that the list is not complete. We intend to add to this list in the future, according to new developments and demands.

In this article, we discussed assets and limitations of the presented declustering codes. However, because of the non-unique nature of seismicity declustering, we did not make an absolute judgment on the quality of seismicity declustering. Even though great progress has been made in the last decade, there are still many open questions, i.e., starting with the physical triggering of earthquakes (aftershocks), effects of uncertainties in the catalog on the results of declustering, or the effect of censored data (selection in time, space and magnitude range) on the outcome. In summary, care should be taken when interpreting results of declustering or results that depend on a declustered catalog, because these results cannot reflect the exact nature of foreshocks, mainshocks and aftershocks; indeed the exact nature of these events may not exist at all!

7 Software

The [online supplement to this article](#) provides various codes, or if not available, contact information for the authors of the codes.

Acknowledgements This article benefited very much from thoughtful and constructive two anonymous reviewers and J. Zechar. Many figures in this work are made using GMT ([Wessel and Smith 1991](#)).

References

- Aki, K. (1956), Some Problems in Statistical Seismology, *Zisin*, 8, 205–228. [5](#)
- Baiesi, M., and M. Paczuski (2004), Scale-free networks of earthquakes and aftershocks, *Phys. Rev. E*, 69(066106). [6](#), [22](#), [23](#)
- Bottiglieri, M., E. Lippiello, C. Godano, and L. de Arcangelis (2009), Identification and spatiotemporal organization of aftershocks, *Journal of Geophysical Research*, 114(B03303). [7](#), [20](#)
- Davis, S. D., and C. Frohlich (1991), Single-link cluster analysis, synthetic earthquake catalogues, and aftershock identification, *Geophys. J. Int.*, 104, 289–306. [17](#)
- Frohlich, C., and S. Davis (1985), Identification of aftershocks of deep earthquakes by a new ratios method, *Geophys. Res. Lett.*, 12, 713–716. [20](#), [21](#)
- Frohlich, C., and S. D. Davis (1990), Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogues, *Geophys. J. Int.*, 100, 19–32. [17](#)
- Gardner, J. K., and L. Knopoff (1974), Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?, *Bull. Seis. Soc. Am.*, 64(5), 1363–1367. [3](#), [4](#), [5](#), [7](#), [8](#), [9](#), [10](#), [11](#), [18](#)

- Hainzl, S., F. Scherbaum, and C. Beauval (2006), Estimating Background Activity Based on Intervent-Time Distribution, *Bull. Seismol. Soc. Am.*, *96*(1), 313–320, doi:10.1785/0120050053. [6](#), [18](#), [19](#), [20](#)
- Kagan, Y., and D. Jackson (1991), Long-term earthquake clustering, *Geophys. J. Intern.*, *104*, 117–133. [13](#)
- Kanamori, H., and D. L. Anderson (1975), Theoretical basis of some empirical relations in seismology, *Bull. Seism. Soc. Am.*, *65*(5), 1073–1095. [12](#)
- Knopoff, L. (1964), The Statistics of Earthquakes in Southern California, *Bull. Seism. Soc. Am.*, *54*(6), 1871–1873. [5](#)
- Knopoff, L., and J. Gardner (1972), Higher Seismic Activity During Local Night on the Raw Worldwide Earthquake Catalogue, *Geophys. J. R. astr. Soc.*, *28*, 311–313. [9](#), [10](#), [11](#)
- Marsan, D., and O. Lengline (2008), Extending Earthquakes’ Reach Through Cascading, *Science*, *319*(5866), 1076–1079, doi:10.1126/science.1148783. [6](#), [8](#), [15](#), [18](#)
- Marsan, D., and O. Lengline (2010), A new estimation of the decay of aftershock density with distance to the mainshock, *Journal of Geophysical Research*, in press. [17](#)
- Molchan, G., and O. Dmitrieva (1992), Aftershock identification: methods and new approaches, *Geophys. J. Int.*, *109*, 501–516. [6](#), [9](#), [11](#)
- Reasenber, P. (1985), Second-order moment of central California seismicity, 1969–82, *J. Geophys. Res.*, *90*, 5479–5495. [3](#), [4](#), [5](#), [7](#), [8](#), [10](#), [11](#), [12](#), [18](#)
- Savage, W. U. (1972), Microearthquake Clustering near Fairview Peak, Nevada, and in the Nevada Seismic Zone, *J. Geophys. Res.*, *77*(35), 7049–7056. [10](#)
- Schorlemmer, D., and M. Gerstenberger (2007), RELM testing center, *Seismol. Res. Lett.*, *78*(1), 30. [7](#), [11](#)
- Uhrhammer, R. (1986), Characteristics of Northern and Central California Seismicity, *Earthquake Notes*, *57*(1), 21. [9](#)
- Wessel, P., and W. H. F. Smith (1991), Free software helps map and display data, *Eos Trans. AGU*, *72*, doi:10.1029/90EO00319. [24](#)
- Woessner, J., A. Christophersen, J. Zechar, and D. Monelli (2010), Building self-consistent, short-term earthquake probability (step) models: improved strategies and calibration procedures, *Annals of Geophysics*, *53*, doi:10.4401/ag-4812. [8](#)
- Zaliapin, I., A. Gabrielov, V. Keilis-Borok, and H. Wong (2008), Clustering Analysis of Seismicity and Aftershock Identification, *Phys. Rev. Lett.*, *101*(1), 1–4. [6](#), [22](#), [23](#)
- Zhuang, J. (2006), Multi-dimensional second-order residual analysis of space-time point processes and its applications in modelling earthquake data, *J. Royal Stat. Soc.*, *68*(4), 635–653. [15](#)
- Zhuang, J., Y. Ogata, and D. Vere-Jones (2002), Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.*, *97*, 369–380. [6](#), [8](#), [13](#), [14](#), [15](#)
- Zhuang, J., Y. Ogata, and D. Vere-Jones (2004), Analyzing earthquake clustering features by using stochastic reconstruction, *J. Geophys. Res.*, *109*(B05301), doi:10.1029/2003JB002879. [15](#)