THE UNIVERSITY *of* EDINBURGH

Complexity, aftershock sequences, and uncertainty in earthquake statistics

Sarah Touati

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2011

Acknowledgements

There are a number of people whose help and support I would like to acknowledge.

First and foremost, it has been a privilege to be supervised by Ian Main and Mark Naylor. From the very beginning they have given me the freedom and the encouragement to make my own discoveries and to publish them, and enough guidance and support to feel assured that I was heading the right way. I have learnt so much from their combined expertise, enthusiasm, and clarity of thought in this subject area.

Thanks are due to Mike Christie whose input on numerical modelling and other advice was much appreciated. I was also helped by the constructive criticism and stimulating discussion in the formal feedback I've received over the course of the PhD, such as my first year Confirmation panel and the reviews of my papers.

I have been lucky to have a group of very bright colleagues besides my supervisors here at Edinburgh working on earthquake problems: Andy Bell, Abhey Bhansal, Caroline Graham, John Greenhough, Yusuke Kawada, and Sabine Lennartz. I have learnt much from our discussion meetings and have greatly appreciated their feedback on conference presentations and manuscripts.

Funding from the Engineering and Physical Sciences Research Council has made this PhD possible. My 6-month break in studies to work for the EU 'NERIES' and 'TRIGS' projects in 2009 allowed me to gain much-valued further experience and maturity. The seismology group at ETH Zurich hosted me for a short visit and invited me to present my work as a seminar, which was a rewarding and useful experience for me also.

On a personal note I was encouraged to do this PhD and given moral support along the way by family members and close friends to whom I am also very grateful.

Declaration

I declare that this thesis has been composed by myself and, except where otherwise acknowledged, is entirely my own work. It has not been submitted, either in whole or in part, in any previous application for a degree.

Sarah Touati
July 2011

Abstract

Earthquake statistics is a growing field of research with direct application to probabilistic seismic hazard evaluation. The earthquake process is a complex spatio-temporal phenomenon, and has been thought to be an example of the self-organised criticality (SOC) paradigm, in which events occur as cascades on a wide range of sizes, each determined by fine details of the rupture process. As a consequence, deterministic prediction of specific event sizes, locations, and times may well continue to remain elusive. However, probabilistic forecasting, based on statistical patterns of occurrence, is a much more realistic goal at present, and is being actively explored and tested in global initiatives.

This thesis focuses on the temporal statistics of earthquake populations, exploring the uncertainties in various commonly-used procedures for characterising seismicity and explaining the origins of these uncertainties. Unlike many other SOC systems, earthquakes cluster in time and space through aftershock triggering. A key point in the thesis is to show that the earthquake inter-event time distribution is fundamentally bimodal: it is a superposition of a gamma component from correlated (co-triggered) events and an exponential component from independent events. Volcano-tectonic earthquakes at Italian and Hawaiian volcanoes exhibit a similar bimodality, which in this case, may arise as the sum of contributions from accelerating and decelerating rates of events preceding and succeeding volcanic activity. Many authors, motivated by universality in the scaling laws of critical point systems, have sought to demonstrate a universal data collapse in the form of a gamma distribution, but I show how this gamma form is instead an emergent property of the crossover between the two components. The relative size of these two components depends on how the data is selected, so there is no universal form.

The mean earthquake rate—or, equivalently, inter-event time—for a given region takes time to converge to an accurate value, and it is important to characterise this sampling uncertainty. As a result of temporal clustering and non-independence of events, the convergence is found to be much slower than the Gaussian rate of the central limit theorem. The rate of this convergence varies systematically with the spatial extent of the region under consideration: the larger the region, the closer to Gaussian convergence. This can be understood in terms of the increasing independence of the inter-event times with increasing region size as aftershock sequences overlap in time to a greater extent. On the other hand, within this high-overlap regime, a maximum likelihood inversion of parameters for an epidemic-type statistical model suffers from lower accuracy and a systematic bias; specifically, the background rate is overestimated. This is because the effect of temporal overlapping is to mask the correlations and make the time series look more like a Poisson process of independent events. This is an important result with practical relevance to studies using inversions, for example, to infer temporal variations in background rate for time-dependent hazard estimation.

Contents

Chapter 1

Introduction

1.1 Self-organised criticality and crackling noise

Crackling noise is a widespread phenomenon associated with a wide variety of complex systems. It occurs when a system releases energy in discrete events on a wide range of sizes. Unlike 'snapping' (for example, breaking a piece of chalk), which happens in one single large event, and 'popping' (for example, popcorn), which consists of small random events, things 'crackle' when energy is released in discrete events on a wide range of sizes: the distribution of sizes is an inverse power law, and in the theoretical limit of an infinite system, there is no characteristic or finite mean size. The system is thus scale-free and looks the same at all length scales. Crackling noise is evident in systems as diverse as avalanches, superconductor and superfluid dynamics, stock market fluctuations, and earthquakes (Sethna et al., 2001).

We may think of the events as cascades of individual perturbations which occur at the level of the elements composing the system; for example, the tumbling of a grain of sand or rice, the slipping of a tectonic block, or the flipping of a magnetic dipole. These perturbations trigger others due to some kind of coupling between nearby elements. Crackling occurs when this coupling between elements is at a critical level that is just strong enough to allow the possibility of an event size spanning the space of the system, and this is the case when each perturbation triggers one further perturbation on average (Sethna et al., 2001). This special level of coupling makes the system scale-invariant; renormalisation by coarse-graining causes the fluctuations neither to be smoothed out nor to pervade but rather to remain similar on all scales. This self-similarity gives rise to the power-law size distribution (Sethna et al., 2001; Wilson, 1979). A power law distribution of sizes is special because renormalising the size scale by some factor results in a distribution that is simply proportional to the original. Zooming in or out essentially leaves you with just the same distribution of sizes.

This kind of behaviour is also known as criticality, in language derived from the physics of phase transitions: there is a critical point in a fluid (of temperature and pressure) beyond which the distinction between a liquid and a gas phase ceases to exist, and fluctuations in density (bubbles and drops) of all sizes are observed.

Barkhausen noise, for example, is a form of crackling noise observed in ferromagnets at a critical point. It can be reproduced by the simple two-dimensional Ising model, in which a ferromagnet is represented by a lattice of individual magnetic dipoles, each existing in one of two states—up or down. These dipoles tend to align with their neighbours, but can also

flip independently and randomly due to thermal fluctuations. The disorder caused by thermal fluctuations thus moderates the effective level of coupling between the elements. At the extreme of low temperature, the dipoles are strongly coupled, 'snapping' into alignment. At very high temperatures, the disorder is so large that the dipoles behave independently, 'popping' from one state to the other randomly. If the temperature T is tuned to a critical intermediate value T_c—the Curie point—where the coupling becomes just strong enough for a transition to propagate through the whole system, one observes avalanches of magnetic transitions of all sizes up to the system size (Wilson, 1979). It is rather remarkable that simple models such as the Ising model can produce such rich complex output.

Around the Curie point, power laws describe the response of various properties of the system to changes in temperature. Distance from the critical point may be expressed in terms of a parameter $t = \frac{T-T_c}{T_c}$. The magnetisation of the system increases away from the critical point as $|t|^{\beta}$, while the magnetic susceptibility and the correlation length diverge towards the critical point as $|t|^{-\gamma}$ and $|t|^{-\nu}$ respectively. β, γ and ν are known as critical exponents. They can be measured for real systems, and calculated by various methods for models. Remarkably, different physical systems may have the same exponents when equivalent properties are compared. For example, the three-dimensional Ising model has the same exponents as a critical fluid (Wilson, 1979). It turns out that the exponents depend only on (1) the dimensionality of space and (2) the number of degrees of freedom in the perturbations (e.g. magnetisation or density). Systems that have these key properties in common are said to be in the same universality class. Other changes, such as the structure of the lattice in the Ising model, make no difference to the exponents.

In some systems the critical behaviour is only observed when particular conditions are met, such as a critical temperature in the Barkhausen effect. In most complex crackling systems in nature there is no external fine-tuning; the system apparently organises itself so it sits near the critical point (Bak et al., 1987). This mechanism is known as self-organised criticality (SOC). A classic model for SOC is the sandpile, to which grains of sand are added and, after the pile has built up sufficiently, avalanches of all sizes are observed (Bak et al., 1988; Paczuski and Boettcher, 1996). The heterogeneity in the sandpile slope moderates the potential for a tumbling grain to cause further tumbles, in much the same way that thermal fluctuations moderate the propagation of magnetic transitions in the Ising model. This heterogeneity is then self-maintaining as events of all sizes occur, never smoothing the surface out but forever regenerating the complexity, keeping the coupling between elements at a critical level.

The conditions for SOC seem to be that a system is dissipative and has many metastable states (Bak et al., 1988). Slow driving is also an important ingredient, as the steady input of energy can then be released in discrete quantities (if the driving speed becomes comparable to the event duration, turbulence is observed instead of SOC (Bak et al., 1988)). The crackling associated with SOC—catastrophic large events being rare and small events being very common—may be described as punctuated equilibrium: large intermittent bursts of activity punctuating longer periods of relatively little activity. The complex, contingent nature of various systems and processes, such as climate, biological evolution, and the economy, suggests that SOC processes may completely pervade the natural world. "Things happen by revolutions, not gradually, precisely because dynamical systems are poised at the critical state." (Bak, 1996)

Extreme events appear as freak accidents: grains of sand on the slope of a sandpile happen to be poised in just the right way to tumble like dominoes at the addition of a single grain,

for example. Per Bak in "How Nature Works" (Bak, 1996) argues that seeing these events in the context of self-organised critical dynamics makes better sense of them than the backward-looking narrative accounts given by the "soft" sciences, which, although accurate, cannot easily provide accurate forward predictions nor real insight into the process; the dynamics of the whole system and its evolutionary history are the real underlying cause of extreme events, whatever the specific local conditions and mechanisms supporting an event might have been. Analytical solution of anything more complex than the two-body problem is "exorbitantly difficult", even if the important dynamical details of a system were able to be fully known. Complexity science may have a contribution to make in predicting general statistical patterns rather than specific localised details. The sandpile model is very simple but the mathematics is "prohibitively difficult". It is only possible to calculate the exponents of the scaling relations involved for even more simplified models.

On a similar note, Stephen Wolfram in "A New Kind of Science" (Wolfram, 2002) asserts that nature probably works mostly by following rules (even if traditional physics often models nature as satisfying constraints), which, combined with the observation of complexity in the output from even simple rules (demonstrated principally through the extensive exploration of one-dimensional binary cellular automata in his book), explains why complexity and randomness is so prevalent in nature. He conjectures that there is probably no shortcut to simplify the complex evolving behaviour of systems like these in order to simulate it and predict future behaviour: the amount of computation required to predict the future is equivalent to the amount of 'computation' performed by the system itself. While unproven, this principle certainly concurs with the unpredictability of complex critical dynamics and with the need for statistical forecasts rather than specific predictions.

1.2 Earthquakes as an SOC system

Earthquakes—sudden releases of built-up tectonic stress in the Earth's crust—have long been known to follow the empirical Gutenberg–Richter frequency-magnitude relation:

$$\log N = a - bm \tag{1.1}$$

where N is the number of events of magnitude greater than or equal to m, and a and b are constants. The moment magnitude is usually used; the magnitude was previously calculated based on the amplitude of seismic waves, but, among other issues, such magnitude scales saturate at the large end of the scale (Kagan, 1991). The scalar seismic moment is defined as $M = \mu A \overline{u}$, where μ is the shear modulus, A is the rupture area, and $\overline{u}$ is the average displacement throughout the rupture. It has units of energy and is proportional to the energy radiated. It can be converted into a (dimensionless) moment magnitude using:

$$m_w = \frac{\log M}{1.5} - 10.73 \tag{1.2}$$

The seismic moment distribution $p(M)$ is given by:

$$p(M) = \beta M_c^{\beta} M^{-1-\beta} \tag{1.3}$$

where M_c is a lower moment cutoff, and β is a constant related to b in equation 1.1.

Adherence to this self-similar power law over a wide range of event sizes means that the behaviour is independent of the microscopic or macroscopic dynamics (Sethna et al., 2001). This, along with evidence such as triggering at a distance (Turcotte, 1991) and the very small stress increases needed to induce seismicity, indicate that the Earth's crust is near failure everywhere (Turcotte, 1999), self-organising into a critical or near-critical state.

The question of whether earthquakes truly consitute a critical process has important implications: in this case a time-independent seismic hazard evaluation would be the only viable approach to prediction (Turcotte, 1991). This is consistent with the general absence of reliable earthquake precursors (Wyss and Booth, 1997). In a truly critical system event size is extremely sensitive to fine details in the dynamics (Main, 1999), and is not related to the history of the process in a direct, reliably predictable way. The effect of past seismicity on the magnitude of an event is unclear—some studies report rather weak magnitude correlations, e.g. Lippiello et al. (2008), but it is usually assumed for modelling purposes that magnitudes are independent of history. Earthquake magnitude may even be essentially unpredictable, i.e. an emergent property of the detailed evolution of the rupture process rather than a pre-determined entity, for example as expected by the appealing (but so far unproven) notion of a 'preparation zone' within which anomalous precursory behaviour might be expected.

Describing seismicity as a critical or slightly sub-critical branching process would lead to an exponent $\beta = \frac{1}{2}$ for the power-law segment of the size distribution; this value is indeed found, regardless of the area or depth, when whole main shock–aftershock sequences of earthquakes are considered as the 'events' (Kagan, 1991). The spatial occurrence of earthquakes exhibits the same kind of scale-invariant complexity seen in the event size distribution: events are clustered spatially around faults, which are hierarchical fractal structures (Helmstetter and Sornette, 2002a). This fractal clustering also strongly relates the earthquake phenomenon to criticality and percolation models (Shcherbakov et al., 2006). On the other hand, although the Gutenberg–Richter law is present regardless of tectonic setting, the discovery that there are regional variations in the statistical properties of seismicity could be taken to indicate that the crust is slightly below the critical point (Al-Kindy and Main, 2003).

An infinite seismic moment is clearly not physically possible in a finite-sized crust. An exponential tail-off to the size distribution occurs in critical systems with a finite size (Bak et al., 1988). A statistical physics approach of maximising entropy also leads to a gamma distribution—a power law with an exponential tail (Al-Kindy and Main, 2003; Kagan, 1991):

$$p(M) = C^{-1} M^{-1-\beta} \exp(\frac{-M}{M_x}) \tag{1.4}$$

where C is a normalising coefficient and M_x is the 'maximum' moment. An exponential tail in the distribution could also indicate slight sub-criticality. The maximum possible earthquake size is not known, either theoretically or practically; there are not enough recorded events to determine the maximum value for the Earth. A single extreme event in 2004 changed the best-fit moment distribution function from a gamma to a power law (Main et al., 2008).

Although the crust may be below the critical point globally, it can still display critical behaviour locally. Jaume and Sykes (1999) and Sammis and Sornette (2002) suggest that fault systems only reach criticality briefly, and when the large events occur that span the fault, these move the system away from criticality until long-range stress correlations build up again. Criticality is the point at which a large event spanning the fault becomes possible (although

not immediately inevitable). In support of this idea is the observation of an extension of the Gutenberg–Richter power-law regime upon the occurrence of large events (Jaume and Sykes, 1999). Some have claimed that large events are often preceded by accelerating seismic moment release; rather than occurring 'out of the blue' as expected in truly critical systems, large events often seem to result from a positive feedback of the event rate on itself, making monitoring and forecasting plausible. Sammis and Sornette (2002) demonstrate that this positive feedback mechanism is predicted by many dynamical models of the process, including crack growth, percolation, damage mechanics, and stress shadow models. Whatever the theoretical merits of the models, definitive proof for accelerating seismicity or moment release in real earthquake data has proven elusive, and the statistical methods used to search for it have been questioned, not least the use of curve fits to cumulative (and hence correlated) data (Hardebeck et al., 2008; Greenhough et al., 2009). In contrast accelerating event rate is a common feature prior to many volcanic eruptions, e.g. Chastin and Main (2003), and is almost ubiquitous prior to rock failure in the laboratory (Ojala et al., 2004). Interestingly Ojala et al. (2004) also show the predictability of the dynamic failure time diminishes systematically as the laboratory strain rates are reduced towards the natural ones in the Earth, which is consistent with the notion that low forcing rates are a key ingredient of self-organised criticality.

Some of the dynamics of earthquake faults, including the Gutenberg–Richter law, can be reproduced using slider-block models (see figure 1.1). These typically consist of two parallel plates, in between which weighted blocks rest on the bottom plate and are connected to the top plate by springs. The blocks are also connected to each other via springs. The top plate is moved slowly and the blocks slip when the spring force reaches a critical level. Nearest-neighbour interactions through the springs allow 'rupture' to propagate across several blocks. Unlike the binary system of ferromagnets, stress is a continuous variable, but crackling noise is also produced. Power-law event size distributions are observed. It can also be implemented as a cellular automaton, with block positions being updated sequentially.

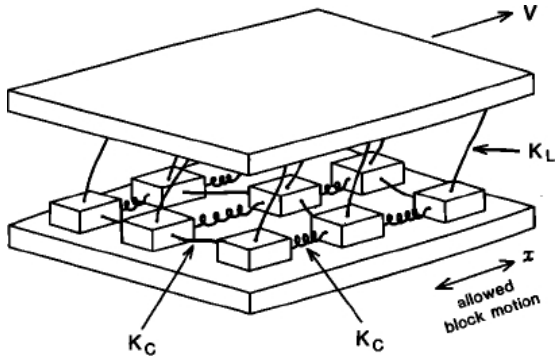


Figure 1.1: Graphic representation of the Burridge–Knopoff slider-block model, from Main (1996).

Two key parameters are the ratio of static to dynamic friction used, and the spring stiffness ratio (in the block-to-block springs versus the block-to-plate springs) (Turcotte, 1999). Different values of parameters such as driving velocity and heterogeneity can produce sub- or super-critical behaviour (Main, 1996).

A benchmark model in the slider-block category is Olami Feder and Christensen's (OFC) non-conservative model (Olami et al., 1992). In this model, blocks return to equilibrium on

slipping, but not all of the force on the slipping block is transferred to its neighbours. This setup is completely deterministic. Introducing noise, by allowing slipping blocks to settle with a small (random) force on them rather than going to a position of zero force, was not found to change the exponent, whose value depends on the spring constant ratio (which also determines the level of conservation of local stress) (Olami et al., 1992). For isotropic spring constants representative of the Earth's crust, which results in 80% conservation, exponents in the range observed for earthquakes are obtained (Olami et al., 1992).

Despite their ability to produce power-law distributed event sizes, slider-block models are a crude approximation to a fault. Earthquakes in reality occur on complex fault networks consisting of tectonic 'blocks' of a power-law size distribution. This fractal structure is the result of dynamic grinding processes over the course of time and is continually evolving (Turcotte, 1999). It thus appears remarkable that the overly-simplified slider-block models can reproduce the Gutenberg–Richter law. However Kagan (1994) points out that there are many different types of models that produce power laws, and that this feature alone does not make them a good reflection of earthquakes. In general power-law exponents are smaller in slider-block models than in earthquakes (Turcotte, 1999).

1.3 Earthquake statistics and aftershocks

Empirical statistical observations constitute a rich phenomenology by which the earthquake process is characterised. Understanding of the statistical properties of earthquakes has important applications in seismic hazard evaluation (Vere-Jones et al., 2005), and also provides important benchmarks against which attempts to model the dynamics of the process can be validated. Deterministic physical modelling and stochastic modelling without physics are two end-members of a spectrum of approaches; both aspects are surely important. A stochastic modelling approach "accepts that some aspects of the physical process are out of range, at least for pratical purposes, and must be replaced in the model by some unknowable and hence random process" (Vere-Jones, 2010). In light of the possibility of earthquakes being a self-organised critical process, it would seem likely that this approach will be vital.

The Gutenberg–Richter law, equation (1.1), is an example of a well-established empirical statistical law, describing the relative average frequencies at which earthquakes of different magnitudes occur. The distribution of earthquakes in time, however, is far from either periodic or Poissonian end-member models, in which occurrence of events of a certain size is respectively completely predictable or completely unpredictable. In most self-organised critical systems, events occur in time as a Poisson process; in earthquakes, however, clustering of events in time occurs due to the ability of events to trigger further events through a variety of physical mechanisms.

The movement along a fault causes a static Coloumb stress change which drops off rapidly with distance from the fault; in the short term, this causes aftershocks through the time-dependent visco-elastic relaxation of the Earth's crust, perhaps through stress corrosion, frictional mechanisms, pore pressure changes due to fluid movement, and other contributing factors (Lindman et al., 2006). In the longer term, the Coulomb stress change is thought to alter the time until the next event, potentially triggering subsequent events through bringing nearby parts of the crust closer to failure. Studies have pointed to a weaker, long-term clustering of large events or whole aftershock sequences, e.g. Kagan (1994), but while we have been able to

observe many short-term aftershock sequences and establish the statistics of this phenomenon, the timescales involved in the recurrence of larger events—those that are of human interest—are too long to enable us to discern from our limited records whether Coulomb stress triggering plays a part in altering that timescale. There is currently no conclusive evidence for or against it.

Dynamic perturbations of the crust also occur in the wake of an event through radiated seismic waves, and the effect of this can reach further in space. These waves can trigger events on faults that were near failure as they pass by, which is fairly easy to detect. Whether they can weaken faults for the longer term and reduce the time until the next event, it is not possible to discern at this point.

The magnitude of an event understandably influences the number of aftershocks triggered by it. The productivity law indicates that the number of aftershocks produced by an event depends on the magnitude m of that event as $e^{\alpha m}$—the relation is exponential in magnitude, or power-law in the seismic moment. The parameter α is typically small for swarm-type activity, and large for clear primary aftershock sequences (Ogata, 1992).

The rate of aftershocks—that is, short-term and short-range triggered events—following a main shock decays in time according to the (empirical) modified Omori law, which is a power law in time:

$$n(t) = \frac{K}{(c+t)^p} \tag{1.5}$$

where t is the time since the main shock. The exponent p is typically around 1. The overall productivity is represented by the constant parameter K, which in reality depends on the magnitude of the parent event as noted above. The parameter c has a small value and may be either a mathematical formalism to avoid a singularity at $t = 0$, a reflection of temporary catalogue incompleteness immediately after the main shock (Kagan and Houston, 2005), or a genuine effect of the physical mechanisms involved (Lindman et al., 2006). The history-dependence expressed by the Omori law gives earthquake rates a degree of time-dependent predictability not present in the classic SOC models such as sandpiles, and coupled with the Gutenberg–Richter law, it can lead to constraints on the expected magnitude range in a given time period.

Aftershock occurrence also decreases with distance from the main shock as an inverse power law (Felzer and Brodsky, 2006), with an exponential tail (Huc and Main, 2003) or perhaps with no cutoff (Baiesi and Paczuski, 2005). This can be thought of as a kind of spatial Omori law. Long-distance triggering is detectable statistically up to around 150km worldwide (Huc and Main, 2003), with aftershocks being reported at up to 1250km in some extreme cases (Helmstetter and Sornette, 2002a). The spatial extent of aftershock occurrence is generally considered to depend on the magnitude of the main shock, which follows straightforwardly from the power-law decay with distance combined with the productivity law; if a larger number of events are produced, the largest expected distance will be greater. In fact the length of the aftershock zone may vary with magnitude as $e^{\alpha m}$ (Ogata, 1998) which resembles the productivity law. Kagan (2002) equivalently formulates the relation in terms of a power law between seismic moment and aftershock zone length.

Correlated events thus tend to be close both in time and in space. Davidsen and Paczuski (2005) demonstrate that the length of a waiting time between events tells us nothing about the

likely distance to the next event. This result means that the temporal and spatial distributions of correlated (triggered) events can be modelled as independent.

Foreshocks are a less-common but similar phenomenon to aftershocks. The magnitudes of foreshocks and aftershocks are smaller than the main shock, by definition. The main shock is always the largest event in a sequence, and as such it can only be decided upon in retrospect. There is no evidence that foreshocks, main shocks and aftershocks are categorically different types of dynamical events in themselves. One therefore has to assume that all events are capable of triggering further events following the Omori law, regardless of whether they themselves were triggered by other events or occurred spontaneously. This could perhaps explain the observation of long-term correlations in the earthquake time series as well as short-term ones. This has been detected through a power-law decaying autocorrelation function (Altmann and Kantz, 2005); using detrended fluctuation analysis (DFA), Lennartz et al. (2008) also show that earthquake catalogues are comparable with long-term correlated synthetic data.

Critical or self-organised critical systems—including simple earthquake models—exhibit $1/f$ noise which is a power-law spectral density in the time series (Bak et al., 1987, 1988; Feder and Feder, 1991; Turcotte, 1999). This usually does not refer to temporal clustering of events, however, but to the distribution of lifetimes of the events themselves, and follows directly from the power-law size distribution of events (Bak et al., 1987). It basically means that the disturbance caused by events—the number of blocks slipping, the number of dipoles switching their orientations, the number of sand grains falling, and so on—varies with time in a self-similar manner giving rise to 'bumps' of all durations (Bak, 1996). Temporal clustering of events and history-dependence is generally not a feature of self-organised criticality, but is not incompatible with it, and some studies have successfully added 'memory' into SOC models to reproduce some temporal statistics of seismicity, e.g. Lippiello et al. (2005).

Temporal correlations (foreshocks and aftershocks) can be introduced into the OFC model through a time-dependent relaxation mechanism (Hainzl et al., 1999). A number of authors in recent years have also reported finding temporal clustering of events in slider-block models without this extra mechanism (Castellaro and Mulargia, 2002; Hergarten and Neugebauer, 2002; Helmstetter et al., 2004; Weatherley, 2006; Hasumi, 2007). Hergarten and Neugebauer (2002) simulated 10^9 events in a lattice of 512 by 512 blocks, and found that the distribution of time intervals between all events was slightly non-exponential at short intervals implying a slight temporal clustering at short inter-event times. Looking closer, they found the biggest events tended to occur in clusters, with the magnitudes also appearing to be correlated. This effect depended on having a non-conservative parameterisation, i.e. a dissipative system, and the temporal clustering showed an increase as conservation level is decreased. Helmstetter et al. (2004) confirmed the importance of this finding and explored its relation to real seismicity, and found that the number of aftershocks produced is smaller than for real earthquakes. Not only this but the number and size of foreshocks was correlated with the main shock size, which is not the case for real seismicity. The magnitudes of large events are thus more predictable in the OFC model than in nature, and foreshocks in the model may be viewed as part of the preparation process towards a large event, or perhaps—along with the aftershocks—simply the spreading out of a large avalanche over several timesteps due to the dissipation, an effect which can produce limited-range power-law increase and decay in event occurrence (Hainzl et al., 2003). Perhaps the weak magnitude correlations reported in real seismicity (section 1.2) are produced by a similar effect, but it would seem to be small. Helmstetter et al. (2004) also

showed that it is not a simple increase in stress that causes aftershocks in the OFC model. Interfering with a running model by increasing the stress on a single block or even a group of adjacent blocks did not produce aftershocks. Instead, they observed that aftershocks result from an interaction between asperities that develop in a complex pattern over time, which seems to link it to the complex self-organisation common to many other SOC models and not so much to real aftershocks.

1.3.1 ETAS model

Point-process modelling is a tool used to represent statistical patterns in time series of discrete events, and is well-suited to modelling earthquake occurrence. Earthquakes can be thought of as a marked point process occurring in space and time: 'marks' can refer to the magnitude of each event, which can be reliably measured, and also the family tree structure, which is generally much harder to determine.

A conditional intensity function $\lambda(t|H_t)$ characterises the behaviour of a point process by giving the probability rate as a function (generally) of time and of the history of the process. Constant λ for example would represent a stationary Poisson process with no dependence on time or history. Many features of earthquake occurrence in time and space can be helpfully modelled using point processes. Examples include detection of seasonality in seismicity related to precipitation levels by the introduction of a cyclic term in the conditional intensity (Ogata, 1999).

The epidemic-type aftershock sequences (ETAS) model is one particularly well-known application of point-process modelling to earthquakes. It incorporates the established empirical relationships described above into a parsimonious parametric representation which can be calibrated to different seismic regions and patterns. It is often used to investigate seismic anomalies as, once calibrated, it offers a null hypothesis representing the 'standard' seismicity of a region.

In epidemic-type models, the number of individuals alive at a given time t is controlled by an immigration rate, a birth rate and a death rate. In the language of earthquakes, "immigration" refers to independent background events occurring, while "birth" corresponds to triggering of further events by past events. The ETAS model is a self-exciting process in which there is no death; instead, the birth rate is kept stable the by the decrease in time of the triggering probability from each past event according to the Omori law (to eventually become negligibly small).

A stationary Poisson seeding rate emulates the constant tectonic loading rate. All past events—constituting the history of the process—contribute to the probability of a future event occurring. The result is independent background events triggering aftershocks, which in turn trigger their own aftershocks, and so on. Triggering is implemented according to the modified Omori law (equation (1.5)), with the overall productivity weighted by the magnitude m of the triggering event according to the productivity law, so K becomes $K\exp(\alpha(m-m_0))$ where m_0 is a lower cutoff magnitude (Ogata, 1988). The parameters c and p are constant with respect to magnitude. All magnitudes are independently taken from the Gutenberg–Richter distribution.

The conditional intensity function can thus be written simply as the sum of the immigration rate μ and the probabilities of 'births' being generated from all past events (Ogata, 1988):

$$\lambda(t|H_t) = \mu + K \sum_{i:t<t_i} \exp(\alpha(m_i - m_0))(t - t_i + c)^{-p} \tag{1.6}$$

where t_i are the times of the past events and m_i are their magnitudes.

In fitting the ETAS model to an earthquake catalogue, the ETAS parameters are commonly inverted from the data using the point-process maximum likelihood method, in which the log likelihood is given by:

$$\log L(T_1, T_2) = \sum_{i:T_1 \leq t_i \leq T_2} \log \lambda(t_i|H_{t_i}) - \int_{T_1}^{T_2} \lambda(t|H_t)dt \tag{1.7}$$

where T_1 and T_2 specify the time period over which the log likelihood is evaluated. As part of my literature search, I carried out a survey of the ETAS parameter values published in a range of studies. Figure 1.2 shows histograms resulting from the collection of this data (Console et al., 2007; Enescu et al., 2006; Faenza et al., 2007; Guo and Ogata, 1997; Hainzl and Ogata, 2005; Harte and Vere-Jones, 2005; Liu and Ma, 2006; Lombardi et al., 2006; Ma and Zhuang, 2001; Ogata, 1992; Ogata et al., 2003). Clearly there is significant variability although typical ranges of values can be determined (it should be noted that because μ varies with geographical area size and magnitude threshold, and so is not an absolute measure of background rate, the values should not be expected to form any particular distribution).

The value of m_0 is usually arbitrarily chosen to be the minimum magnitude for catalogue completeness. Sornette and Werner (2005b) point out that the completeness threshold has no relation to the minimum magnitude capable of triggering other events, which is what m_0 really represents in ETAS. They show that choosing a higher m_0 can cause the parameter α and the branching ratio (average number of daughters per parent) to be reduced. In other words more events appear to be independent. In effect the ETAS parameters obtained with an incorrect m_0 are a renormalised version of the 'true' values.

The ETAS model can also be extended to describe the spatial distribution of events (Ogata, 1998; Ogata and Zhuang, 2006). The conditional intensity becomes a function of x and y coordinates as well as of time t, but its integration over all space recovers the temporal-only ETAS model. As discussed previously, temporal and spatial triggering may be treated as independent and separable, so the triggering term is a product of the temporal term and a spatial term $f_i(x - x_i, y - y_i)$. The form of the spatial dependence most often used is an inverse power law in the distance and an exponential in the magnitude of the parent event, with random direction (Ogata, 1998), although there are various other models. Hainzl et al. (2008) made the improvement of considering the parent event as a line source rather than a point source.

The background spatial organisation is more difficult to model and poses a serious challenge in space-time point-process modelling. Background occurrence is highly non-homogeneous in space due to the complex fault structure. The way the background pattern is often inverted for an earthquake catalogue is by smoothing the pattern of events using a Gaussian kernel on all events (or perhaps just the larger ones (Schoenberg, 2003)). Choosing an appropriate bandwidth for the smoothing is not trivial: too little smoothing results in capturing noise; too much smoothing may not capture fine details of the fault network. Zhuang et al. (2002) set the bandwidth for each event according to how densely populated the area around the event is. However inverting a background pattern in space from a set of recorded events remains difficult.

The event rate in the ETAS model is capable of 'exploding'. The key quantity governing stability in the event rate is the branching ratio n, the average number of daughters per parent, obtained as the expectation of the triggering term $\phi_m(t)$ (the Omori and productivity laws

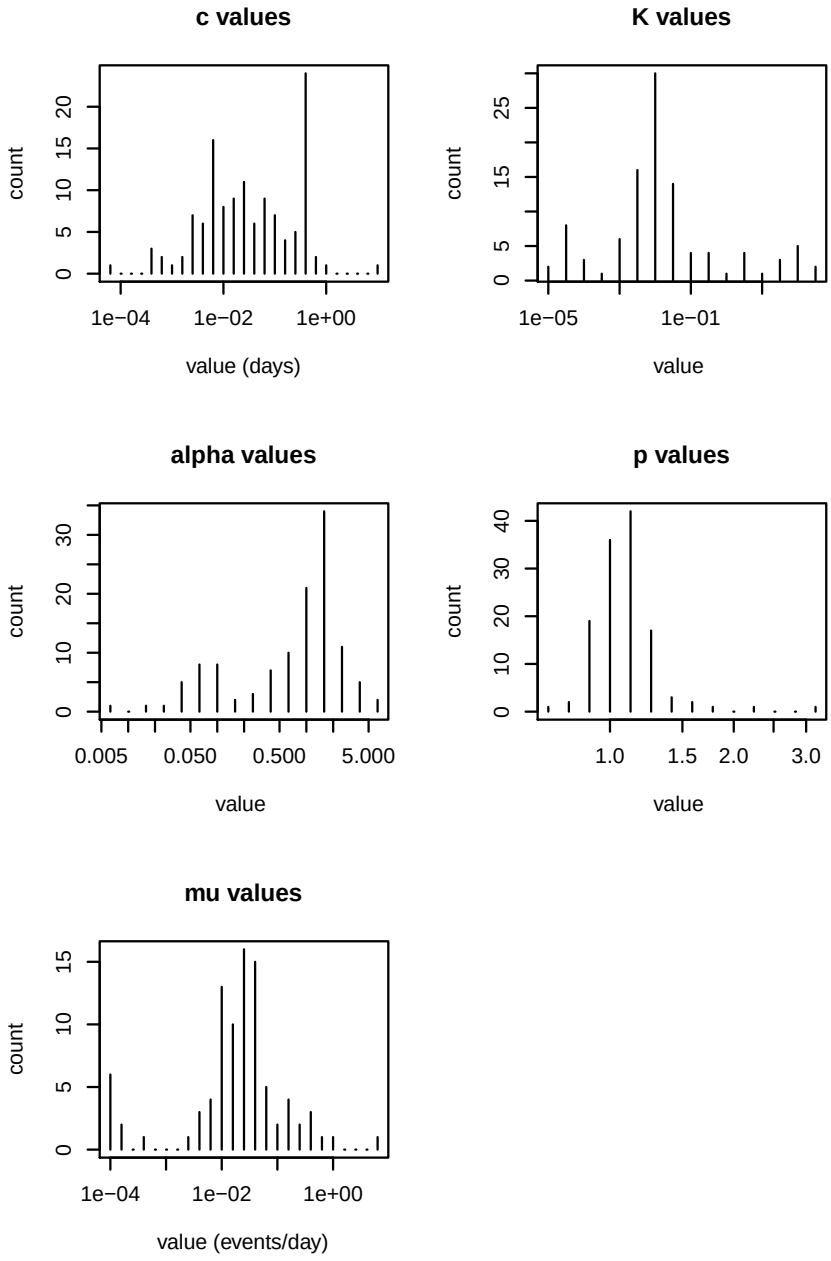


Figure 1.2: Histograms of ETAS parameter values used in literature.

combined) integrated over all time:

$$n \equiv \int_0^\infty dt \int_{m_0}^{m_{max}} p(m)\phi_m(t)dm \tag{1.8}$$

where $p(m)$ is the magnitude distribution, the Gutenberg–Richter law. For $p > 1$ this is given by (Sornette and Werner, 2005a):

$$n = \frac{Kc^{1-p}}{p-1}\frac{\beta}{\beta-\alpha}\frac{1-\exp(-(\beta-\alpha)(m_{max}-m_0))}{1-\exp(-\beta(m_{max}-m_0))} \tag{1.9}$$

where here β is the exponent of the Gutenberg–Richter distribution (b in equation 1.1) expressed with base e, i.e. $p(m) = \beta e^{-\beta m}$ with $\beta = b \ln 10$. For a non-truncated Gutenberg–Richter law with $m_{max} = \infty$, as is usually used in the absence of a value for m_{max}, this expression reduces to:

$$n = \frac{Kc^{1-p}}{p-1}\frac{\beta}{\beta-\alpha} \tag{1.10}$$

n is infinite when $p < 1$ or $\alpha > \beta$, unless there is a finite maximum magnitude (Sornette and Helmstetter, 2002); simulation of the model with these parameters would result in a finite-time singularity in the event rate.

There is a critical point at $n = 1$ (Helmstetter and Sornette, 2002b): when each event triggers one further event on average, the seismicity is just barely able to propagate itself indefinitely. Both the term "critical point" and the meaning in terms of a minimum level of coupling that could produce a limitless cascade are reminiscent of the critical point concept in fluids and ferromagnets. In the latter, the coupling operates between the smallest elements of the system to produce events as cascades of small perturbations; in the case of aftershocks, we seem to have a second-order coupling that produces cascades of the events themselves, perhaps also as a critical process.

It is illuminating to consider the average 'global' aftershock sequence; that is, the average event rate as a function of time following a seed event of random (Gutenberg–Richter distributed) magnitude, taking into account all cascades of further generations. For $n < 1$—the sub-critical regime (Sornette and Helmstetter, 2002)—the average event rate following a seed event can be shown analytically to be similar to the individual Omori power law, but with a smaller (negative) exponent at shorter times due to the further generations of aftershocks (Helmstetter and Sornette, 2003a); a break of slope in the event rate exists at some characteristic time t^* (a function of c, n and p), beyond which the usual Omori exponent applies, as shown in figure 1.3. For $n > 1$, the super-critical regime (Sornette and Helmstetter, 2002), an exponential growth regime follows the same t^* (Helmstetter and Sornette, 2002b), reflecting self-sustaining seismicity, as demonstrated in figure 1.4. At the critical value $n = 1$, presumably there is a single power law in the event rate without a change in slope; indeed the expression for the characteristic time in Helmstetter and Sornette (2003a) becomes infinite at $n = 1$.

Given a Poisson seeding rate, the sub-critical regime will exhibit stationary seismicity in the long term. It has been said to represent a dissipative system with some aseismic release of energy (Helmstetter and Sornette, 2002b). When calibrating the ETAS model to represent the normal seismicity of a region, a branching ratio less than 1 is the only result that makes sense, because in the long-term the tectonic forcing is very stable, and provides a physical constraint on any upward fluctuation in the event rate.

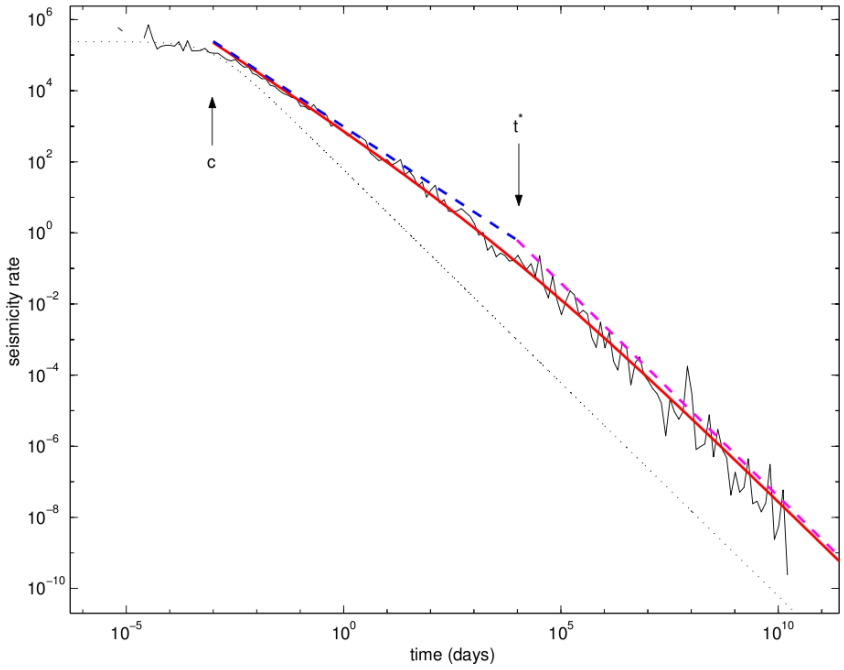


Figure 1.3: Sub-critical branching ($n < 1$) in the ETAS model: event rate as a function of time following a seeding event. A characteristic time t^* separates an initial power law decay from a later steeper power law decay. From Helmstetter and Sornette (2002b)

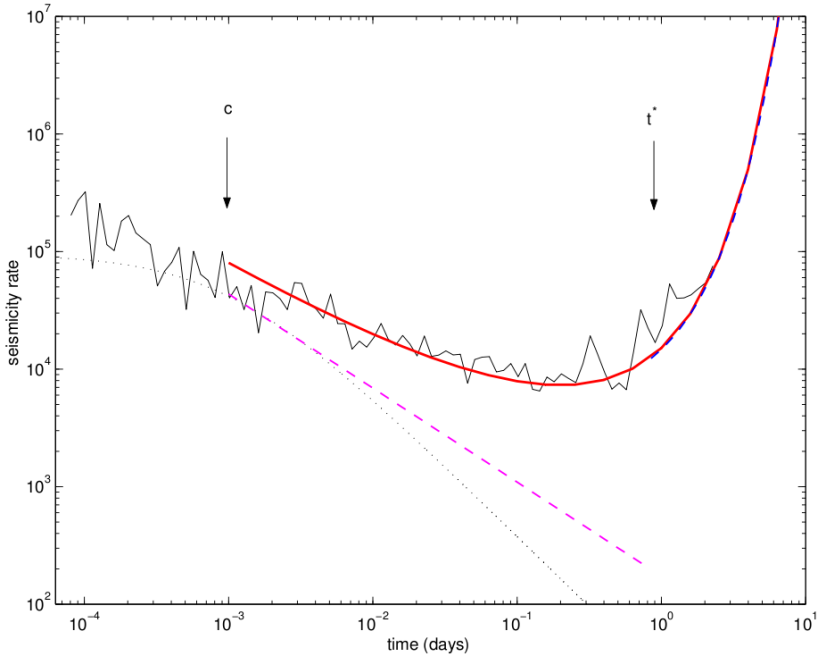


Figure 1.4: Super-critical branching ($n > 1$) in the ETAS model: event rate as a function of time following a seeding event. A characteristic time t^* separates an initial power law decay from a later exponential growth. From Helmstetter and Sornette (2002b)

The super-critical regime represents a situation where energy is flowing into the region (Helmstetter and Sornette, 2002b); simulations in this regime will show a steady increase in the event rate over time. Although divergent seismicity cannot continue indefinitely, in the short term the super-critical and singular regimes may provide a useful and accurate description of phenomena such as foreshock sequences (Helmstetter and Sornette, 2002b) and the acceleration of seismicity prior to a large event (Jaume and Sykes, 1999; Sammis and Sornette, 2002). Other precursory phenomena such as quiescence, swarms and aftershock bursts have also been suggested (Helmstetter and Sornette, 2002b).

However, super-critical branching may not be needed to explain these; Helmstetter and Sornette (2003c) show that accelerating sequences of what turn out to be foreshocks are present in sub-critical ETAS simulations. Just as in real data, when several of these foreshock sequences are stacked together they form a temporal power law with an exponent smaller than the Omori p, and their absolute rate does not appear to depend on the magnitude of the main shock (there is no equivalent of the productivity law for foreshocks).

Certain other empirical observations, besides foreshocks, arise from those basic empirical laws encoded in the ETAS model too. For example, the size of the largest aftershock in any sequence is consistently around 1 unit smaller than the main shock. This is known as Båth's law (Shcherbakov and Turcotte, 2004), and may at first seem to imply history-dependent magnitudes. Helmstetter and Sornette (2003b) show, however, using simulations of the ETAS model, that Båth's law arises spontaneously from a combination of the Omori law, the productivity law, and independent Gutenberg–Richter distributed magnitudes, coupled with the procedure of picking out main shocks retrospectively.

A further example is the spatial diffusion of aftershock activity away from the main shock. This is sometimes observed (Tajima and Kanamori, 1985), but it is very slow and far from normal (Corral, 2007), with the correlation length and mean triggering distance increasing as t^H with $H = 0.06$ (Huc and Main, 2003). Helmstetter and Sornette (2002a) show formally that this can be explained by the idea that aftershocks trigger their own aftershocks as in the ETAS model.

In summary the ETAS model, with the random selection of magnitude from the parent Gutenberg–Richter law coupled with Omori aftershock triggering, can explain empirical observations such as Båth's law, spatial diffusion of aftershock activity, and—occasionally—foreshock sequences, which in real data may be assigned false significance through data selection and/or stacking.

There are other point-process models for earthquakes along similar lines, the most prominent of which is the Branching Aftershock Sequences (BASS) model (Turcotte et al., 2007), which implements magnitude-dependent productivity by utilising Båth's law in the aftershock production rather than the productivity law. These types of models stand as a powerful standard description of earthquake occurrence and a null hypothesis against which to test for further effects.

1.4 Earthquake inter-event times

The distribution of waiting times between seismic events has generated much interest and discussion over the last decade because of claims, drawing on critical-point analysis, that it has a universal form. This terminology is reminiscent of the universality classes of SOC (section

1.1), but the meaning here is somewhat different. At a critical point, a quantity's relationship with various different variables takes on a form that unifies these variables. A data collapse—the collapse of different sets of data onto a single curve—under rescaling by quantities involving those variables, can indicate criticality because it demonstrates this unification of the different variables. Bak et al. (2002) were the first to propose a scaling law to unify earthquake properties in time, space and magnitude; that is, to unify the Omori law, the fractal spatial distribution of epicentres, and the Gutenberg–Richter law into a single relationship. They divided California into square spatial cells of length L, computed the inter-event times for each cell, mixed data from different cells of the same size to increase the count, and produced a histogram of the inter-event times; this was repeated for different values of L and minimum magnitude. They found that if they rescaled the x-axis (inter-event times T) by an expression $S^{-b}L^{d_f}$ (where S is the logarithm of the minimum cutoff magnitude, b is the Gutenberg–Richter b-value (b in equation 1.1), and d_f is the spatial fractal dimension), which is linked to the average event rate, and rescaled the y-axis (probability) by T^{α}, where α is some exponent, the histograms all appeared to collapse onto a single curve (figure 1.5).

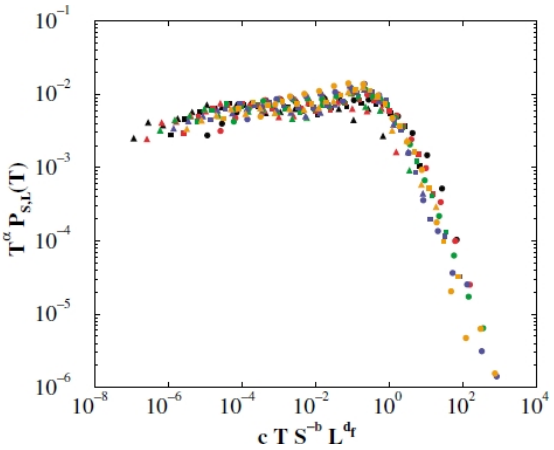


Figure 1.5: Figure 4 from Bak et al. (2002). Each histogram is a subset of California data with different magnitude threshold and different spatial area size, rescaled as shown by the axes and explained in section 1.4 (c is a constant equal to 10^{-4}), and showing only inter-event times > 38s.

The form of the distribution—and thus the function that the curves collapsed onto—appeared to be a power law at short times, followed by a 'kink' and a more rapid exponential decay at longer T. An exponential decay would seem to indicate uncorrelated events, so they interpreted the kink as separating correlated and uncorrelated event pairs: it represents the value of T, for a given L and S, at which event pairs start to be independent.

Then in 2003 Corral published a paper (Corral, 2003) endorsing the importance of Bak et al.'s result, which he said reveals "scaling in the spatiotemporal occurrence of earthquakes". He considerably simplified this universal scaling law by proposing that the form of the distribution depends only on the event rate: he both multiplied the inter-event times, and divided the histogram counts, by the earthquake rate. This rate changes with space and time and so a histogram of data aggregated over different spatial cells, or gathered over a prolonged period, such as in Bak et al. (2002), gives a mixed distribution—the result of mixing different rates. Corral avoids this mixing by selecting stationary periods from catalogues, and finds it is a

universal gamma function, agreeing in the end with Bak et al. on the main result (without touching on the interpretation in terms of correlated and uncorrelated events). In a later paper (Corral, 2004), he finds that stationary data from a wide variety of catalogues and vastly different spatial areas collapses in this way; he also then makes the innovation of rescaling the non-stationary data by the time-varying event rate in order to get the same result yet again.

This form of the inter-event time distribution is consistent with the simple dynamics captured by slider-block models. One study on the cellular automaton slider-block model for earthquakes found that a gamma distribution of inter-event times was obtained by tuning the conservation parameter to around 0.2 (Weatherley, 2006). A study on the fully dynamic slider-block model found that stiff connections between the blocks resulted in a Weibull distribution similar to real earthquakes (Abaimov et al., 2007).

There is some discussion in many of these studies about the value of the exponent for the power-law segment of the distribution, and how it relates to parameters of the seismicity such as the Omori p value. The value of the exponent is variously believed to be around the value of the Omori parameter p (Bak et al., 2002), $\sqrt{p}$ (Lindman et al., 2005) or $2 - 1/p$ (Jonsdottir et al., 2006).

The distribution of spatial distances between sequential events has been a topic of some scrutiny also and is clearly a related phenomenon. Davidsen and Paczuski (2005) find that the distribution of distances between successive earthquakes is a power law with exponent -0.6, irrespective of magnitude. The distribution of distances between random event pairs, by contrast, is a power law with exponent +0.14; a (different) power law in this case exists purely because the events occur in a fractal space. Events adjacent in time therefore tend to be close in position, which is not surprising since they are likely to be correlated events. Corral (2006a, 2007) suggests not just one but two regimes of power-law decay in the distribution of spatial jumps, hinting at bimodality; steep decay at shorter distances, shallow decay for longer jumps. A dynamical triggering effect is responsible for the short-distance behaviour, while longer jumps are said to represent uncorrelated occurrence. The crossover point is independent of magnitude and would seem to represent a maximum triggering distance. However Corral states that its value scales with the size of the region considered, being around 15km for Southern California, and 200km for the worldwide case.

Bak et al. and Corral are the main two proponents of the idea of a universal scaling law for inter-event times. Yet they are vague about the precise meaning of this scaling. Bak et al. do not explain the process by which they arrived at their particular rescaling, but say, "Only critical processes exhibit this type of data collapse, known as scaling in critical phenomena". Their observation of a data collapse is provided in support of the argument that earthquakes are a self-organised critical process. At a critical point, there are physical properties that can be shown to scale with various different variables in a unified way, so that one of these variables may be altered in compensation of a change in another to produce the same value in the physical property. Data collapse implies criticality, therefore, because it shows this unification of the different variables.

However, aside from this very general inference, nothing much can be said as there is no theory of inter-event times in complex systems. As Corral puts it, "The understanding of this, however, is still far beyond us" (Corral, 2004). Davidsen et al. say, in regard to finding the same scaling function for acoustic emission (AE) events in rock compression studies, that the scaling of these intervals "indicates the existence of a nontrivial universal mechanism in the AE

generation process" (Davidsen et al., 2007). This statement does little to shed any light on the practical meaning and import of these findings.

The universal gamma distribution has not gone uncontested. Davidsen and Goltz (2004) tested the unified scaling law of Bak et al. on California and Iceland data and found an additional power-law regime for small waiting times with a smaller exponent. The two exponents appear to be common across datasets and potentially universal; however, they found no universal scaling function to fit all three segments of the curve: the smaller intervals do not scale in the same way as the longer ones. Incidentally, for Iceland as well as California data, their plots (figure 1.6) show clear double-peaked bimodality, although this is not commented on by the authors. Bimodality can also be seen in the inter-event distance distribution in (Corral, 2007) and is fit by two power laws in that study. Lindman et al. (2005) suggest a peak rather than a kink in the distribution, based on simulations of Poisson process background events together with triggered Omori aftershocks. They also address Bak et al.'s interpretation, demonstrating that this does not mark a sharp transition from a correlated to an uncorrelated regime in their simulations by quantifying the relative proportions of these two types of intervals throughout the range of the distribution. Instead, they argue, it reflects limited-duration sampling of the aftershock sequences: the intervals lengthen as the sequence progresses and so the distribution is truncated at a point corresponding to the rate the sequence had decayed to when it was truncated. They also show that the scaling with the magnitude threshold is a simple consequence of the Gutenberg–Richter law.

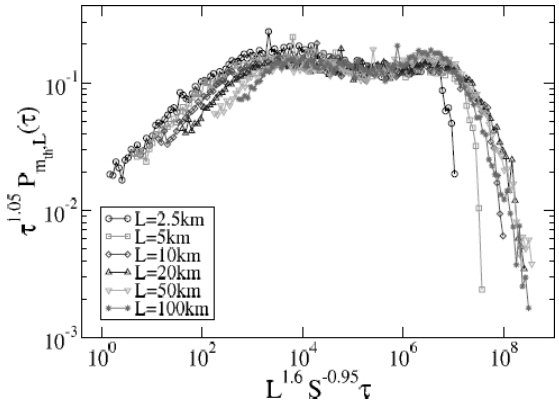


Figure 1.6: Figure 2 from Davidsen and Goltz (2004): Subsets of California data with different spatial area sizes as shown in the legend, rescaled in the same way as Bak et al.

Saichev and Sornette (2006, 2007) carried out an extensive analytical treatment to determine the shape of the inter-event time distribution based on the Epidemic-Type Aftershock Sequences (ETAS) model, and they deduce that it is only approximately universal and of a gamma form assuming that the Omori power-law decay of the aftershock rate has an exponent close to 1, and the branching ratio of seismicity, the average number of events each event triggers, is around 0.7–1. The nature of their formula with two asymptotic power laws again suggests a degree of bimodality, and they state that the power-law intermediate asymptotics in Corral's analysis occurs as a "crossover" between their steeper power law at short intervals and the exponential decay at longer intervals. Prior to that, Molchan (2005) had published an analytical study on the earthquake inter-event time, and showed that if the distribution was universal, it necessarily had to be an exponential—clearly not what is observed. He explains the apparent universality

in the form of a power law with exponential tail as being due to the asymptotics. This would suggest that data collapse was rather trivial and not necessarily related to criticality.

In response to these criticisms, Corral (2009) acknowledges that the apparent scaling law for inter-event times is puzzling. He proposes that a universal distribution may not have to take an exponential form where there are long-range correlations in the time series. He also questions the relevance of results based on the ETAS model as this model is not fully self-similar. This argument depends on such strict scale invariance being an axiomatic principle, whereas ETAS has been shown empirically to be an extremely robust feature of seismicity world-wide.

Using a cluster model for earthquake occurrence, Molchan (2005) also identifies the rate of exponential decay in the distribution at long intervals as being equal to the fraction of main events in the catalogue, a potentially important metric. Hainzl et al. (2006) take this latter point further and test the theory using synthetic catalogues, and find a significant correlation between the exponential rate and the fraction of main shocks. Traversa and Grasso (2010) use this technique to estimate main shock fractions for earthquake catalogues at volcanoes during inter-eruptive phases, which they find to be gamma-distributed. In terms of other practical applications of the universal gamma distribution idea, Bottiglieri et al. (2009a) similarly find that catalogues at volcanic areas collapse onto the same gamma function as tectonic catalogues, inferring from this that they are generated by similar processes. Corral claims to have "taken advantage of the scaling properties of the recurrence time distribution to improve the statistics of occurrence of large earthquakes over small regions" (Corral, 2005) by mixing data from different spatial cells to increase histogram counts.

Thus, following the initial claim of data collapse by Bak et al. (2002) which was intended to argue for earthquakes being a self-organised critical phenomenon, a pattern now seems to be emerging in the literature of attaching deep (if vague) significance to gamma distributions and data collapse in the inter-event times as a result of this universality hypothesis and related ideas. The cause of this gamma form and how it arises from the underlying processes has remained an open question until now, particularly in how it relates to the aftershock triggering process, and there is a need for clarification of what exactly it means if inter-event time histograms from different catalogues and tectonic/non-tectonic settings appear to rescale and collapse onto a single gamma form. Does it imply self-organised criticality, a universal mechanism for the events, or anything else?

1.5 Thesis overview

This thesis answers these questions, freely exploring the inter-event time histograms without the assumption of a gamma form and in particular investigating the bimodality suggested by some of the plots in the reviewed literature. The gamma form, data collapse and universality are shown to be an artifact of data selection procedures such as the restriction on using only 'stationary' data. The true form of the distribution is determined by the balance between correlated and independent successive events in the series, which in turn depends on the extent of temporal overlap of independent aftershock sequences—a function of the spontaneous rate or, equivalently, regional extent of the dataset. This new way of thinking about earthquake time series and inter-event times affords important insight into other problems concerning uncertainty in statistical measures, such as the mean inter-event time and the maximum-likehood parameters for statistical models such as ETAS.

Chapter 2 describes the methods used.

Chapter 3 looks at the uncertainty in the mean inter-event time of a catalogue, and how the mean converges with respect to catalogue length. A sequential sample of inter-event times does not resemble a random sample from the underlying distribution, because of correlations in the form of aftershocks. This means that the event rate converges slowly to its true long-term value. However, catalogues from larger regions converge faster; later chapters shed more light on the reason for this.

In chapter 4, I explore the inter-event time distributions of real catalogues, and find both bimodality and unimodality (and, therefore, non-universality). I explain these forms in terms of the ETAS model. There are two categories of interval: those between related (co-triggered) events belonging to the same aftershock sequence, and those between unrelated events belonging to different sequences. These are distributed differently and this potentially leads to a bimodal overall distribution—but whether it does depends on the extent to which the separate sequences overlap in time. Significant overlapping, which occurs when there is a high spontaneous rate e.g. in catalogues spanning large regions, causes the inter-sequence intervals to dominate the population and the overall distribution becomes unimodal and exponential.

In chapter 5, I confirm the connection between spontaneous rate in the ETAS model and region size in real data, showing that when these are respectively increased, the increased overlap of aftershock sequences that results causes the IET distribution to go from bimodal to an exponential shape. My main point is that the effect of aftershock triggering is masked in this case. The shape of the distribution becomes invariant with respect to the other (triggering) parameters and is the same except for a change in the mean value. This indistinguishability is reflected in poor inversion success for the ETAS parameters: increased frequency of failure to converge, bigger errors, less accurate error estimation, and a systematic overestimation of the spontaneous rate and underestimation of the branching ratio, which is entirely consistent with the observation that the intervals are overwhelmingly dominated by independent ones, rendering the underlying aftershock processes basically invisible.

In chapter 6, I go on to explore the inter-event time distribution for earthquakes in volcanic settings, looking to use a similar categorisation of intervals to explain its form and infer something about the properties of volcanic earthquake time series.

Chapter 7 provides a general discussion bringing the work together and setting it in context, exploring ideas of how aftershocks fit into the SOC paradigm, and discussing applications and future plans arising from the results. Chapter 8 then concludes the thesis.

Chapters 3 to 6 contain papers published or submitted, preceded (and in the case of chapter 5, succeeded) by additional text that sets the context. The details of each paper, in chapter order, are as follows:

- M. Naylor, I. G. Main, and S. Touati. Quantifying uncertainty in mean earthquake interevent times for a finite sample. *Journal of Geophysical Research-Solid Earth*, 114(B01316), 2009.
- S. Touati, M. Naylor, and I. G. Main. Origin and Nonuniversality of the Earthquake Interevent Time Distribution. *Physical Review Letters*, 102(16), 2009.
- S. Touati, M. Naylor, I. G. Main, and M. Christie. Masking of earthquake triggering behavior by a high background rate and implications for epidemic-type aftershock sequence inversions. *Journal of Geophysical Research-Solid Earth*, 116, 2011.

- A. F. Bell, S. Touati, M. Naylor, and I. G. Main. The nature and origin of interevent time distributions for volcanic earthquakes. *Submitted to JGR*, 2011.

Chapter 2

Methods

In this chapter I present and explain the computational methods by which the data analysis, parameter inversions, and simulations were performed for the work presented in the thesis. Codes are included so that the results of the thesis can be easily reproduced: self-contained functions and longer code examples are in the Appendix, while shorter code examples are included in this chapter.

All of the work was done in R (R Development Core Team, 2010), which is a free statistical and graphical programming environment built on the S language. The Statistical Seismology Library (SSLib) (Harte, 2007a), an R package, was also used where stated. SSLib includes earthquake catalogues as data objects, provides functions (`subset.rect`, `subset.polygon` and so on) for subsetting these catalogues according to chosen spatial, temporal and magnitude limits, and provides various functions for statistically analysing the data including fitting and simulating stochastic models such as ETAS.

Earthquake catalogues report times, magnitudes, depths, and latitude and longitude coordinates of the events. When dealing with catalogues, a completeness magnitude must be determined—a threshold above which all the events in the time–space window under consideration can be assumed to have been recorded. Below this level only those events occurring closer to the seismic stations may be recorded. The value of this threshold depends on the network and can change over time as well as more stations are added.

2.1 Visualising the events as a function of time

I use various plotting methods to visualise the progression of events as a function of time in real and synthetic earthquake catalogues.

One way is to simply plot event times and magnitudes as a 'comb' plot, with a cumulative event count (an example is shown in figure 2.1). The comb plot is essentially a time–magnitude scatter plot drawn as vertical lines rather than points. It is easily produced in R by specifying `type="h"` in the plotting function (line 2 in the code below). The cumulative count may be easily achieved by plotting each event's chronological number against its time of occurrence (lines 5–6).

```
1   par(mar = c(5, 4, 4, 4) + 0.3)
2   plot(times, magnitudes, type="h", xlab="Time (days)", ylab="Magnitude")
```

```
3   par(new = TRUE)
4   event.indices <- 1:length(times)
5   plot(times, as.numeric(event.indices), type="l", axes=FALSE, bty = "n",
6       xlab = "", ylab = "")
7   axis(side=4, at = pretty(range(as.numeric(event.indices))))
8   mtext("Cumulative events", side=4, line=3)
```

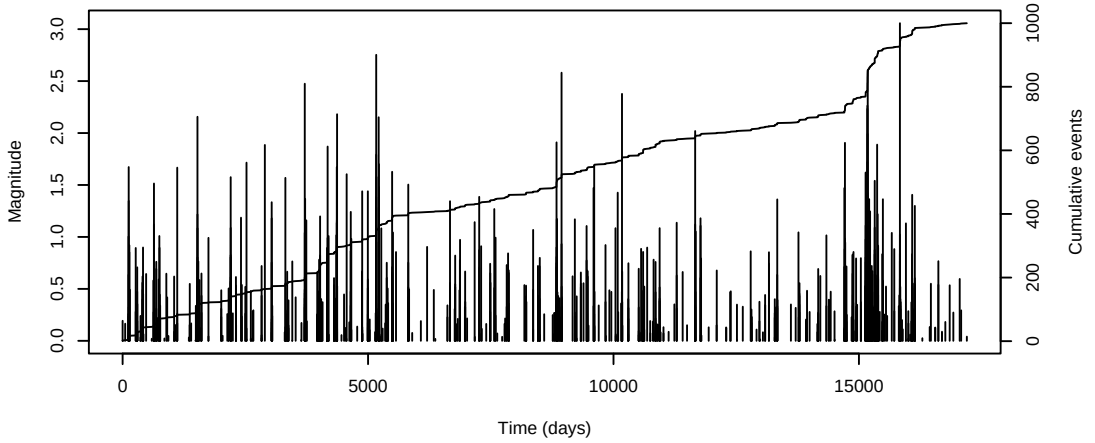


Figure 2.1: An example of a comb plot of event magnitude versus time with a cumulative event count curve overlaid.

An ongoing mean event rate may be calculated (and plotted as a function of time, e.g. figure 2.2) to give a visual indication of its rate of convergence to a steady value. This is done by producing a sequence of average rates, calculated from the start of the catalogue to the current time, at the time of each event:

```
1   first.time <- times[1]
2   event.time <- times[2:length(times)]
3
4   mean.rate <- NULL
5   event.number <- NULL
6   for(i in 1:(length(event.time)))
7   {
8       event.number <- c(event.number, i)
9   }
10  mean.rate <- event.number/(event.time-first.time)
```

Mean event rates can be calculated for each month of the catalogue to show how this event rate fluctuates over time:

```
1   date.from <- julian(1,1,1970) # start of catalogue
2   date.to <- julian(1,1,2100) # end of simulation
```

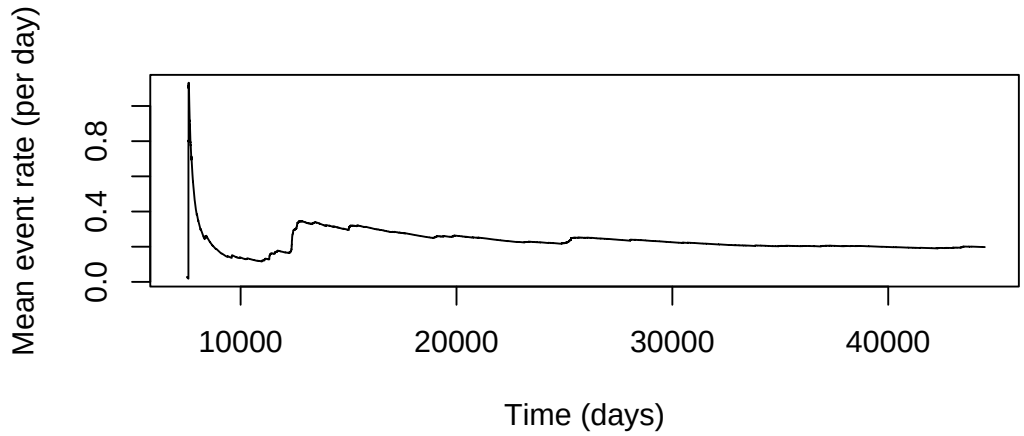


Figure 2.2: An example of a plot of the ongoing average event rate as a function of time for a catalogue.

```
3   monthlytime <- seq.dates(from=date.from, to=date.to, by="months")
4
5   monthlyevents <- 0
6   monthlydays <- 0
7   for(i in 1:(length(monthlytime)))
8   {
9       # monthly days
10      start <- monthlytime[i]
11      end <- monthlytime[i+1]
12      if(is.na(monthlytime[i+1])) {end <- date.to}
13      if(start==end) {break}
14      monthlydays <- c(monthlydays, (end - start))
15
16      # monthly events
17      month.subset <- subset.rect(catalogue, minday=start, maxday=end)
18      monthlyevents <- c(monthlyevents, length(month.subset$indices))
19  }
20
21  monthlymean.rate <- monthlyevents/monthlydays
```

An example of this type of plot is shown in figure 2.3.

2.2 Inversion of ETAS parameters

The code for carrying out ETAS parameter inversions makes use of SSLib functions and largely follows the examples given in the SSLib manual (Harte, 2007b). It uses the minimiser `nlm`, built into R, that follows a Newton-type algorithm, to minimise the negative of the log likelihood

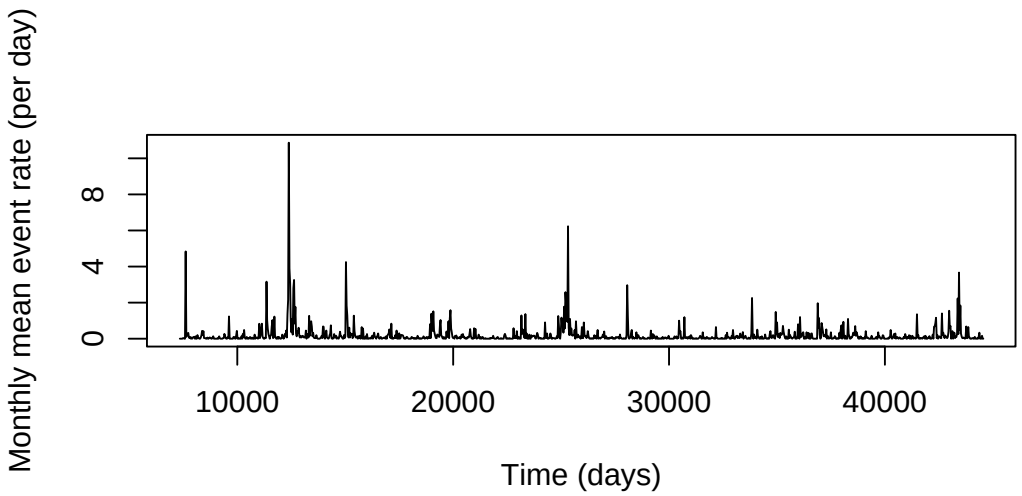


Figure 2.3: An example of a plot of the month-by-month mean event rate against time for a catalogue.

(equation (1.7) in Introduction).

When inverting parameters for real data, the magnitudes are rescaled by the threshold value prior to inversion, for example:

```
1   NZnorth$magnitude <- NZnorth$magnitude - 4.0
```

If simulation using the inverted parameters is required, this will then make the simulation, which uses 0 as the minimum magnitude by default, straightforward. The threshold value is then added back on to the magnitudes afterwards. When inverting from synthetic catalogues this is not necessary.

Given a catalogue `sim` and a `start` and `end` time in days, the following code will invert parameters over the specified time period (but taking the full history at each timestep into the conditional intensity calculation). The vector `params`, passed into `nlm`, contains the initial parameter values from which the inversion will be started. No more than 1000 iterations will be performed (`iterlim=1000`), the Hessian matrix giving the final likelihood surface gradients will be returned (`hessian=TRUE`), and the suggested step sizes for the parameters are given in the `typsize` argument.

```
1   TT <- c(start, end)
2   posterior <- make.posterior(sim, etas.cif, TT=TT)
3   neg.posterior <- function(params) (-posterior(params))
4   try(z <- nlm(f=neg.posterior, p=params, iterlim=1000, hessian=TRUE,
5       typsize=c(0.1, 10, 1, 0.01, 1), print.level=2))
```

The following example inverts only μ while keeping the other parameters fixed at their true values. A uniform ('non-informative') prior distribution from $-\infty$ to ∞ is specified for μ, whereas Dirac functions are used as priors for all other parameters.

```
1   y <- prior.info(density=c("NIprior", "Dirac", "Dirac", "Dirac", "Dirac"),
```

```
2       par1=c(-Inf, params[2], params[3], params[4], params[5]),
3       par2=c(Inf, params[2], params[3], params[4], params[5]))
4   posterior <- make.posterior(sim, etas.cif, TT=TT, prior.info=y)
5   neg.posterior <- function(params) (-posterior(params))
6   try(z <- nlm(f=neg.posterior, p=params[1], iterlim=1000, hessian=TRUE,
7       typsize=0.1, print.level=2))
```

To recover the standard error from the Hessian matrix after the inversion, the following code is used:

```
1   covariance <- solve(z$hessian)
2   stderr <- sqrt(diag(covariance))
```

Where visualisation of the log-likelihood surface over two dimensions is needed, SSLib provides a function `pp.contours` which is helpful. Its arguments are the earthquake data, the parameter values to be used (most likely the maximum-likelihood inversion result), the conditional intensity function to be used, the time limits over which to evaluate the likelihood, the indices of the two parameters which are to be used as the x and y coordinates, and a list of values for each of these parameters providing the evaluation grid. An example of its usage is as follows:

```
1   w <- pp.contours(NZnorth, params, etas.cif,
2       TT=c(catalog.start, catalog.end), param.index=c(2, 1),
3       steps.x=seq(2.9, 3.3, 0.04), steps.y=seq(0.18, 0.22, 0.004))
```

The returned object contains the coordinates of the evaluation grid and a matrix of the values of the log likelihood at each point in the grid. The result may be plotted as contours using the R function `contour` or as a grid of coloured points; see the example in figure 2.4.

2.3 Temporal ETAS simulation algorithms

SSLib has a simulation algorithm that works in conjunction with a choice of several conditional intensity functions provided, of which ETAS is one. It uses the thinning method which works as follows. The time to the next event is provisionally calculated based on the current conditional intensity (at the current event); but because the conditional intensity decreases with time (as the Omori contributions die down), this would produce too many events, so 'thinning' is needed. The event therefore may or may not actually be created, with a probability that is the ratio of the conditional intensity at that time to the conditional intensity at the previous event.

SSLib uses an alternative parameterisation to the one I presented in the Introduction (section 1.3.1). In place of K there is a parameter $A = K/c^p$. The SSLib simulation code is included in section A.1 of the Appendix (`pp.sim.default`). An example of its usage, to synthetically extend the New Zealand North Island dataset (`NZnorth`) using the result of a parameter inversion (`z`), is as follows:

```
1   sim <- pp.sim.default(NZnorth, z$estimate, etas.cif,
2       c(catalog.end, simulation.end), output=TRUE, seed=90, magn.sim=bvalue)
```

The data passed in is used as the initial history of the process, with its triggering effects accounted for in the conditional intensity calculation. The times specified—`catalog.end` and

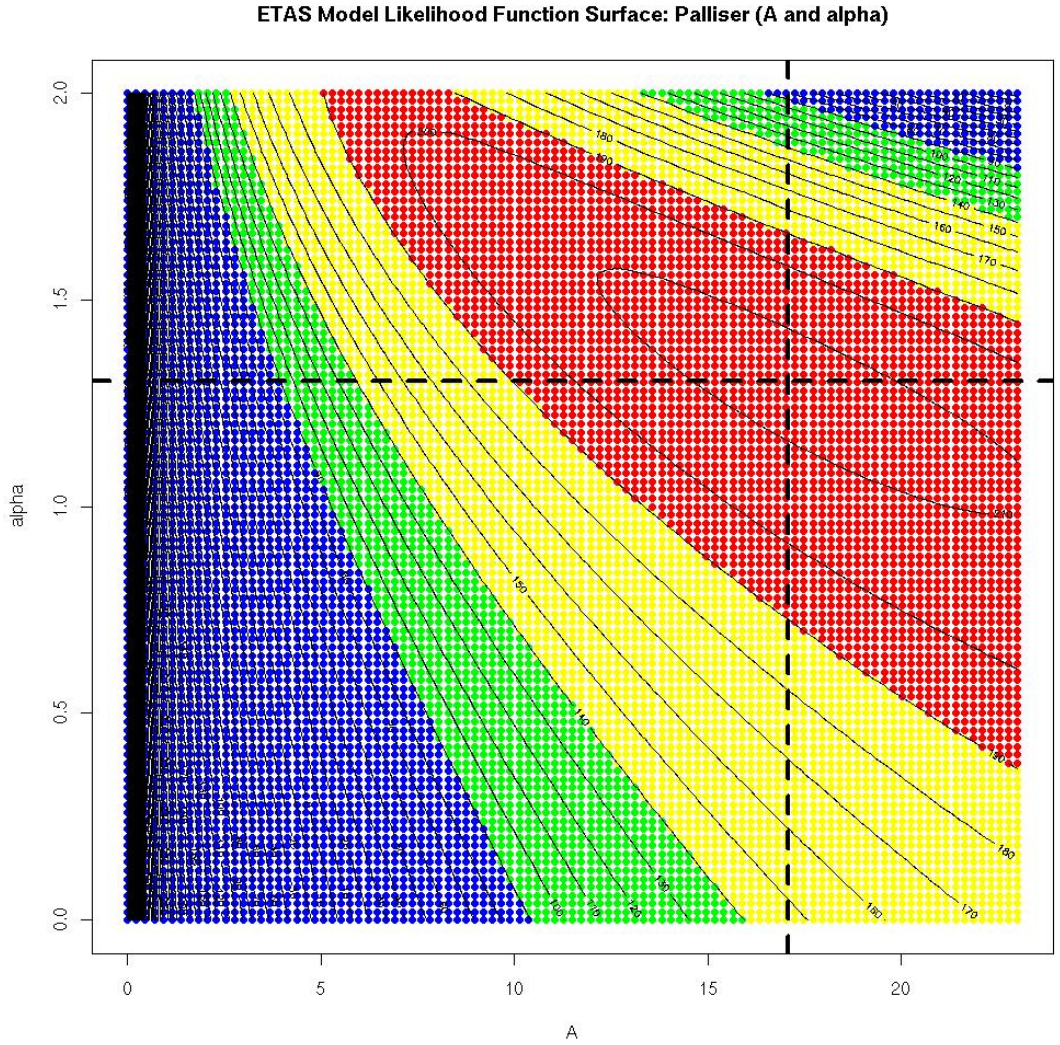


Figure 2.4: An example of a cross-section through the ETAS log-likelihood function in the plane of the parameters $A = K/c^p$ and α (see section 1.3.1 and 2.3), for a catalogue complemented with a grid of coloured points.

`simulation.end`—are the temporal limits of the simulation, `bvalue` would be the Gutenberg–Richter b-value obtained for the dataset (using the SSLib function `freq.magnitude`), and the returned object will contain both the real data passed in and the synthetic data created in the simulation.

The SSLib function was unsuitable for most of my purposes, however, as it does not record the 'family structure' of triggered generations. It calculates the conditional intensity as the sum of μ and the triggering term from past events (without considering these as two separate sources of events). Thus it progresses steadily through time, collecting both background events and aftershocks with no distinction between the two.

I therefore used the code from the SSLib function as a basis for writing my own simulation algorithm using the thinning method and the alternative ETAS parameterisation with $A = K/c^p$. My function creates the Poisson sequence of background events first, and then calls a

recursive function to create the aftershocks for each background event, and the aftershocks for each aftershock, and so on. It records a coded index for each event to indicate its relationship with other events, which is returned as part of the output along with the times and magnitudes. The index system is best illustrated with an example: 436.3.1 would indicate the first aftershock triggered by the third aftershock of the 436th background event.

This simulation procedure turned out to be much faster to run for long catalogues, as the summing of conditional intensity contributions from past events is only performed within each sequence and not for the catalogue as a whole.

A further problem with the SSLib ETAS simulation is that if there is little or no 'history', that is, past events passed in to the simulation, the event rate can take a long time to increase to its steady value as the triggering term increases with the number of recent past events. One solution is to allow for a run-in period and remove this from the result. However due to the way that aftershocks occur in cascading sequences, an event can still be producing indirect aftershocks for a long time; the duration required for the run-in period would be larger than intuitively expected (and choosing its value appropriately is not trivial). The problem is especially acute for simulations with high μ, as a high temporal density of events means that the computational time per unit of simulation time is too high to make the creation of a discarded run-in period practical.

The most effective way of achieving a stable event rate—and also a realistic ratio of spontaneous to triggered events in the catalogue—was to simulate each Omori sequence for a fixed duration (see section 2.3.1 for a derivation of this duration), even if they overshoot the end of the catalogue, and then wrap aftershocks occurring after the end time back to the start (as many times as necessary). These aftershocks' indices also had to be altered to break the connection with their parent events, as they should now be considered to be offspring of events occurring before the start of the simulation.

My simulation code is in sections A.2 (`etas.sim`) and A.3 (`create.aftershocks`) of the Appendix and I describe and explain the procedure in the following sections 2.3.1 to 2.3.4. The arguments to `etas.sim` are as follows: `params` is a vector containing the ETAS parameters; `max.events` is the required number of events to be simulated; `m.min` is a magnitude threshold, below which the events are removed from the catalogue; `seed` is a seed for the random number generators; and `bvalue` is the Gutenberg–Richter b-value.

2.3.1 Sequence duration

Since we are wrapping aftershock sequences back to the start, we need to know how long to run them for. Lines 12–17 of `etas.sim` handle this and are explained as follows. The duration of an aftershock sequence according to the Omori law is infinite, but it is possible to define a finite time period T over which say 95% of the aftershocks should have occurred in an average (direct) sequence and use that as an effective sequence length.

The average rate of aftershocks as a function of time is obtained by integrating the ETAS triggering term multiplied by the magnitude distribution $p(m) = \beta e^{-\beta m}$ over all magnitudes:

$$\lambda_{\langle m \rangle}(t) \quad = \quad \int_0^\infty \lambda_m(t) p(m) dm$$

$$
\begin{aligned}
&= \int_0^\infty e^{\alpha m} \frac{K}{(c+t)^p} \beta e^{-\beta m} dm \\
&= \frac{\beta K}{(c+t)^p} \int_0^\infty e^{(\alpha-\beta)m} dm \\
&= \frac{\beta K}{(c+t)^p} \left[\frac{e^{(\alpha-\beta)m}}{\alpha-\beta} \right]_0^\infty \\
&= \frac{\beta}{\beta-\alpha} \frac{K}{(c+t)^p}
\end{aligned} \tag{2.1}
$$

The integration of this over all time would give the number of aftershocks in an average (direct) sequence, that is, the branching ratio n. We need the time T such that integration from 0 to T will give us some fraction k of the branching ratio (e.g. 95%).

$$
\begin{aligned}
kn &= \int_0^T \frac{\beta}{\beta-\alpha} \frac{K}{(c+t)^p} dt \\
&= \frac{\beta}{\beta-\alpha} K \left[\frac{(c+t)^{1-p}}{1-p} \right]_0^T \\
&= \frac{\beta}{\beta-\alpha} K \frac{(c+T)^{1-p} - c^{1-p}}{1-p} \\
&= \frac{\beta}{\beta-\alpha} K \frac{c^{1-p} - (c+T)^{1-p}}{p-1}
\end{aligned} \tag{2.2}
$$

Then identifying the branching ratio $n = \frac{Kc^{1-p}}{p-1} \frac{\beta}{\beta-\alpha}$ and using $\theta = p - 1$:

$$
\begin{aligned}
kn &= n - nc^{p-1}(c+T)^{1-p} \\
k &= 1 - c^{p-1}(c+T)^{1-p} \\
(c+T)^{-\theta} &= (1-k)c^{-\theta} \\
1-k &= \left(\frac{c}{c+T} \right)^\theta
\end{aligned} \tag{2.3}
$$

For $T \gg c$ (the typical aftershock sequence has duration much larger than c),

$$
\begin{aligned}
1-k &\approx \left(\frac{c}{T} \right)^\theta \\
T &\approx c(1-k)^{-\frac{1}{\theta}}
\end{aligned} \tag{2.4}
$$

I used this formula for the effective sequence length in my simulation algorithm, with $k = 0.95$.

In order to avoid excessive wrapping of sequences, the catalogue created needs to be of considerable length, so the simulation creates a catalogue of at least 100000 events (lines 7–8 of `etas.sim`) and truncates the result if fewer events were needed (lines 87–92). For the same reason, the sequence length T is reduced if it is estimated that this will result in more than 1000 wraps (lines 20–27).

2.3.2 Estimating the number of background events

In line 38 of `etas.sim`, the required number of background events is estimated by dividing the total number of required events by $(1+n)$ (and multiplying by $10^{bm_{min}}$ to account for the fact that events with magnitudes below m_{min} will later be removed). If direct aftershocks alone were created, and no further cascades, the total number of events N would approximately equal $\mu + \mu n$, meaning that the number of background events would roughly equal $N/(n+1)$. This therefore is an over-estimate of the required number of background events, given that we are creating cascades of sequences. A more accurate estimate taking the cascades into account would be $N(1-n)$, because n is also the fraction of the catalogue that is background events (Helmstetter and Sornette, 2003d), but creation of background events is not time-consuming and it is safer to err on the side of creating too many.

The calculation of the branching ratio in line 5 is done by calling the function `calc.n`, which is contained in section A.4 of the Appendix.

2.3.3 Creating the background events

Having made these initial calculations, the next task is to create the background events (lines 39–57). The time to the next event is sampled from an exponential distribution with rate μ and the magnitude is a sample from the Gutenberg–Richter distribution. The coded index given to each background event is simply the chronological event number.

2.3.4 Creating the aftershocks and tidying up

Having completed the background contribution to the catalogue, the background rate is then set to zero and the background events are passed into the function `create.aftershocks` (discussed below). What is returned consists of those background events and the cascades of aftershocks. The events are then put into chronological order, truncated to the required number of events if necessary, and returned from the function.

The function `create.aftershocks` is recursive and is the most complex part of the whole procedure. It consists of a loop starting on line 10 and ending on line 142 in which it takes each of the events passed into the function in turn and processes the sequence creation.

The first command in the loop sets the current time to the time of the selected parent event. It then examines the coded index of the parent event to find out if it is a background event; if so, it checks whether enough events (of magnitude $\geq m_{min}$) have been created at this point—that is, the background events prior to the current time plus all the aftershocks created for them. If the number of these events exceeds the required catalogue size, some tidying up is performed (described later in this section) before returning. This way, the excess background events are simply removed at this point and time is not wasted creating sequences for them.

If the required catalogue length has not been reached yet, the next step in the loop is to create the direct aftershocks. This is done in lines 84–122 and implements the same thinning method as the SSLib function `pp.sim.default` (section A.1 of the Appendix). The function `conditional.intensity` is called, which is contained in section A.5 of the Appendix. A coded index is given to each aftershock which is identical to its parent index but appended with a '.' followed by the number of the aftershock within the sequence. The sequence is stopped when its duration reaches the pre-determined length of the sequence.

Having created direct aftershocks, the next generation (and so on) of aftershocks are then created by calling the `create.aftershocks` function from within itself, passing in the direct aftershocks to be used as the new set of parent events. If the end of the loop is reached at line 142, the cascades created are concatenated with the parent events passed in and returned.

I will now describe what happens when the cascades have been completed for a number of background events and it is found that the total number of events (above m_{min}) has exceeded the required number (lines 24–76). All events below m_{min}, and background events later than the current time (which will be the end time), are first removed (lines 33–40). Then it wraps the aftershocks occurring later than the end time back to the start. The times are adjusted in lines 63–64. Lines 54–62 deal with the changes required to the coded indices of these events (see the start of section 2.3 for why this needs to be done). Essentially the number before the first '.'—that is, the number indicating the background event from which this aftershock was directly or indirectly triggered—is increased. The index of the final background event is added to this number each time the event is wrapped back to the start.

In lines 68–75 the background and triggered events are all concatenated together and returned.

This concludes the simulation algorithm for the temporal ETAS model. An alternative version with a single seeding event is shown in section A.6 of the appendix, `etas.sim.agas`. This is used for simulating a global aftershock sequence. A single background event of a random magnitude (from the Gutenberg–Richter distribution) is created. The desired sequence length may be specified through the argument `sequence.length`, or if not, it will be calculated in the same way as for general ETAS simulations. The average gobal aftershock sequence is obtained by running it repeatedly with different values of the `seed` argument, concatenating the created catalogues and dividing the event count by the number of catalogues.

2.4 Spatial ETAS simulation algorithm

Simulation of space–time ETAS catalogues was needed for the nearest-neighbour plot in figure 5.11 of chapter 5. The functions for doing this are in sections A.8 (`etas.sim.spatial`) and A.9 (`create.aftershocks.spatial`) of the Appendix. They are largely the same as the temporal-only simulation described in section 2.3, but they require the extra spatial ETAS parameters, and they record x and y coordinates for each event.

The argument `coords` in `etas.sim.spatial` allows one to specify a vector of coordinates for the spatial region over which the catalogue is to be created. If two coordinates are input, they are taken to be the centre and radius of a circular region. If four coordinates are specified, they are taken to indicate the two x-coordinates followed by the two y-coordinates bounding a rectangular area.

I used the following spatial density function for aftershocks:

$$f(r,\theta) = \frac{q-1}{\pi}\frac{d^{2(q-1)}}{(r^2+d^2)^q} \tag{2.5}$$

obtained from Console et al. (2003), in which r is the distance from the parent event, θ is the direction, and d and q are the two additional spatial ETAS parameters. It is isotropic and an inverse power law in the distance.

Its implementation in a simulation algorithm requires inverse transform sampling to obtain

a value of r. This involves obtaining the cumulative distribution function and then rearranging to get an expression for r in terms of a probability. Uniform random values of probability from 0 to 1 can then be fed in to sample from the distribution.

The cumulative distribution is:

$$\begin{aligned} F(R) &= \int_0^{2\pi}\int_0^R f(r,\theta) r\,dr\,d\theta \\ &= (q-1)d^{2(q-1)}\int_0^R \frac{2r}{(r^2+d^2)^q}dr \\ &= (q-1)d^{2(q-1)}\left[\frac{(r^2+d^2)^{1-q}}{1-q}\right]_0^R \\ &= -d^{2(q-1)}\left[(r^2+d^2)^{1-q}\right]_0^R \\ &= -d^{2(q-1)}\left[(R^2+d^2)^{1-q} - d^{2(1-q)}\right] \\ &= 1 - \frac{d^{2(q-1)}}{(R^2+d^2)^{q-1}} \end{aligned} \tag{2.6}$$

The distribution is sampled from by setting it equal to u, a uniform random value between 0 and 1. Rearranging for R gives:

$$R = d\sqrt{(1-u)^{-\frac{1}{q-1}} - 1} \tag{2.7}$$

This formula is used in lines 160–1 of `create.aftershocks.spatial` to choose the distance from the parent event when creating an aftershock. (The direction is chosen uniform-randomly in line 164.) Also, if u is set equal to 0.95, then the expression (2.7) gives the expected radius for the circle containing 95% of the aftershocks—a kind of spatial equivalent of the effective sequence length. This is calculated in line 20 of `etas.sim.spatial`, because the area of the simulation is extended at the boundaries by twice this distance while simulating, and cropped back to its proper size at the end. This is to ensure that the spatial density of events at the boundaries is reasonably similar to that away from the boundaries, by including events that have been triggered by events outside the boundaries.

For spatial positioning of background events, in the case of a rectangular region, the x and y coordinates are chosen as uniform random variables within the simulation area. In the case of a circular region, coordinates are assigned for the square region of side length equal to the diameter of the circle and centred on the circle's centre. If the coordinates fall outside the circle, the sampling is repeated until they fall within the circle.

Aside from the extra code to position the aftershocks spatially, `create.aftershocks.spatial` differs from its temporal-only counterpart in the way that it counts up the events. Because it is simulating over a larger area than required, when checking if the required catalogue length has been reached yet, it has to count only the events that are within the proper spatial bounds.

2.5 Inter-event time histograms

Inter-event time histograms are usually plotted with logarithmic axes due to the range of values involved and to show up the presence of power laws as straight lines. A good reference

illustrating the different ways of binning the values for both linear and logarithmic histograms is Bonnet et al. (2001).

The most straightforward way of creating a logarithmic histogram is to put the data in logarithmic bins and plot the logarithm of the count. The count in each bin is then equal to the probability density multiplied by the bin width and multiplied by the total count ($\pm$ an error). To create an empirical probability density function (PDF), therefore, the counts must be normalised by both the total count and the bin widths. This results in decreasing the slope of the curve by 1 (Bonnet et al., 2001).

My R function for creating histograms of inter-event times, `create.iet.hist`, is in section A.7 of the Appendix. It takes a vector of inter-event times as an argument, and a boolean argument `normalise` for indicating whether to normalise the counts by the bin widths. There is an optional argument `breaks` to specify the breaks between the bins; if this is not specified the function creates a sequence of 80 values spanning the range of inter-event times (lines 5–6).

2.5.1 Plotting correlated and uncorrelated subsets

The following code example shows how to plot a histogram of inter-event times from an ETAS simulation, including the correlated and uncorrelated subsets (see figure 2.5 for an example of the result). To decide whether an interval is correlated or not, the coded indices of the events are examined. All we are interested in is whether the events have the same first-generation number, meaning that they belong to the same global aftershock sequence. The sets of intervals are then binned (using the same set of bins for all three) and plotted.

```
1  times <- sim$time
2  gas.indices <- sim$gas.indices
3
4  # get all the iets
5  times1 <- as.numeric(times[1:length(times)-1])
6  times2 <- as.numeric(times[2:length(times)])
7  iets <- times2-times1
8
9  # get the correlated and uncorrelated iets separately
10 indices1 <- gas.indices[1:(length(gas.indices)-1)]
11 indices2 <- gas.indices[2:length(gas.indices)]
12 indices1.sub <- as.numeric(sub("[.].*", "", indices1))
13 indices2.sub <- as.numeric(sub("[.].*", "", indices2))
14 corr <- (indices2.sub-indices1.sub)==0
15 correlated.iets <- iets[corr]
16 uncorrelated.iets <- iets[!corr]
17
18 # plot histograms
19 breaks <- seq(from=min(log10(iets)),to=max(log10(iets)),length.out=80)
20 histogram <- create.iet.hist(iets, breaks=breaks, normalise=FALSE)
21 plot(histogram$x, histogram$y, log="xy", type="o", pch=1, lty=1,
22     xlab=expression(paste(tau," (days)")), ylab="Count")
23 if(length(correlated.iets)>0)
```

```
{
    histogram <- create.iet.hist(correlated.iets, breaks=breaks,
        normalise=FALSE)
    points(histogram$x, histogram$y, col="red", type="o", pch=2, lty=1)
}
if(length(uncorrelated.iets)>0)
{
    histogram <- create.iet.hist(uncorrelated.iets, breaks=breaks,
        normalise=FALSE)
    points(histogram$x, histogram$y, col="green", type="o", pch=3, lty=1)
}
legend(x="topleft", legend=c("All","Same-sequence","Inter-sequence"),
    pch=c(1,2,3), col=c("black","red","green"))
```

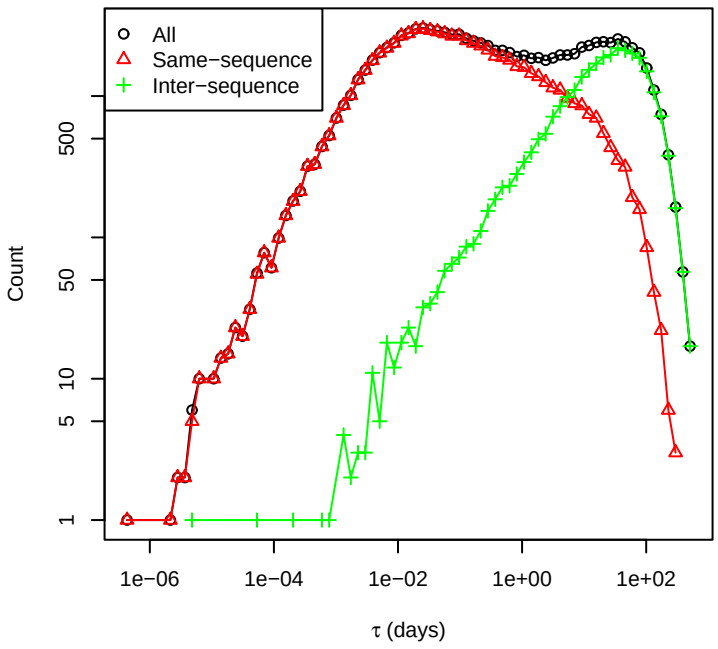


Figure 2.5: An example of a histogram of inter-event times τ for an ETAS simulation (black circles), with the correlated and uncorrelated subsets shown in red triangles and green crosses, respectively.

Having presented the methods of data visualisation, ETAS inversion, simulation and inter-event time distribution analysis, the following chapters will now present the work carried out using these methods and the results obtained.

Chapter 3

Quantifying Uncertainty in Mean Earthquake Inter-Event Times

This paper looks at the rate of convergence of the mean inter-event time in earthquake catalogues as the catalogue length is increased. It frames the slow (non-Gaussian) rate of convergence of catalogues in terms of the degree of correlation of successive inter-event times, providing evidence from ETAS model simulations as well as the autocorrelation of real inter-event time sequences to support this.

My contribution to this paper was to provide the work for section 3.1.4.3 by running ETAS simulations which were used to make figures 3.13 and 3.14 (the R code for the convergence plot is presented and explained in Appendix section A.10). I produced synthetic earthquake catalogues from ETAS model simulations to show that the error in the mean event rate decays with sample size N more slowly than $1/\sqrt{N}$, just as for real earthquakes, and to show that more specifically, the rate of convergence varies systematically with the parameters p and α of the model.

3.1 Paper

Seismic activity is routinely quantified using means in event rate or inter-event time. Standard estimates of the error on such mean values implicitly assume that the events used to calculate the mean are independent. However, earthquakes can be triggered by other events and are thus not necessarily independent. As a result, the errors on mean earthquake inter-event times do not exhibit Gaussian convergence with increasing sample size according to the central limit theorem. In this paper we investigate how the errors decay with sample size in real earthquake catalogues and how the nature of this convergence varies with the spatial extent of the region under investigation. We demonstrate that the errors in mean inter-event times, as a function of sample size, are well estimated by defining an effective sample size, using the autocorrelation function to estimate the number of pieces of independent data that exist in samples of different length. This allows us to accurately project error estimates from finite natural earthquake catalogues into the future and promotes a definition of stability wherein the autocorrelation function is not varying in time.

The technique is easy to apply, and we suggest that it is routinely applied to define errors on mean inter-event times as part of seismic hazard assessment studies. This is particularly important for studies that utilise small catalogue subsets (fewer than ~ 1000 *events) in time-dependent or high spatial resolution (e.g., for catastrophe modelling) hazard assessment.*

3.1.1 Introduction

The earthquake record of the digital age is not yet sufficiently well temporally sampled to define a long-term average in recurrence rates, primarily due to the extreme events having recurrence times greater than the digital record era of ~ 35 years. For example, the Sumatran earthquake was sufficient to modify the best fit statistical model of the global frequency-moment distribution from a gamma distribution before the event to a pure power law Gutenberg–Richter fit after the event (e.g. Main et al., 2008). Here we explore the effect of such limited temporal sampling on the distribution of inter-event times and specifically examine its rate of convergence to a central limit as the sample window is increased.

Statistical convergence and the reduction of errors with increasing sample size are essentially the same phenomenon. The accuracy of any mean estimate derived from a finite sample size N relies on the central limit theorem which tells us that in the limit of infinite sample size, for uncorrelated data, the distribution of sample means will tend to a normal distribution independent of the form of the 'parent' distribution (e.g. Laplace, 1812; Bouchaud and Potters, 2001; Ross, 2003). As a corollary the rate at which the sample mean converges to the parent distribution mean (from an infinite sample) is $1/\sqrt{N}$. In practice, for uncorrelated data, $1/\sqrt{N}$ Gaussian convergence is observed for relatively small sample sizes ($N > 10$) for a wide range of parent distributions provided the moments are finite. However, the sample mean will converge slower than $1/\sqrt{N}$ if correlations exist in the data.

The applicability of such analysis also requires the parent distributions to be stationary. On a timescale of a few Ma, much greater than the repeat times of the largest earthquakes, the far field drivers of plate motion are remarkably stationary (DeMets, 1993). As a consequence, time-independent hazard assessments are made by assuming the long-term behaviour is stationary. How the resulting strain is accommodated locally will depend upon the structural setting, for example in collisional mountain belts the continuous driving of plate convergence results in punctuated deformation of individual thrusts within a fold and thrust belt $\sim 10^6$–10^7 years (Naylor and Sinclair, 2007). The short-term nature of the catalogues (~ 35 years of digital data) means that the degree of stationarity is impossible to assess, and introduces significant uncertainty in estimates of long-term hazard.

We demonstrate here how three separate effects limit the statistical convergence of earthquake event rates and inter-event times. These are (1) correlations in the form of aftershocks, (2) the skewness of the underlying distribution, and (3) finite sample size. First we demonstrate how the statistical properties of mean inter-event times are relatively more statistically stable than event rate as a metric for characterising background activity.

3.1.2 Non-Gaussian convergence of event rates and inter-event times

In this section we discuss the nature of convergence observed in three different earthquake catalogue subsets for New Zealand, Southern California and global PDE (table 3.1.2) that we

Table 3.1: Ranges and properties of the earthquake catalogues used here.

Catalogue property	PDE	NZ	SCEC
Date range	12 January 1971–12 January 2005	12 January 1969–12 January 2006	12 January 1968–12 January 2006
Lower magnitude threshold, m_c	5.0	4.0	3.0
Total number of events available, N_{events}	49,196	14,745	12,128
Number of samples of size 4096 that can be drawn from catalogue subset, N_{4096}	12	3	2

will be analysing further in this study. Main et al. (2008) demonstrated that the running monthly mean of earthquake event rate is steadily increasing in the global CMT catalogue; we extend this analysis here. Monthly event rates, $\dot{N}_{monthly}$ are shown for each catalogue in figures 3.1a–3.1c; they all confirm that lower event rates are more frequently observed. Global monthly event rates (figure 3.1c) do not look like a scaled up version of the regional rates (figure 3.1a and 3.1b) because triggering dominates in the regional catalogues, where the spatial extent of the analysis region is on the order of the spatial earthquake triggering correlation length ($\sim$ 100km) (e.g. Huc and Main, 2003). Since high monthly event rates occur less frequently, i.e., there is significant negative skew in the event rate probability density function, lower event rates are sampled more frequently and the running mean monthly event rate will, on average, converge to the mean from below (figures 3.1d–3.1f) in agreement with Main et al. (2008). Jumps in cumulative mean event rate are coincident with the infrequent extreme high monthly event rates. Similarly, the standard deviation of the monthly event rates used to generate the running mean (figures 3.1g–3.1i) increases dramatically when these high event rates are sampled. This effect is so strong for the small spatial sample regions (New Zealand and SCEC) that the standard deviation of event rates becomes greater than the mean (figures 3.1j and 3.1k). In contrast, the global sample has a coefficient of variation less than 1 (figure 3.1l). Thus convergence of event rates will occur when such extreme event rates no longer significantly affect either the mean or standard deviation of the event rate. This will only occur once the variance in the tail at high event rates is well sampled; in other words we must wait for the largest events or use a declustering algorithm to remove extreme fluctuations to produce a well-behaved metric. Due to the effect of these extreme events on the running monthly event rate, convergence is punctuated and highly non-Gaussian ($N \gg 10$ for convergence).

An alternative way to view the same data is to convert it to a mean monthly inter-event time measured in days $\bar{\delta t}_{monthly} = 30.44/\dot{N}_{monthly}$, where 30.44 is the average monthly duration measured in days (figures 3.2a–3.2c). This transformation changes the weights of the events such that the frequently occurring low rates now carry the most weight and the extreme events carry the least. The difference in weights between the mean event rate and mean inter-event time is related to the difference between arithmetic means and harmonic means. In calculating the running mean the occurrence of an extreme event only increments the sum of past inter-event times by a small amount and the number of events by 1 resulting in a small change in the mean. This is in contrast to running mean event rate where an extreme event added a

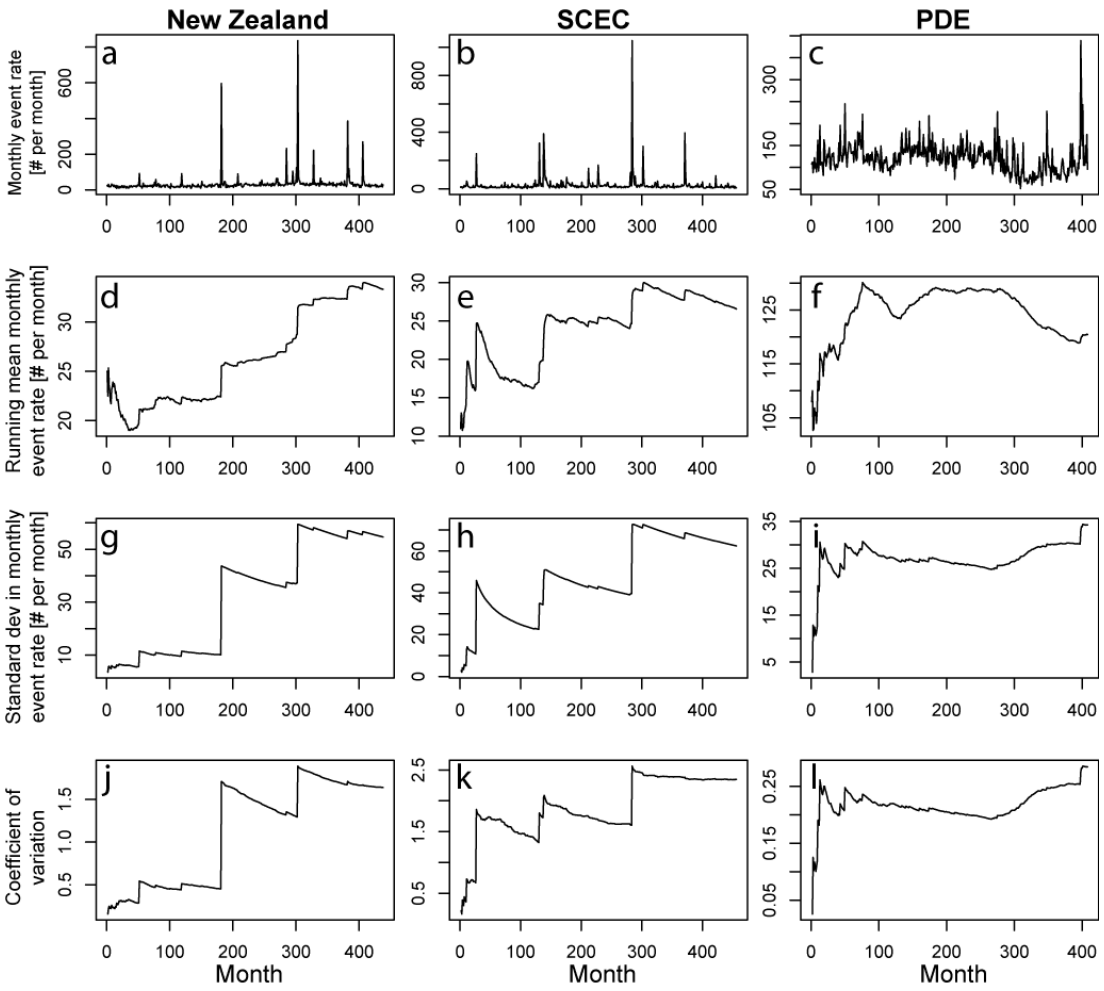


Figure 3.1: Examples of earthquake data that demonstrate the non-Gaussian convergence of mean event rates for New Zealand, Southern California, and a global catalogue in agreement with the analysis. (a–c) The incremental monthly event rates. (d–f) Steadily rising running mean event rates with the most rapid changes in event rate coincident with intermittent spikes in the monthly event rate. (g–i) Non-Gaussian convergence of event rates characterised by a series of intermittent jumps that correlate with high monthly event rates. (j–l) The coefficient of variation for the running mean monthly event rates.

very large number to the sum of past event rates resulting in step changes in the mean. Thus mean inter-event times tend to converge from above, rather than below, and also to mitigate the impact of the extreme events on the running mean (figures 3.2d–3.2f). This is reflected in the coefficient of variation which is much more stable and less than 1 for all catalogues (figures 3.2j–3.2l). Thus the uncertainties on earthquake inter-event times derived purely from the underlying parent distribution are much lower than for event rates, even though one is completely derived from the other.

In summary, inter-event times mitigate the effect of extreme event fluctuations, a desirable property for a metric characterising background activity. Thus in the remainder of this paper we investigate convergence of earthquake mean inter-event times in order to quantify errors on the mean.

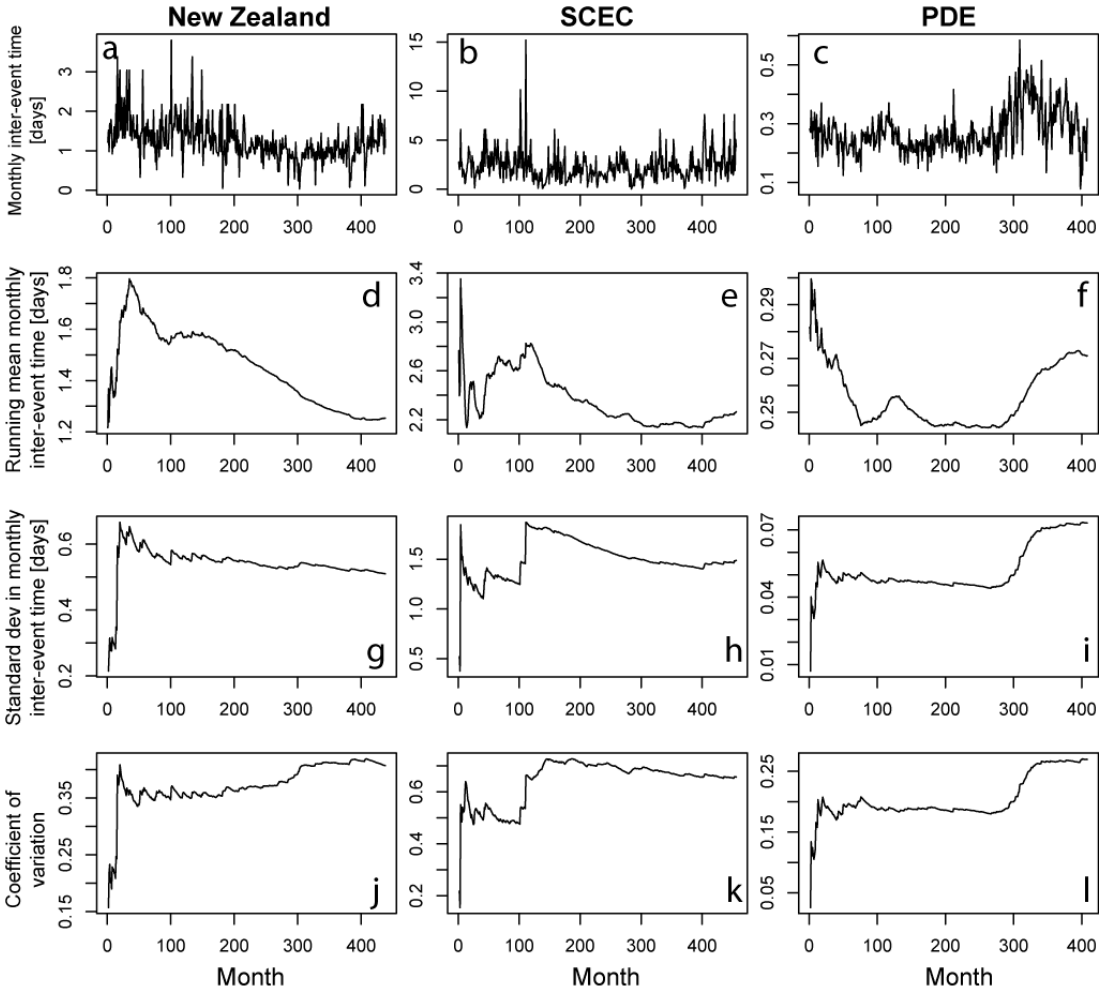


Figure 3.2: Interevent time. (a–c) Monthly, (d–f) running monthly mean, and (g–i) standard deviation of inter-event times making up the running monthly mean. (j–l) Coefficient of variation for the running monthly mean inter-event times.

3.1.3 Theory

In this section we review and develop the theory required to understand convergence of mean earthquake inter-event times. First, we define some terminology. A parent distribution is a probability density function from which values of a random variable are drawn, which we assume to be stationary over very long timescales in our application. A collection of random variables drawn from the parent distribution is a sample. The sample mean is the mean of the random variables in a sample. The sample length, N is the number of values used to calculate the sample mean. The distribution of sample means is a histogram of many sample means taken from the parent distribution. Here our time series is a sequence of inter-event times. The relation between these definitions and the earthquake problem which we apply it to in this study is illustrated in figure 3.3, where the known parent distribution described above is replaced by the measured earthquake time series.

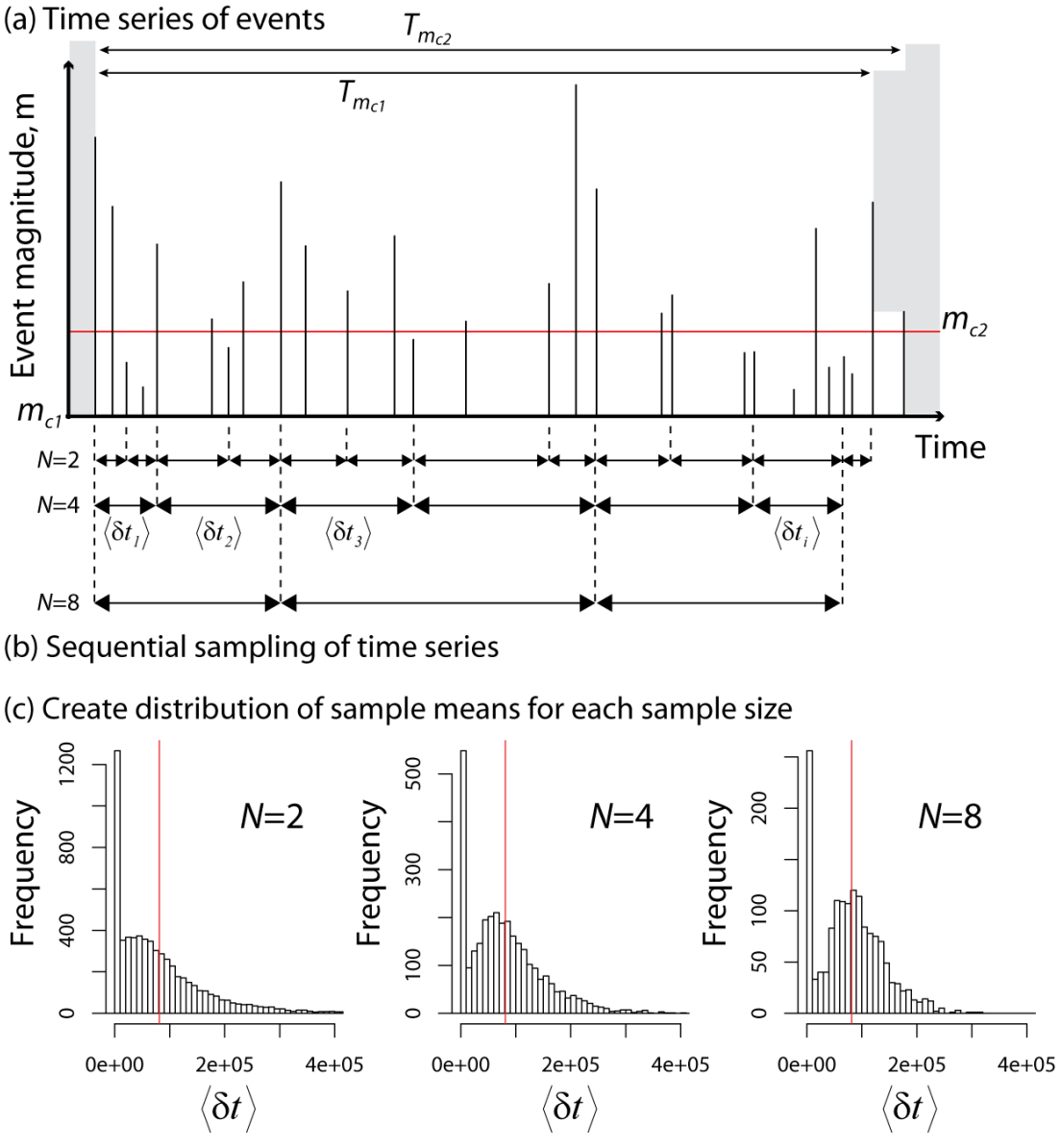


Figure 3.3: Illustration of how the distribution of sample means is created from a time series of earthquake events for the case of sequential sampling.

3.1.3.1 Null hypothesis for convergence without correlations: random sampling from the gamma distribution

Two key results of probability theory underlie this study.

(1) The "Strong Law of Large Numbers" (e.g. Poisson, 1837; Ross, 2003) gives the intuitive relation that for a sequence of independent random variables $x_1, x_2, ...$ with a common stationary parent distribution, and an expectation of the mean $\langle x_i \rangle = \mu$, then with probability 1,

$$\frac{x_1 + x_2 + x_3 + ... + x_N}{N} \rightarrow \mu \tag{3.1}$$

as $N \rightarrow \infty$. In other words, the mean of a randomly sampled distribution tends to the mean of the parent distribution as $N \rightarrow \infty$; it is this result that allows us to expect that we gain a better estimate of the mean inter-event time with a longer earthquake record.

(2) The "central limit theorem" (Laplace, 1812) gives the less intuitive relation that the distribution of sample means drawn randomly and independently from a parent distribution, provided its first and second moments are finite, tends to a Gaussian distribution as $N \rightarrow \infty$, independent of the parent distribution (e.g. Ross, 2003). This drives Gaussian convergence. Formally, the distribution of

$$\frac{x_1 + x_2 + x_3 + ... + x_N - N\mu}{\sigma\sqrt{N}} \tag{3.2}$$

tends to the standard normal as $N \rightarrow \infty$. That is,

$$P\left\{\frac{x_1 + x_2 + x_3 + ... + x_N - N\mu}{\sigma\sqrt{N}} \leq a\right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx \tag{3.3}$$

as $N \rightarrow \infty$.

Note that in this formal definition, the sample mean multiplied by the sample length is removed from the numerator, centring the normal about the origin. The power of this result is that it holds independent of the parent distribution, provided the first and second moments are finite, and that in practice this limiting behaviour can be seen even at relatively small N for many real systems. If the central limit theorem, including the assumption of independent events, holds then

$$\sigma_{sample} \approx \frac{\sigma_{parent}}{\sqrt{N}} \propto N^{\eta} \sigma_{parent} \tag{3.4}$$

with $\eta = -0.5$.

The generalised central limit theorem describes the sum of random variables from parent distributions with power law tails, and hence infinite second moment or variance. These converge to an alpha-stable Levy distribution (Gnedenko and Kolmogorov, 1968). Gaussian convergence and the regular central limit theorem are special cases of this generalised central limit theorem for finite second moment.

Combining the central limit theorem and the Law of Large Numbers, the distribution of sample means in real data converges to a normal distribution centered about the mean at a rate of $1/\sqrt{N}$ as $N \rightarrow \infty$.

We demonstrate this with an example which defines our null hypothesis for the convergence of earthquake inter-event times. The distribution that best describes the distribution of earthquake inter-event times for stationary periods, $f(\delta t)$ is the gamma distribution (Corral,

2004).

$$f(\delta t) = C\delta t^{\gamma-1}\exp(-\delta t^{\delta}/B) \tag{3.5}$$

Corral (2004) noted that the parameters describing this distribution, when rescaled by the mean inter-event time and normalised by the bin width in linear space, are the shape parameter, γ and scale parameter, B and the power $\delta \approx 1$. Using the analytic values for the mean, $\mu = \gamma B$ and variance, $\sigma = \gamma B^2$ of the gamma function, these values imply a coefficient of variation in the mean inter-event times of $CV = \sigma/\mu \approx 1.2$, i.e., the range of inter-event times is significantly greater than the mean inter-event time. Note that this coefficient of variation is calculated for the raw data which is different to that presented in figure 3.2 which was calculated for monthly averages. Even though this distribution is unlikely to be universal in the most general sense (Hainzl et al., 2006) it is adequate for our purpose because it captures the first order mean, variance and skew of the data which is the information that modifies the central limit theorem convergence.

Using this parent gamma distribution we randomly generated a distribution of 1000 sample means for each of the different sample lengths, N (figure 3.4). We plot the sample length and standard deviation of the distribution of sample means on a log-log scale to test for power law dependence in the rate of convergence with sample size (figure 3.5a). In agreement with the central limit theorem, the distribution of sample means quickly tends to a normal distribution as $N \to \infty$ at a rate of $1/\sqrt{N}$ (i.e., $\eta \approx -0.5$ in figure 3.5b). This constitutes our null hypothesis for the rate of convergence of randomly sampled inter-event times from a parent gamma distribution which we compare with sequential earthquake inter-event times that might be more strongly affected by temporal correlations such as aftershock sequences.

3.1.3.2 Convergence with correlations: autocorrelation and effective sample length

The autocorrelation function (ACF) provides an empirical measure of the strength and range of correlations in a data set. The strength of interactions between time series data spaced at a lag k can be quantified using the autocovariance coefficient which is defined as the average product of departures at times t and $t + k$,

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x}) \tag{3.6}$$

The autocovariance with no lag ($k = 0$), c_0, returns the variance of the data. For non-zero lags, persistence is indicated by positive values of c_k and anti-persistence is indicated by negative values of c_k. For non-zero lags, a lack of correlation is indicated by $c_k = 0$ in an infinite sample and c_k less than some limit defined by the 95% confidence intervals for finite samples in which counting errors must be taken into account (e.g. Zivot and Wang, 2006).

The autocorrelation function, at lag k, is a variance normalised version of the autocovariance at lag k,

$$r_k = \frac{c_k}{c_0} \tag{3.7}$$

The expected variance for a sample of length N can be calculated for data with correlations by summing the autocorrelation function over the first N terms and multiplying by the variance

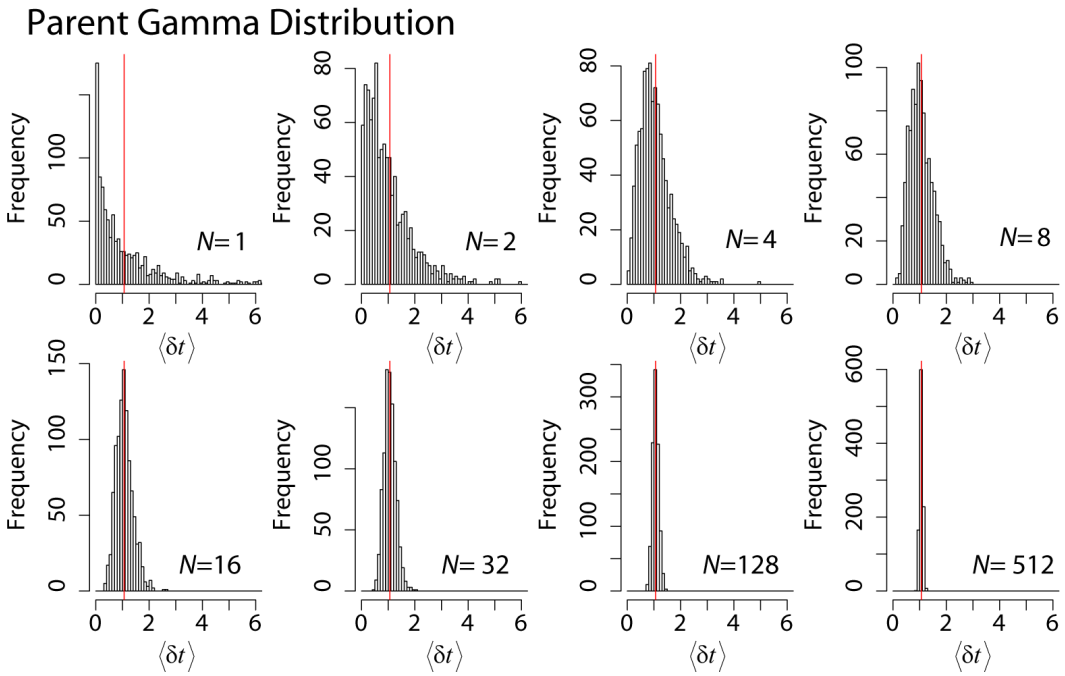


Figure 3.4: Demonstration of the central limit theorem using sample of means from a parent gamma distribution. The distribution of sample means tends to a normal distribution centered about the mean of the parent distribution with $\sigma_{sample} = \sigma_{parent}/\sqrt{N}$. This random sampling of the gamma distribution represents the null hypothesis for the rate of convergence of earthquake inter-event times.

in the data (Bouchaud and Potters, 2001). Modifying the central limit theorem to include correlations produces

$$\sigma_N^2 = c_0 \left(\frac{1}{N} + \frac{2}{N} \sum_{k=1}^{N} \left(1 - \frac{k}{N} \right) r_k \right) \tag{3.8}$$

Note that in the absence of correlations ($r_k = 0$ for $k \neq 0$) this relation returns the central limit theorem rate of convergence. It is possible to devise sampling schemes for nonindependent variables, e.g., mixing processes, which in principle could pass the null hypothesis of Gaussian convergence. Our analysis shows that this is not the case for earthquake inter-event times.

Using this information about the statistical nature of the correlations we can define an effective sample length, N_0 which estimates how many independent pieces of data there are in the inter-event time catalogue and is dependent upon the length of the sample under investigation. This effective length is defined by setting equation (3.8) equal to c_0/N_0 and rearranging. The process is non-Markovian in that future events are dependent upon more than just the current event, thus the effective sample size needs to be a function of autocorrelation coefficients over a range of lags that define the decay of the statistically significant correlations in figures 3.6a–3.6c.

$$N' = \frac{N}{\left(1 + 2 \sum_{k=1}^{N} \left(1 - \frac{k}{N} \right) r_k \right)} \tag{3.9}$$

We will demonstrate that combining this effective sample size with the central limit theorem

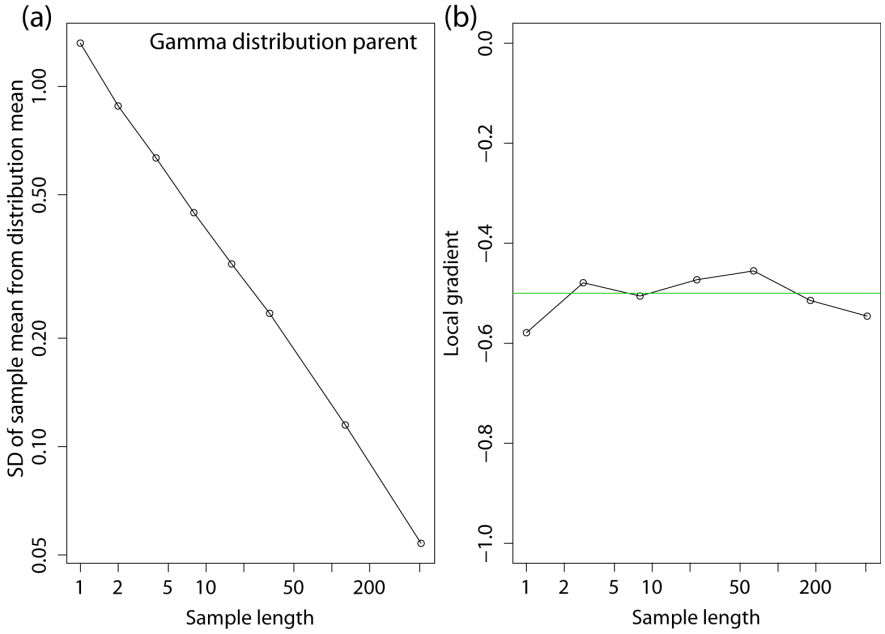


Figure 3.5: Demonstration of the central limit theorem prediction that the distribution sample mean converges as $\sqrt{N}$ using the sample mean distributions in figure 3.4. The solid line shows the standard deviation of the sample mean distribution from the mean of the parent distribution derived from figure 3.4 with varying sample length N. The gradient of the line gives the power for the rate of convergence with sample size and compares well with the central limit prediction of -0.5 (green).

provides an appropriate correction for estimating the errors on earthquake mean inter-event times. For positive correlations $N' \leq N$, and the true error estimate is bigger than that expected from the standard deviation based on N.

3.1.4 Analysis

The analysis will be carried out on three earthquake catalogue subsets (table 3.1.2). There are two regional catalogues, New Zealand and Southern California, and one global catalogue, PDE. All of the analysis is performed using R (R Development Core Team, 2010) and SSLib (Harte, 2007b). We also demonstrate that the results are relatively insensitive to the precise choice of magnitude cutoff.

3.1.4.1 Sequential and random sampling from earthquake catalogues

The central limit theorem assumes that the events in a sample are independent. However, aftershocks and foreshocks are known to generate correlations in the earthquake record. This has been utilised in stochastic computational models of earthquake time series, e.g., the Epidemic Type Aftershock Sequence (ETAS) model (Ogata, 1988), which simulate the earthquake record as a random background Poisson process from which aftershocks are stochastically produced in a branching process until no more aftershocks are produced. Such schemes imply that non-nearest neighbour events in the earthquake record may be correlated; consequently inter-event times may also be correlated. In order to investigate the effect of these correlations on the convergence of mean inter-event time we analyse the sequential catalogue and a randomised

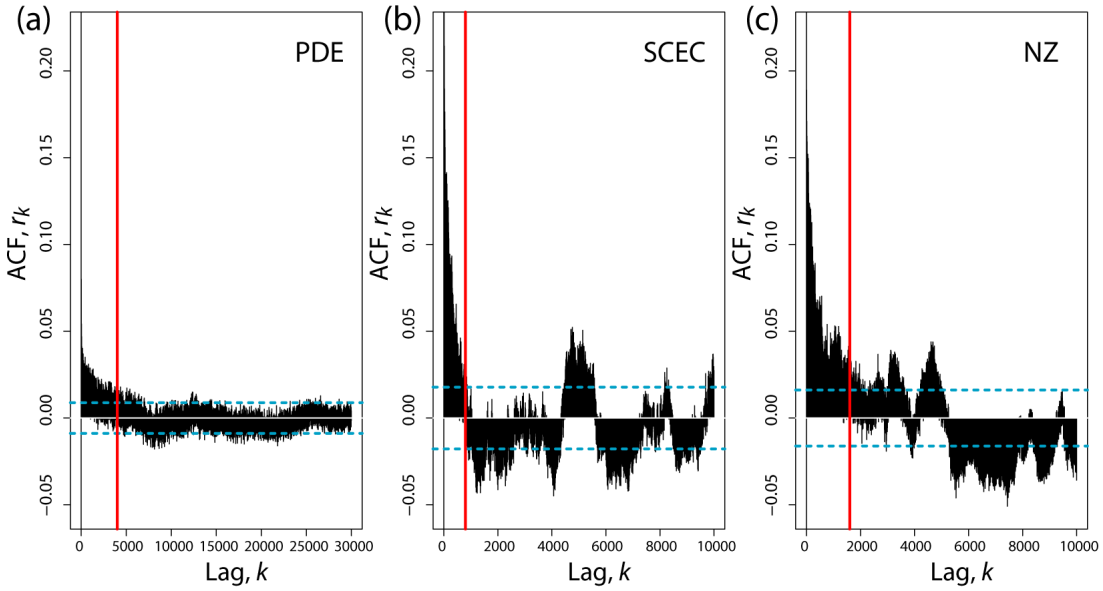


Figure 3.6: Autocorrelation function for the PDE, Southern California, and New Zealand catalogues, plotted as histogram-style lines. The horizontal dashed blue lines show 95% confidence limits for a random process. The vertical red line shows where we have chosen to have the correlations decayed; the results are insensitive to the precise choice of this point. The correlations in the PDE catalogue are weaker but longer-lasting than those in the Southern California or New Zealand catalogues.

version in which the inter-event times are randomly shuffled to remove temporal correlations.

In order to compare the sequential and random inter-event times across different catalogues, we must define appropriate length subsets of the main catalogues such that the subset length is a multiple of all of the sample lengths. We therefore choose to analyse subsets that are multiples of 2 from an initial sample size of 1 up to 4096, i.e., $N \in (1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096)$. Table 3.1.2 lists how many times samples of size 4096 can be drawn from each catalogue, N_{4096}. We then extract a sequential subset from the catalogue of length $N_{Sample} = 4096 \times N_{4096}$ from which we calculate random and sequential sample mean distributions. The disadvantage of this approach is that we are not using all available data in the time series. The advantage is that we ensure that we can make a fair comparison of the different catalogues in order to substantiate the conclusions of this paper and develop a technique for calculating an effective error, and predicting its convergence.

For the case of the sequential catalogue, sample means are calculated for each sample length N using consecutive inter-event times (figure 3.3). The number of events in each distribution varies with the sample size and is given by N_{Sample}/N. Our choice of N and N_{4096} also ensures that this is an integer.

For the case of the random catalogue, sample means are calculated for each sample length N using the same catalogue subset as was used in the sequential analysis but the sample selection is now random without replacement, a process equivalent to randomly shuffling the catalogue. This ensures that the mean of the random and sequential distributions is the same.

Typical variations in the sample mean distributions for the random and sequential cases for representative values of N are shown for the NZ, SCEC and PDE catalogues in figures 3.7,

3.8 and 3.9, respectively. The randomly processed sample means visually agree well with the predictions of the central limit theorem (compare figures 3.7–3.9 with figure 3.4). However, the sequential sample mean histograms converge more slowly than $1/\sqrt{N}$, and do not reach this limit even at $N = 4096$. This is very far from the $N > 10$ 'rule of thumb' for Gaussian convergence referred to earlier.

We quantify the rate of convergence to the Gaussian, as we did in figure 3.5, by taking the standard deviation of the distribution of sample means about the distribution mean as a function of sample size (figure 3.10a). The sequential processed rates (figure 3.10a, red lines) converge significantly slower than the randomised rates (figure 3.10a, black lines). The local gradients of figure 3.10a are shown in figures 3.10b–3.10d to estimate the observed rate of convergence. The randomised data consistently fluctuates about the null hypothesis rate of $\eta = -0.5$. In contrast to the randomised case, the sequential rates of convergence are significantly slower (shallower slopes) and vary with sample size. In some, there is a tendency toward the central limit (steepening slope) for the largest sample sizes.

In relative terms, the sequential PDE catalogue converges faster than either the sequential New Zealand or Southern California earthquake catalogues for a given value of N (figure 3.10a) most likely because it is dominated by a greater proportion of independent events (Huc and Main, 2003). However, it appears that for the largest samples the local gradient for the PDE catalogue is still rising (figure 3.10b) in contrast to the NZ and SCEC catalogues where the rate has stabilised or is falling toward the value predicted by the central limit theorem (figures 3.10c and 3.10d). Thus Gaussian convergence ($\eta \to -0.5$) need not occur at a fixed value of N for different regions.

We interpret this difference between catalogues as a finite size effect that is dependent on the geographic extent of the catalogue subset being analysed. The PDE catalogue has global coverage; therefore in the short term there are fewer correlations than the regional catalogues. This increased independence of the events in the PDE catalogue would be expected to produce a faster convergence compared to regional catalogues that contain a larger fraction of dependent events.

3.1.4.2 Dependence of convergence rate on magnitude cutoff

We perform the analysis using subset catalogues with different magnitude cutoffs for the random and sequential sampling. The randomly sampled models show no statistically significant variation in the rate of convergence for any of the catalogues as the magnitude cutoff is varied (figure 3.11). The growing errors at longer sample lengths derive from the increase in counting errors.

The sequentially sampled New Zealand and Southern California catalogue show little variation in convergence rate with magnitude cutoff (figures 3.12a–3.12d). In contrast, the global PDE catalogue experiences faster convergence as the magnitude cutoff is raised (figures 3.12e and 3.12f). This finite size effect arises because a significantly higher proportion of events are correlated in the regional catalogues (see sections 3.1.2 and 3.1.4.1). Where a smaller proportion of events are correlated in the PDE catalogue, increasing the magnitude threshold significantly reduces the number of correlated pairs since the larger events are more likely to be main shocks.

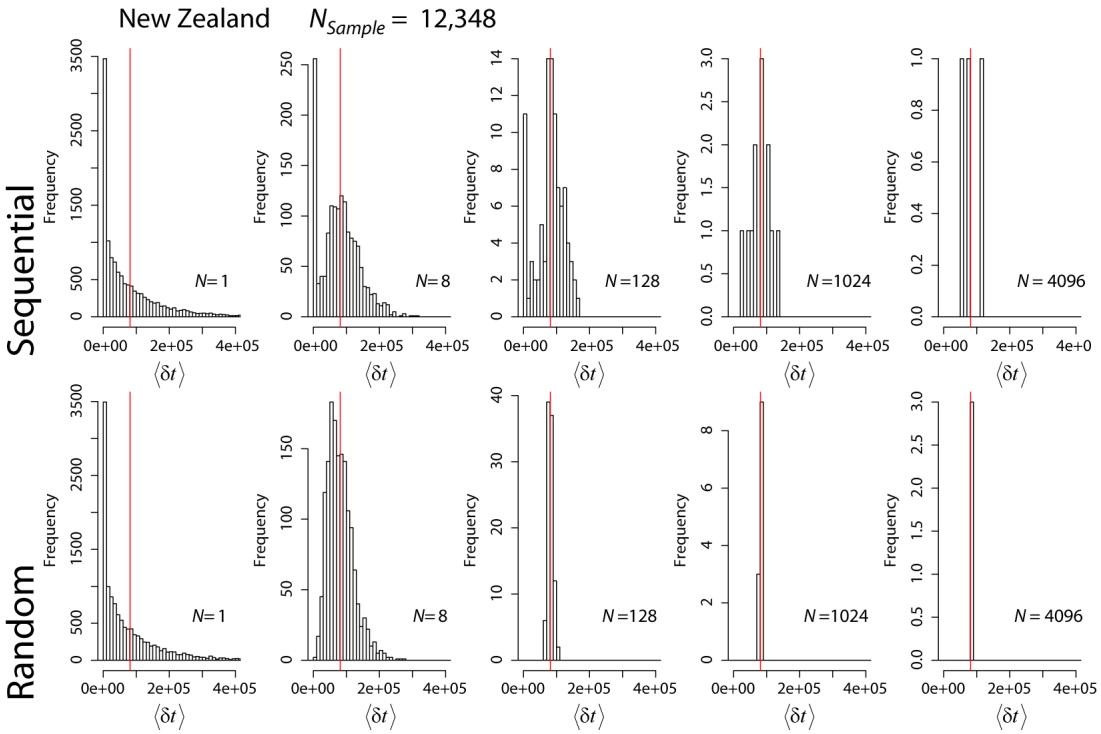


Figure 3.7: Examples of how the distribution of sample means varies as the sample length N tends to infinity for the (top) sequentially and (bottom) random processed data in New Zealand.

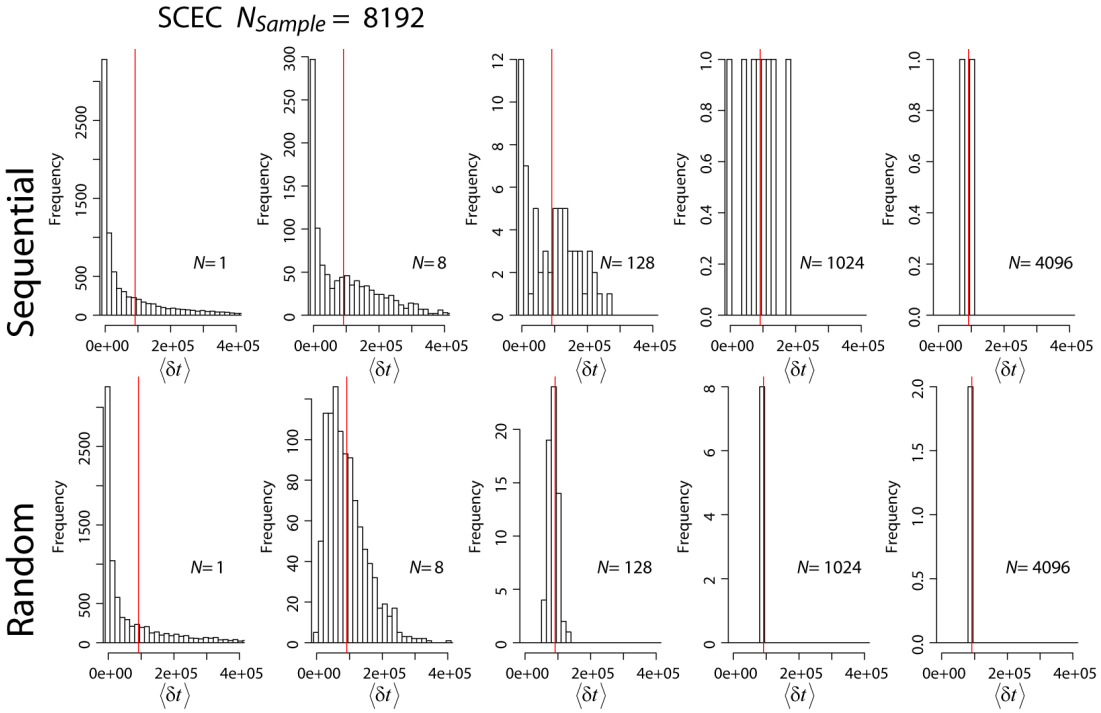


Figure 3.8: Examples of how the distribution of sample means varies as the sample length N tends to infinity for the (top) sequentially and (bottom) random processed data in Southern California.

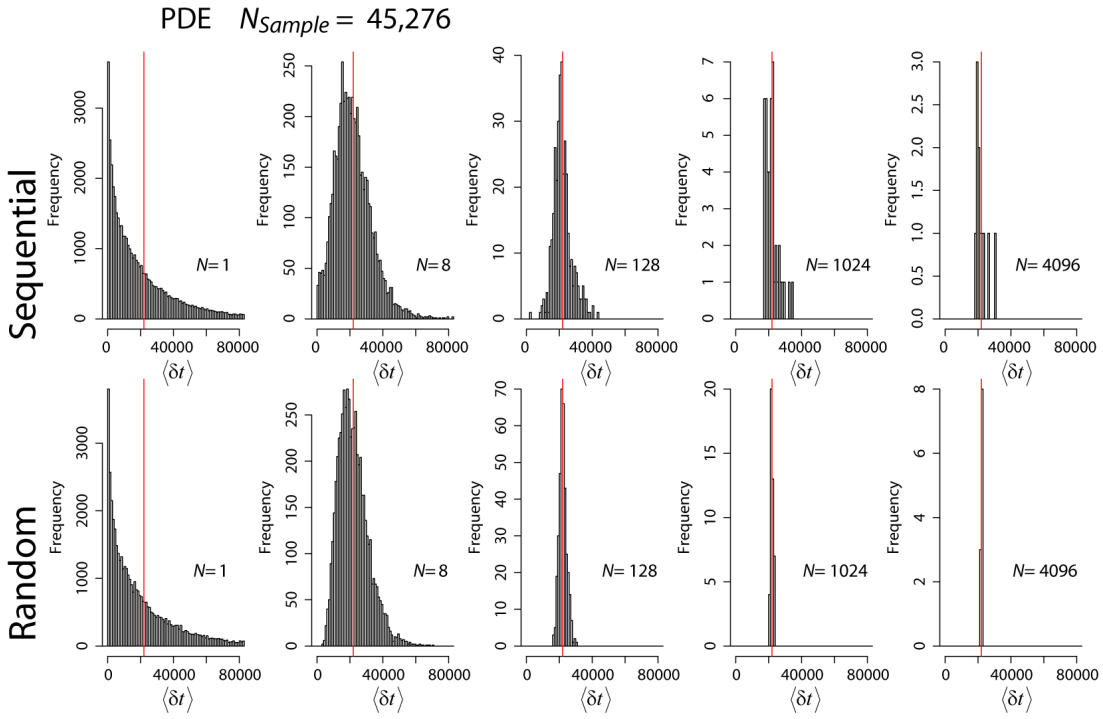


Figure 3.9: Examples of how the distribution of sample means varies as the sample length N tends to infinity for the (top) sequentially and (bottom) random processed data in the global PDE catalogue.

3.1.4.3 Investigating the effect of correlations using the ETAS model

In order to test whether typical stochastic earthquake event models capture the non-Gaussian convergence of mean inter-event times observed in natural catalogues, we also investigated convergence in the Epidemic Type Aftershock Sequence (ETAS) model (Ogata, 1988). The ETAS model applies a conditional intensity function to seed aftershock events. The key two components of the model are a background Poisson rate, μ and the generation of aftershocks conditional on past events. The aftershock model incorporates several empirical seismological relationships:

1. Gutenberg–Richter law through the b-value which describes the distribution of observed seismic moment as a power law.

2. The modified Omori's law which empirically describes the number of triggered events following a main shock as a power law in time,

$$n(t) = \frac{K}{(c+t)^p} \tag{3.10}$$

   where K and c are constants. Increasing p increases the rate at which aftershocks decay. This defines the parameter $A = n(t=0) = K/c^p$.

3. The productivity law weights the triggering contribution of any event by its magnitude as $e^{\alpha m}$.

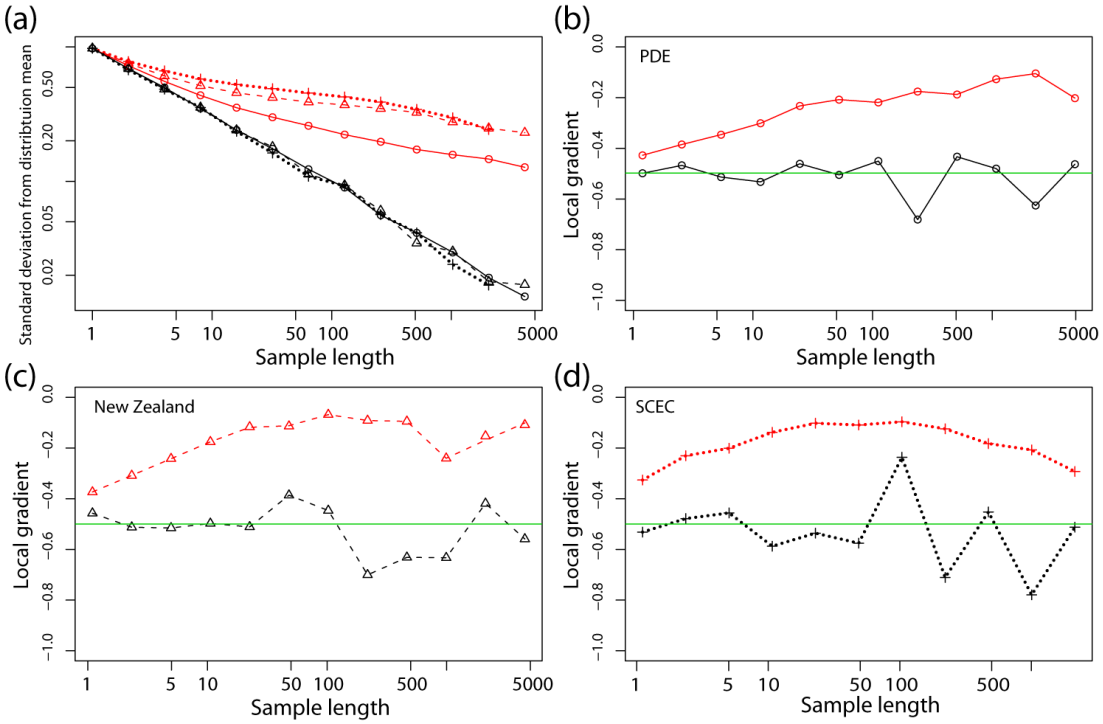


Figure 3.10: Standard deviation of the sample mean distribution from the mean of the parent distribution with varying sample length for (black) randomly sampled and (red) sequentially sampled earthquake data for (circle) PDE, (triangle) New Zealand, and (cross) SCEC. The randomly sampled data agree with the null hypothesis presented by the central limit theorem of convergence (see figure 3.4). All of the sequential data converge slower than predicted by the central limit theorem.

These relations are combined in the conditional intensity function by summing over the history of past events, H_t at times t_i as

$$\lambda(t|H_t) = \mu + K \sum_{i:t_i<t} \exp(\alpha(m_i - m_0))(t - t_i + c)^{-p} \tag{3.11}$$

where m_i is the magnitude of the past event and m_0 the lower magnitude cutoff for the generated sequence.

Stability in the event rate is governed by the branching ratio, the average number of aftershocks per event, which is given by (Sornette and Werner, 2005a)

$$n = \frac{Ac}{p-1}\frac{\beta}{\beta-\alpha}\frac{1-\exp(-(\beta-\alpha)(m_{max}-m_0))}{1-\exp(-\beta(m_{max}-m_0))} \quad \text{for } p>1 \tag{3.12}$$

where $\beta = b\ln(10)$ and m_{max} is a maximum magnitude cutoff. If we do not impose a magnitude cutoff, we get

$$n = \frac{Ac}{p-1}\frac{\beta}{\beta-\alpha} \quad \text{for } p>1, \beta>\alpha \tag{3.13}$$

The condition for stability is that n should be finite and less than 1. Firstly, to have a finite n we require $p > 1$ if $\beta > \alpha$ (in the absence of a magnitude cutoff); secondly, the parameters A, c, p, α and β must have values that combine to give $n < 1$ (Sornette and Helmstetter, 2002).

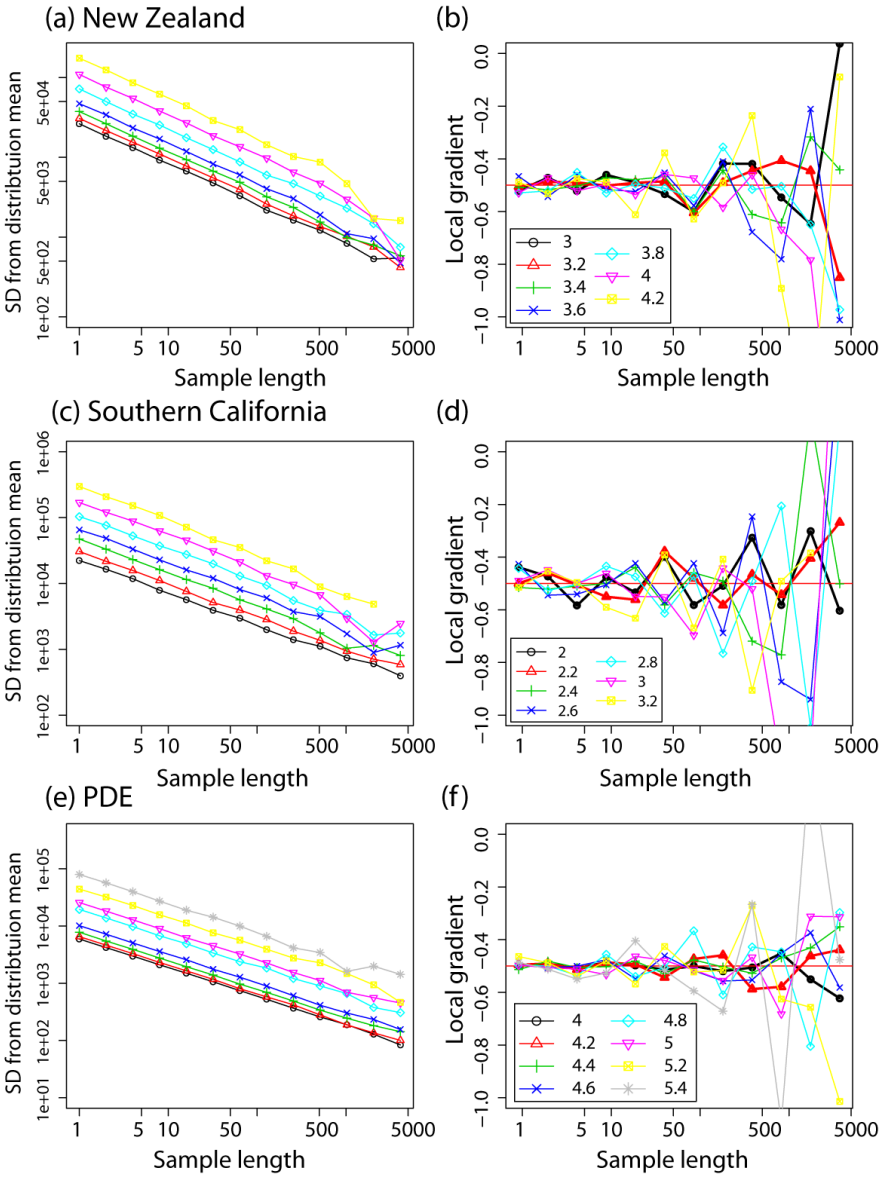


Figure 3.11: Rates of convergence of mean inter-event times for randomised catalogues with varying magnitude cutoff.

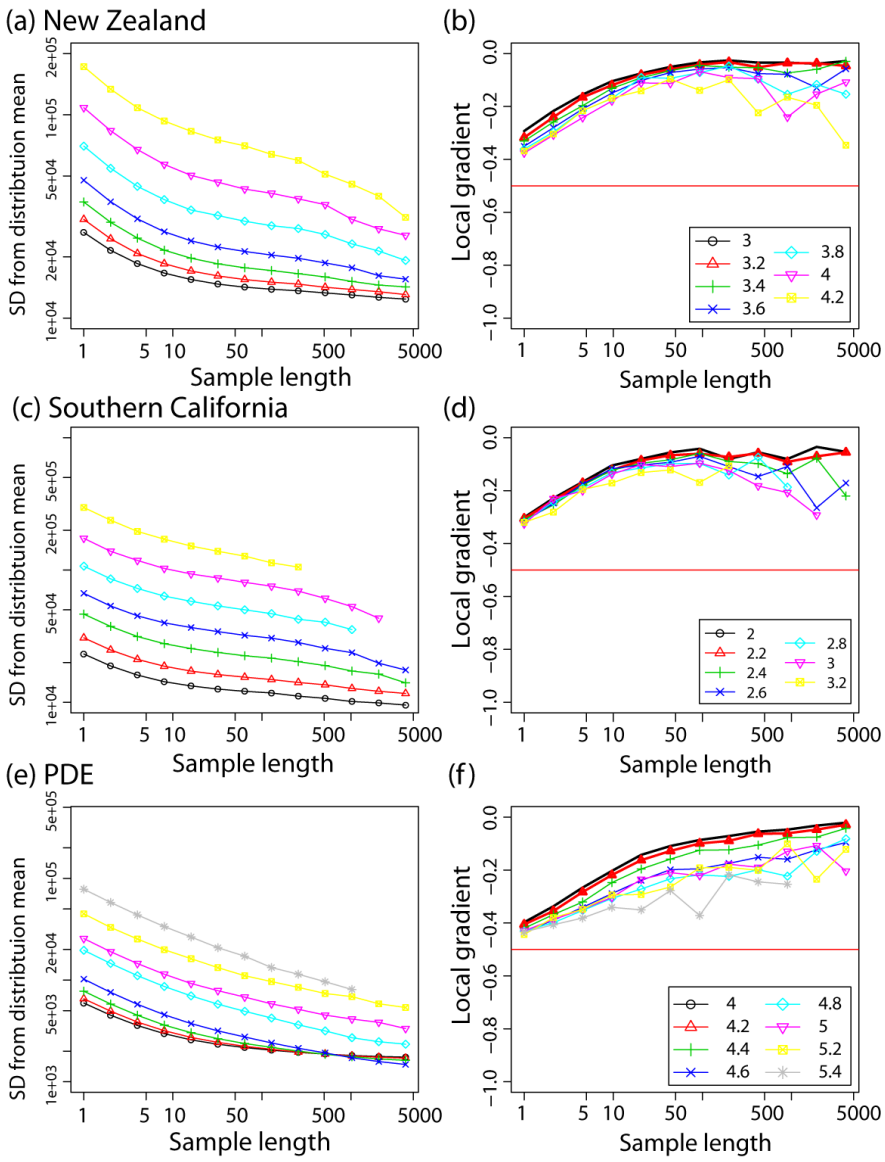


Figure 3.12: Rates of convergence of mean inter-event times for sequentially sampled catalogues with varying magnitude cutoff.

Extrapolating the record

We generated a set of 5 ensemble long synthetic catalogues using the ETAS model to investigate how earthquake inter-event times converge for longer records than are available in current earthquake catalogues, where the parameterisation is chosen such that the inter-event time distribution is known to be stationary over the time frame investigated. Analysis was only performed on data after initial transients associated with the model run-in phase had decayed. Since we are interested in the generic behaviour of earthquake-like inter-event correlations we choose a representative parameterisation of the ETAS model, rather than trying to simulate a specific catalogue, with $b = 1, \mu = 0.5, A = 10, \alpha = 0.9, c = 0.01, p = 1.2$. These values represent typical ones in the middle of the range of reported values in the literature, and correspond to a branching ratio of $n = 0.821$, thus the ETAS model is stable. The synthetic catalogues generated here correspond to a 430-year simulation length. On average these contain 363747 events, much more than in any real catalogue. These events range over 6 orders of magnitude above an arbitrary magnitude threshold. Since this record is substantially longer, we now investigate the behaviour up to and including sample means containing $N = 32768$ events.

The qualitative variation of the histograms produced using the randomly and sequentially sampled synthetic ETAS time series (e.g., figure 3.13) are comparable to that for the real earthquake catalogues analysed (e.g., figures 3.7 and 3.9).

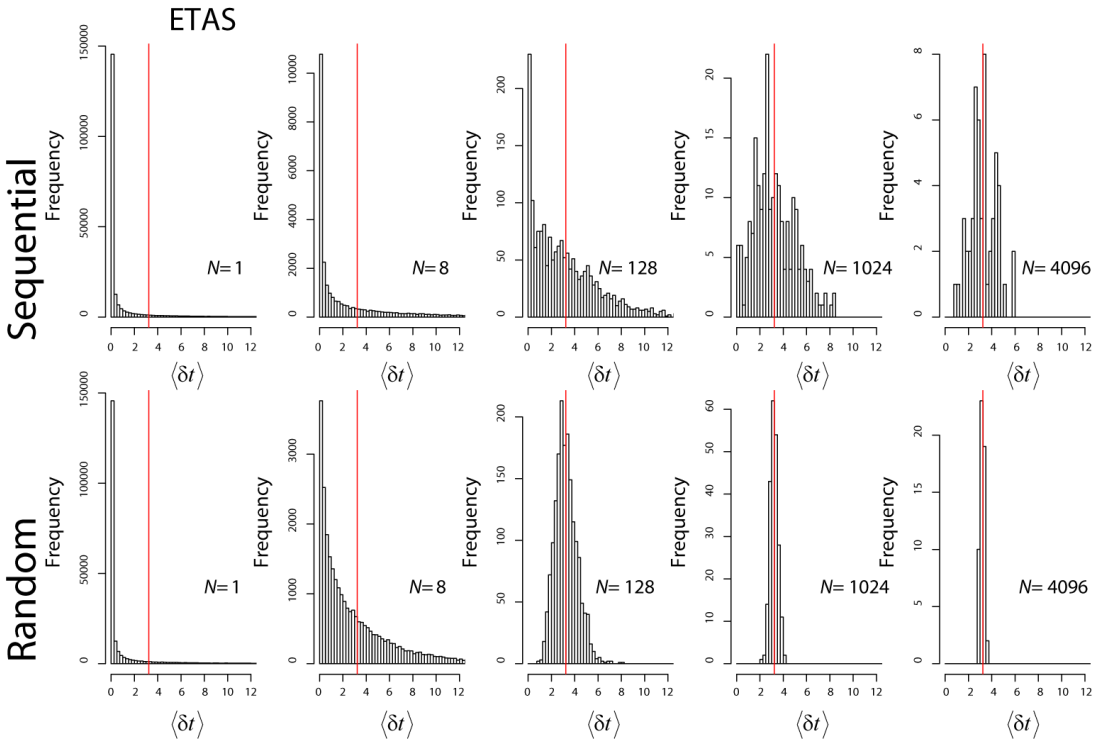


Figure 3.13: Examples of how the distribution of sample means varies as sample length N tends to infinity for (bottom) random and (top) sequential data derived from the ETAS model.

The rates of convergence of the sample mean are shown in figures 3.14a and 3.14b for five realisations of the ETAS model. The ensemble runs clearly demonstrate the effects of growing counting errors as the sample size increases, making it hard to discern the actual rate of convergence for the largest sample sizes. The ensemble mean of these runs is added

to mitigate the effects of counting errors. Once again, the random model agrees well with the central limit theorem prediction of convergence at a rate of $1/\sqrt{N}$, i.e., the local gradient fluctuates about $\eta = -0.5$ (figure 3.14b). The rates of convergence are slower for the sequential sampling.

Effect of varying the strength of correlations

The ETAS model also allows us to investigate the effect of varying the strength of correlations in the inter-event time series.

Increasing p increases the rate at which aftershocks decay and hence decreases the ratio of dependent to independent events. Thus decreasing p slows the rate of convergence (figures 3.14c and 3.14d). The longer-lived correlations associated with lower p increase the duration of the non-Gaussian convergence as longer event records become necessary in order to increase the proportion of uncorrelated to correlated events to the same degree.

Further, in the model the number of aftershocks produced by an event depends on the magnitude of that event as $e^{\alpha m}$. For higher values of α, there will be a larger number of correlated events (aftershocks) following an event of a given magnitude, and the temporal duration of the sequence is therefore longer assuming the decay rate p is kept constant. This effect increases the average correlation length, but also makes it more variable with event magnitude.

Figures 3.14c–3.14f referred to above all show ensemble runs for each of the model parameterisations. Fluctuations between the ensemble runs at a given p or α are less than the differences as p or α are varied, as shown in figures 3.14c and 3.14e. Thus the convergence rate and evolution for a given parametrisation is robust against statistical variability in the ETAS model and provides a technique for validating regional studies that have been fitted to the ETAS model.

Summary of ETAS model convergence

Overall, the ETAS model does a good job in reproducing the convergence observed in real data. It converges at a slower rate than the random case due to the presence of many correlations in the shorter samples. As the sample size increases, the ensemble rate of convergence increases toward $1/\sqrt{N}$. Such convergence depends critically on the sample period being sufficiently long compared to the average time between extreme events that generate many aftershocks, and may not occur at a particular universal value of N. However, simulation using the ETAS model is not very good for predicting errors on the mean inter-event times for large sample sizes because of computational constraints on running enough large-scale simulations to reduce counting errors to a suitable size (see growing fluctuations with sample size for sequential sampling in figure 3.14b, dashed blue lines).

3.1.5 Error prediction using autocorrelation function

The correlogram (plot of the autocorrelation function, r_k defined in equation (3.7) as a function of lag k) for each of the catalogues is plotted in figure 3.6 to quantify the strength and range of inter-event time correlations. In the time domain, r_k identifies that the global PDE catalogue contains weaker but longer range correlations than the NZ and SCEC catalogues (figure 3.6).

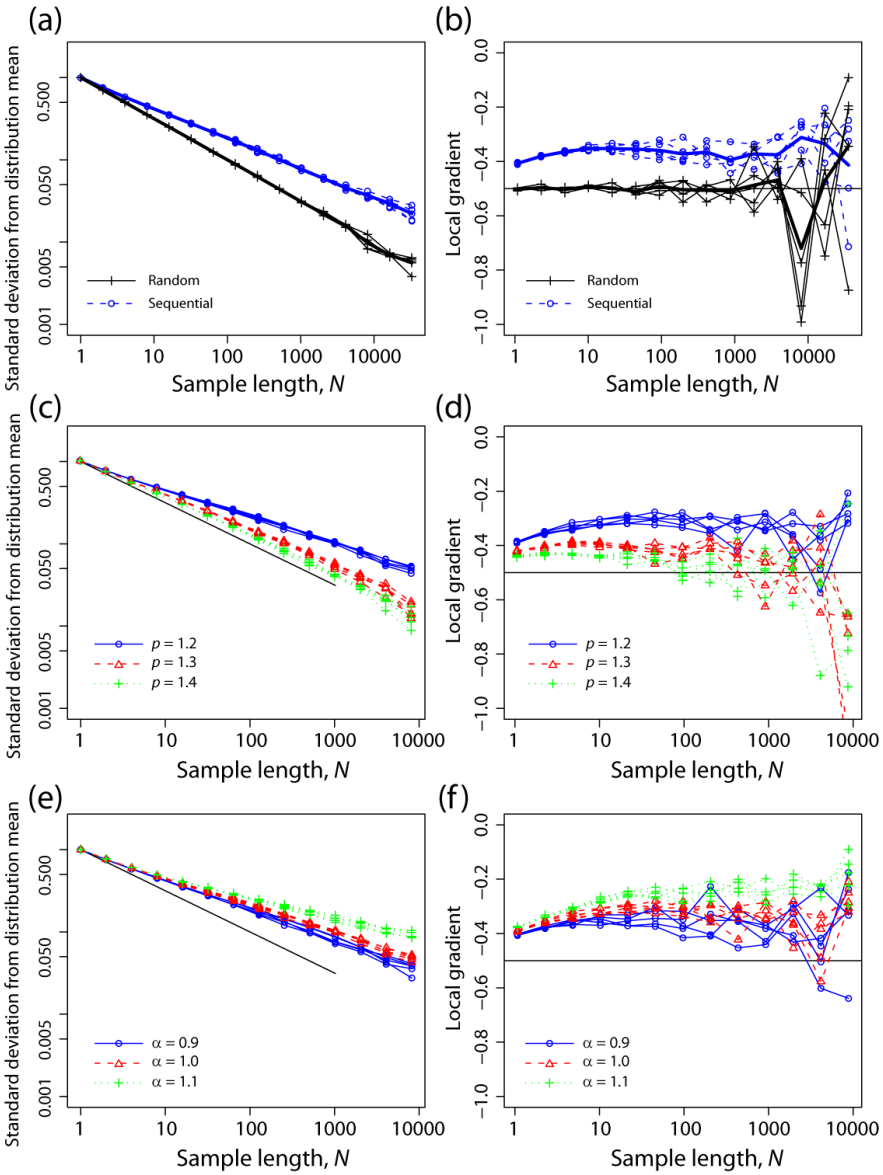


Figure 3.14: (a and b) Rate of convergence of sample mean using five ensemble runs of the ETAS model with the same parameters. The ensemble mean has been added for each set of runs as a thicker line. (c and d) The rate of convergence varies as the Omori decay rate is increased. (e and f) The rate of convergence varies as the dependence of aftershocks on magnitude is increased.

Thus r_k can be used quantitatively to investigate temporal finite size effects. The horizontal dashed blue lines show the 95% confidence limits about 0 based on the assumption that the data is random and uncorrelated (e.g. Zivot and Wang, 2006), thus decaying correlations above the dashed blue line at short lag times can be accepted with this level of confidence. Below these confidence limits the correlations are not statistically significant. As the lag is increased counting errors in the autocorrelation coefficients grow, generating spurious fluctuations in the moving average, so we must also crop the useful data, at a lag of length k_c, where the correlations first decay below the confidence intervals. The position of the cropping is slightly arbitrary, due to high-frequency fluctuations in r_k, and is marked on figure 3.6 as the vertical red line. The longer range, almost periodic, moving average diversions above the confidence limits to the right of the cropping point (e.g., figure 3.6b shows several clear excursions above the confidence limits) are artifacts arising from a finite sample size. The autocorrelation function should be used to verify that a correlation length measured from data is real and not just one of these sampling effects.

Summing the autocorrelation function over the significant range of non-zero lags (i.e., summing r_k from $k = 1 : k_c$) gives the total number of events that each event is correlated with, on average. For the catalogues investigated here, each event is on average correlated with 74.0 events for New Zealand over a range of ~ 1600 events, 48.6 events for southern California over a range of ~ 800 events and 49.8 events for PDE over a range of ~ 4000 events. It is important to stress that the results are insensitive to the precise choice of the position of this cutoff. This defines the range of the correlations to be on the order of a few thousand events.

For approximately Gaussian-distributed data with correlations, the effective sample length N' defined by equation (3.8) can be used to define convergence in the Central Limit form as $1/\sqrt{N'}$, which is plotted in figures 3.15a and 3.15b. The standard deviation predicted by the effective sample length compares well with that for the real data (see figures 3.10, 3.15b, and 3.15c). Errors are largest and deviate most from $1/\sqrt{N}$ convergence for small sample sizes in regional catalogues. The errors for all catalogues tend to converge for larger sample sizes. Thus we have demonstrated that we can reproduce the observed rate of convergence using only knowledge of the variance and autocorrelation function defined by the data.

This use of the autocorrelation function to generate an effective sample size as a function of sample length reproduces the observed errors well and allows us to predict how the errors decay at larger sample sizes than are currently available. Further, this analysis motivates a complementary definition of stationarity, i.e., where the autocorrelation function is temporally stable.

3.1.6 Discussion and potential applications

The arithmetic mean inter-event time, proportional to the inverse of the harmonic mean event rate, provides the most stable measure of earthquake activity and converges smoothly. This should be preferred as a measure of earthquake activity over the arithmetic mean event rate which does not converge cleanly due to its sensitivity to extreme events. Skewness in the inter-event time parent distribution makes it more likely that the mean inter-event time will converge from above (or the event rate from below). Thus on average we will underestimate earthquake activity from catalogue samples which have not yet converged. This could be a significant systematic error in many current estimations of time-independent seismic hazard.

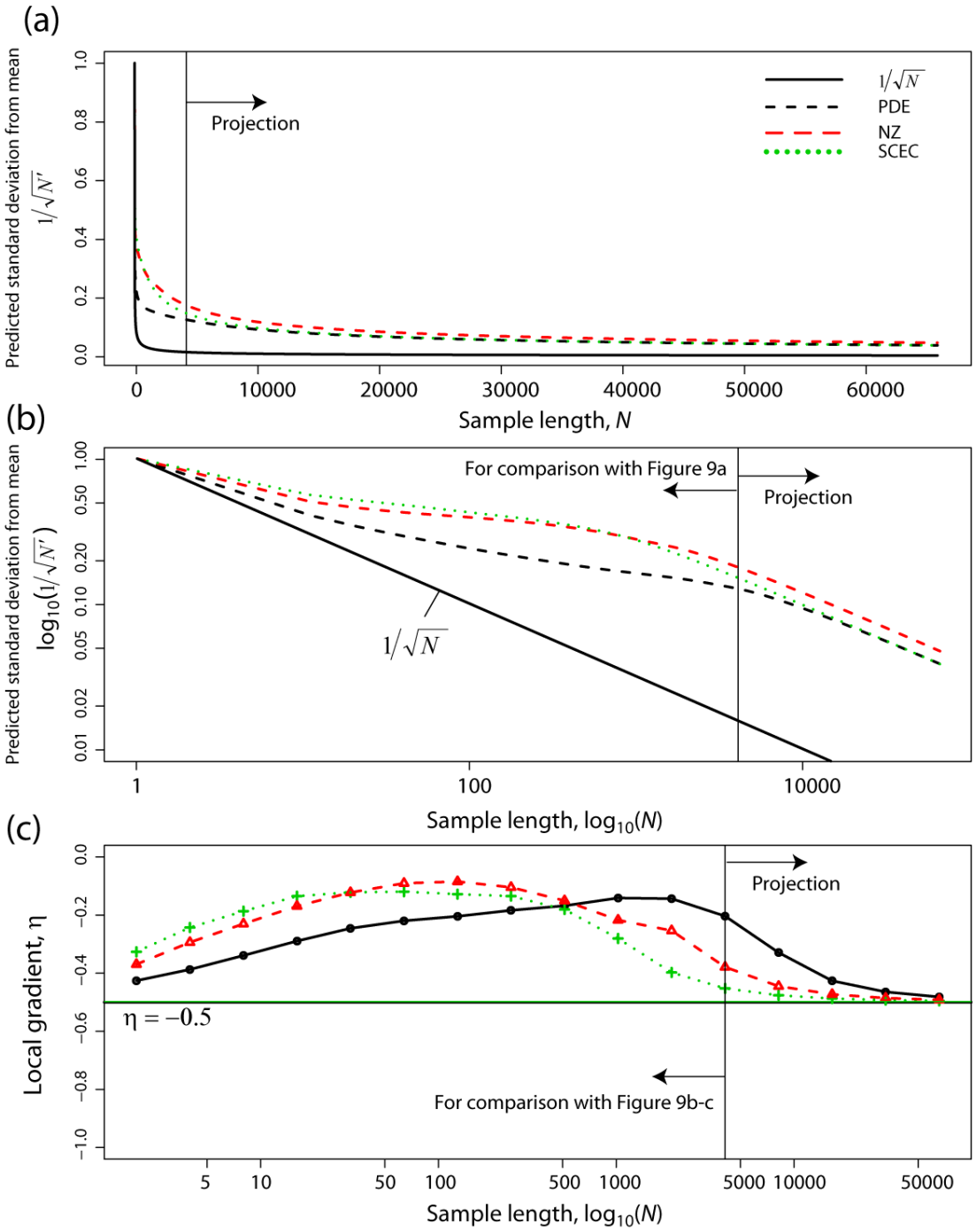


Figure 3.15: (a) Plot comparing $1/\sqrt{N}$ central limit convergence in the absence of correlations with $1/\sqrt{N'}$ autocorrelation-corrected convergence for the PDE, NZ, and SCEC catalogues. To convert these normalised values to actual standard deviations in the mean requires that the plotted values be multiplied by the standard deviation of the underlying data. The presence of persistence in the inter-event time series generates slower convergence. (b) The same as for (a) but plotted on a log-log scale. The vertical black line shows the transition between data that can be directly compared with figure 3.9a and the projection of the effect of the correlations on larger sample sizes. (c) The local gradients of (b) for direct comparison with figures 3.9b and 3.9c. The horizontal line at $\eta = -0.5$ shows the theoretical rate for central limit convergence in the absence of correlations.

The underestimate of random errors on the mean are largest for small sample sizes. Time-dependent hazard assessments typically use small subsets of a catalogue, and so are very vulnerable to this source of statistical noise. Our results suggest time-dependent studies need to demonstrate that observed 'anomalous' trends in activity lie outwith the error bounds expected from sampling a correlated time series. This sampling effect presents an additional source of difficulty in posing statistical models involving aftershock sequences as a null hypothesis to be rejected in evaluating earthquake forecasting power at a higher level, for example, from candidate precursors (Main, 1999).

The autocorrelation function shows that geographically smaller, regional catalogues, such as New Zealand and Southern California, have stronger correlations over a shorter range of lags than the global PDE catalogue which has weaker but longer lasting correlations. Generally, this will translate into larger errors for more localised geographic studies. This result has important implications for high spatial resolution seismic hazard studies, for example those currently being developed for application in catastrophe ("cat") modelling in the earthquake reinsurance market (Grossi and Kunreuther, 2005). Fortunately equation (3.8) can be used to estimate the proportion of correlated events and the effective error in such small samples. It can also be used a posteriori to validate the hypothesis of stationarity in time-independent seismic hazard, i.e., the autocorrelation function is not varying significantly in time.

Many studies use declustering algorithms to remove aftershock events from earthquake catalogues. The technique presented here could be used as an a posteriori check on the quality of the declustering, with ideal declustering defined by a filtered catalogue (of events identified as independent by the algorithm) exhibiting Gaussian convergence, with slower rates η defining a less effective declustering algorithm. Finally the stability of the analysis with respect to varying magnitude cutoff implies a degree of self-similarity with respect to the strength and range of correlations in the inter-event time series, and is an indication of the robustness of the technique.

3.1.7 Conclusions

Hazard assessments should try to quantify errors, but rarely do explicitly. We have shown that the rate of convergence of the mean earthquake inter-event time is slower than the central limit theorem prediction of $1/\sqrt{N}$, primarily due to the presence of correlations in the inter-event time series. As a consequence many studies of seismic hazard (time independent or time dependent) currently underestimate the true uncertainty in mean and standard deviation of parameters such as event rate or inter-event time.

We have presented a simple technique to quantify errors on earthquake mean inter-event times using only the variance of the data and the autocorrelation function of the inter-event time series. Specifically, the autocorrelation coefficients, r_k can be used to define an effective sample size which corrects the sample size for the number of events that it is likely to be correlated to, which changes as a function of sample size, such that $N' = N/(1 + 2\sum_{k=1}^{N}(1 - \frac{k}{N})r_k)$. This effective sample size can be used to estimate the rate of convergence of the earthquake inter-event times as a function of sample size as $1/\sqrt{N'}$.

Notation

Statistical variables:

$x_1, x_2, ...$	A sequence of random variables
N	Number of random variables/events in sample
$\mu, \bar{x}$	Mean of a sample
σ	Variance of a sample
CV	Coefficient of variation
η	Local rate of convergence

Earthquake analysis specific variables:

δt	An inter-event time
$\langle \delta t \rangle$	A sample mean inter-event time
N_{events}	Number of events in an earthquake catalogue subset
N_{4096}	Whole number of samples of length 4096 that fit into catalogue subset

Gamma function properties:

$f(\delta t) = C\delta t^{\gamma-1} \exp(-\frac{\delta t^{\delta}}{B})$	The generalised gamma function
γ	Shape parameter
B	Scale parameter
C	Normalisation factor

Autocorrelation and effective sample size:

k	Lag between events
c_k	Autocovariance at lag k
r_k	Autocorrelation at lag k
N'	Effective sample size estimating the number of independent pieces of data in a sample

ETAS parameters:

b	Gutenberg–Richter b-value
μ	Background rate
α	The exponent relating the production of aftershocks as a function of magnitude
p	Omori's law power, describing the decay rate for aftershocks
$A = n(t=0) = K/c^p$	Occurrence rate of earthquakes in the Omori's law at zero lag from an event in equation 3.10

Chapter 4

Origin and Non-Universality of the Earthquake Inter-Event Time Distribution

This chapter describes one of the most important component parts of my thesis. In the paper that follows, I draw attention to the bimodality in natural tectonic inter-event time distributions, which was alluded to in the Introduction (chapter 1), and clarify its origin in terms of the two categories of inter-event times—correlated and uncorrelated. I show using the end-member case of the whole world that the distribution does not take on a simple universal form; the variation evades a data collapse under rescaling by the mean event rate as done by Corral.

I developed the idea of correlated and uncorrelated inter-event times through thinking about how aftershock sequences overlap temporally, while working on the paper in the previous chapter. I wrote a simulation algorithm for the ETAS model which records the family structure of triggered events (see Methods, section 2.3) and used it in carrying out all the work for the paper, with the exception of figure 4.5 which was done by Mark Naylor. I acted as lead author in the paper.

The plotting of the histograms is described in Methods, section 2.5; in particular section 2.5.1 describes the procedure for plotting correlated and uncorrelated subsets of data, in figure 4.7. I plot the histograms in two different ways: as a simple histogram with logarithmic bins, and as a normalised probability density function created by dividing the counts by the bin widths; there is a difference in slope of 1 between these two types of plot (see figure 4.4). The latter method was used by Bak et al. (2002), Corral (2003), and Davidsen and Goltz (2004); it is worth noting, though, that due to the different ways of rescaling the y-axis (using the variable inter-event time versus the average), Bak's and Davidsen's distributions (figures 4.1 and 4.2 respectively) have a slope of around 1 less than Corral's (figure 4.3).

4.1 Paper

Many authors have modelled regional earthquake inter-event times using a gamma distribution, whereby data collapse occurs under a simple rescaling of the data from different regions or time periods. We show, using earthquake data and simulations,

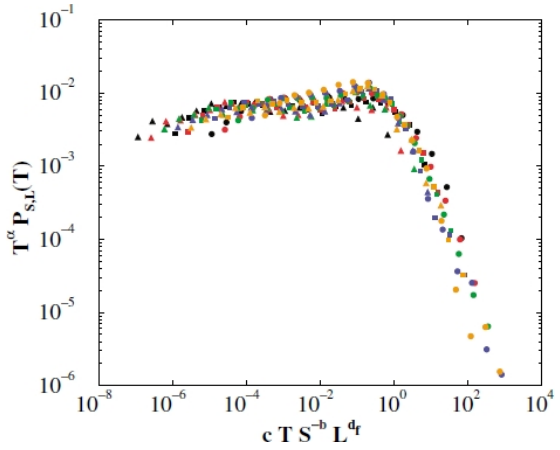


Figure 4.1: Figure 4 from Bak et al. (2002). Each histogram is a subset of California data with different magnitude threshold and different spatial area size, rescaled as shown by the axes and explained in section 1.4, and showing only inter-event times > 38s.

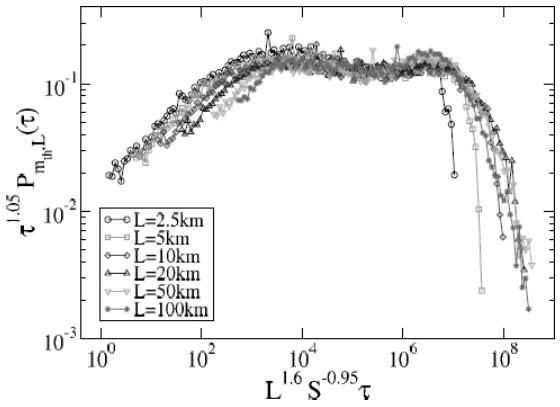


Figure 4.2: Figure 2 from Davidsen and Goltz (2004). Subsets of California data with different spatial area sizes as shown in the legend, rescaled as in figure 4.1.

that the distribution is fundamentally a bimodal mixture distribution dominated by correlated aftershocks at short waiting times and independent events at longer times. The much-discussed power-law segment often arises as a crossover between these two. We explain the variation of the distribution with region size and show that it is not universal.

Over the past decade, much scientific attention has been focused on the distribution of waiting times between earthquake events within a region or catalogue (Bak et al., 2002; Corral, 2003, 2004; Davidsen and Goltz, 2004; Shcherbakov et al., 2005; Molchan, 2005), whose form and origin have great importance for the development of physical or statistical models of earthquake dynamics. Most authors have explored these inter-event times by fitting the empirical histograms to a gamma distribution, which has led to the suggestion of universality (Bak et al., 2002; Corral, 2003, 2004; de Arcangelis et al., 2008; Lennartz et al., 2008), even down to the scale of laboratory rock fracture experiments (Davidsen et al., 2007). The idea that has enthusiastically been pursued is that a rescaling involving region size and magnitude cutoff (Bak et al., 2002), or simply the mean event rate (Corral, 2003), produces data collapse

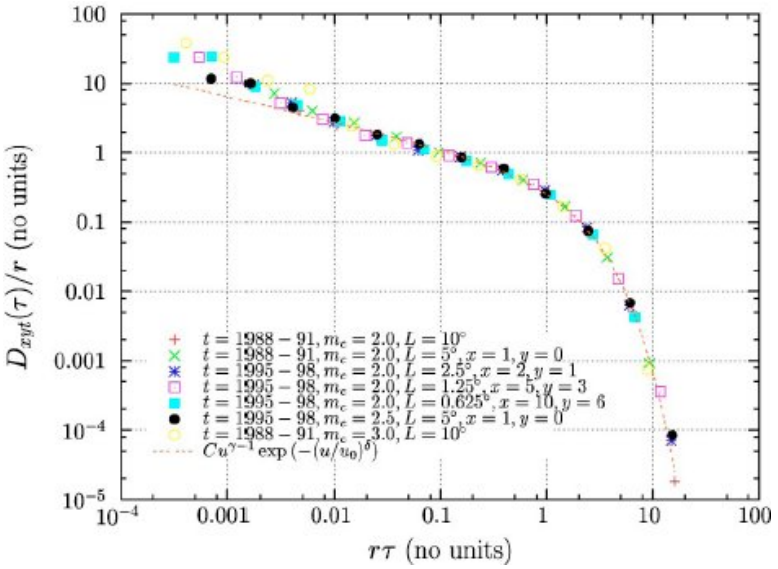


Figure 4.3: Figure 2 from Corral (2003). Subsets of California data as described in the legend, filtered to obtain 'stationary' periods, rescaled by the event rate, and showing only inter-event times > 38s.

onto a universal gamma distribution. Much of this analysis, however, has been restricted to the use of data with an apparently stationary event rate (e.g. Corral (2003); Lennartz et al. (2008)), since mean event rates are otherwise poorly constrained; there is some evidence that non-stationary data analysed in the method of Bak et al. do not collapse (Carbone et al., 2005). Further, it is common also to omit the shortest inter-event times from the analysis (Bak et al., 2002; Corral, 2003); if these inter-event times are included, the approximate data collapse is broken (Davidsen and Goltz, 2004). These common restrictions tend to filter out high rates of aftershocks and thereby introduce a strong selection bias in the data, leading us to question the proposed universality. Recent analytic studies on the Epidemic-Type Aftershock Sequences (ETAS) model, a process-based stochastic earthquake occurrence model (Ogata, 1988), have indicated that the inter-event time distribution is not universal, but may be approximately universal under some circumstances (Saichev and Sornette, 2006, 2007).

We first explore afresh the structure of global and regional earthquake inter-event time series, then demonstrate how the observed range of inter-event time distributions arise from well-known empirical laws of seismicity by presenting comparative results from simulations of the ETAS model. We show that the form of the inter-event time distribution in both earthquake catalogues and simulations is generally bimodal, and is best described as a mixture distribution, formed by the different patterns associated with correlated and uncorrelated event pairs. The frequently cited power-law segment in the distribution (Bak et al., 2002; Corral, 2003; Davidsen and Goltz, 2004) arises as a crossover between the two peaks of these distributions. Support for bimodality in the distribution can also be found in published literature involving real data (Davidsen and Goltz, 2004; Molchan and Kronrod, 2007), simulations (Lindman et al., 2005), and analytic studies on the ETAS model (Sornette et al., 2008), although no comment has, up to now, been made on the bimodal shape. We draw attention to it and present a physically-motivated, intuitive explanation for its origin.

Simulations also provide a way around the stationarity problem and allow us definitively to

reject the hypothesis of universality. While on short timescales the earthquake rate is inherently non-stationary, earthquakes in the long term are a stationary, albeit non-linear process; their event rate converges slowly but definitely onto a well-defined value (Naylor et al., 2009). Using lengthy ETAS simulations we are able to include strong aftershock activity in our inter-event time sequence and still confidently define a mean event rate. This is akin to using a small region (such as Southern California) but recording data for a very long time. As expected from the bimodality, the distributions only approximately rescale with the mean event rate when including a wider range of realistic seismic patterns in the analysis than is available currently in real catalogues (Naylor et al., 2009; Main et al., 2008).

Inter-event time histograms for a global and a regional (Southern California) catalogue are shown in figure 4.4. Our preferred way for plotting the data is shown in the left hand figures in which a clear distinction can be seen between the global data, which forms a single-peaked distribution, and the regional distribution, which appears more bimodal. More commonly, inter-event time distributions are plotted in the form of the right hand figures where each frequency has been normalised by the bin width, which tends to smooth out the two bumps so they cannot readily be distinguished as such (figure 4.4(b) and (d)). They can however be seen in some rescaled plots in the literature (e.g. Davidsen and Goltz (2004)). It is important to highlight the evidence that these are real features and that the much-discussed gamma distribution is in fact only an approximate description of the inter-event time histogram.

We will now explain these observations using synthetic catalogues generated by the ETAS model. This is a stochastic point-process model in which independent seeding events (sometimes described as background events) occur as a Poisson process in time with constant rate μ, and all past events above a threshold magnitude M_0 may produce aftershocks. The magnitudes of all events are picked independently from the power-law Gutenberg–Richter distribution,

$$\log N = a - bm, \tag{4.1}$$

where N is the number of events in a given time period with magnitude $\geq m$, and a and b are constants. The model incorporates the empirical observation that events tend to cluster in time due to the time-dependent relaxation of the crust by the release of triggered aftershocks, whose rate $n(t)$ decays as a power law in time after a sizeable event according to the Omori law,

$$n(t) = \frac{K}{(c+t)^p}, \tag{4.2}$$

where K, c and p are constants and t refers to time. Aftershock productivity depends exponentially on the magnitude of the parent event, so that larger events trigger a greater number of aftershocks. Combining these relations defines the conditional intensity function λ for the ETAS model,

$$\lambda(t|H_t) = \mu + A \sum_{i:t<t_i} \exp\left(\alpha(m_i - M_0)\right) \left(1 + \frac{t - t_i}{c}\right)^{-p}, \tag{4.3}$$

where t_i are the times of the past events and m_i are their magnitudes. Thus, the five ETAS parameters are μ, the independent event rate, c, p and $A = K/c^p$, the Omori parameters, and α, the productivity parameter. The model does not make an arbitrary distinction between

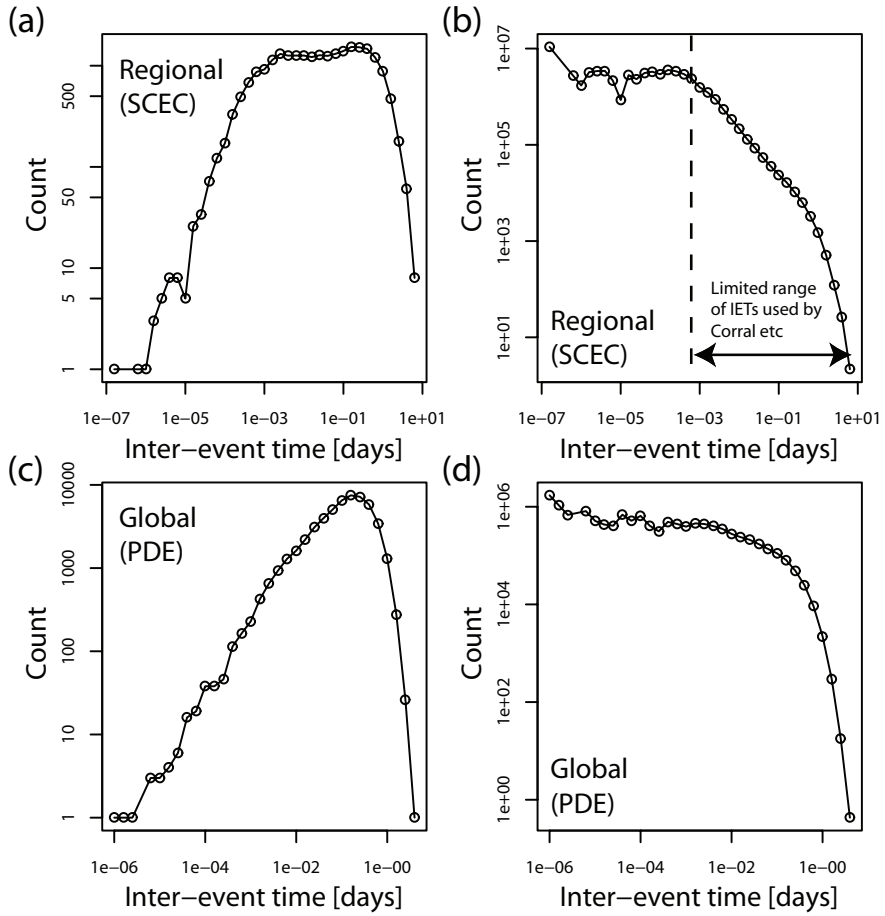


Figure 4.4: Inter-event time histogram for the Southern California Earthquake Center (SCEC) catalogue, plotted (a) without and (b) with normalisation by the bin widths. Events larger than magnitude 2.4 between 1984 and 2000 were used. Inter-event time histogram for the worldwide Preliminary Determination of Epicenters (PDE) catalogue, plotted (c) without and (d) with normalisation by the bin widths. Events larger than magnitude 5.0 between 1970 and 2006 were used.

foreshocks, main shocks and aftershocks, but rather regards all events as capable of triggering further events according to these simple rules. Thus, each independent event may result in a cascade of nested aftershock sequences. We refer to such a cascade as a global aftershock sequence, following Helmstetter and Sornette (2002b).

In order to make comparisons with global and regional earthquake data, we concentrate on the rate of independent events, μ, which is effectively the average frequency for which global aftershock sequences are initiated. This parameter can be considered as a proxy for region size; increasing μ increases the extent to which global aftershock sequences overlap in time (figure 4.5), with the effect that a smaller proportion of the events that follow a given event within a certain time period are correlated with it, as is known to be the case for larger regions (Huc and Main, 2003). Because of the spatial heterogeneity of earthquake occurrence, the actual relation between region size and effective μ will be non-linear, but this is beyond the scope of this paper.

The effect of μ on the synthetic ETAS inter-event time distribution is shown in figure 4.6. For large μ, the shape of the distribution is similar to that of a Poisson process. For

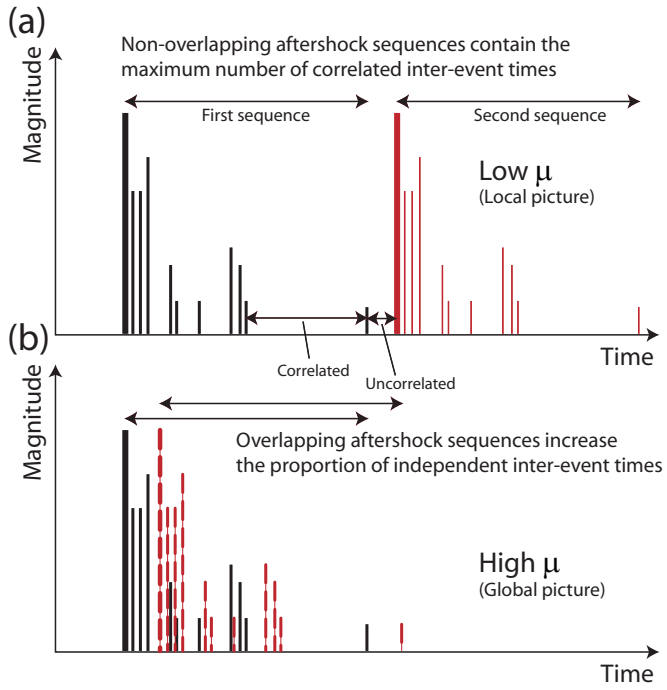


Figure 4.5: Cartoon illustrating how varying the independent event rate changes the proportion of correlated to independent inter-event times as a consequence of the degree of overlap of triggered aftershock sequences. (a) Regional catalogues tend to resolve aftershock sequences; (b) Because of the increased independent event rate, global catalogues tend to contain more temporal overlapping and thus more independent inter-event times.

intermediate values of μ, the peak of the distribution flattens out and we see the familiar gamma distribution. When μ is made very small, however, the straight segment becomes peaked in two places (figure 4.6(a)). Thus, as we decrease μ, we progress from a unimodal to a bimodal distribution, comparable to the progression from global to regional scale in real data (compare figure 4.6(a) with figure 4.4(a) and (c)).

The ETAS model allows us to perform further analysis since we have knowledge of whether each event is independent or triggered. We may therefore categorise each inter-event time as either correlated, defined as occurring between two events belonging to the same global aftershock sequence, or uncorrelated, occurring between events of different aftershock sequences (see figure 4.5). Figure 4.7 shows these correlated and uncorrelated subsets superimposed onto the histograms, for various different values of μ. It is clear that the complicated distribution arises as the sum of two physically motivated distributions which have much simpler forms. The uncorrelated waiting times are exponentially distributed, as expected for independent events. The distribution of correlated waiting times, shown most clearly in figure 4.7(d), has three segments: an exponential increase towards a peak at short inter-event times (which is omitted from many published distribution plots), a power-law decay segment, and an exponential tailoff at longer times.

The variation of μ produces a variation in (i) the position of the tailoff in the correlated distribution, and (ii) the relative sizes of the two subsets. This effect arises from the interference of temporally overlapping aftershock sequences (figure 4.5). As μ is increased, fewer aftershocks are allowed to occur before each sequence is 'interrupted' by the onset of a new global sequence;

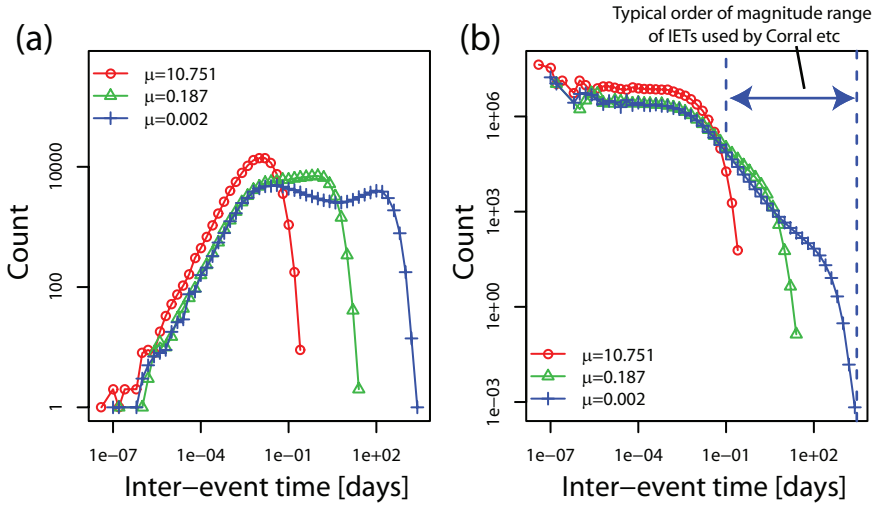


Figure 4.6: Synthetic inter-event time histograms for three ETAS simulations with μ values as indicated, plotted (a) without and (b) with normalisation by the bin widths. These few examples demonstrate the range of behavior observed in real earthquake inter-event time distributions (figure 4.4). The other ETAS parameters used were $A = 10$, $\alpha = 1$, $c = 0.01$ and $p = 1.2$.

the sequence of course continues, but its power-law signature in the inter-event time series does not. The range of shapes arising from combinations of these two simpler forms means that we are able qualitatively to describe the distribution of earthquake inter-event times as a function of a single reduced parameter: the ratio of consecutive independent to dependent events.

The approximate gamma function used to model the overall inter-event time distribution is not directly related to the gamma distribution of correlated events; the effective power-law exponent for the former would depend on the relative heights of the two distributions. This exponent is therefore not directly related to the Omori p parameter, but would be expected to change with the other parameters as well in a complicated way.

Since it has become clear that the distribution is not strictly a gamma distribution, the scaling relation proposed by Corral obviously breaks down (figure 4.8). Note that we have only considered variations of the seeding rate μ without changing the branching ratio. Given that the rate of exponential decay at long inter-event times is equal to the fraction of independent events in the catalogue (Hainzl et al., 2006), and that the power-law slope for correlated events depends on the parameter p (Lindman et al., 2005; Hainzl et al., 2006; Jonsdottir et al., 2006), changing the other parameters would be expected to further degrade the data collapse. Raising the lower magnitude cutoff would also alter the shape of the distribution somewhat.

Using the ETAS model and varying the rate of independent events, μ, has allowed us to explore the range of temporal seismicity patterns that result from changing the region size. We have shown that the inter-event time distribution is best described as a mixture distribution. It arises as the sum of two contributions: gamma-distributed waiting times between correlated event pairs, those belonging to the same aftershock sequence, at short waiting times; and exponentially distributed times between uncorrelated events at longer times. The larger the region

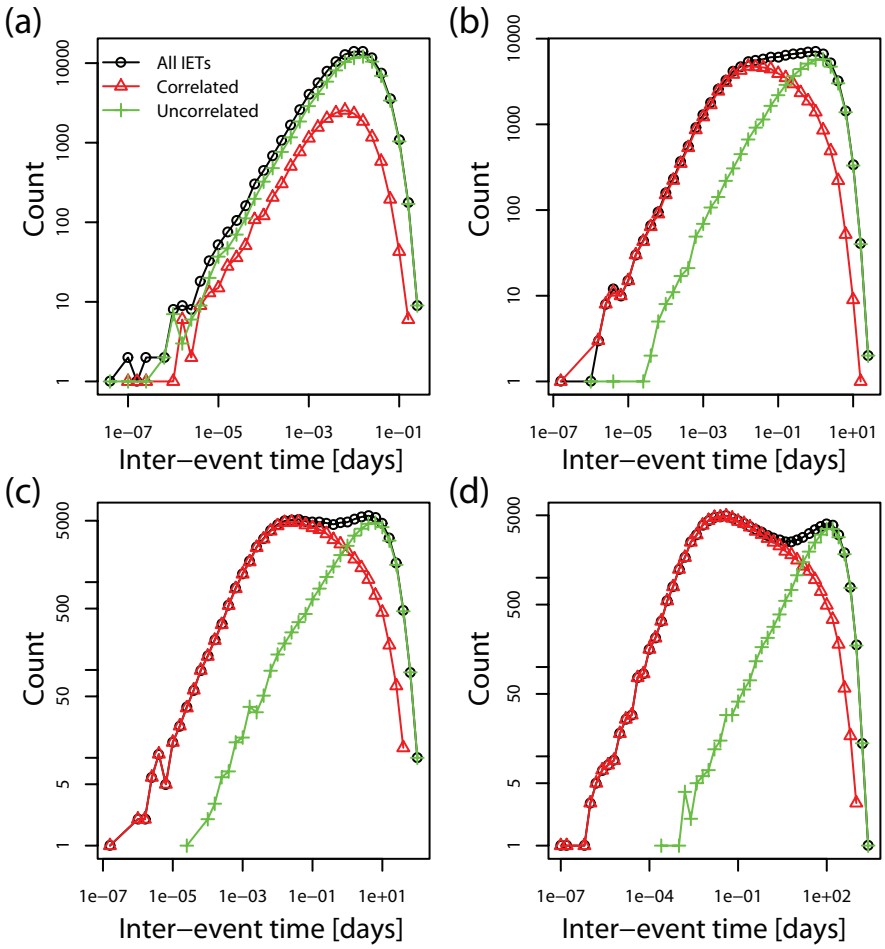


Figure 4.7: Inter-event time histograms for four ETAS simulations with μ values (a) 10.751, (b) 0.187, (c) 0.047, and (d) 0.002, plotted without normalisation by the bin widths. Correlated (triangle) and uncorrelated (cross) inter-event times are shown together with their sum, all inter-event times (circle). The other ETAS parameters used were $A = 10$, $\alpha = 1$, $c = 0.01$ and $p = 1.2$.

considered, the higher the initiation rate of independent sequences, and so the deeper the interleaving of separate aftershock sequences within the earthquake time series; sequential events are thus more likely to be independent of each other in catalogues from larger regions, resulting in a more exponential distribution. Conversely smaller regions have highly non-random time series and show two distinct bumps in their inter-event time distributions. For intermediate values of μ, the crossover between the correlated and uncorrelated curves can result in an apparent power law in the overall distribution, whose exponent does not have a simple relationship to any of the ETAS parameters. Artificially selecting stationary periods from a time series that is fundamentally non-stationary on the time scale considered introduces a strong sample bias that takes the form of an apparently universal gamma distribution. The true dependence of the distribution just on region size—ignoring different effective branching ratios from one region to another—is too complicated to fulfil a simple unified scaling law.

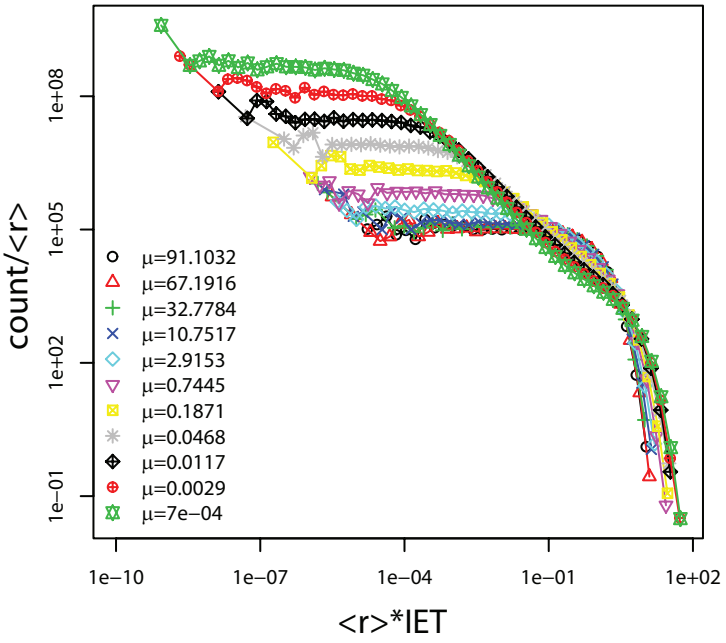


Figure 4.8: Inter-event time (IET) histograms for 11 ETAS simulations with μ values as indicated, plotted with normalisation by the bin widths, rescaled by the mean event rate r in the manner of Corral. Data collapse is not observed. The other ETAS parameters used were $A = 10$, $\alpha = 1$, $c = 0.01$ and $p = 1.2$.

Chapter 5

Masking of Earthquake Triggering Behaviour by a High Background Rate

A key insight contained in the previous chapter is that the spontaneous rate of earthquakes (parameter μ in the ETAS model) controls the extent of *temporal overlapping* of separate aftershock sequences. Higher μ means greater overlapping, which reduces the proportion of correlated inter-event times, thereby 'hiding' the aftershock sequences in each other. This turns out to have far-reaching consequences in terms of what can be discerned about triggering in catalogues with a high effective μ value, consequences which are explored in this chapter's paper (section 5.3) by testing standard ETAS inversion techniques on synthetic ETAS catalogues to see if these recover the known parameters accurately. I found that at high seeding rates μ, overlapping aftershock sequences significantly bias the inversion. The general idea of exploring parameter inversion accuracy was initially suggested by Mike Christie, and I carried out all of the work programme and acted as lead author. By way of introduction to the paper here, I review current procedures and issues with inversion of model parameters (section 5.1), and present the problematic results of my initial ETAS inversion trials which led into this work (section 5.2).

5.1 Parameter inversion

Calibrating models requires the adjustment of parameter values in order to improve the agreement of the model's calculations with certain benchmarks. The process of accepting a computational model of a real physical system involves verification (testing for the absence of bugs), validation (checking that the model correctly describes the physical system, i.e. that the equations are correct), and calibration (obtaining the correct parameters for the model) (Trucano et al., 2006).

In fitting the ETAS model to an earthquake catalogue, the ETAS parameters are commonly inverted from the data using the point-process maximum likelihood method, in which the log likelihood is given by:

$$\log L(T_1, T_2) = \sum_{i:T_1 \leq t_i \leq T_2} \log \lambda(t_i|H_{t_i}) - \int_{T_1}^{T_2} \lambda(t|H_t)dt \tag{5.1}$$

where λ is the ETAS conditional intensity function, evaluated for the data in the catalogue. T_1 and T_2 specify the time period over which the log likelihood is evaluated.

Veen and Schoenberg (2008) developed a technique based on an Expectation Maximisation (EM) method. The idea is that if the branching structure of the catalogue is known, this knowledge can be incorporated into the inversion and the performance is improved. For real catalogues the branching structure cannot be known, but a stochastic reconstruction can be performed (Zhuang et al., 2002) in order to approximate it. An iterative procedure of maximum likelihood estimations and stochastic reconstructions is used in order to converge on a solution.

In Bayesian inference, if approximate parameter value ranges are known, prior distributions can be applied to the likelihood function to produce a posterior probability density function (PDF). For example, for calibrating the ETAS model, the commonly used ranges could be used as priors. In this case gradient-based optimisation algorithms could perhaps be used successfully as the correct solution should be enhanced in the likelihood function (Sambridge and Mosegaard, 2002). In other cases, Monte Carlo algorithms can be employed which sample the parameter space more exhaustively than gradient methods and so are less likely to get 'lost' in false solutions (Sambridge and Mosegaard, 2002).

There are many types of sampling algorithm. Uniform sampling as a basic example chooses all parameters at random at each iteration. This is inefficient as it does not concentrate the search in the areas of interest. Better methods include simulated annealing and genetic algorithms (Sambridge and Mosegaard, 2002). Simulated annealing uses an analogy with statistical mechanics, whereby the slow cooling of a substance results in the minimum energy state being reached. The acceptance of a new point in parameter space depends partly on a 'temperature' term. Higher temperatures make acceptance more likely, which enables the transition out of local minima in the misfit function. If cooled slowly, it should settle into the region of the global minimum. Simulated annealing relies on Markov Chain Monte Carlo methods, of which examples include the Metropolis–Hastings algorithm and the Gibbs sampler. These effectively simulate the likelihood function (which generally has no explicit mathematical form and is explored through forward-modelling) to efficiently sample the most probable regions of the parameter space (Sambridge and Mosegaard, 2002). Genetic algorithms 'evolve' solutions in a similar way to the evolution of genes in biology. A set of model parameterisations is initially randomly generated. The object is to compare them and attempt to produce 'fitter' parameterisations through three operations: selection (replication), crossover (swapping parts), and mutation (Sambridge and Mosegaard, 2002). Finally, ensemble inference is an approach which compares the ensemble of acceptable solutions found in the posterior PDF. It is possible to be confident about any properties that they have in common, or similar prediction outcomes (Sambridge and Mosegaard, 2002).

Inversion of model parameters is not trivial. Carter et al. (2004, 2006) illustrate the dangers of trusting the apparent best-fit parameters. They constructed a simulated physical scenario involving an oil reservoir, and fit their model to it using a genetic algorithm. By simulating their physical system they had complete understanding of it and therefore knew what parameter values the inversion should deliver. They found that for the case where the model they were fitting was the same one used to generate the physical system, the highest peak in the likelihood

	North Island	South Island	Cape Palliser
μ	0.195 ± 0.021	0.072 ± 0.003	0.014 ± 0.007
$A = K/c^p$	3.101 ± 0.454	0.759 ± ?	17.078 ± 8.284
α	0.649 ± 0.055	2.906 ± ?	1.304 ± 0.150
c	0.021 ± 0.005	0.006 ± ?	0.006 ± 0.003
p	1.065 ± 0.019	1.087 ± 0.009	1.214 ± 0.060
b	0.99	0.95	1.07
n	1.440	∞	0.978

Table 5.1: Parameter values obtained for each catalogue through maximum-likelihood inversion

function occurred at the correct parameter values, although even in this case some algorithms would settle on a false solution. In the case where the physical system was altered slightly such that the model could no longer perfectly describe it (as is the case in real modelling scenarios), the most likely parameter values were those of the false solution, which if adopted would lead to very wrong predictions.

5.2 Initial inversion work

Initially the intention was to fit the ETAS model to three subsets of data from the earthquake catalogue of New Zealand, and extend these catalogues forward in time through simulation, with a view to using this technique to explore the convergence of the mean inter-event time beyond the current earthquake records (as was done in chapter 3 for purely synthetic ETAS catalogues). However, this procedure resulted in a poor match when comparing the real data with its synthetic extension based on the maximum-likelihood ETAS parameters. No previous studies had drawn attention to such failures despite the ETAS model being widely used in analysis of seismicity. This provided the motivation for the inversion tests presented in the paper (section 5.3.5).

The inversion procedure is laid out in Methods, 2.2. The three sets of New Zealand data to which the ETAS model was calibrated covered the North Island, South Island, and Cape Palliser, a small region in the south-east of the North Island. North Island consisted of events from the New Zealand catalogue above magnitude 4 between 1/1/1970 and 31/12/1999, north of 41°S. South Island data contained events above magnitude 4 between 1/1/1970 and 31/12/1999, in the area encompassed by the points (43°S, 172.5°E), (41°S, 167.5°E), (47°S, 159°E), (52°S, 167°E), and (45°S, 176°E). The Cape Palliser set, supplied by SSLib, is a subset of events from the New Zealand catalogue within a 36km radius of 41.684°S and 175.503°E, with magnitude ≥ 2.5 and depth ≤ 40km between 1/1/1990 and 31/12/1991 Harte (2007b).

The parameter values obtained are listed in table 5.1, along with standard errors. The b-value for each catalogue, and the calculated branching ratio, are also given. The standard errors for three of the South Island parameters were undefined, due to negative variances occurring (negative diagonal entries in the inverse of the Hessian matrix). This would seem to be symptomatic of a serious problem in the inversion.

Examples of the results of simulating forward are shown in figures 5.1 and 5.2, and excerpts of the R code for these plots is given in Methods section 2.1. A vertical dotted line indicates the transition-point between the real data and the synthetic continuation of it in time. The synthetic data generated for Cape Palliser superficially appears consistent with the short amount

of real data (figure 5.2). Figure 5.1 by contrast shows a marked difference, visually, between the real and synthetic data for the North Island.

The branching ratio was greater than 1 for the North Island, which resulted in a steadily increasing event rate in the synthetic data (figure 5.1). For the South Island, the best-fit α was greater than $\beta(= b \ln 10)$ giving an infinite branching ratio; as a result, the simulation rapidly reached an unphysical singularity at a finite timestep. A singularity is unphysical since the long-term loading is stationary. Only the Cape Palliser inversion resulted in a branching ratio less than 1, despite the fact that it is based on a catalogue of much shorter duration and smaller geographical area and therefore would be the least likely to contain a representative sample of long-term stationary seismicity. Obviating such counter-intuitive parameter combinations would require suitably constrained solutions not yet available for the purely statistical ETAS model.

The Akaike Information Criterion is a widely-used comparative measure of the goodness of fit, taking into account the log likelihood and promoting parsimony by penalising the use of larger numbers of parameters. I use AICc, which is a corrected AIC for smaller sample sizes and is given by:

$$AICc = 2k - 2 \times LL + \frac{2k(k+1)}{n-k-1} \tag{5.2}$$

where k is the number of parameters, LL is the log likelihood, and n is the sample size (Burnham and Anderson, 2002). The AICc values for the three ETAS model fits used are 10243.398 for the North Island, 5383.982 for the South Island, and -413.943 for Cape Palliser. Clearly there is huge variation from poor fit for the North Island to good fit for Cape Palliser.

The most obvious first line of inquiry into the issues with the inversion would be to investigate whether the inversion algorithm is settling on false local maxima in likelihood, resulting in incorrect parameters. The first step in exploring whether other maxima exist was to start the inversion process off at different initial parameter values, to see whether a single solution is reached from a variety of starting points. This is indeed what happened. It is therefore safe to assume that the algorithm is finding the global maximum. To investigate whether this maximum is well-defined or whether there may be other elevated areas in the log-likelihood surface that could be explored as alternative acceptable solutions, the log-likelihood surface was plotted across the range of parameter space deemed acceptable (see Methods section 2.2). Figures 5.3 to 5.5 show a 2D slice in the A-α plane, with the other parameters held at their maximum likelihood values. These two parameters were chosen because they (along with p) have the strongest effect on the triggering behaviour. In each case it seems that there is only one area of elevated likelihood.

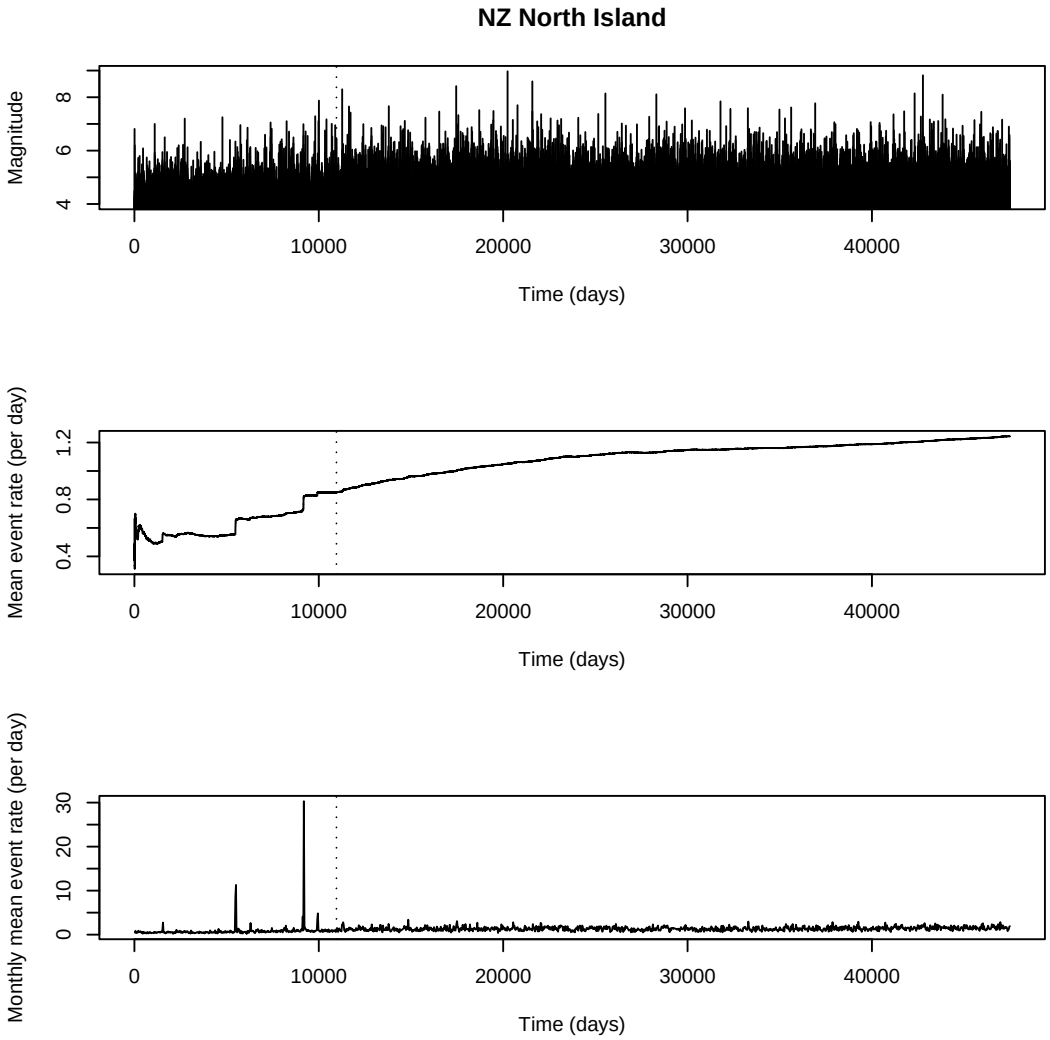


Figure 5.1: North Island real and synthetic data, to the left and right of the dashed line respectively. Events (top); ongoing mean event rate per day (middle); mean event rate per day for each month (bottom).

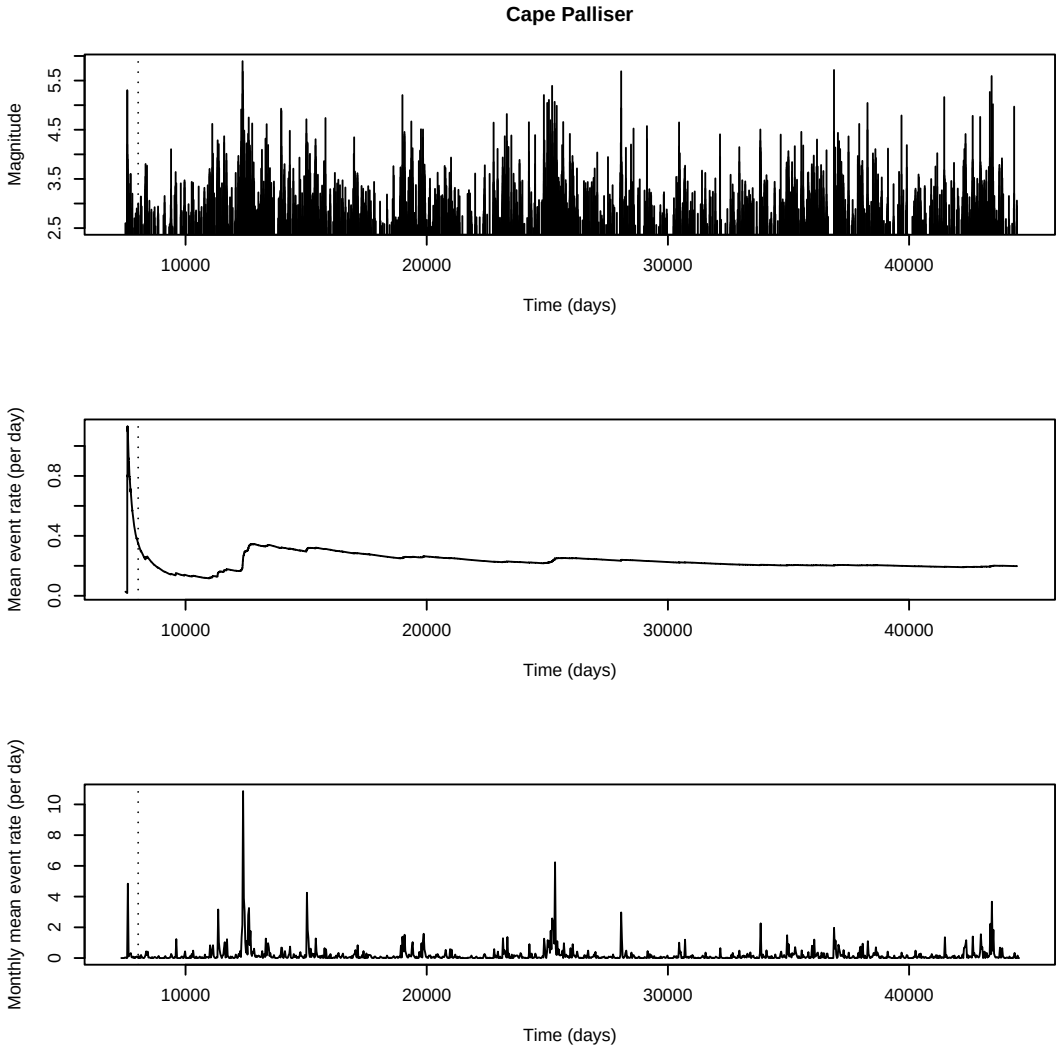


Figure 5.2: Cape Palliser real and synthetic data, to the left and right of the dashed line respectively. Events (top); ongoing mean event rate per day (middle); mean event rate per day for each month (bottom).

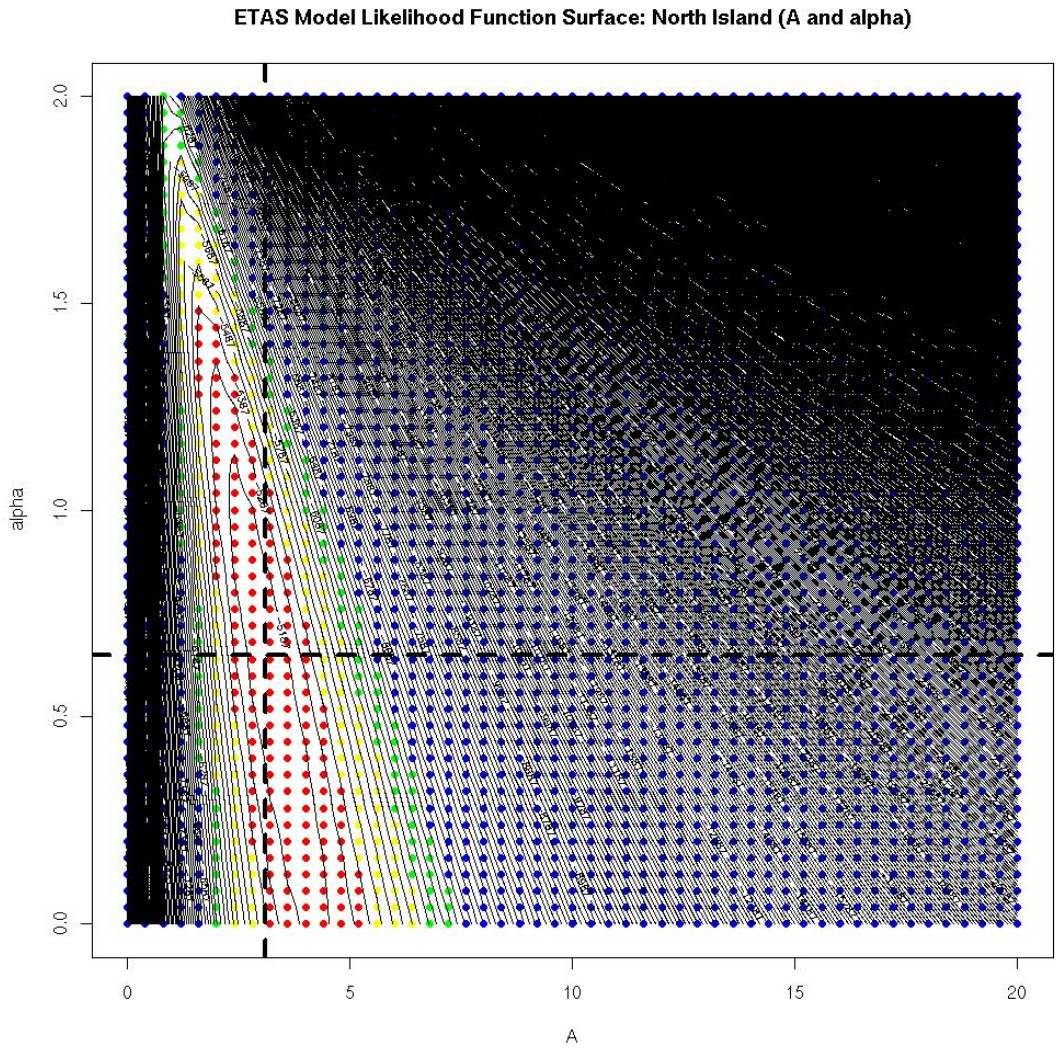


Figure 5.3: Likelihood surface for North Island—contours, and colour-coded dots for the points at which the log likelihood was evaluated.

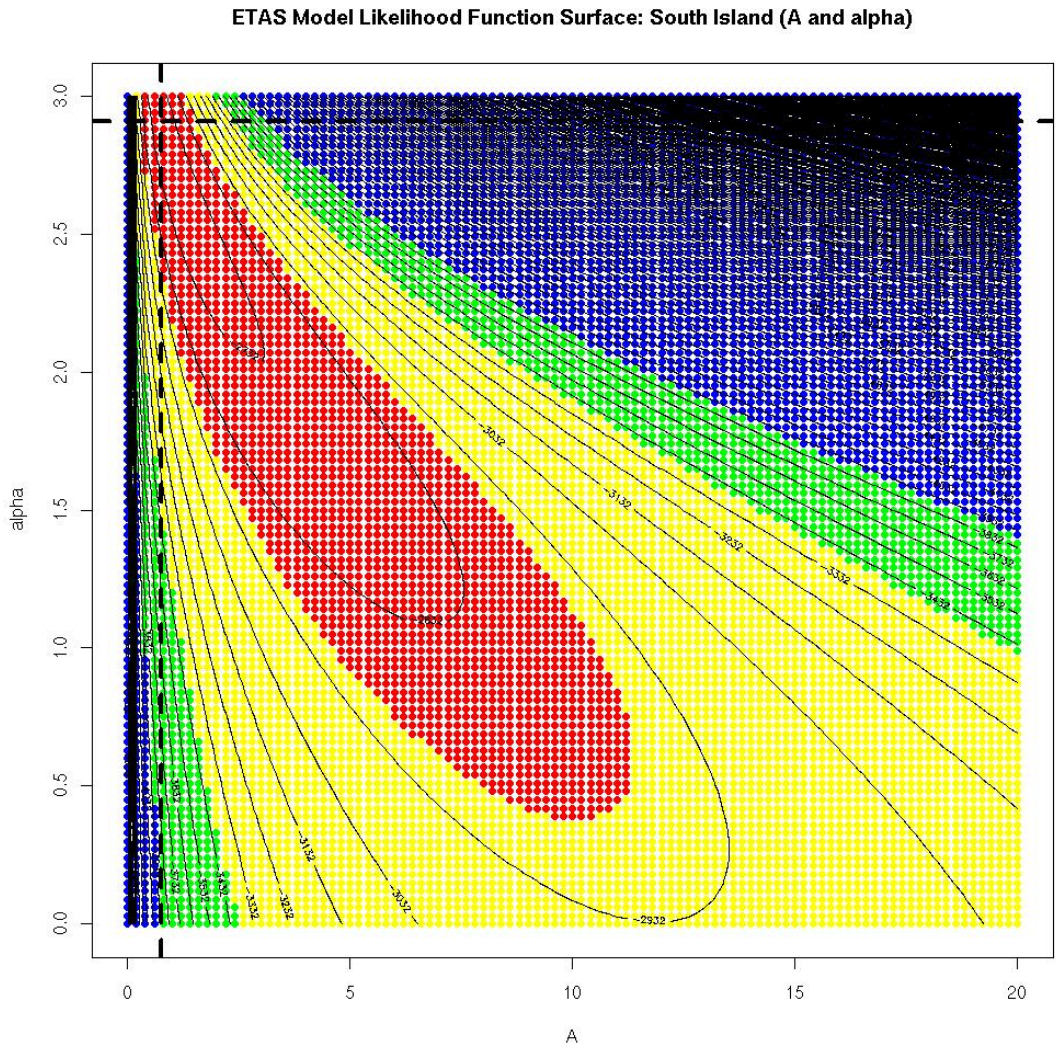


Figure 5.4: Likelihood surface for South Island—contours, and colour-coded dots for the points at which the log likelihood was evaluated.

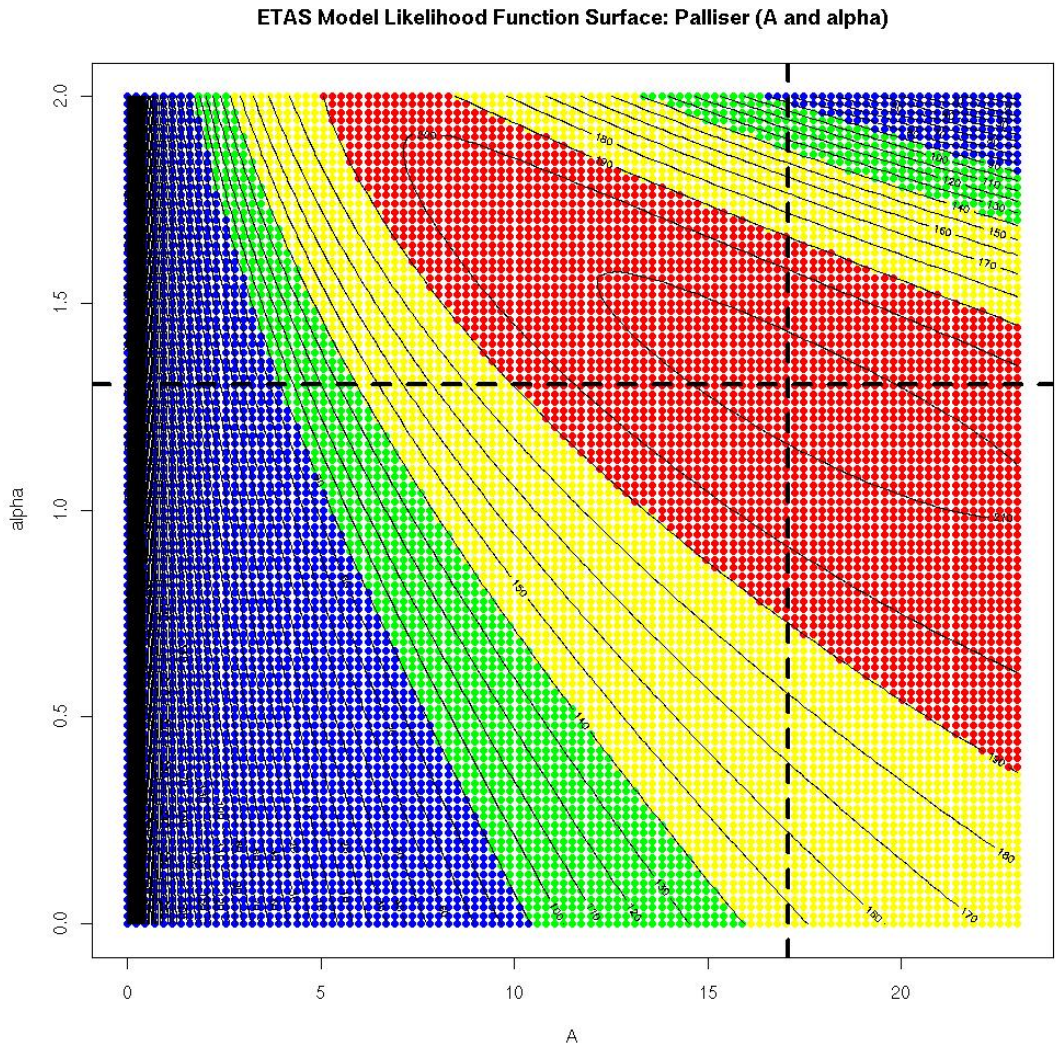


Figure 5.5: Likelihood surface for Cape Palliser—contours, and colour-coded dots for the points at which the log likelihood was evaluated.

This motivated larger-scale testing of the inversion algorithm on synthetic ETAS catalogues to determine how successful it is generally and what factors that success might depend on. The results of such testing form part of this paper (section 5.3.5). From the earlier work on inter-event time distributions (chapter 4), it seemed conceivable that a high μ value and hence a high level of temporal overlap of aftershock sequences might make the true parameters more difficult to invert. That is, there may be a well-defined global maximum, but its parameters may be biased due to overlapping aftershock sequences, and hence systematically offset from the parameters of the underlying process. Evaluating the conditional intensity involves looking at the time gap elapsed from past events, and we have seen that inter-event times become dominated by inter-sequence, random, exponentially-distributed values at high μ. Indeed the paper demonstrates that this conjecture was accurate. The relative success of the Cape Palliser inversion (at least according to a qualitative comparison of real and synthetic data and the AICc value) in the initial study may possibly be explained by its smaller area and lower effective μ as compared to North Island and South Island.

Methods section 2.5 details the procedure for inter-event time histograms shown in several figures of the paper. Figure 5.6 (right-hand column) uses Methods section 2.1. The specific method for fitting Saichev and Sornette's analytical distribution in the central column of figure 5.6 is presented in the Appendix section A.11. For figure 5.9, the specific procedure for putting error bars on the histogram and fitting an underlying exponential is shown in the Appendix section A.12. Dividing the PDE catalogue into spatial cells for figure 5.10 is explained in the Appendix section A.13. Figure 5.11 involves a spatial ETAS simulation, described in Methods section 2.4, and the nearest-neighbour analysis for the plot is explained more fully—with R code—in the Appendix section A.14. Figure 5.12 involved simulating the average global aftershock sequence, which is explained in Methods at the end of section 2.3. Figures 5.15 to 5.18 are backed by the inversion technique presented in Methods section 2.2.

5.3 Paper

We examine the effects of the spontaneous background event rate and aftershock triggering characteristics on the temporal statistics of seismicity in the epidemic-type aftershock sequence model. Recent work has shown that the earthquake inter-event time distribution is generally bimodal: a superposition of a gamma component from triggered aftershocks at short time intervals and an exponential component at longer intervals from spontaneous events and the overlapping of independent aftershock sequences. The relative size of these two components varies between catalogues, so there is no simple, universal scaling; at the extreme of high spontaneous rate, e.g., in large regions, the high probability of temporally overlapping aftershock sequences causes the exponential component to dominate. Here we further explore the effects of both the spontaneous rate and the aftershock triggering parameters. We show that the analytical theory of Saichev and Sornette (2007), although valid under their assumptions, gives the impression of a more 'universal' behaviour if used outside its stated range of applicability. We also show that within the high-overlap (high-spontaneous rate) regime, a maximum likelihood inversion of the model's temporal parameters is both less accurate and biased; specifically, the background rate is systematically over-estimated. This has implications on the suitable range of region

sizes for which parameter inversion may be reliable and must therefore be taken into account in any inversion for temporal variations in background rate in time-dependent hazard calculation.

5.3.1 Introduction

Over the past decade, much literature has been published on the distribution of time intervals between successive earthquakes in a recorded catalogue (Bak et al., 2002; Corral, 2003, 2004; Davidsen and Goltz, 2004; Hainzl et al., 2006; Shcherbakov et al., 2005; Molchan, 2005). It has become commonplace to model these inter-event times using a single gamma distribution, which has led some authors to propose universality, extending from tectonic earthquakes (Bak et al., 2002; Corral, 2003, 2004; de Arcangelis et al., 2008; Lennartz et al., 2008)) to rock fracture (Davidsen et al., 2007). This idea follows from the observation that rescaling the data by region size and magnitude cutoff (Bak et al., 2002) or the mean event rate (Corral, 2003) causes the distribution curves to collapse on top of one another. Analytical investigations by Saichev and Sornette (2006, 2007) confirmed an approximately universal form over the range of parameters considered typical. However, our recent work (Touati et al., 2009) generalises these observations beyond the data selection criteria used by these authors, and shows that the long inter-event times do not scale in a 'universal' way from the short ones. If the distribution is to be described as universal, that term is only applicable in a different sense than originally proposed (Bottiglieri et al., 2010).

We start from the point of view that earthquakes are in two broad categories: Firstly, a spontaneous category, which is usually an approximate Poisson process comprised mainly of 'background' tectonically-driven events, but likely also includes contributions from other driving forces e.g. fluids as well as events that are triggered by other events below the detection threshold. These events may be considered independent of each other. Secondly, a triggered category, consisting of events triggered by others through the time-dependent relaxation of the Earth's crust, and whose rate follows the empirical Omori aftershock law. When considering inter-event times, we are looking at temporal intervals between pairs of events, which—it follows—are also of two types: the events are either causally related—that is, part of a common sequence of triggered events—or else, they are unrelated and independent of each other. We refer to the former as same-sequence pairs and the latter as inter-sequence pairs of events. In our previous paper (Touati et al., 2009) we showed that inter-event time distributions can take on a range of shapes from an exponential form to a double-peaked curve, depending on the balance between aftershock production and the rate of spontaneous background events; the gamma distribution is thus an approximation, and universality does not strictly hold. We presented a new understanding of the earthquake inter-event time distribution as an essentially bimodal mixture distribution, comprising the contributions from precisely the two categories of event pairs described at the start of this paragraph. Those same-sequence events that happen to occur consecutively in the catalogue produce gamma-distributed inter-event times, since the temporal decay of the aftershock rate follows a power law. (If each sequence were left uninterrupted by further spontaneous events, the inter-event time distribution would also be a power law.) The occurrence of inter-sequence events consecutively, on the other hand, produces an exponential inter-event time distribution, as long as the different spontaneous seeding events from which they arise are themselves a Poisson process—a reasonable assumption in the case

of stationary tectonic loading conditions. Thus the observation of bimodal distributions in earthquake data can be explained in terms of the peaks of these two contributions being well resolved.

In this paper we extend this analysis of inter-event time distributions. We start by describing the Epidemic-Type Aftershock Sequences (ETAS) model (section 5.3.2), a stochastic model utilised throughout the paper in which aftershock triggering is viewed as a multi-generational branching process. We then show the effect of the spontaneous event rate on the temporal statistics in section 5.3.3. We test our results against the analytical model for earthquake inter-event times of Saichev and Sornette (2007) and confirm that the latter is applicable within its stated assumptions, i.e. only at higher inter-event times; at low inter-event times inappropriate extrapolation of their model would produce an overestimate of occurrence, and give a false impression of broader-band universal scaling. In section 5.3.4 we demonstrate the relevance of our paradigm to real data and show the relationship between seeding rate in the ETAS model and region size in real data. We then explore the interaction between the effects of the triggering parameters and the seeding rate in section 5.3.5. We show that the effect of any changes in the branching parameters is diminished as the seeding rate is increased; in other words, seismicity that is characterised by an exponential inter-event time distribution has a high redundancy in terms of the effective triggering parameters. This simple observation sparked by our understanding of the inter-event time distribution turns out to have deep consequences for the inversion of temporal ETAS parameters, in that significantly less accurate results are obtained for higher seeding rates (as in larger or more tectonically active regions). We point out the implications for the possibility of detecting a varying background rate.

5.3.2 ETAS Model

The ETAS model is a stochastic point-process model based on well-known empirical laws of earthquake occurrence. The process is seeded by spontaneous—commonly called 'background'—events, occurring as a Poisson process in time with constant rate μ, to represent steady-state tectonic driving. Aftershocks may then be triggered by all events; the aftershock rate $n(t)$ decays as a power law in time after a sizeable event according to the Omori law,

$$n(t) = \frac{K}{(c+t)^p}, \tag{5.3}$$

where K, c and p are constants and t refers to time. The number of aftershocks produced depends exponentially on the magnitude of the parent event. Each event's magnitude is selected independently from the Gutenberg–Richter distribution,

$$\log N = a - bm, \tag{5.4}$$

where N is the number of events in a given time period with magnitude $\geq m$, and a and b are constants. There is a lower threshold magnitude m_0; there may or may not be an upper magnitude cutoff. The ETAS conditional intensity function λ is a function of time and of the history of the process H_t at time t:

$$\lambda(t|H_t) = \mu + K \sum_{i:t<t_i} \exp\left(\alpha(m_i - m_0)\right)(t - t_i + c)^{-p}, \tag{5.5}$$

where t_i are the times of the past events and m_i are their magnitudes. Thus the five ETAS parameters are μ, the spontaneous event rate; c, p and K, the Omori law parameters; and α, the productivity parameter. Because we cannot distinguish a priori between foreshocks, main shocks and aftershocks, the model regards all events as capable of triggering further events. Thus, each spontaneous event may result in a cascade of nested aftershock sequences known as a 'global' aftershock sequence (Helmstetter and Sornette, 2002b).

An alternative parameterisation which we will use is obtained by making the substitution $A = K/c^p$:

$$\lambda(t|H_t) = \mu + A \sum_{i:t<t_i} \exp\left(\alpha(m_i - m_0)\right) \left(1 + \frac{t - t_i}{c}\right)^{-p}, \tag{5.6}$$

For the spatio-temporal ETAS model, the conditional intensity varies with spatial coordinates x and y in such a way that the total conditional intensity for the entire region is equal to the temporal-only ETAS conditional intensity (equation 5.6). Spatial clustering of aftershocks is accounted for by a spatial kernel $f_i(x - x_i, y - y_i)$, one implementation of which is:

$$f_i(x - x_i, y - y_i) = \frac{(q-1)d^{2(q-1)}}{\pi[x^2 + y^2 + d^2]^q} \tag{5.7}$$

where d and q are parameters (Hainzl et al., 2008).

A useful quantity which we will make use of in this paper is the branching ratio, defined as the average number of aftershocks generated by each event, which is obtained by integrating $A\exp(\alpha m)(1 + \frac{t}{c})^{-p}$ over both times and magnitudes from 0 to ∞, and is given by:

$$n = \frac{Ac}{p-1}\frac{\beta}{\beta - \alpha} \tag{5.8}$$

The parameter μ, being the average frequency of spontaneous events, has no effect on aftershock generation; it instead represents the average rate at which aftershock sequences are initiated and so determines the temporal overlap extent of separate sequences. It represents the roughly stationary component due mainly to tectonic loading, and we show in this paper that it may be identified as a proxy for region size. We first show the effect of μ on the inter-event time distribution and then consider the effects of the other parameters.

5.3.3 Bimodality and the seeding rate

To demonstrate the effect of the seeding rate parameter μ on the inter-event time distribution for the ETAS model, we simulate synthetic earthquake catalogues of 100000 events each, for four different values of μ, and plot histograms of the inter-event times τ (figure 5.6). We plot both the logarithm of the count in logarithmic bins, which optimally shows the bimodality, and the more usual probability density function (PDF) obtained by normalising the counts by the bin widths, which tends visually to mask the bimodal form of the data (Touati et al., 2009). The effect of plotting a simple histogram is merely to increase the slope by 1 with respect to the equivalent PDF plot (Bonnet et al., 2001, figure 2(b)).

The origin of the histogram shape and its variation with μ becomes clear when we superpose the two subsets of data arising from the two categories of event pairs (figure 5.6, left-hand column). We henceforth refer to same-sequence and inter-sequence intervals, following our stated convention for the event pair categories giving rise to these intervals. In increasing

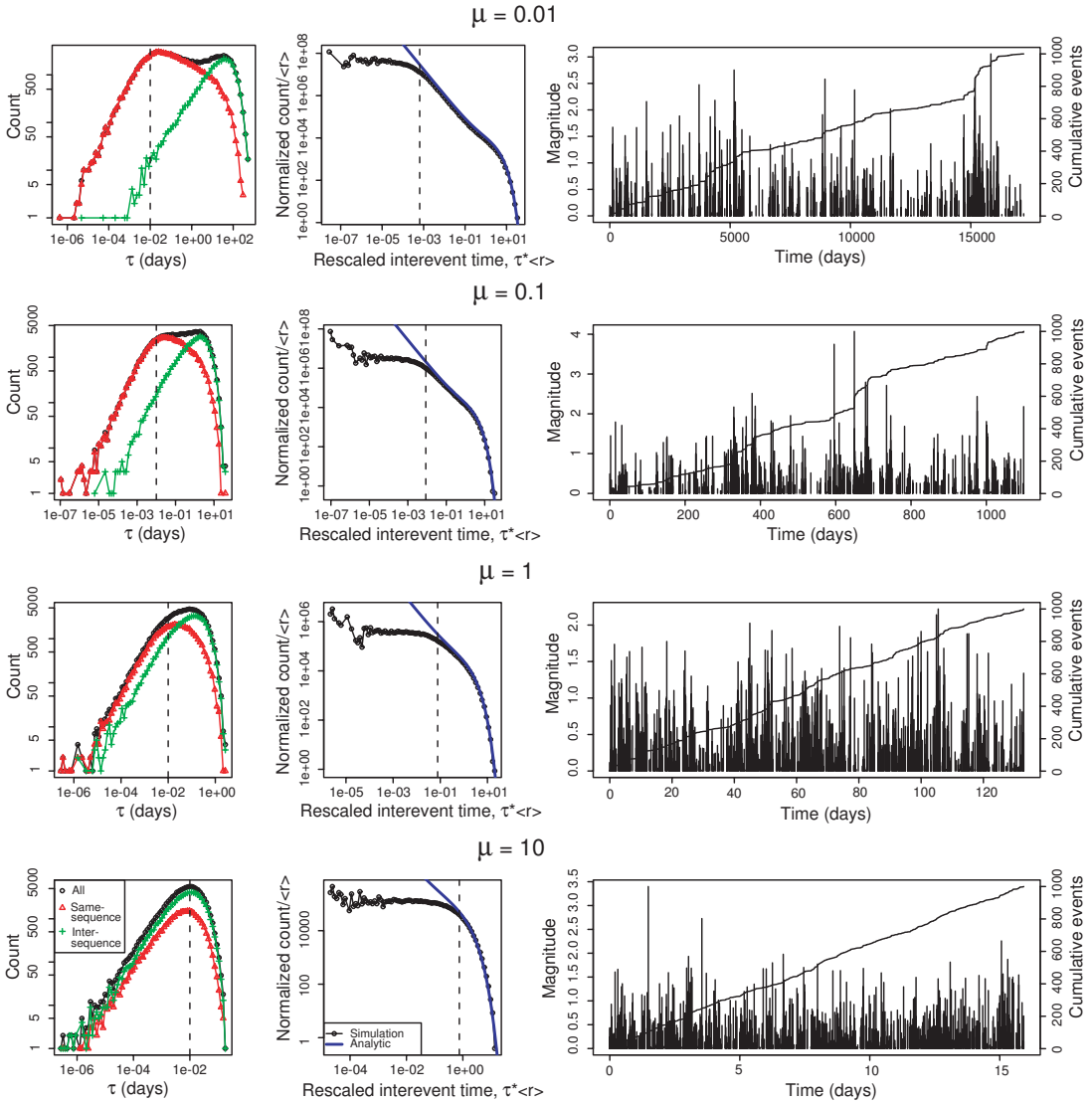


Figure 5.6: The effect of varying the ETAS parameter μ on: the inter-event time histogram (left column), with sub-distributions for correlated (red triangle) and uncorrelated (green cross) event pairs superposed; the normalized PDF rescaled by the mean event rate, plotted along with the analytic solution of Saichev and Sornette Saichev and Sornette (2007) (central column); the time series and cumulative event count for a sample of 1000 events (right column). The value of c ($c\langle r\rangle$) is indicated by a dotted vertical line in the histogram (PDF) plots, which represents a lower limit of Saichev and Sornette's solution's range of applicability. The higher μ is, the more exponential the inter-event time distribution, and the smoother the cumulative event count increase. Values of μ used were as shown; other ETAS parameters were held fixed at $A = 10$, $\alpha = 1$, $c = 0.01$, $p = 1.2$; 100000 events were simulated in each case.

the seeding rate, more earthquake sequences overlap within the dataset; the inter-event times thus become dominated by those exponentially-distributed ones arising from inter-sequence event pairs (see the illustration, figure 5.7), and the overall distribution becomes unimodal and exponential-like. The same-sequence inter-event time distribution becomes truncated at shorter inter-event times due to the typical sequence being interrupted earlier in its progression (note the change in scale of the τ-axis between the different plots).

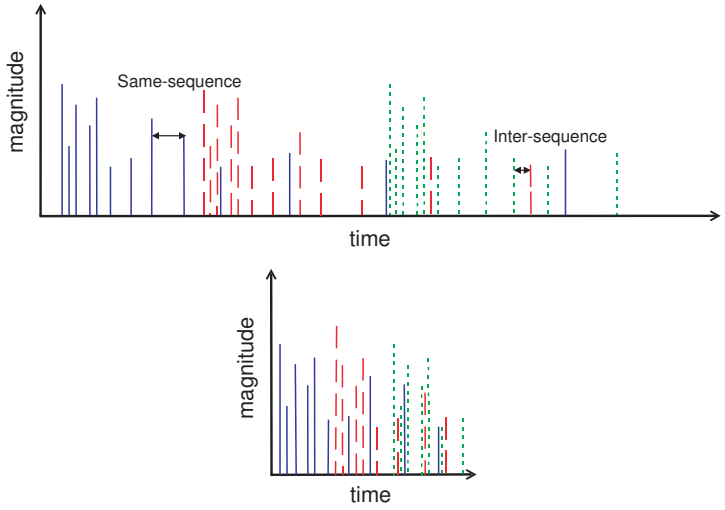


Figure 5.7: Illustration of the effect of seeding rate: when changing the seeding rate from low (top) to high (bottom), distinct aftershock sequences—shown in different colours and line patterns—are initiated more frequently; this is effectively a translation of the sequences along the time axis, without compression. (The bottom figure spans a shorter time to avoid the need for further sequences to be introduced). Thus, a higher overlap between the aftershock sequences causes more of the inter-event times to result from inter-sequence event pairs.

The right-hand column in figure 5.6 shows comb diagrams of the events in a 1000-event subset of each catalogue, along with a cumulative count of the events, giving further visual illustration of the effect. As μ is increased there is clearly less variability in the temporal occurrence of events, and the cumulative count becomes a smooth increasing function of time. This is in line with an exponential-like distribution of inter-event times. The stepped cumulative distribution at low μ, on the other hand, is characteristic of well-resolved aftershock sequences.

To put this in the context of previous reported results, firstly we note that several studies motivated by the work of Corral (2003) have selected, from real earthquake datasets and experimental datasets, periods showing short-term stationarity in the event rate. (Short-term 'stationary' periods emerge by chance in any dataset, when there is a local absence of large events which would cause the event rate to change rapidly. In the longer term, of course, all tectonically-driven regimes can be considered stationary.) This requirement allows a straightforward rescaling of the data by the mean event rate, but it appears that as a side effect, the data selection procedure generally results in a histogram similar to the $\mu = 1$ case in figure

5.6, which approximates a 'universal' gamma distribution. Thus an apparently benign and otherwise quite reasonable filter for apparently stationary periods tunes the search to regimes where there is sufficient overlap of aftershock sequences to blur the underlying bimodality.

Secondly, we consider the analytical PDF of Saichev and Sornette, who report that the previously-discussed universality is merely approximate and relies on a value of p close to (but greater than) 1. From our point of view, a smaller p value would certainly reduce the appearance of bimodality by reducing the power-law decay exponent of the same-sequence distribution (see section 5.3.5), although we would suggest that there are additional alternative ways of observing a single power law, such as by controlling the effective overlap level through data selection. We compare our simulations with Saichev and Sornette (2007, equation 36) in the middle column of figure 5.6, which shows the data in the more commonly-used format of a normalized PDF rescaled by the mean event rate. We note that there is a divergence between the theoretical and simulated curves at inter-event times comparable to c and smaller, due to the clearly-stated key assumption $\tau \gg c$ in their derivation. At high μ this assumption becomes more limiting due to the average inter-event time shortening towards c, reducing the bandwidth over which the distribution function is valid, and giving the false impression of a more universal form if the range of applicability is mistakenly overlooked. Thus we present here a generalisation of the work by these authors and Corral, extending the analysis beyond datasets that appear stationary in the short term, or have p close to 1, and extending the bandwidth of inter-event times below c.

We can further elucidate the relation between Saichev and Sornette's theory and our understanding by plotting its two components separately: firstly, $\phi(x, m)$, which is the exponential probability of observing no events in the rescaled time interval $x = \lambda(m)\tau$; and secondly, the other component, which comes from the twice differentiation of $\phi(x, m)$ in obtaining the inter-event time PDF. Figure 5.8 shows that $\phi(x, m)$ corresponds to the exponential inter-sequence component of the distribution. This follows from the fact that $\phi(x, m)$ is concerned with the longest observable inter-event times, contingent on both the spontaneous rate and aftershock production pattern. In differentiating twice with respect to time, the constant background rate component of the exponent is eliminated and the effect of the aftershocks is revealed in the other component: the same-sequence power-law decay is captured along with a shallower power law leading towards the inter-sequence peak.

Finally, we note that the universality of Bottiglieri et al. (2010) relies on being able to compensate for the variation of μ through a corresponding variation in c, thereby—as we understand it—maintaining a roughly constant overlap extent between the separate aftershock sequences in the catalogue. This is generally not the case in selecting real data, particularly in catalogues of different spatial extent, as we show next. To summarise our contribution so far, we replace the universality paradigm, which has been shown both by Saichev and Sornette and by Bottiglieri et al. to be of limited scope, with an intuitive, physically-motivated paradigm based on the classification of inter-event times as either same-sequence or inter-sequence intervals.

5.3.4 Seeding rate and region size

We now turn our attention to real data from the worldwide Preliminary Determination of Epicentres (PDE) catalogue, and demonstrate that the ETAS model is—in principle—a valid description of seismicity even on large spatial scales, with the parameter μ acting as a proxy

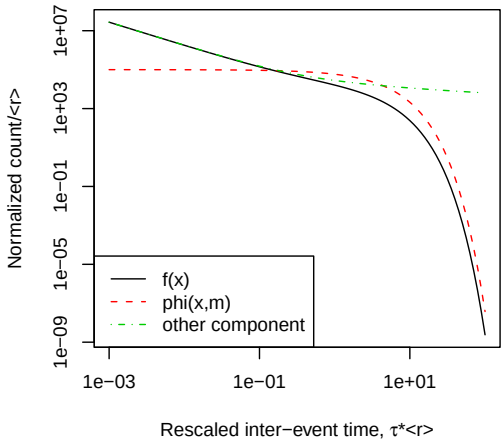


Figure 5.8: Saichev and Sornette's analytical PDF (equation 36 of Saichev and Sornette (2007)), along with its two components: the exponential $\phi(x, m)$ and the other term which results from twice differentiation of $\phi(x, m)$. ETAS parameters were $\mu = 0.1$, $A = 10$, $\alpha = 1$, $c = 0.01$, $p = 1.2$.

for the size of the region.

We choose data between 1 Jan 1969 and 1 Jan 2005, and apply a minimum cutoff magnitude of 5.0. The catalogue is complete over this time period and magnitude range, as indicated by the absence of any roll-off in the magnitude–frequency relation towards smaller magnitudes. We first select events belonging to a circular region centred on 130° longitude and 0° latitude. This position is arbitrary but is in the most active seismic region, giving us a maximal amount of data. By varying the radius of this circle, we can tune the number of events included to correspond to the variation in μ in figure 5.6; that is, approximately a tenfold increase in the event rate at each radius increase. (We cannot simply increase the area tenfold to create an analogy with figure 5.6, since earthquake occurrence is highly inhomogeneous.)

We must remove all events below magnitude 5.0 due to incompleteness of the catalogue. Figure 5.9 shows the result; the error bars shown are based on 95% confidence limits of a binomial error distribution. Since the apparent branching ratio decreases as the magnitude threshold is raised (Sornette and Werner, 2005b), the relatively high magnitude threshold would be expected to significantly reduce the same-sequence component and the appearance of bimodality in the inter-event time distribution. However, we still find clear indication of bimodality at smaller region sizes, with a transition to a unimodal form as the region size grows, just like in an ETAS model with increasing μ.

It would be an over-simplification for most modelling purposes to regard global seismicity as a single ETAS parameterisation, and we are not promoting such an idea. However, the use of the whole world presents a clear end member case that catalogues of progressively smaller regions generally tend away from, and figure 5.9 shows that this end member case is at least qualitatively comparable to simply using a high μ in the ETAS model. The effective parameters of seismicity are not constant with respect to space, yet we can say that the global averaging of the effective parameters through this inter-event times analysis produces identifiable bimodality

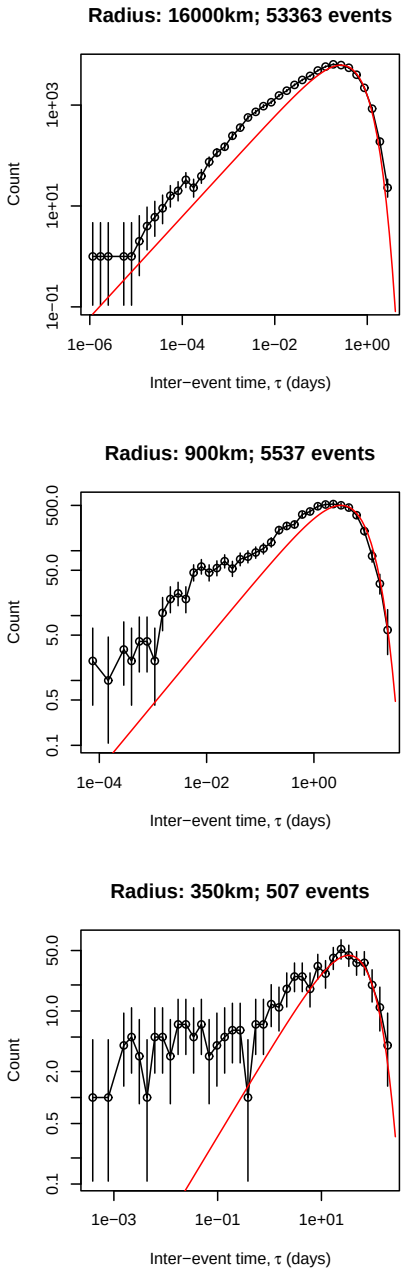


Figure 5.9: The effect of varying the spatial window size on the inter-event time histogram, for the PDE catalogue from 1 Jan 1964 to 1 Jan 2005. Red line shows the exponential fit for the background event pairs, based on the peak value, as described in the text. The smaller the region size is, the greater the deviations from this exponential distribution at short inter-event times, indicating a greater number of dependent event pairs occurring sequentially. Circular regions centred on 180° longitude and 0° latitude were used. A minimum magnitude cutoff of 5.0 was applied to ensure catalogue completeness.

as expected by the simple ETAS model. Stacking together seismicity of different Poissonian seeding rates through the concatenation of smaller regions should not worry us because the sum of several Poisson processes will be yet another Poisson process. The point here is to illustrate our paradigm for examining inter-event times, and to demonstrate that the transition from bimodality to unimodality with an increasing μ has some genuine physical meaning in terms of the region size, and hence that the often subjective choice of what constitutes a 'region' can strongly condition the model parameters.

We can use curve-fitting to demonstrate that the decay in the inter-event time distribution at large τ is exponential at all region sizes, which further supports the comparison with ETAS. It is convenient to note two points: firstly, when plotting exponentially-distributed data as a logarithmic histogram without normalisation, there is a peak; analytically, the x-value at which this peak occurs is $1/\lambda$, where λ is the exponential parameter. Secondly, in the case of ETAS inter-event times, the peak in the underlying inter-sequence component coincides with a peak in the overall distribution (figure 5.6). We therefore assume that we can fit an exponential distribution to the latter portion of the histogram simply by reading off the peak value of inter-event time. The red curves in figure 5.9 result from this fitting; in each case, the fit is visually convincing, and so we consider these fitted exponentials to represent the inferred contribution from inter-sequence events, with a possible amplitude adjustment in the larger region cases. (Note however that we cannot infer the spontaneous event rate from this component of the distribution because the inter-sequence component arises from both aftershocks and spontaneous events; see section 5.3.5.) It is also clear from figure 5.9 that the additional contribution from same-sequence event pairs exists, and grows more prominent as the region size is decreased.

We may question whether the unimodal, exponential-like distribution observed for the largest region can be explained by the fact that in such a large spatial area, we are combining data from different seismic regions, with different effective triggering parameters and spontaneous rates, resulting in a Poisson process through this mixing. A second approach to check this is to repeat the analysis with mixed data at all region sizes. This additionally allows us to use the events from the whole globe and avoid such large counting errors in the histograms.

We split the Earth into two-dimensional spatial cells of equal size. It is most convenient to use, as cell boundaries, only those great circles that are lines of constant longitude or latitude; that is, the equator, and various lines of longitude. The inter-event times are generated for each cell separately, and then a histogram is produced after concatenating data from all cells, with the procedure repeated for different cell sizes. The results of this are shown in figure 5.10 (left). There is remarkably clear bimodality at smaller region sizes, despite the inevitable smoothing due to the mixing of data from different cells. This confirms that mixing of data alone cannot explain the exponential-like distribution of inter-event times coming from large regions. Incidentally, figure 5.10 (right) shows a similar plot for a regional catalogue—Southern California Earthquake Center (SCEC), using events between 1 Jan 1984 and 1 Dec 2000, up to longitude 245°, and magnitudes above 2.2—broken into rectangular cells of (approximately, on the curved Earth surface) equal area; for the whole region (i.e. 1 cell) the form is similar to that for the PDE catalogue split into 64 cells, and so the plot shows a kind of continuation of the spatial shrinking with the result that the same-sequence peak grows more dominant.

We can conclude, then, that in a large region, the vast majority of inter-event times come from unrelated, inter-sequence event pairs, and infer that the effective μ increases with region

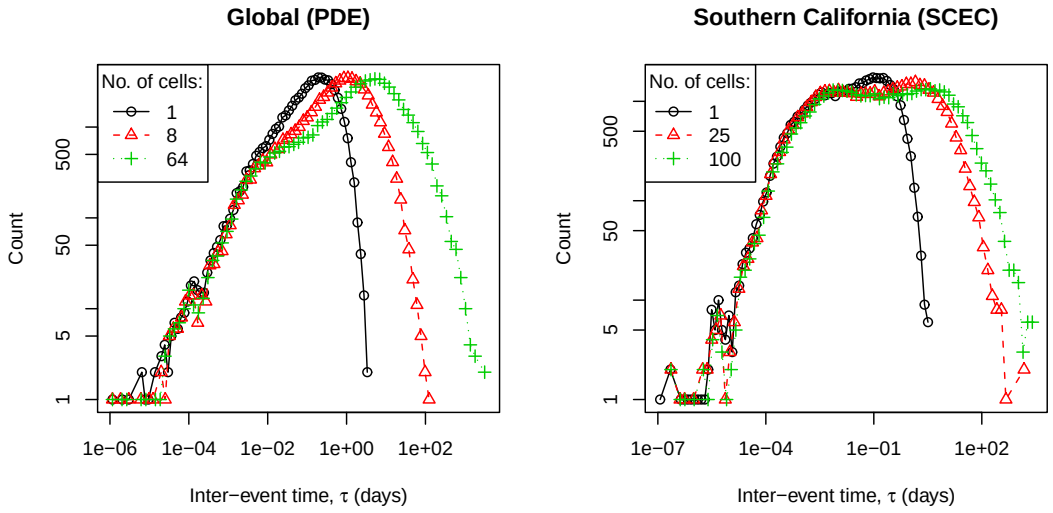


Figure 5.10: The effect of varying the spatial window size on the inter-event time histogram, for the PDE catalogue from 1 Jan 1964 to 1 Jan 2005 (left) and the SCEC catalogue from 1 Jan 1984 to 1 Dec 2000 (right). In the case of more than one spatial cell, the plots were generated by mixing inter-event times from the cells as described in the text. A minimum magnitude cutoff of 5.0 (2.2) was applied to ensure catalogue completeness for the PDE (SCEC) data.

size. This means μ is in effect not a material parameter at the regional scale, since an arbitrary choice of region size can reduce or increase its value. In principle its spatial density function may be more representative as a material parameter that varies with tectonic region, but less so with region size.

Although the inter-event time distribution becomes more exponential-like as region size is increased, this is not to say that large regions have a lower effective branching ratio. It would seem to be obvious that the branching ratio in a large region cannot be lower than that of the smaller spaces of which it is comprised; certainly in our ETAS analogue, we do not vary the branching ratio in altering μ—we know that the clustering in the catalogue has merely become 'hidden' due to an increased proportion of inter-sequence event pairs occurring sequentially. However, as we will see in section 5.3.5, inversion of ETAS parameters can give the false impression that the branching ratio decreases with increasing region size.

Spatial information can be used to overcome the temporal overlapping of aftershock sequences in declustering techniques, since triggering probability decreases rapidly with distance (Huc and Main, 2003). These techniques provide a way of categorising intervals as same-sequence or inter-sequence for real data, to further investigate the relevance of our ETAS-based paradigm. Zaliapin et al. (2008) presented a method for analysing the spatio-temporal clustering within a catalogue and using this information to decluster. They define a relationship ('distance' n_{ij}) between one event and a later event in terms of the time and space interval between the events; an event's nearest neighbour is then defined as the event for which this n_{ij} value is minimal. They find that for ETAS catalogues, the values fall naturally into two clusters; one for same-sequence aftershocks, which occur close in space and time to each other, and another for inter-sequence events which are more separated in space and time. This obser-

vation of bimodality—in the clusters and their marginal distributions—complements very well our inter-event times analysis above.

Figure 5.11 (top left) shows space and time components of the simple nearest-neighbour distance defined as $n_{ij} = r\tau$, for a spatial ETAS simulation (using the kernel (5.7)). On the right, we show a histogram of the time components—that is, the time interval to the nearest neighbour for each event. We can label each interval as same-sequence or inter-sequence based on which cluster it falls into, as determined by setting a dividing line as shown in the left-hand plot, and superpose histograms for these categories. The inter-event time histogram is shown as a dashed line for comparison. Note that the intervals to the nearest neighbours are in general different from the inter-event times, as the nearest neighbour may not be the next event in time if there is a later event that is closer in space. The intervals to the nearest neighbours therefore contain a higher proportion of same-sequence relationships, which is evident from the extended power-law segment, particularly at longer intervals for which there are no inter-event times. The division between the clusters is somewhat arbitrary, and inevitably overlap of the clusters will lead to mis-labelling of some intervals, but a clear power law is recovered nonetheless.

The bottom two plots in figure 5.11 repeat this analysis for the Southern California Earthquake Center (SCEC) catalogue. The two clusters overlap to a greater extent—likely due to variations in the seismicity parameters throughout the region and also the inappropriateness of the hypocentral distance for aftershocks of large events—and so mis-labelling of intervals is a bigger problem, but a clear power law can nevertheless be seen in the histogram of intervals labelled as same-sequence. This further confirms that our interpretation of the inter-event time distribution and the origins of bimodality are relevant for real data. A possible practical application of the technique would be to obtain the power-law exponent in the histogram as a means of inverting p through its relationship with this exponent (see Helmstetter and Sornette (2002b) for details of the relationship).

Having established the effect of the spontaneous rate μ and its connection with region size, we now proceed to explore the effects of the triggering parameters and the interaction between this and the seeding rate.

5.3.5 Distinguishability as a function of seeding rate

While μ determines the occurrence rate of 'global' aftershock sequences as whole entities and clearly plays an important role in the temporal statistics, different aftershock generation parameters also make an impact on the statistics. The relative proportions of same-sequence and inter-sequence intervals in the series, and the bimodality or otherwise of the inter-event time distribution, depend on both these aspects.

The global aftershock sequence has been shown analytically to exhibit a double power-law decay in the event rate, with a smooth transition between these two regimes (Helmstetter and Sornette, 2002b), as opposed to the single power-law for the 'local' aftershock sequence (the non-recursive Omori law). The exponents of the two power-law segments are $1 - \theta$ for early times and $1 + \theta$ for later times, where $\theta = p - 1$. The effect of changes in p on the event rate of the global aftershock sequence is quite dramatic as shown in the left-hand plot of figure 5.12, which is produced by stacking simulated aftershock sequences.

This effect on the power-law exponent propagates through to the inter-event times, altering the exponent in the same-sequence component of the inter-event time distribution. Figure

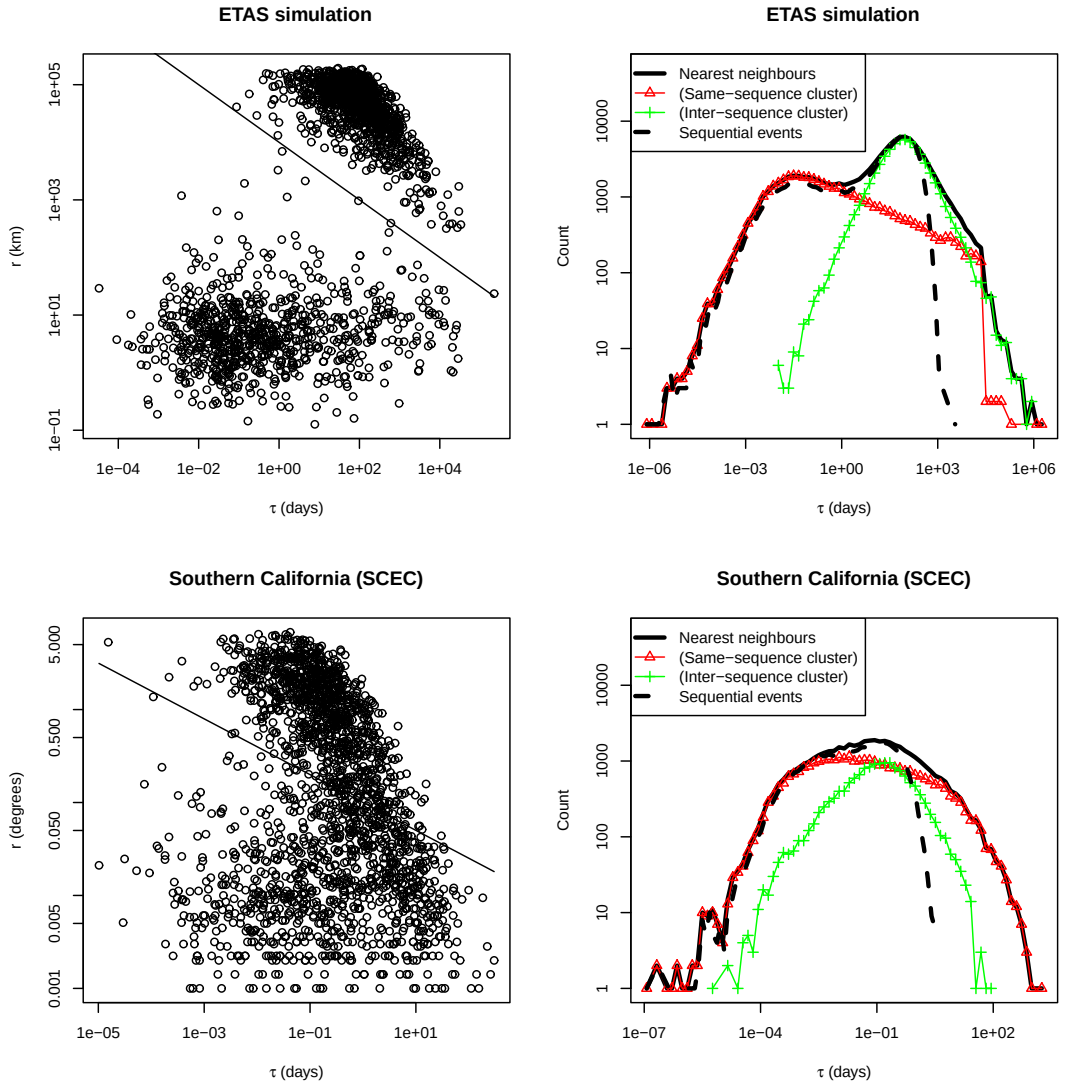


Figure 5.11: Left: space (r) and time (τ) distances to the nearest spatio-temporal neighbour for each event of a spatial ETAS simulation (top) and a Southern California catalogue (bottom). The first 2000 events of each catalogue only are shown. Right: histogram of time intervals to the nearest neighbour (solid black line). Same-sequence and inter-sequence event pairs are superposed, in red triangles and green crosses respectively, classified according to which cluster they fall into as defined by the dividing lines shown in the left figures. Also shown as a dashed black line is the straightforward inter-event time histogram. Note that the intervals are not identical, as the spatial information allows more same-sequence events to be selected as nearest neighbours despite not being sequential in time alone. ETAS parameter values used are $\mu = 0.01$, $A = 5$, $\alpha = 1$, $c = 0.01$, $p = 1.2$, $d = 2$, $q = 1.5$; SCEC events between 1984 and 2000 of magnitude 2.2 and above were selected.

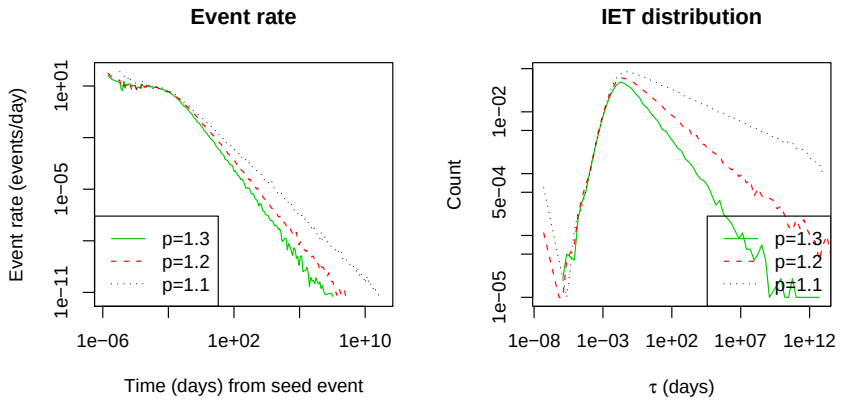


Figure 5.12: The effect of varying the ETAS parameter p on: the average global aftershock sequence (event rate against time, left); and the distribution of inter-event times τ within the sequence (average histogram, right). Other parameters were held fixed at $\mu = 0.01$, $\alpha = 1.1$, $c = 0.01$ and $A = 5$.

5.12 also shows a histogram of the inter-event times generated from the simulated aftershock sequences in the right-hand plot. Note that in the absence of further seeding events, the power-law decay in this distribution component has—in principle—no upper cutoff; in practice, it is truncated at a point dependent on the temporal length of each simulation run, which was arbitrarily set. We can see here that the requirement for p to be small in order to observe an apparently universal gamma distribution, from Saichev and Sornette (2007), works by ensuring a shallow power law and thus reducing the appearance of bimodality.

When ETAS parameters are inverted for a real catalogue, events below the completeness threshold are first removed, and then the unjustified assumption is implicitly made that the completeness threshold m_c coincides with the smallest possible event m_0. Triggering relationships that exist between events on either side of the detection threshold are thus hidden by the removal of the small events. Sornette and Werner (2005b) have shown that this unavoidable false assumption results in a renormalised set of parameter values and an underestimation of the true branching ratio. The upshot of this is that the magnitude threshold m_c can be thought of as a tuning dial for the effective branching parameters; the higher it is set, the lower the apparent branching ratio is. (It of course alters the effective spontaneous event rate also.)

We now examine the effects of these two aspects—seeding and aftershock generation—together. We combine a selection of different branching parameterisations with firstly, a low μ value and secondly, a high μ value, and simulate full catalogues to determine the combined effect.

The histograms of inter-event times for the low μ catalogues (figure 5.13, left-hand plot) are all bimodal. We see that the variability between global aftershock sequences carries through to the full catalogues, causing the shapes of the curves for low μ to differ significantly from each other. Perhaps counter-intuitively, the position and height of the second (exponential) peak also varies with the branching parameters, even for constant μ; this is because it reflects the event rate within the average sequence at the time where overlap occurs, and so does not depend only on the spontaneous rate.

The histograms on the right in figure 5.13 are derived from the high-μ simulations and

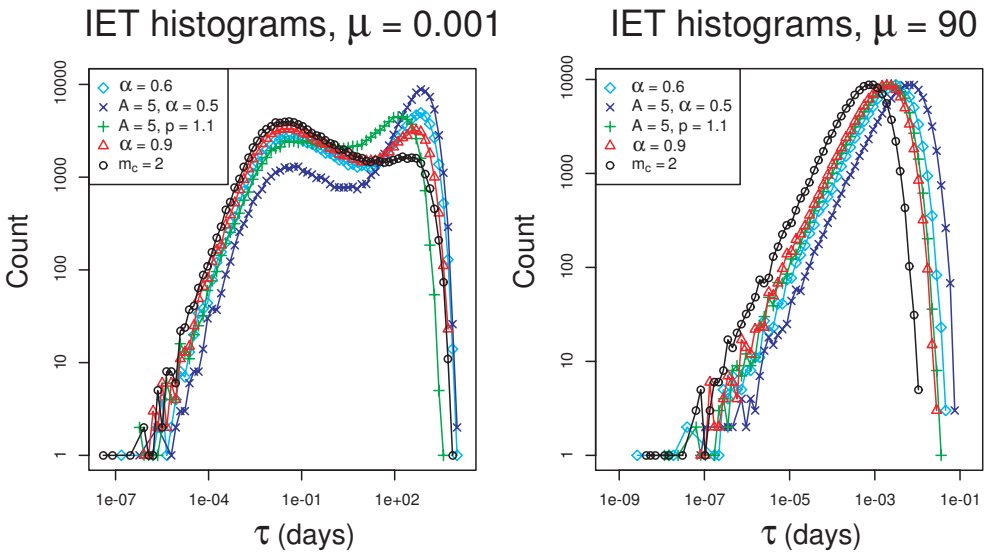


Figure 5.13: Histograms of inter-event times τ from ETAS simulations of 100000 events each, created using five different parameterisations, coupled with a low μ (left) and a high μ (right); different shapes at low μ become very similar shapes at high μ. Other ETAS parameter values were as stated in legend, with unstated values defaulting to $A = 10$, $\alpha = 1.1$, $c = 0.01$, $p = 1.2$, $b = 1$ and $m_c = 0$.

represent extreme examples where the power law is no longer visible. This is a consequence of the significant overlapping of aftershock sequences as discussed in section 5.3.3. By contrast with the low-μ case, it is hard to distinguish between the sets of parameters used for the high-μ simulations besides a shift of the distribution along the τ-axis. Each curve could conceivably be created by a wide range of sets of parameter values, because the effect of the branching is 'hidden'.

This point is underlined when we look at normalised and rescaled versions of these histograms (figure 5.14), as is typically done when looking for universal data collapse (e.g. Corral, 2003). Strong data collapse is observed at high μ, confirming that in that regime, the different parameterisations of the model result in inter-event time distributions that are described by the same underlying function (an exponential). No such data collapse is observed at low μ, though visually this may seem an attractive interpretation from noisier (real) data.

This strongly suggests that the uncertainty in all parameters varies systematically with the actual value the parameter μ takes relative to the others. We will now proceed to confirm this transition towards indistinguishability in branching characteristics as μ increases, through measurements of the accuracy of ETAS parameter inversions.

It is common to analyse earthquake catalogues by inverting ETAS parameters from the catalogue using the maximum-likelihood method. The log-likelihood for a point process is given by

$$\log L = \sum_i \log \lambda(t_i|H_t) - \int \lambda(t|H_t)dt \tag{5.9}$$

where λ is the conditional intensity function, t_i refer to times of events, and H_t is the history of the process at the time t (Harte, 2007b). Optimisation is typically performed by minimising

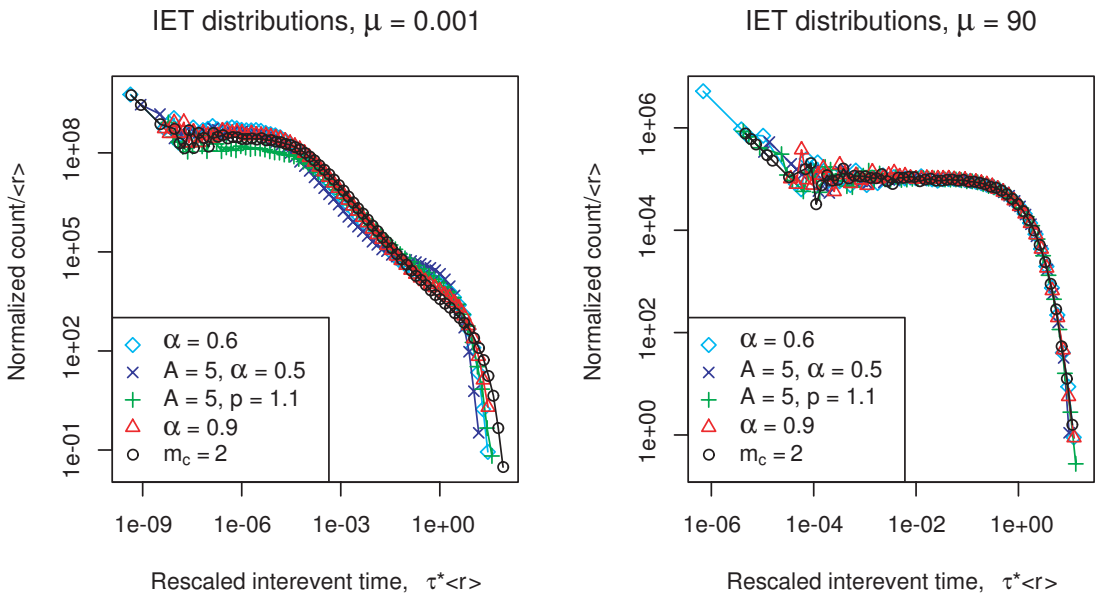


Figure 5.14: Normalised PDFs rescaled by the mean event rate, using the data from the previous figure. Variation at low μ becomes a data collapse at high μ.

the negative log-likelihood with a Newton-type algorithm.

In this study, we invert ETAS parameters for synthetic catalogues generated by the ETAS model; unlike with real data, we can use our knowledge of the input parameter values to evaluate how well the inversion algorithm performs under various circumstances. We first test the algorithm on a synthetic catalogue of 50000 events created using typical ETAS parameters. Rather than invert parameters for the whole catalogue, we divide it into samples of a particular number of successive events, and invert for each sample. This can then be repeated using different sample sizes, giving a spread of values and corresponding standard errors from which we can determine the sample size required to give an accurate inversion.

Figure 5.15 (a) shows, for each sample size, two types of mean error: firstly, the standard error, which is the square root of the appropriate diagonal entry in the approximate covariance matrix, averaged across all inversions; and secondly, the mean absolute difference between the inverted value and the true value, $|x - x^*|$ (where x represents a parameter). The mean of $|x - x^*|$ is generally just below the mean standard error, indicating that on average the standard error is a reasonable estimate of the actual error. Throughout much of the range of sample sizes, a power law relationship between error and sample size is apparent. The exponent is similar for all parameters; the curves are vertically offset in proportion to the values of the parameters. However, for c and p, both types of error are highly elevated for small sample sizes. The inversion procedure is failing to converge for some few samples and getting 'lost', which is greatly influencing the average error. In practice these inversion results would of course be rejected, but we include them here to show that they exist, and clearly *don't* exist at larger sample sizes. This is presumably because, in a small sample of events, the temporal decay of the aftershock rate is highly uncertain and could be fit by a wide range of pairs of values in these parameters which have a clear covariance. At large sample sizes, variability in the error curve increases, simply due to obtaining fewer samples from the catalogue over which to compute the

average.

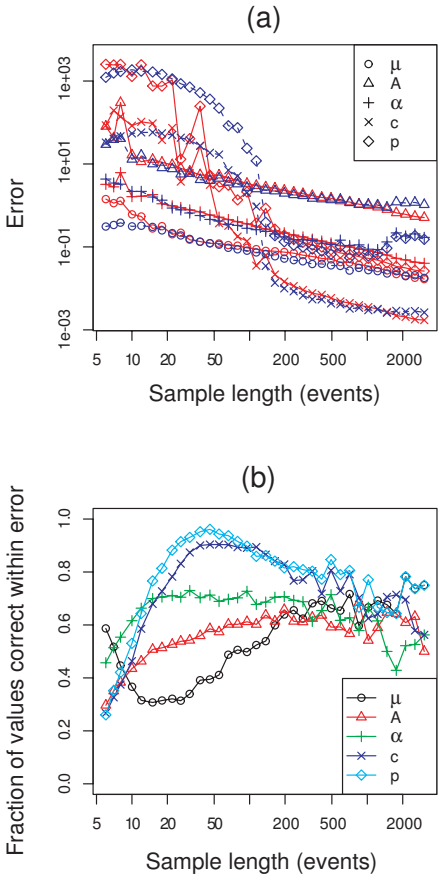


Figure 5.15: (a) Comparison of mean standard error e (red, solid line) with mean absolute difference between the true value x and its inverted estimate x^* (blue, dashed line) for each parameter x as a function of sample size, for inversions of ETAS parameters from a synthetic ETAS catalogue. (b) Fraction of inverted values for which the true parameter value is within the bounds of the inverted value $\pm$ the standard error, as a function of sample size. The ETAS parameters used to create the catalogue were $\mu = 0.07$, $A = 5$, $\alpha = 1.1$, $c = 0.01$, $p = 1.1$.

Figure 5.15 (b) shows the fraction of inverted values that are correct within their standard error as a function of sample size. This is best understood as a measure of how good the standard error estimates are. At large sample sizes, for all parameters, around two thirds of the inversions produce values with error bounds which contain the true value. The algorithm is therefore producing a standard error comparable to one standard deviation, which is the 67% confidence limits. At smaller sample sizes however, the curves separate out, with the error estimate being particularly incorrect for μ (too low) and c and p (too high). This separation of the parameters is consistent across similar inversions carried out with other synthetic catalogues (not shown), and can be understood in terms of under-sampling of background events (for the inversion of μ) and of limited-duration aftershock sampling (since c and p concern the temporal dimension of aftershock occurrence).

Both of these plots, figure 5.15 (a) and (b), indicate that a sample size of 1000 events is able to support the inversion of parameters with acceptable error estimates. We use this sample size to now explore the effect of the bimodality or otherwise of the inter-event time distribution on

the invertibility of parameters, by performing inversions on synthetic catalogues with different μ values. Each catalogue is created with 100000 events, which are then divided into 100 samples of 1000 successive events each; parameter values are inverted for each sample. Rather than start each inversion off at some common set of initial parameter values, we eliminate the possible effect of initial proximity to the solution—which would not otherwise be constant—by starting each inversion off at the true solution.

Figure 5.16 (a) shows the two types of error again, this time as a function of μ, at constant sample size. They are again averaged over the 100 samples for each catalogue. Both types of error increase with μ, particularly steeply in the case of the error in inverted μ. The error in c and p also increases dramatically for larger μ, similarly to the case for small sample sizes in figure 5.15.

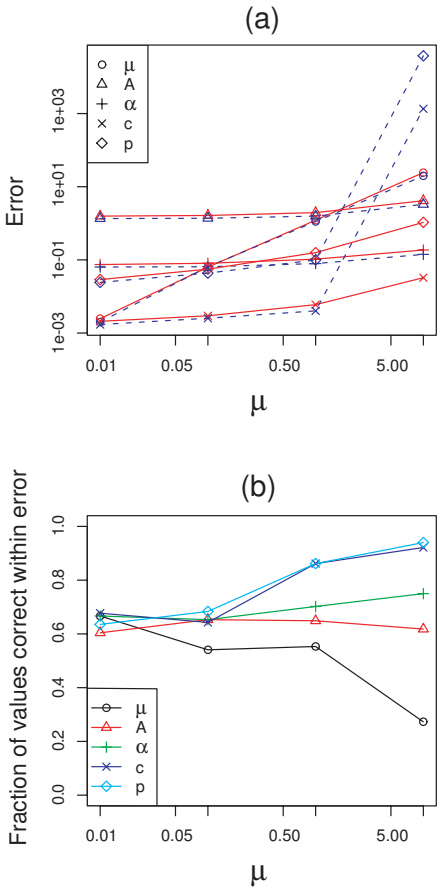


Figure 5.16: (a) Comparison of mean standard error e (red, solid line) with mean absolute difference between the true value x and its inverted estimate x^* (blue, dashed line) for each parameter x as a function of the true μ value, for inversions of ETAS parameters from four synthetic ETAS catalogues of different μ (see the corresponding inter-event time distributions in figure 5.6). (b) Fraction of inverted values for which the true parameter value is within the bounds of the inverted value $\pm$ the standard error, as a function of the true μ value. Other ETAS parameters were held fixed at $A = 10$, $\alpha = 1$, $c = 0.01$, $p = 1.2$. Inversions were started at the true solution.

In figure 5.16 (b) the fraction of inverted values that are correct within the standard error, for all parameters, is around two thirds for low μ as our previous inversion exercise led us to expect for a sample size of 1000. As μ is increased, however, the fraction of values correct

within error starts to deviate from this value, indicating that just as for small sample sizes, large μ values produce inversions with inaccurate error estimates. The pattern of separation across the parameters is identical to the small sample sizes case, with the estimated error being too low for μ and too high for c and p.

In order to establish beyond doubt that the parameter estimations are truly converging, we recorded the values of the estimates at each iteration, along with the gradients of the log likelihood function, for 12 of the samples at each value of μ. Figure 5.17 (a) shows an example for the lowest value of μ, which is representative of the pattern observed in every inversion for that μ value. The log-likelihood gradients go towards zero and the parameters—which start out at their true values—do not wander far from those values. This confirms that the iteration procedure has converged.

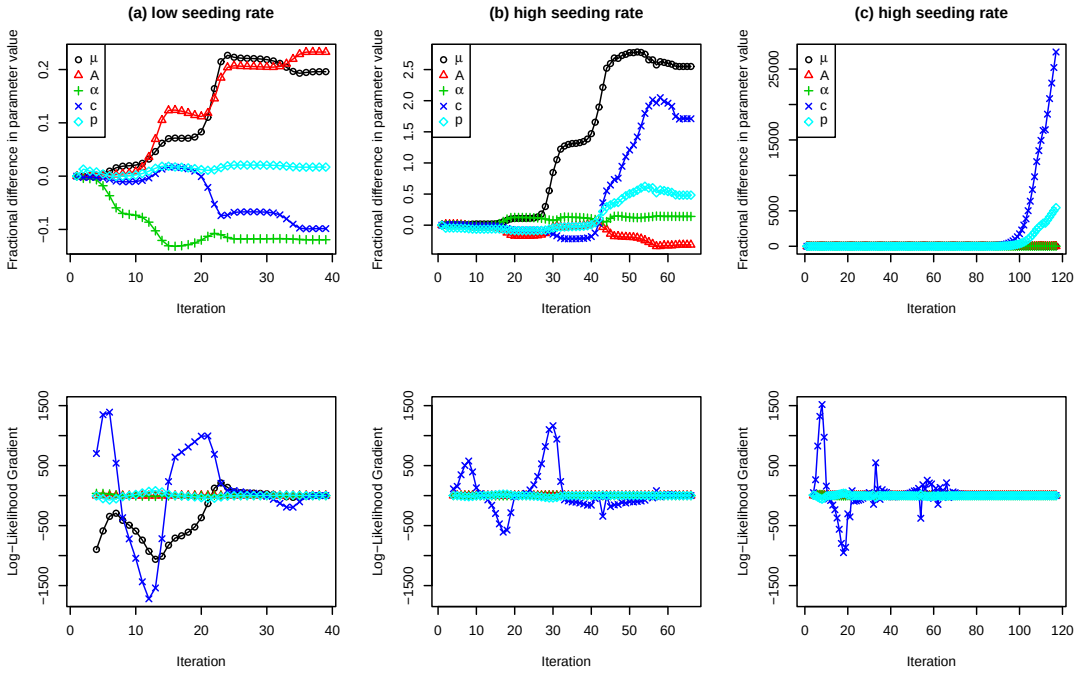


Figure 5.17: Parameter estimates (as a fractional difference from the initial, true value) and log-likelihood function gradient at each iteration of an inversion for (a) one of the samples of the lowest μ value from figure 5.16, and (b,c) one of the samples of the highest μ value. Likelihood gradient is not shown for the first three iterations where it fluctuates strongly. Towards the final iterations, parameter estimates are stable and gradients go to zero for (a) and (b), although in (b) the parameter values have moved further from their (true) starting values. In (c) the inversion terminated with excessively large steps being taken in one parameter, so this inversion failed to converge, although the final gradients are near zero.

At high μ, there were two distinct scenarios in the examples. The first and most common scenario, occurring in 9 out of 12 inversions, was similar to the situation for low μ and is shown in figure 5.17 (b). The log-likelihood gradients similarly tend towards zero, and the parameter values take only small steps towards the end of the inversion, although they have moved far from their starting points in some cases. This situation also represents convergence, albeit towards a false solution. The second scenario is where the inversion terminates due to the maximum step size being exceeded on one parameter (figure 5.17 (c)), although the log-likelihood gradients are

also very low. We can infer that the likelihood surface is nearly flat in one dimension, leading that parameter rapidly away from the true value. This cannot be classed as convergence; however, it is a situation that only occurs at high μ, which is an important observation.

The differences in the inversion outcomes as μ is varied are consistent with the fact that as μ is increased, the overlapping of aftershock sequences causes earthquake occurrence in the catalogue to look more like a Poisson process in time (figure 5.6), which hides the effect of the branching parameters on the time series (figure 5.13). We suggest that this makes the origin of each event more ambiguous, and profoundly, we see here that this is reflected in poor inversion success. Even with a catalogue produced by—and thus perfectly described by—the ETAS model, and with the ability to start the inversion at the true solution, luxuries that real data does not afford us, the error is still significant for high μ and the error estimates produced by the inversion are significantly incorrect.

The situation may be improved with new more accurate inversion methods such as the Expectation Maximisation (EM) method ofVeen and Schoenberg (2008), but our primary concern is to test current standard practice and point out where and why it fails. These issues may also apply to alternate techniques, since—as we have illustrated—in principle, finding independent events in data where many aftershock sequences overlap is inherently problematic, so any improvement is unlikely to be first order.

To illustrate the potential consequences of underestimating the error, we consider the following example. Hainzl and Ogata (2005) inferred fluid signals in Vogtland earthquake data in terms of a varying seeding rate, by performing a series of ETAS inversions on the data using a moving time window. Here we test the alternate hypothesis that an apparent variation in μ may be observed in synthetic ETAS catalogues with constant seeding rate, purely as a consequence of overlapping aftershock sequences.

We repeat their analysis using synthetic catalogues, comparing the lowest and highest μ values used in our inversion accuracy analysis. Like Hainzl and Ogata, we invert only for μ while holding all other parameters fixed at their true values—a procedure which might reasonably be expected to produce very reliable estimates of μ. We use, in each case, a window size (in days) calculated to give an average of 400 events per window, which is the approximate mean number of events in Hainzl and Ogata's windows; although the full seismic history is included in $\lambda(t|H_t)$ in equation 5.9, the sum and integration are performed only over the time window. Hainzl and Ogata do not specify the step size, but we make the step size equal to the window size so that each point in the graph represents a completely different sample with no overlap. Figure 5.18 shows the results of this series of inversions. The error bars shown are, in this case, two standard errors (i.e. two standard deviations, or 95% confidence limits).

At both values of μ, a variation in the inverted μ values occurs, which mirrors the variation in the overall event rate as in figure 7 of Hainzl and Ogata (2005). The actual number of spontaneous background events in each time window—divided by the window length—is shown as a dotted line, which (unsurprisingly) remains in the vicinity of the true μ value and does not vary to the same extent as the inverted values. We would therefore interpret these type of inversion results with caution, and accept that while there may be good reason to reject a single ETAS parameterisation for the Vogtland swarm data based on the additional analysis of Hainzl and Ogata, particular parameter values for an ETAS inversion cannot be regarded as too reliable.

In addition, we notice that for our high μ simulation, the inverted value is consistently

and significantly above the true value, showing that a bias is introduced in the inversion due to highly overlapping aftershock sequences. This is all the more surprising given that it is a one-dimensional inversion with all other parameters forced to be correct. We understand this as further confirmation that aftershocks are being mistaken for spontaneous seeding events due to the overlapping of sequences and the consequent destruction of their temporal signature. This effect naturally entails an underestimation of the branching ratio. In figure 7 of Hainzl and Ogata (2005), the inverted μ values for the case where all parameters are allowed to vary is consistently higher than those values where only μ is inverted. This too is suggestive of the overestimation of μ we have shown can happen when aftershock sequences overlap.

In summary, we have demonstrated clearly that the ETAS maximum likelihood inversion scheme, under specific (high) values of the seeding rate, produces a bias in the inverted seeding rate (and hence in the branching ratio also), and an underestimate of its uncertainty. For a finite recorded period, catalogue size can be adjusted by varying the size of the included region; this however is akin to varying the effective μ and so it must be borne in mind that while increasing the amount of data tends to improve the inversion success, the effect of increased temporal overlapping of aftershock sequences has the opposite effect. We suggest that the absence of discernible bimodality in the inter-event time distribution is likely to preclude accurate inversion of temporal ETAS parameters. Studies which seek to demonstrate a varying background rate must establish that the variation lies outwith the substantial error bounds caused by overlapping aftershock sequences; triggering is a difficult null hypothesis to reject.

5.3.6 Conclusions

We have explored, within the framework of the ETAS model, the effects of the spontaneous event rate and the aftershock triggering parameters on the earthquake inter-event time distribution. We have shown, using data from the PDE and SCEC catalogues, that this is fundamentally a bimodal distribution, and that the degree of observable bimodality depends on the region size of the data set. Our ETAS analysis demonstrates that the bimodality can be explained in terms of two populations of inter-event times: those originating from same-sequence and inter-sequence event pairs, respectively. These populations depend on the extent of temporal overlap of separate aftershock sequences, and so the dependence of the distribution shape on region size is essentially a dependence on seeding rate. This observation forces us to reject the hypothesis proposed in earlier literature of a universal scaling law for earthquake inter-event times: the small inter-event times, which are dominated by same-sequence aftershocks, do not generally scale to the large ones, although the distribution may be seen to approximate a universal gamma form under the constraints of visually stationary periods, low p values and inter-event times much greater than c. Our analysis thus represents a generalisation of Corral's and Saichev and Sornette's work, taking it beyond these constraints.

The statistics of the ETAS model are a realistic representation of real data, even for large region sizes—which are outside of the model's typical range of application—although inversion of parameters for large regions is problematic. Large regions occupy a high-spontaneous-rate subset of parameter space, in which the inter-event time distribution becomes unimodal and exponential-like due to the temporal overlapping of aftershock sequences; parameterisations within this high-overlap regime have a high redundancy in terms of the shape of the distribution. As a result of this, maximum likelihood inversion of parameters from the simulated

catalogues becomes less successful as the spontaneous event rate is increased, with mean parameter estimates across realisations becoming further away from the true value, and standard errors becoming less reliable and on average larger. The spontaneous rate is also systematically over-estimated within this high-overlap regime. This introduces new challenges in determining both the reliability of parameter inversions from real data and the suitable range of region sizes for which temporal ETAS parameter inversion may be successful, as well as in the ability to detect significant temporal variations in parameters such as the spontaneous rate, for example in time-dependent hazard calculations.

5.4 Supplementary details

Figure 5.17 in the paper presents the progression of the inversion algorithm for a selection of samples from the catalogues with the lowest and highest μ values, to indicate typical outcomes of the inversion. These plots were introduced to address a reviewer query over whether the inversions had really succeeded in finding the maximum likelihood solution. They show that the number of iterations in each case is large enough to preclude the possibility that starting at the true solution was causing the algorithm to fail immediately. The gradients approach zero in every case, which further supports the conclusion that the inversion is reaching a solution. In figures 5.19 to 5.24 below, I show the results for all samples at each of these two extreme μ values, for the sake of further completeness than was possible in the published paper.

In terms of the space–time nearest-neighbour plots (figure 5.11) in the paper), alternative space–time distance metrics were experimented with at the suggestion of a reviewer. A term involving magnitude was included as one possibility. It was found that the simple distance $r\tau$ did as good a job as any other at separating the points into two clusters, so this was used.

Another part of the discussion with reviewers involved clarifying that the aggregation of many independent processes, for example the temporal overlapping of aftershock sequences in the case of high μ, produces a Poisson process with an exponential distribution of inter-event times (Palm–Khintchine theorem, e.g. Heyman and Sobel (2004)). This is indeed the case but it is counter-intuitive to many seismologists who think that the correlations in aftershock sequences must be retained in the time series even at high seeding rates. This work therefore has made valuable points and could change the way this process is understood, perhaps leading to novel methods of constrained inversion to correct (to some extent) for the inversion bias identified in the work.

The use of the time-only ETAS model (as opposed to a model with a spatially-varying background rate) in an analogue for the whole Earth was questioned in reviews. As explained in the paper, this is meant as a qualitative analogue only, for theoretical purposes, and not one with practical use. However since space and time components of seismicity are considered separable and independent, the spatial inhomogeneity of the background process does not affect its temporal rate. One can think of the Earth consisting of small sub-regions for which the background process is approximately homogeneous, each sub-region having a different background rate in time. The sum of these Poisson processes of different values will be yet another Poisson process.

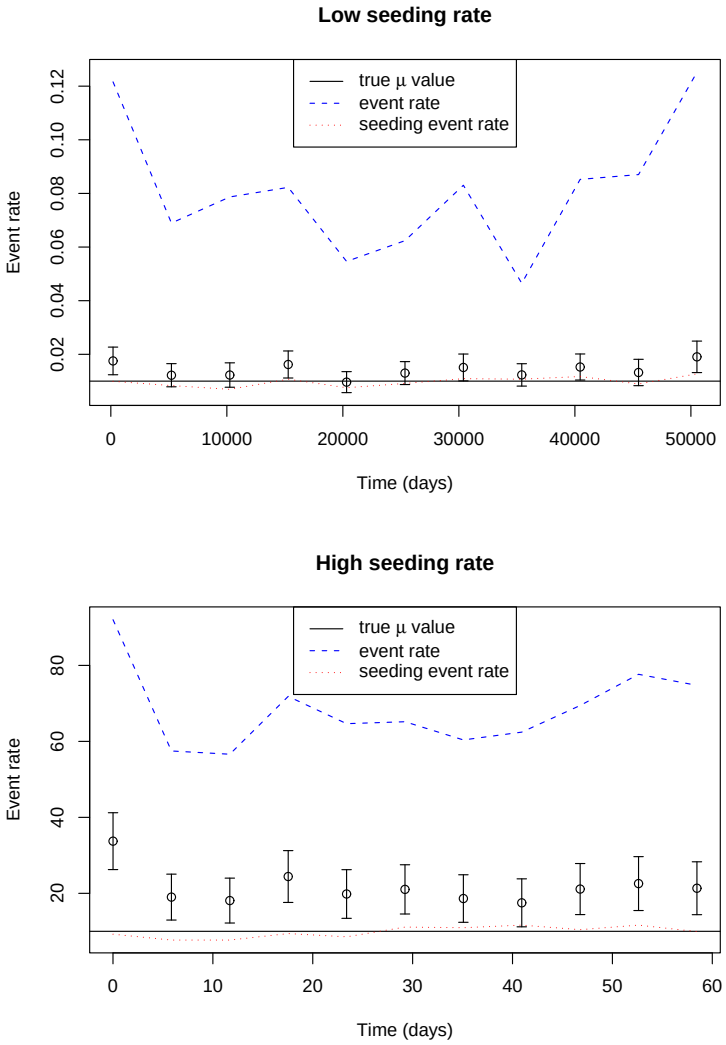


Figure 5.18: Circles with error bars show a series of inversions of μ, holding all other parameters fixed at their true values, from a low μ (top) and a high μ (bottom) synthetic catalogue. In each case the window size was calculated to give an average of 400 events per window, and the windows do not overlap. Variation in the inverted values follows the variation in the event rate, demonstrating that an apparent variation in seeding rate can spuriously occur from a catalogue with a constant seeding rate. Additionally, μ is over-estimated when the true seeding rate is high, due to temporal overlapping of aftershock sequences.

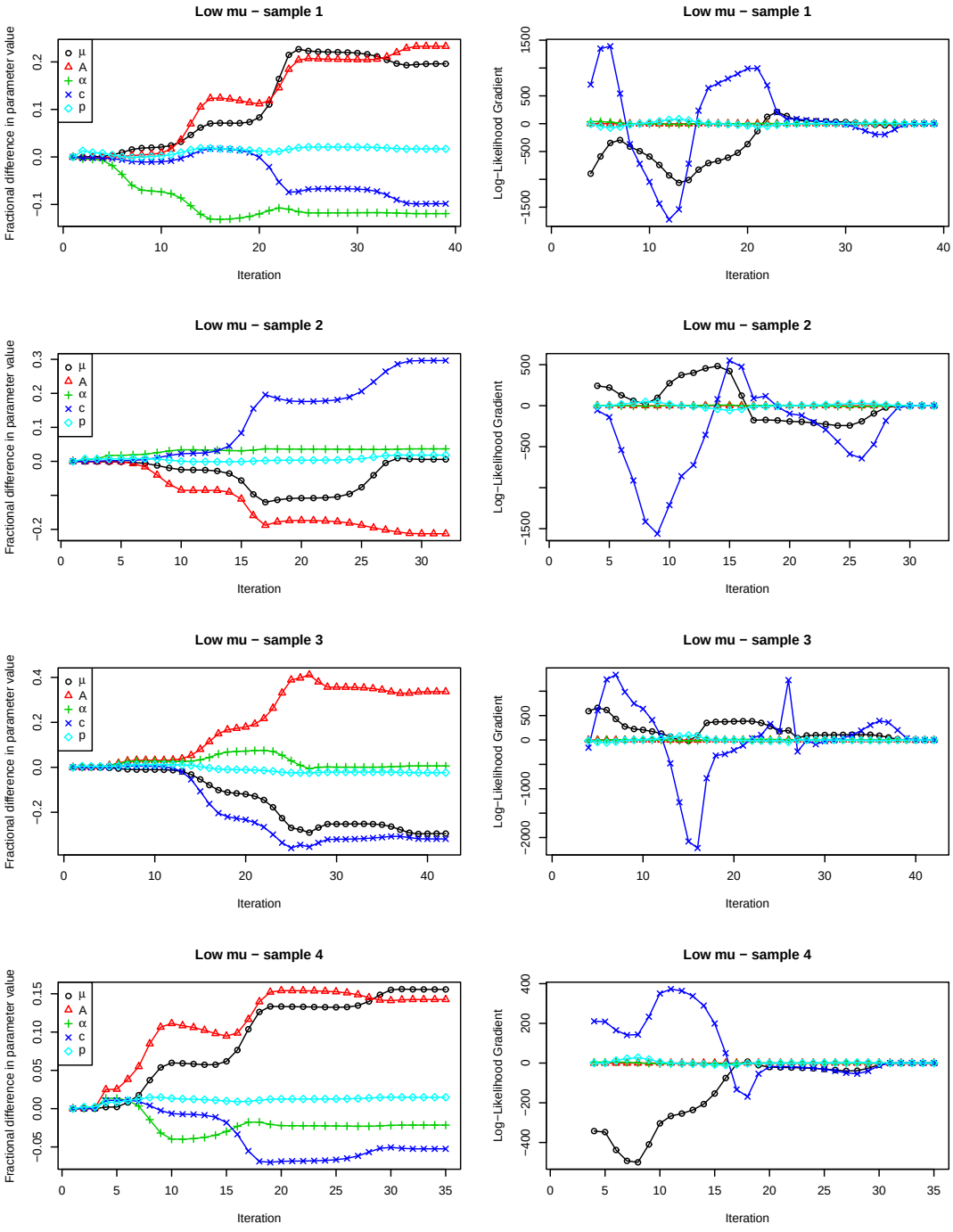


Figure 5.19: Iterations and likelihood gradients for samples 1–4 of the lowest-μ catalogue ($\mu = 0.01$). The gradient goes to zero and the parameter values do not wander far from their starting (i.e. correct) values.

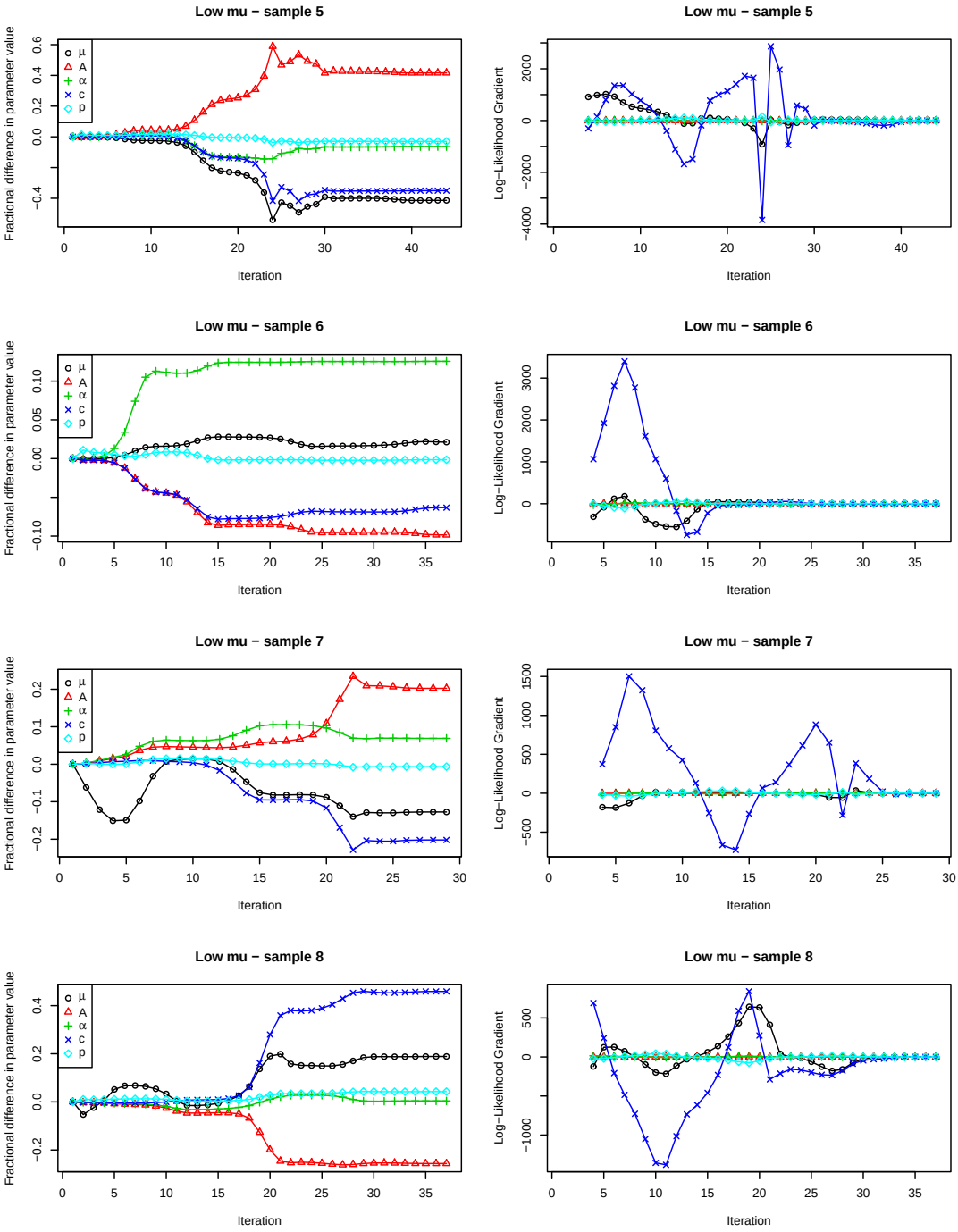


Figure 5.20: Iterations and likelihood gradients for samples 5–8 of the lowest-μ catalogue ($\mu = 0.01$). The gradient goes to zero and the parameter values do not wander far from their starting (i.e. correct) values.

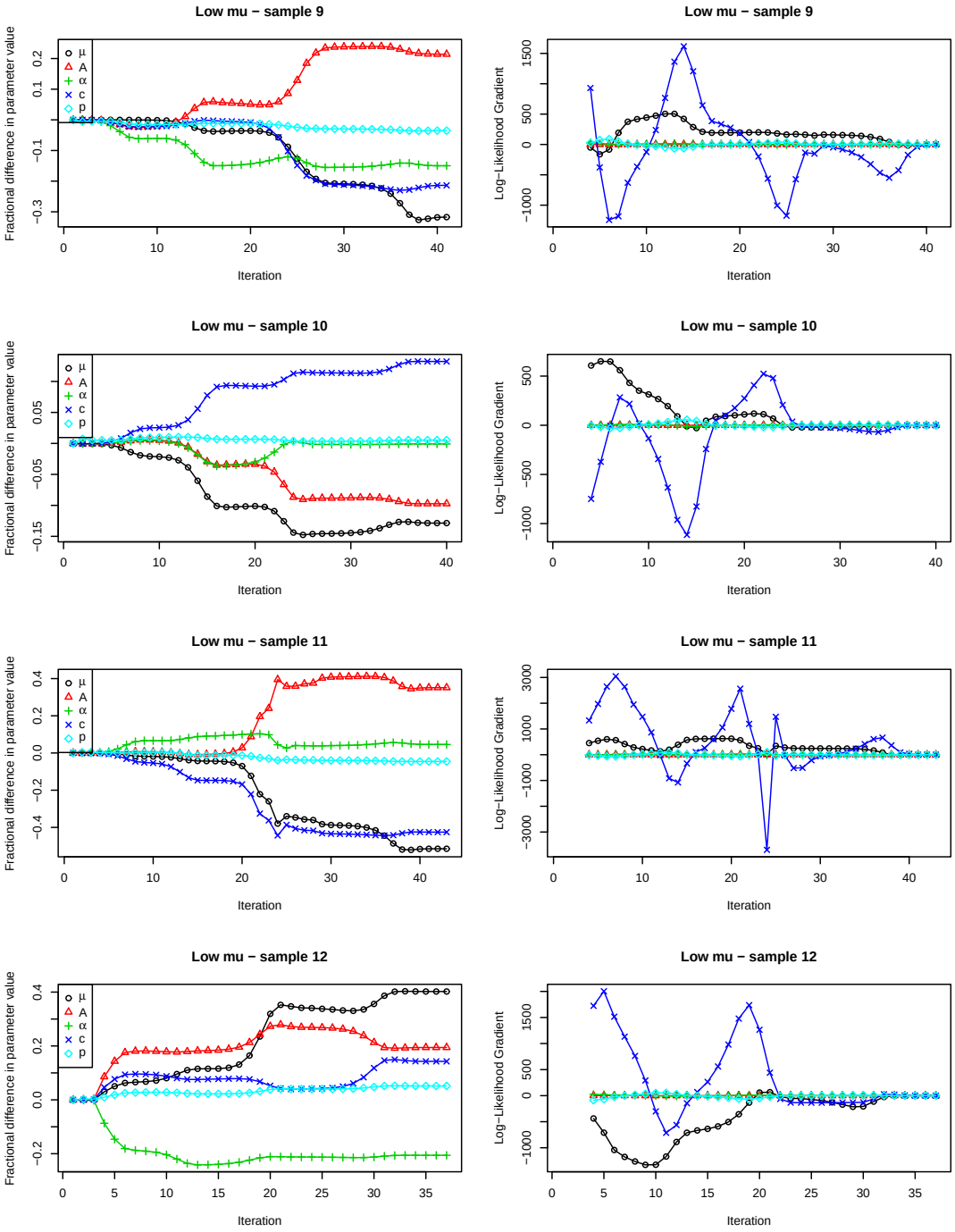


Figure 5.21: Iterations and likelihood gradients for samples 9–12 of the lowest-μ catalogue ($\mu = 0.01$). The gradient goes to zero and the parameter values do not wander far from their starting (i.e. correct) values.

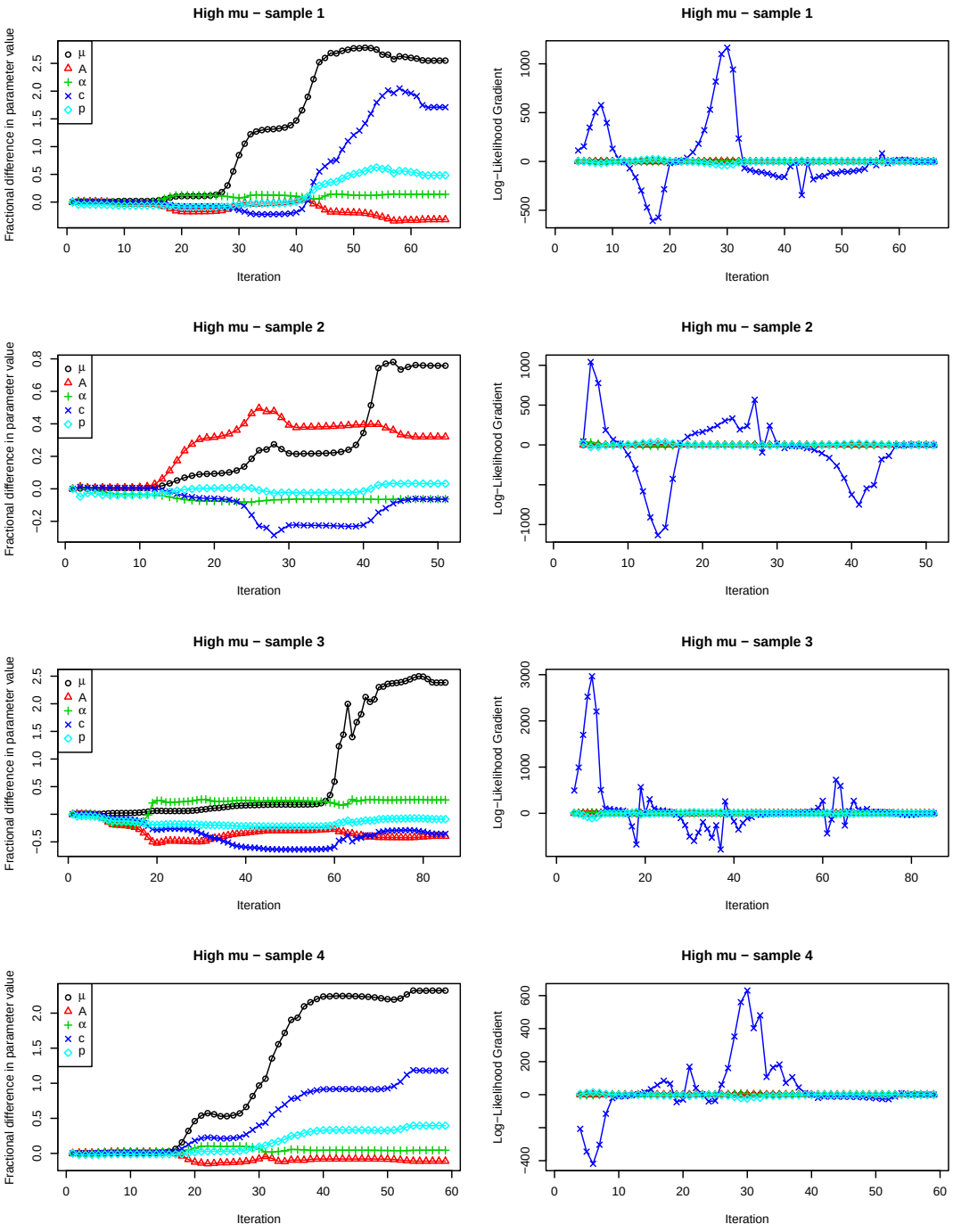


Figure 5.22: Iterations and likelihood gradients for samples 1–4 of the highest-μ catalogue ($\mu = 10$). The parameter values end up much further from their starting (i.e. correct) values.

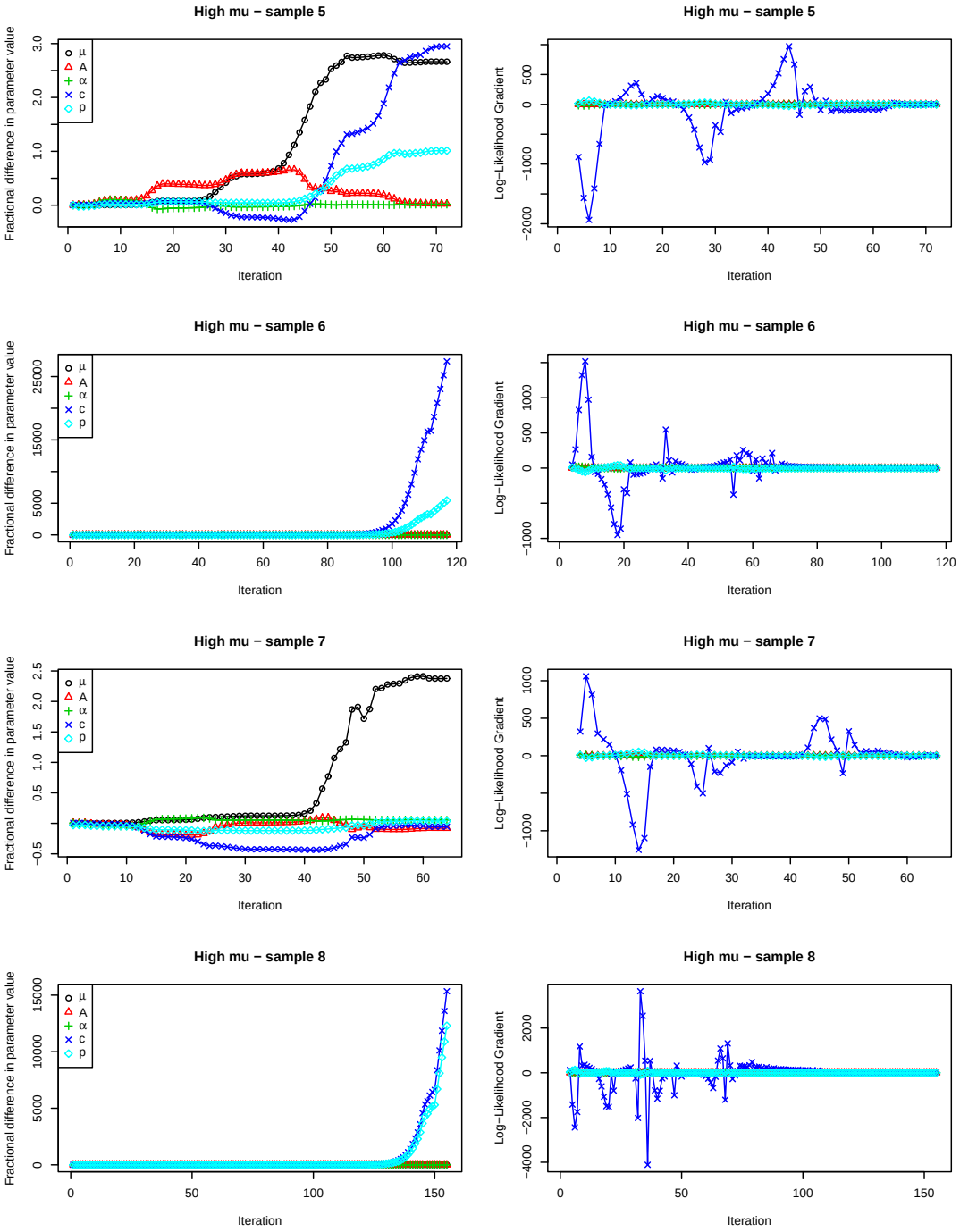


Figure 5.23: Iterations and likelihood gradients for samples 5–8 of the highest-μ catalogue ($\mu = 10$). The parameter values end up much further from their starting (i.e. correct) values, with samples 6 and 8 getting 'lost'.

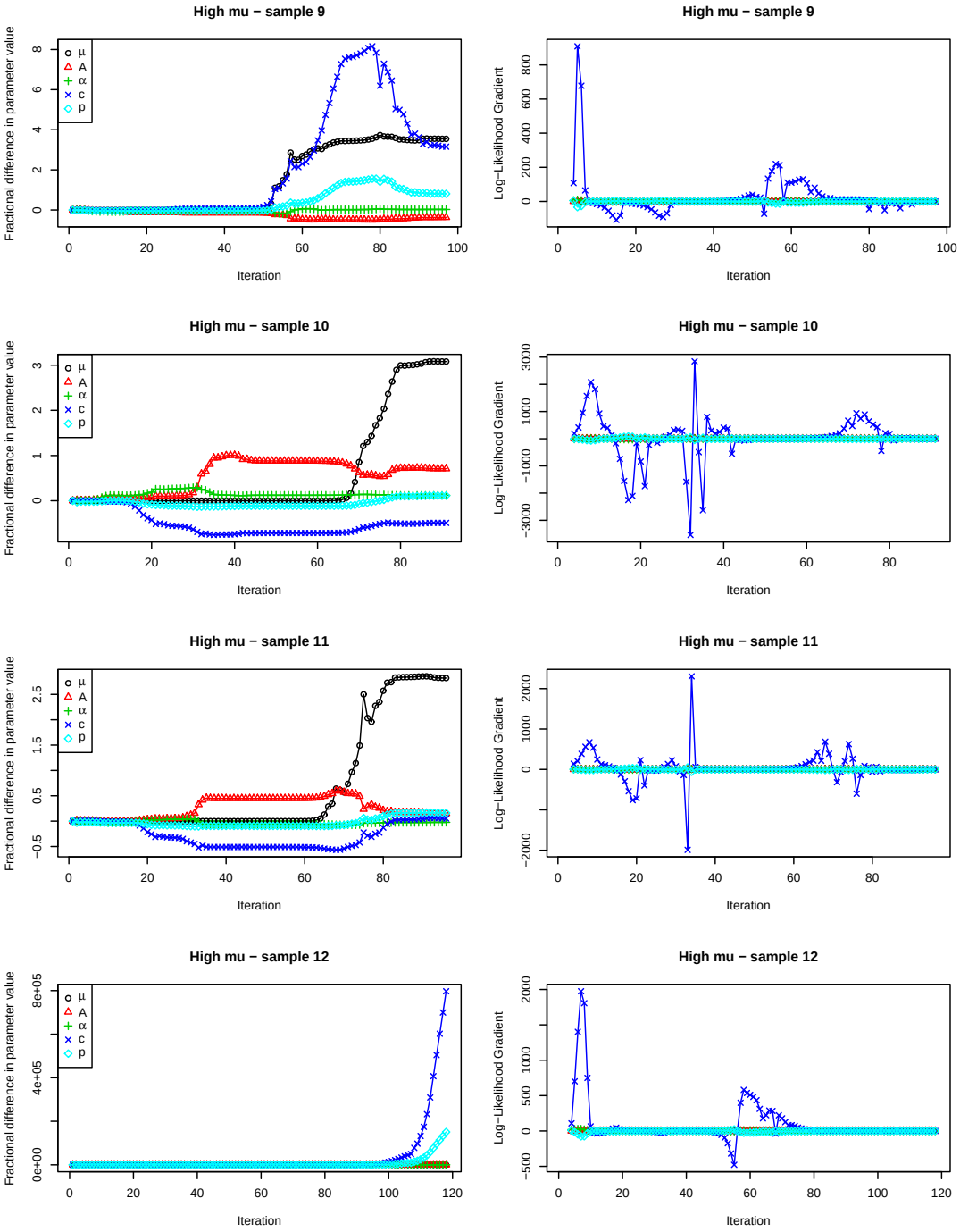


Figure 5.24: Iterations and likelihood gradients for samples 9–12 of the highest-μ catalogue ($\mu = 10$). The parameter values end up much further from their starting (i.e. correct) values, with sample 12 getting 'lost'.

Chapter 6

The Nature and Origin of Inter-Event Time Distributions for Volcanic Earthquakes

In the preceding chapters I presented and justified a theoretical framework for understanding the inter-event times of tectonic earthquakes. In this chapter the ideas are extrapolated into a non-tectonic setting in order to explore whether they are relevant there. Recent literature on volcanic earthquake inter-event times has made similar claims of a universal gamma distribution as have been made for tectonic earthquakes. Given that there are Omori-type correlations and magma-driven triggering correlations in these datasets, it seems likely that the inter-event time distribution will have some similar features to the tectonic case and not be universal in reality. A more general model however needs to include a wider range of processes than just background and aftershocks. In the manuscript reproduced in section 6.1 we use the same philosophy of classifying the intervals according to their type as outlined in chapter 4, in order to derive insight into these processes and their effects on the time series in different component parts of active volcanoes. At the time of writing the manuscript is in review with the Journal of Geophysical Research.

My main role as second author in this paper was to contribute the basic philosophy of inter-event time distribution non-universality and of understanding inter-event times through classifying them as different types.

6.1 Manuscript

There has been much recent debate on the form and physical origin of the inter-event time distribution for earthquakes, which in turn has implications for the interpretation of earthquake interactions and time-dependent seismic hazard. Motivated by empirical studies, the main question that has been addressed is whether inter-event time distributions for different catalogues are consistent with a single 'universal' model, i.e. the distribution of earthquake inter-event times is independent of all other local properties such as aftershock productivity, magnitude-frequency distribution and background rate, and can be represented by a generalised gamma

function. Recently, the argument for a universal inter-event time distribution has been supported by claims that this model is also able to explain the inter-event time distribution of earthquakes at volcanoes. Here we test the hypothesis that the inter-event time distribution for volcanic earthquakes is universal. We analyse the well-characterised seismicity from the volcanoes on the island of Hawaii and demonstrate that the inter-event time distribution is instead both spatially and temporally variable. In particular, in space and time windows where magmatic processes (such as dyke injection and pressurisation of the magma chamber) control the stress field, inter-event time distributions cannot be modelled by a simple gamma distribution and are frequently bimodal. This dependence on space and time clearly indicates that a universal model is not appropriate for volcanic settings. Rather, we can explain the inter-event times with a model consisting of accelerating and decelerating rates of events preceding and succeeding volcanic activity. The same model is able to explain the inter-event time distributions observed at Vesuvius and Campi Flegrei volcanoes in central Italy.

6.1.1 Introduction

The temporal statistics of earthquake populations underpin earthquake recurrence and hazard models and allow us to quantify the physics of earthquake interactions. One approach for characterising the temporal properties of an earthquake catalogue is through the distribution of time intervals between successive events, and there has been much recent interest in the nature and physical origin of the inter-event time distribution for earthquakes (Corral, 2006b; Hainzl et al., 2006; Saichev and Sornette, 2006; Corral, 2007; Saichev and Sornette, 2007; Bottiglieri et al., 2009a; Touati et al., 2009; Bottiglieri et al., 2010; Traversa and Grasso, 2010). In particular it has been debated as to whether inter-event time distributions for different catalogues are consistent with a single universal model. The term "universal" has different connotations in different scientific disciplines. In the case of the earthquake inter-event times, it has been used in the context of observations that the probability density function for different tectonic regions and magnitude thresholds apparently collapse to a single distribution when simply rescaled by the mean earthquake rate (Bak et al., 2002; Corral, 2003, 2004; Davidsen and Goltz, 2004). In this scenario, the distribution of earthquake inter-event times is independent of all other local properties such as aftershock productivity, magnitude-frequency distribution and background rate, and can be represented by a generalised gamma function:

$$f(\tau) = C\tau^{\gamma-1}\exp(-\frac{\tau^{\delta}}{B}) \tag{6.1}$$

where τ is the product of the inter-event time Δt and the mean earthquake rate λ (Corral, 2004; Hainzl et al., 2006; Saichev and Sornette, 2007), γ is the shape parameter, B the scale parameter, and C a normalisation factor. These observations are supported by an analytical formulation of the inter-event time distribution based on the Epidemic-Type Aftershock Sequence (ETAS) model which approximates a gamma function when the Omori p parameter is close to 1 and the branching ratio n is in the range 0.7–1 (Saichev and Sornette, 2006, 2007). The motivation of a universal model is that the results can be applied without a consideration of the details of the underlying process; if true this would have important implications for the approach to hazard forecasting.

However, many other studies argue that the earthquake inter-event time distribution for tectonic regions is not universal (Lindman et al., 2005; Hainzl et al., 2006; Jonsdottir et al., 2006; Touati et al., 2009). In particular, Touati et al. (2009) use earthquake catalogues synthesised with the Epidemic-Type Aftershock Sequence (ETAS) model with a constant background rate to argue that the inter-event time distribution is more generally bimodal, and best described as an aggregate of two contributions: (1) exponential-distributed inter-event times between independent Poisson background or inter-aftershock sequence events; and (2) Omori law waiting times between causally related intra-aftershock sequence events. In typical tectonic catalogues, these component distributions overlap, giving rise to an aggregate distribution that can, to first order, be approximated by a gamma function (Touati et al., 2011). However, as these properties vary from region to region and cannot be accounted for by rescaling according to the mean inter-event time, such arguments mean that the inter-event time distribution cannot be strictly universal.

Surprisingly, the argument for a universal inter-event time distribution has been supported by claims that this model is able to explain the inter-event time distribution of earthquakes at volcanoes (Bottiglieri et al., 2009a). Volcanic regions frequently display extreme end-member earthquake dynamics, characterised by features such as high earthquake rates, aftershock productivities (Klein et al., 2006) and Gutenberg–Richter b-values (Wyss et al., 2001), and rapidly changing earthquake rates due to non-stationary driving stresses (e.g. as a result of magma movement). Consequently, volcanic seismicity may provide a natural laboratory to understand the role of processes that are more subtle in normal tectonic settings.

Bottiglieri et al. (2009a) analyse VT earthquake catalogues from Hawaii, Vesuvius and Campi Flegrei, and argue that on rescaling by the average earthquake rate, the inter-event times collapse to the same gamma function as non-volcanic regions. On this evidence, the authors argue that volcanic and tectonic seismicity has the same temporal organisation and that the occurrence probability of volcanic earthquakes can be determined using a tectonic model.

Traversa and Grasso (2010) compare the inter-event time distributions of earthquakes occurring during inter-eruption phases and dyke injection events at Mt Etna and Vesuvius. They show that during inter-eruption phases, the distribution of inter-event times can be approximated by a gamma distribution, and find that the proportion of background events (20–40%) is similar to that in non-volcanic areas. In contrast, they find that during two dyke injection episodes at Mt Etna, the gamma distribution can be rejected, and argue that this is due to a strong external forcing.

Here we test the hypothesis that the inter-event time distribution for volcanic earthquakes at Hawaii, Vesuvius and Campi Flegrei is universal. We analyse the well characterised seismicity from the volcanoes on the island of Hawaii and demonstrate that the inter-event time distribution is spatially and temporally variable. In particular, in space and time windows where magmatic processes (such as dyke injection and pressurisation of the magma chamber) control the stress field, the inter-event time distribution cannot be modelled by a simple gamma distribution. This dependence on space and time clearly indicates that a universal model is not generally appropriate for volcanic settings. Rather, we can explain the inter-event time distribution by a model consisting of accelerating and decelerating rates of events preceding and succeeding volcanic activity. The same physical models are able to explain the inter-event time distributions observed at Vesuvius and Campi Flegrei volcanoes in central Italy. As the

underlying physical process influences the nature of the inter-event time distributions, we refute the claim of universality. Understanding of the process is important for making the most reliable forecasts.

6.1.2 Data

The main focus of this study is volcano-tectonic earthquakes reported in the Hawaiian Volcano Observatory (HVO) catalogue between 1st January 1969 and 31st December 2004. During this period, subaerial volcanic activity on the island of Hawaii has only occurred at Kilauea and Mauna Loa volcanoes. Consequently, seismicity of these two volcanoes constitutes the majority of the instrumentally recorded earthquake catalogue (figure 6.1). As the volcanic activity and associated seismicity of these systems has been widely studied, it is possible to use this understanding to explore the spatial and temporal dependence of the inter-event time distribution.

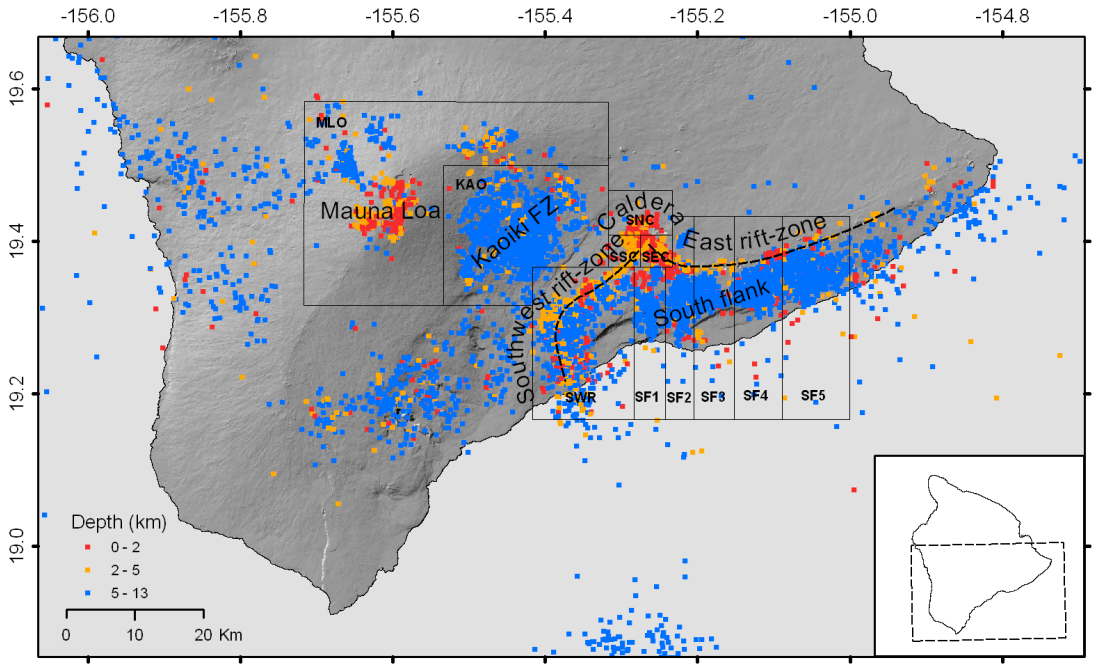


Figure 6.1: Location of shallow (depth < 13km) magnitude 2.5 and above earthquakes in southern Hawaii, 1969–1983. Hawaii volcano observatory seismic regions used for data selection are shown by black polygons (see text for details). Approximate surface expression of Kilauea rift zones are shown by dashed black line.

6.1.2.1 Kilauea

In the years preceding the start of the Pu'u O'o eruption in January 1983, volcanic activity at Kilauea was characterised by frequent, episodic eruptions and non-eruptive dyke intrusion along one or both of two rift zones located on the east and south-west flanks (Klein et al., 1987). These events result from the rapid release of magma from a shallow reservoir located beneath the summit caldera (Tilling and Dvorak, 1993). The rift zones are regions of localised extension that separate the stable interior from the mobile south flank (Delaney et al., 1998),

which accommodates rift-zone opening by seaward movement of up to 40cm/yr (Delaney and Denlinger, 1999). Forty-one eruptions and intrusions are recorded at Kilauea during the study period (Klein et al., 1987). Since January 1983, Kilauea has been in almost continuous eruption, with activity concentrated along the east rift zone. Overall rates of seismicity and deformation rates have fallen (Delaney and Denlinger, 1999).

To investigate the temporal dependence of the inter-event time distribution, we compare two successive time periods at Kilauea: 1st January 1969 to 31st December 1983 and 1st January 1984 to 31st December 2004. 1969 marked the start of consistent earthquake reporting at Hawaii. The division at the end of 1983 approximately coincides with both the change in long-term volcanic behaviour from episodic to continuous and a change in the local magnitude scale (Wyss et al., 2001).

To investigate the spatial variability in the inter-event time distribution, we use spatial filters based on volumes defined by the HVO (Nakata, 2006). The hypocentres of Hawaiian earthquakes are located in segregated 'seismic volumes' that are associated with distinct volcanic processes described above (Klein et al., 1987). At Kilauea, shallow (< 13km) seismicity is dominated by earthquakes within three main volumes (figure 6.1): (1) to depths of 5km beneath the summit caldera, which lies above an enduring shallow magma reservoir, the top of which lies at 3–4km depth (Tilling and Dvorak, 1993); (2) to depths of 5km within the rift zones; and (3) at depths between 5 and 13km beneath the mobile south flank.

For the caldera, we include earthquakes at depths of less than 5km in the shallow east, west and north caldera volumes (SEC, SWC and SNC). For the rift zones, we include earthquakes at depths of less than 5km within the south-west rift zone volume (SWR). Although earthquakes also occur in the east rift zone, they are insufficient in number to analyse separately. For the south flank, we include earthquakes at depths between 5 and 13km within the five south flank volumes (SF1–SF5). Seismicity in these three regions accounts for over 85% of the total earthquakes recorded at Kilauea during the study period. Hypocentre location uncertainty is in the order of 1–2km (Klein et al., 1987), sufficiently small that the vast majority of earthquakes are correctly allocated to their respective sub-volume.

Earthquake frequency-magnitude distributions for selected seismic regions at Kilauea for the periods 1969–1983 and 1984–2004 are shown in figure 6.2. We use these distributions to select cutoff magnitudes of 2.1 for 1969–1983 and 1.7 for 1984–2004. It is likely that at times of very high earthquake rate, for example after large magnitude earthquakes, the true completeness magnitudes are higher than these values. However, we repeat the analysis using the higher cutoff magnitudes of 2.5 and 2.1 respectively and find no significant changes to the results. Although the counting errors associated with these distributions are greater due to the smaller sample sizes, the broad patterns are similar to the lower cutoff magnitudes. As catalogue incompleteness will be correlated with inter-event time, we expect partial incompleteness to systematically result in disproportionately fewer counts at short inter-event times.

6.1.2.2 Mauna Loa

Mauna Loa has experienced two recent eruptions, in 1975 and 1984, although more frequently active prior to the mid-1950s. These eruptions were associated with dyke injection into the southwest and northeast rift zones, with magma supplied from a shallow reservoir located beneath the summit. As with Kilauea, rift-zone opening at Mauna Loa is accommodated by

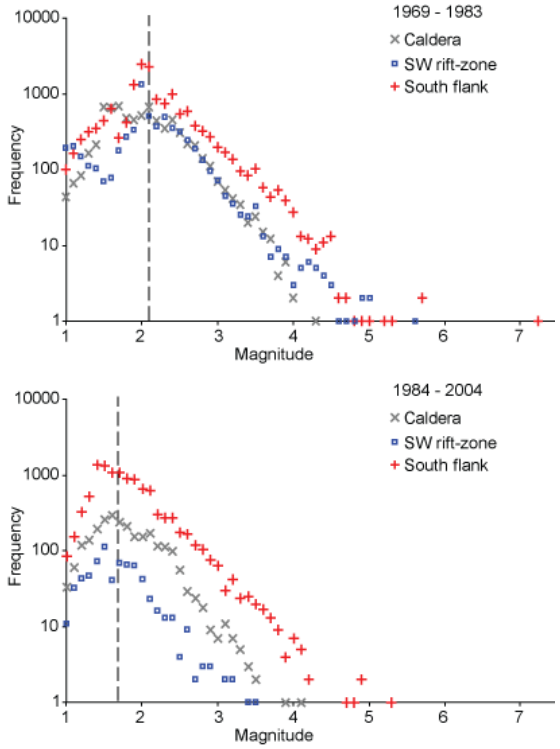


Figure 6.2: Magnitude-frequency distributions for: the caldera, SW rift zone and south flank regions of Kilauea (a) 1969–1983 and (b) 1984–2004. Chosen cutoff magnitudes are indicated by the dashed lines and are 2.1 for 1969–1983 and 1.7 for 1984–2004.

movement of the SE flank of the volcano, and this deformation is marked by earthquakes within the Kaoiki fault zone. Here we analyse earthquakes occurring between 1st January 1969 and 31st December 1984. The extended analysis period with respect to Kilauea includes the seismicity associated with the 1984 eruption.

Similarly to Kilauea, earthquake hypocentres at Mauna Loa are predominantly located within segregated volumes of the edifice. Here we focus on the two regions that have been most seismically active since 1969: (1) below the summit caldera at depths of less than 5km; and (2) within the Kaoiki fault zone at depths between 5 and 13km. These regions are defined according to the HVO classification scheme as the MLO region and the KAO region, respectively.

Earthquake frequency-magnitude distributions for Mauna Loa and the Kaoiki fault zone for the period 1969–1984 are shown in figure 6.3. We use these distributions to select cutoff magnitudes of 2.2 for Mauna Loa (1175 earthquakes) and 2.0 for the Kaoiki fault zone (5133 earthquakes). As for Kilauea, we repeat all analyses with higher cutoff magnitudes of 2.5 and 2.2 respectively and find no significant change in the results.

6.1.2.3 Campi Flegrei and Vesuvius

Campi Flegrei and Vesuvius volcanoes are monitored by the Vesuvius observatory. For Campi Flegrei, data is available between May 1983 and May 1984; a period associated with significant ground deformation and elevated seismicity. For Vesuvius, we use data between 1972 and 2008, during which time the volcano has been in a state of relative quiescence. On the basis of the magnitude-frequency distribution (figure 6.4), we select cutoff magnitudes of 1.9 for Vesuvius

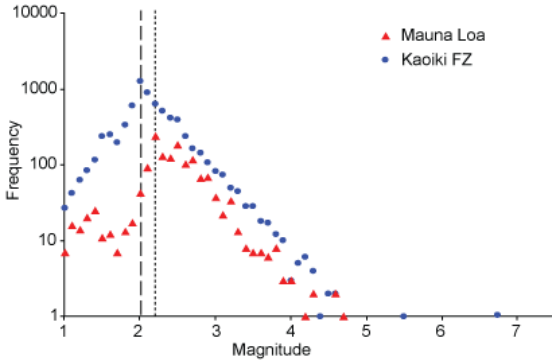


Figure 6.3: Magnitude-frequency distributions for Mauna Loa and the Kaoiki fault zone (1969–1984). Chosen cutoff magnitudes are indicated by the dashed lines and are 2.2 for Mauna Loa and 2.0 for the Kaoiki fault zone.

(1429 earthquakes) and 1.2 for Campi Flegrei (2300 earthquakes).

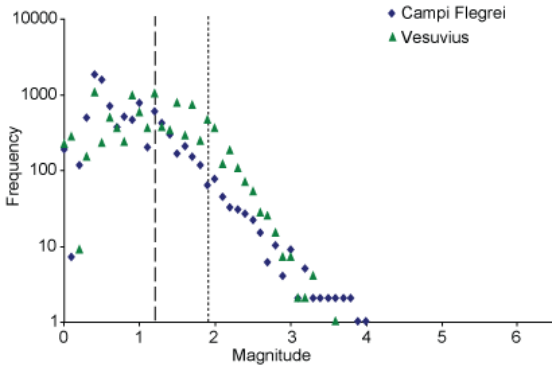


Figure 6.4: Magnitude-frequency distributions for earthquakes at Campi Flegrei (1983–1984) and Vesuvius (1972–2008). Chosen cutoff magnitudes are indicated by the dashed lines and are 1.2 for Campi Flegrei and 1.9 for Vesuvius.

6.1.3 Spatial and temporal variability of the inter-event time distribution at Kilauea

Time series of daily earthquake counts and total earthquakes for the selected seismic regions of Kilauea between 1969 and 1983 are shown in the left panel of figure 6.5. The three regions display distinct temporal patterns of earthquakes. Background rates are low in the caldera and SW rift zone, but are interspersed by episodes with elevated rates. Episodes in the caldera are characterised by accelerating rates of earthquakes, evolving over timescales of weeks to months, which frequently culminate in a dyke-fed eruption or intrusion (figure 6.5a). Episodes in the SW rift zone are characterised by rapid onset of high rates of earthquakes that gradually decrease to background levels after a few hours to days, in a manner similar to tectonic aftershock sequences (figure 6.5b). Sequence initiation often coincides with a dyke-fed eruption or intrusion identified by ground deformation data (Klein et al., 1987). In contrast, rates of earthquakes in the south flank are generally much higher and relatively constant over timescales of years (figure 6.5c). A change in the long-term earthquake rate occurs in November 1975 corresponding to the

magnitude 7.2 Kalapana earthquake.

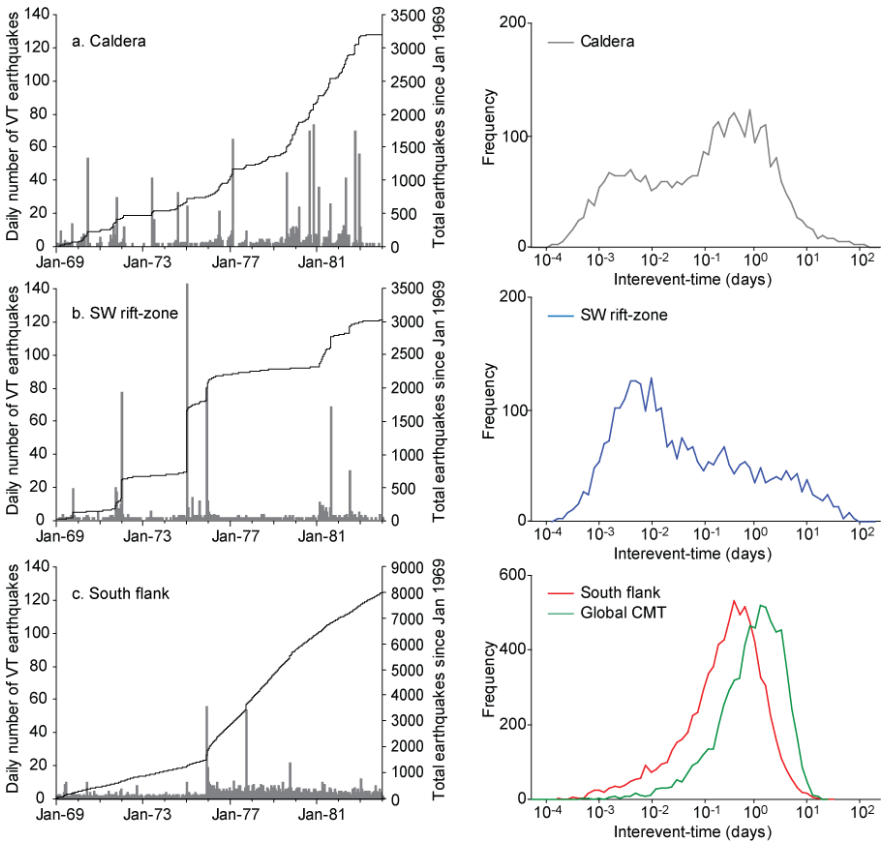


Figure 6.5: Left panel: Time series of the daily number of earthquakes (grey columns) and total number of earthquakes from 1st January 1969 (black line) for magnitude 2.1 and above earthquakes occurring in (a) the caldera, (b) the SW rift zone, and (c) the south flank volumes of Kilauea volcano, 1969–1983. Right panel: Associated frequency distribution of inter-event times counted for 0.1 log-day bins and not normalised according to linear bin width. Green curve in (c) shows inter-event time distribution for moment magnitude 5.8 and above earthquakes in the global Harvard CMT catalogue, 1st January 1976–31st December 2007.

We now consider the inter-event time distributions for these three regions, 1969–1983. We plot these distributions both as histograms of inter-event time counts for logarithmically-spaced bins (figure 6.5, right panel) and as a probability density function (pdf) when normalised by bin width, sample size and mean inter-event time (figure 6.6). The latter plotting method has been widely used by studies arguing for a data collapse to a universal gamma model (Corral, 2004; Davidsen and Goltz, 2004). Each region has a different distribution and these are incompatible with a single functional form (figure 6.5). The south flank histogram has a single maximum at an inter-event time of 0.5 days, and the form of the distribution closely resembles that for typical tectonic seismicity, such as the global Harvard Centroid Moment Tensor catalogue (figure 6.5c). We find that the form of the distribution is indistinguishable before and after the 1975 Kalapana earthquake, despite a change in the mean inter-event time. In contrast, the histogram for the caldera is distinctly bimodal, with two local maxima at inter-event times of 0.0025 days (3–4 minutes) and 0.5 days. The histogram for the southwest rift zone has a single maximum at an inter-event time of 0.005 days (7 minutes). The decrease in frequency

at higher inter-event times is not steady, with a point of inflection with an inter-event time of approximately one day. The distributions for the caldera and southwest rift zone are clearly incompatible with the simple universal model.

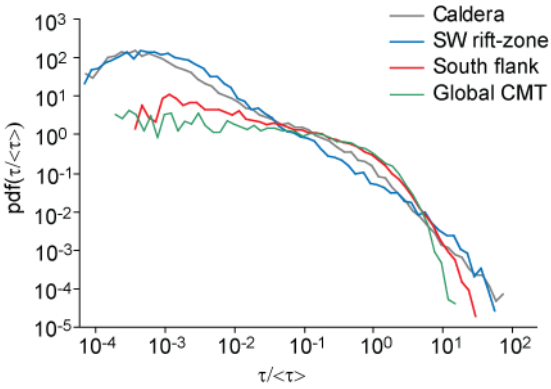


Figure 6.6: Probability densities of inter-event times for the three seismic regions at Kilauea, 1969–1983, and the global CMT catalogue (data is selected as for figure 6.5). Inter-event times τ are normalised according to the mean inter-event time $\langle\tau\rangle$.

The differences in the inter-event time distributions for different regions of Hawaii can be understood with reference to the respective earthquake rate time series. For the caldera and rift zones of Kilauea, background earthquake rates are low and seismicity is dominated by a series of distinct accelerating or decelerating sequences driven by individual magmatic events, each with an Omori law-like distribution of inter-event times. The limited interleaving of separate sequences means that each one will be well-resolved within the time series, thus increasing the relative amplitude of the Omori component (Touati et al., 2009, 2011). The accelerating sequences are associated with pre-eruptive or intrusive deformation, driven by elevated magma pressure, and evolve over timescales of weeks or months. The decelerating sequences are associated with deformation during and following dyke injection and evolve over timescales of hours or days. The time series for Kilauea caldera can be divided into separate accelerating and decelerating sequences and individual inter-event times marked with the appropriate category (figure 6.7a). The two peaks are clearly associated with the contribution due to these two types of sequence with different mode inter-event times (figure 6.7b).

The time series for the southwest rift zone is dominated by the contribution from a few major aftershock-like sequences with decelerating mean earthquake rates. The onset of these sequences closely coincides with the onset of known dyke injection events, and is the likely result of dyke-induced stress changes. These sequences do not overlap in time and so each will contribute an Omori law-like set of earthquake inter-event times to the overall distribution. In contrast, the relatively steady and pervasive deformation experienced by the south flank generates a high background rate of earthquakes with the initiation of many separate aftershock sequences. These sequences overlap temporally to a similar extent to tectonic seismicity over large regions, and result in an inter-event time histogram where the correlated and uncorrelated contributions closely overlap.

We repeat the analysis for the time interval 1st January 1984 to 31st December 2004 where volcanic activity was characterised by almost continuous lava extrusion along the east rift zone. The time series of daily earthquake counts and total earthquakes are shown in the left panel of

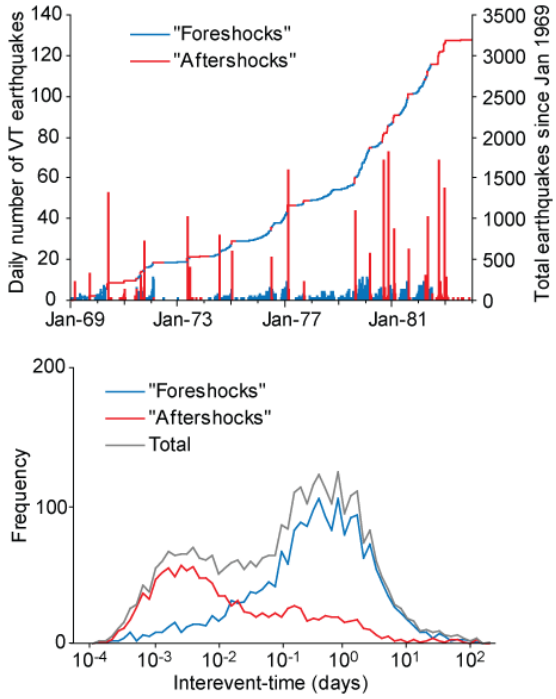


Figure 6.7: (a) Time series of the daily number of earthquakes (columns) and total number of earthquakes from 1st January 1969 (solid line) for magnitude 2.1 and above earthquakes located at Kilauea caldera. Blue and red components represent accelerating (pre-intrusive) and decelerating (post-intrusive) sequences (see text for details). (b) Frequency distribution of inter-event times for caldera earthquakes (as for figure 6.5), marking the contribution of inter-event times within foreshock and aftershock sequences separately.

figure 6.8a–c and highlight the change to less strongly temporally clustered rates of seismicity in all three regions.

The associated inter-event time histograms for seismicity between 1984 and 2004 are shown in the right panel of figure 6.8 and as PDFs in figure 6.9. Although the form of the distribution of south flank inter-event times is unchanged, there is a clear change in the distributions for the caldera and SW rift zone. These distributions are now similar to that for the south flank. This change confirms the controlling role that the style of volcanic activity plays in the volcanic earthquake inter-event time distribution.

6.1.4 Inter-event time distributions at Mauna Loa

Time series of daily earthquake counts and total earthquakes for the selected seismic regions of Mauna Loa are shown in the left panel of figure 6.10. The time series for the summit MLO region shows two distinct episodes with accelerating mean rate of earthquakes, culminating in the 1975 and 1984 eruptions. Outside these episodes, the rates of earthquakes are low. In contrast, the time series for the Kaoiki fault zone is characterised by a higher, steady rate of events, with occasional transient higher rates of earthquakes.

The associated inter-event time histograms and are shown in the right panel of figure 6.10 and as PDFs in figure 6.11. Similar spatial variability in the inter-event time distribution is observed at Mauna Loa. The distribution for the Kaoiki fault zone is very similar to the south flank of Kilauea, with a single maximum at an inter-event time of 1 day. In the summit

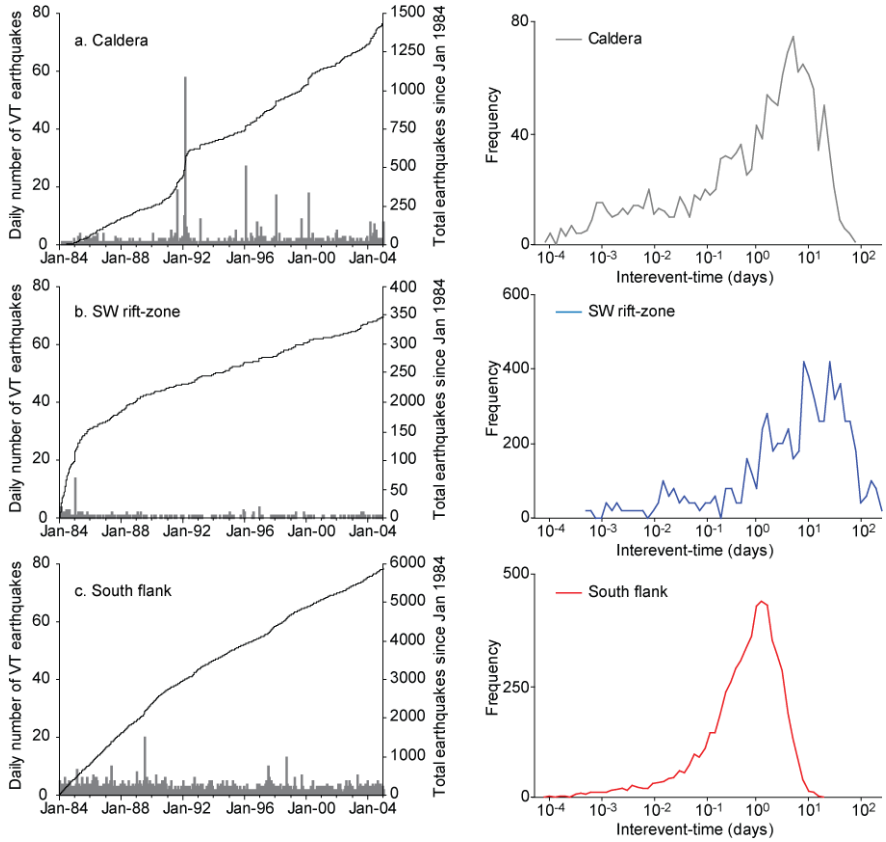


Figure 6.8: Left panel: Time series of the daily number of earthquakes (grey columns) and total number of earthquakes from 1st January 1984 (black line) for magnitude 1.7 and above earthquakes occurring in (a) the caldera, (b) the SW rift zone, and (c) the south flank volumes of Kilauea volcano, 1984–2004. Right panel: Associated frequency distribution of inter-event times counted for 0.1 log-day bins and not normalised according to linear bin width.

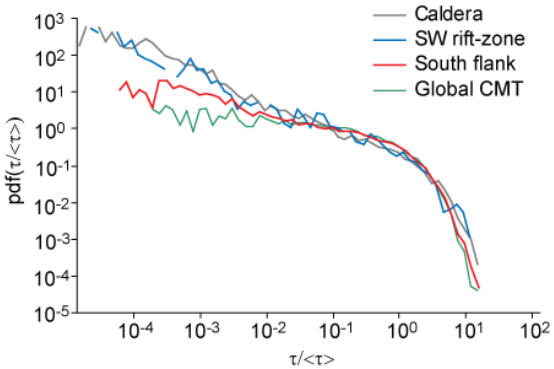


Figure 6.9: Probability densities of inter-event times for the three seismic regions at Kilauea, 1984–2004, and the global CMT catalogue (data is selected as for figure 6.8). Inter-event times τ are normalised according to the mean inter-event time$\langle\tau\rangle$.

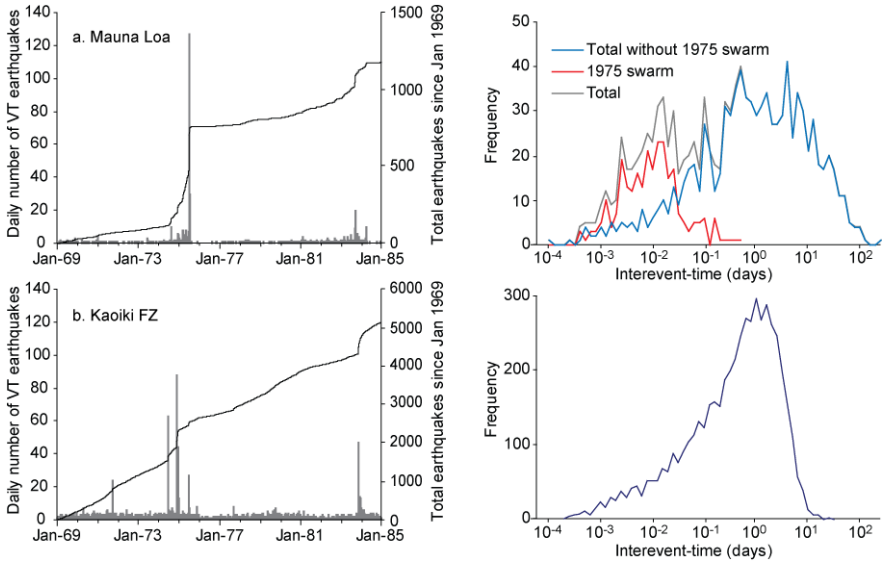


Figure 6.10: Left panel: Time series of the daily number of earthquakes (grey columns) and total number of earthquakes from 1st January 1969 (black line) for (a) magnitude 2.2 and above earthquakes occurring in the MLO region and (b) magnitude 2.0 and above earthquakes occurring in the Kaoiki fault zone of Mauna Loa volcano, 1969–1984. Right panel: Associated frequency distribution of inter-event times. The contribution of inter-event times associated with the 1975 eruptive swarm is marked separately.

MLO region, there are indications that the distribution is similar to the bimodal histogram for Kilauea caldera, although the sample size is small and counting error large.

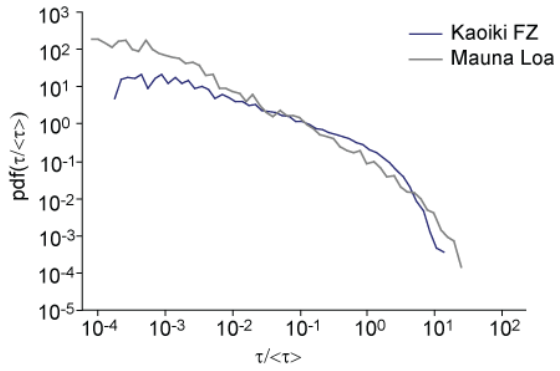


Figure 6.11: Probability densities of inter-event times for (a) magnitude 2.2 and above earthquakes occurring in the MLO region and (b) magnitude 2.0 and above earthquakes occurring in the Kaoiki fault zone of Mauna Loa volcano, 1969–1984 (data is selected as for figure 6.10). Inter-event times τ are normalised according to the mean inter-event time.

For the MLO region of Mauna Loa, background earthquake rates are low and seismicity is dominated by a series of distinct accelerating or decelerating sequences driven by individual magmatic events, each with an Omori law-like distribution of inter-event times.

For the MLO region of Mauna Loa, the origin of the inter-event time distribution can be understood by separating the contribution associated with the 1975 swarm associated with the onset of the eruption and intrusion of the a dyke into the NE rift zone (figure 6.10b). The majority of short inter-event times in the histogram occur within this three-day episode.

6.1.5 Inter-event time distributions at Campi Flegrei and Vesuvius

The time series of daily earthquake counts and total numbers of earthquakes are shown in figure 6.12. The time series for Campi Flegrei clearly shows two episodes of accelerating mean earthquake rates in 1983–4, culminating in major swarms on 13th October 1983 and 1st April 1984 (figure 6.12a), whereas there are no systematic trends in the time series for Vesuvius (figure 6.12b). Earthquakes at Campi Flegrei and Vesuvius are reported with a temporal precision of one minute.

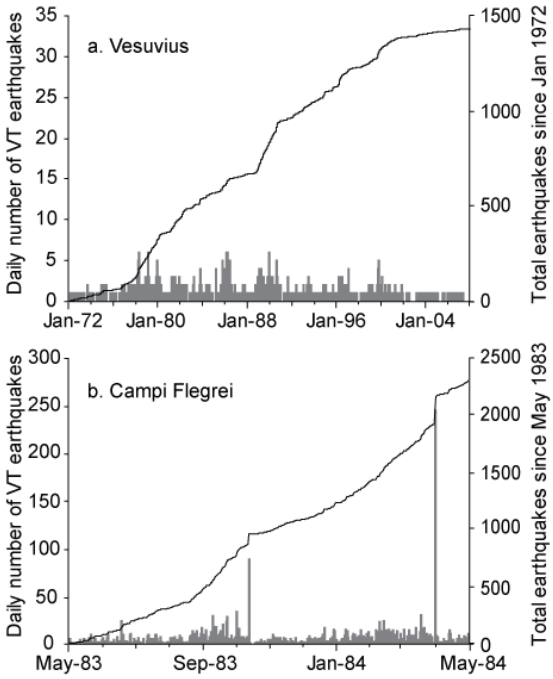


Figure 6.12: Time series of the daily number of earthquakes (grey columns) and total number of earthquakes (black line) for (a) magnitude 1.9 and above earthquakes occurring at Vesuvius, 1972–2008 and (b) magnitude 1.2 and above earthquakes occurring at Campi Flegrei, 1983–1984.

Earthquake inter-event time frequencies and normalised probability density functions are shown in figure 6.13. Despite approximately two orders of magnitude difference in the mean inter-event time, the form of the distributions is similar. This distribution is similar in turn to that for the south flank of Kilauea, the Kaoiki fault zone and the global CMT catalogue. However, there are more short inter-event times at Campi Flegrei than for these other volcanoes, despite the temporal resolution of the earthquake time being restricted to one minute.

6.1.6 Discussion

Our results clearly demonstrate that the characteristics of seismicity (including the inter-event time distribution) for Hawaii between 1969 and 2004 are extremely spatially and temporally heterogeneous. The detailed analysis of spatial and temporal subsets of the catalogue provides insight into the physical origins of volcanic earthquake inter-event time distributions. In contrast, the statistics of the catalogue of earthquakes for the whole of Kilauea, and indeed Hawaii, is dominated by events located in the south flank of Kilauea. Additionally, the amalgamation

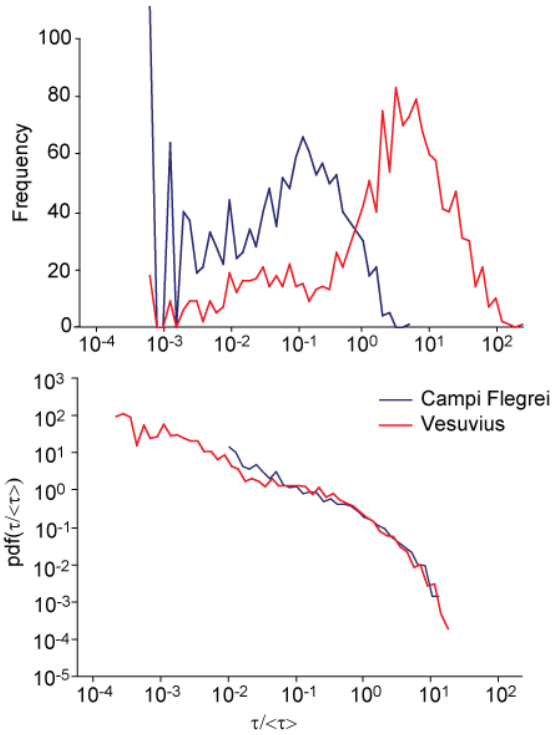


Figure 6.13: (a) Frequency distributions of inter-event times for magnitude 1.2 and greater earthquakes at Campi Flegrei (1983–1984) and magnitude 1.9 and greater earthquakes at Vesuvius, 1972–2008. (b) The probability densities of inter-event times in (a), normalised according to the mean inter-event time $\langle\tau\rangle$.

of seismically distinct regions into a single catalogue increases the proportion of Poissonian uncorrelated earthquakes, resulting in a distribution that more closely resembles the generalised gamma function. The details of underlying processes are less apparent in larger regions because of temporal overlapping (Touati et al., 2011). Consequently, in order to fully characterise the inter-event time distribution for volcanic regions, it is important to consider active and quiescent periods independently (Traversa and Grasso, 2010) and to spatially isolate the volcanoes or active parts of the volcanoes, as the processes involved can be very localised.

The degree of spatial and temporal heterogeneity we describe means that the distribution of volcanic earthquake inter-event times cannot be considered universal. The distributions from Kilauea caldera and south-west rift zone, Mauna Loa and Campi Flegrei are incompatible with either the simple gamma function of Corral (2004) and Saichev and Sornette (2007) or the more general function of Touati et al. (2009). These regions are characterised by temporal variations in the mean rate of earthquakes caused by stress changes associated with magma movement and there is almost no period where the seismicity could be considered to be stationary. Consequently the inter-event time distributions cannot be explained by simple analytical solutions based on a combination of a constant rate of Poissonian background events and Omori-law distributed triggered earthquakes.

We observe inter-event time distributions that are distinctly bimodal, consisting of two components with different mean inter-event times. Bimodal inter-event time distributions in non-volcanic seismicity have been explained as an aggregate of two components consisting of correlated and uncorrelated earthquake pairs (Touati et al., 2009). Correlated earthquake

pairs are between events within the same aftershock sequences and have an Omori power-law distribution. Uncorrelated earthquake pairs are between two background events, a background event and an unrelated aftershock event, or between two events from unrelated aftershock sequences. However, the bimodality in these volcanic scenarios is better explained as resulting from the combination of accelerating and decelerating sequences of earthquakes before and after volcanic activity, which evolve over different timescales.

We present the inter-event time distributions as both simple histograms and as probability density functions normalised by the mean inter-event time. The normalised PDF allows direct inspection of the asymptotes in the gamma function (Corral, 2004). However, the normalisation can mask considerable fluctuations in the raw count data, and this may have contributed to the apparent gamma distribution claimed by Bottiglieri et al. (2009a). We suggest that good practice would be to present both types of plot.

Volcano-tectonic earthquakes result from brittle failure of a volcanic edifice and are primary indicators of volcanic unrest. The spatial and temporal patterns of VT earthquakes provide information about the material response of the edifice to changing stress conditions and are keys to eruption forecasting. However, improved quantification of eruption forecasts requires a better understanding of the dynamics of VT seismicity and in particular the nature of clustering in pre-eruptive periods. Further study of the inter-event time distribution in these periods may prove fruitful. The inter-event time distribution may also allow characterisation of changes in the nature of source driving seismicity, potentially identifying re-supply of magma.

6.1.7 Conclusions

The nature of the distribution of times between successive earthquakes in a catalogue is important for understanding the hierarchical organisation of seismicity and has relevance for seismic hazard assessment. Our analysis of earthquake catalogues for Hawaiian and Italian volcanoes shows that the inter-event time distribution for volcanoes is spatially and temporally variable and is generally not consistent with a gamma distribution. At Kilauea and Mauna Loa, the inter-event time distributions regions of the volcano where seismicity is driven by magmatic processes can be explained as a combination of pre-volcanic accelerating (foreshock-like) earthquake sequences with a long mean inter-event time, and post-volcanic decelerating (aftershock-like) earthquake sequences with a short mean inter-event time. At Kilauea caldera, we are able to distinguish these sequences a priori, and show that they underlie the two peaks of a bimodal inter-event time distribution. At volcanoes where the magmatic driving forces are currently small (e.g. Vesuvius), or in regions of volcanoes where seismicity is primarily driven by tectonic processes, the inter-event time distribution more closely resembles that for tectonic seismicity and can be explained by a simple epidemic model based on independent background events and cascades of triggered daughter events. These observations confirm that, contrary to previous claims, the inter-event time distribution for volcanic earthquakes cannot be considered universal and that forecasting for the different sites requires an understanding of site-specific physical processes. They also suggest that systematic changes in the inter-event time distribution can be used as a diagnostic tool for inferring changes in the processes driving volcanic seismicity, such as re-pressurisation of the magmatic system.

Chapter 7

Discussion

Having established some important results in the previous chapters concerning earthquake inter-event times and the deeper insights on the masking effect of temporal overlapping of aftershock sequences, I will now set the work in context and explore its consequences. In section 7.1 I will explore and clarify the meaning of the data collapse in the inter-event time distribution as reported by Corral, Bak et al., addressing the studies showing universality and data collapse. Section 7.2 provides a general discussion on aftershocks in terms of the SOC paradigm and my opinions on how they might fit into it. In section 7.3 I discuss applications and consequences of my results on other problems, and review the subsequent studies that have cited my work. In section 7.4 I look at the practices of inferring temporal variations in seismicity, and in section 7.5, explore possible future work in this area using the insights from the thesis to inform on the reliability of such techniques under a variety of circumstances.

7.1 Non-universality

Looking for a data collapse in the earthquake inter-event time distribution, particularly by a rescaling involving other parameters of the data as in Bak et al. (2002), was motivated by the SOC paradigm. However, the link between the specific data collapse presented in these studies and SOC is thus far unclear.

Corral's rescaling involves multiplying the inter-event times τ, and dividing the histogram counts, by the mean rate r. This means the x-axis becomes $r\tau$, and the y-axis is $\frac{1}{r}f(\tau) = \frac{1}{r}f(\frac{1}{r}x)$. The mean rate r is the reciprocal of the mean of the distribution, so it means normalising the inter-event times τ by the mean inter-event time, and multiplying the counts by this mean. What is really implied by data collapse under this rescaling?

In the case of an exponential distribution of inter-event times $f(\tau) = \lambda e^{-\lambda\tau}$, the mean rate is equal to the rate parameter λ and the rescaling results in a function $y = e^{-x}$, which is independent of λ; data collapse would thus be achieved exactly (figure 7.1).

Alternatively we can consider a gamma distribution for τ, which has the form:

$$f(\tau) = \tau^{k-1}\frac{e^{-\tau/\theta}}{\theta^k\Gamma(k)} \tag{7.1}$$

where k is the rate parameter and θ is the shape parameter; the mean value of τ is $k\theta$, giving a mean event rate $r = \frac{1}{k\theta}$. The rescaling would then lead to:

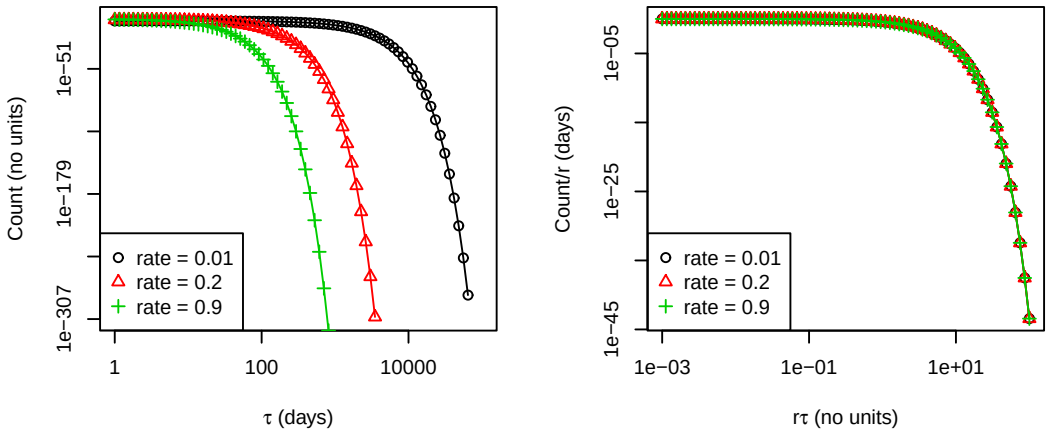


Figure 7.1: Exponential pdfs (left) and data collapse after rescaling by the mean rate r (right). Rate parameter values are as indicated in the legend.

$$\begin{aligned} y &= \frac{1}{r} f(\frac{1}{r}x) \\ &= k\theta(k\theta x)^{k-1} \frac{e^{-kx}}{\theta^k \Gamma(k)} \\ &= k^k \theta^k x^{k-1} \frac{e^{-kx}}{\theta^k \Gamma(k)} \\ &= k^k x^{k-1} \frac{e^{-kx}}{\Gamma(k)} \end{aligned} \tag{7.2}$$

The shape parameter θ has been eliminated, which means that data collapse depends entirely on the the rate parameter k. Thus if (and only if) the power-law exponents are identical across the different histograms, then a complete data collapse will occur (see figures 7.2 and 7.3). More generally this means that it is easy to get a reasonably convincing data collapse under this kind of rescaling with any distribution that is gamma-like and has an exponential decay portion, as the earthquake inter-event time distribution does. I would conclude that the data collapse is simply a property of these types of mathematical functions and not particularly related to criticality.

Corral required a well-defined event rate in order to perform his rescaling and so he argued against (1) aggregating inter-event times taken from different regions where the event rate varies between the regions, and (2) aggregating inter-event times from different time periods, where the event rate varies in time (Corral, 2003). Point (2) effectively means that the duration of the catalogue used to create the inter-event time distribution must span only a period in which the event rate appears "stationary" i.e. roughly constant. However, short-term filtering for "stationary" periods is highly subjective: such data selection requires many additional free parameters to define the spatio-temporal window, leading to the possibility of bias in the search, and exaggerated significances if the effect of the additional model parameters is not taken into

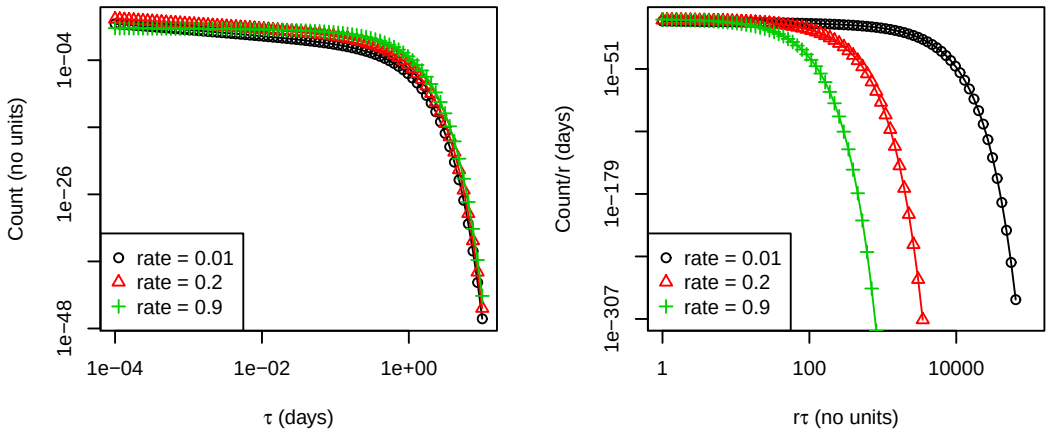


Figure 7.2: Gamma pdfs with shape parameter equal to 0.1 and different rate parameters as indicated in the legend. In the pdf plot (left), the slope of the power-law segment changes but the position of exponential decay does not. Under rescaling by the mean rate r (right), data collapse is not observed.

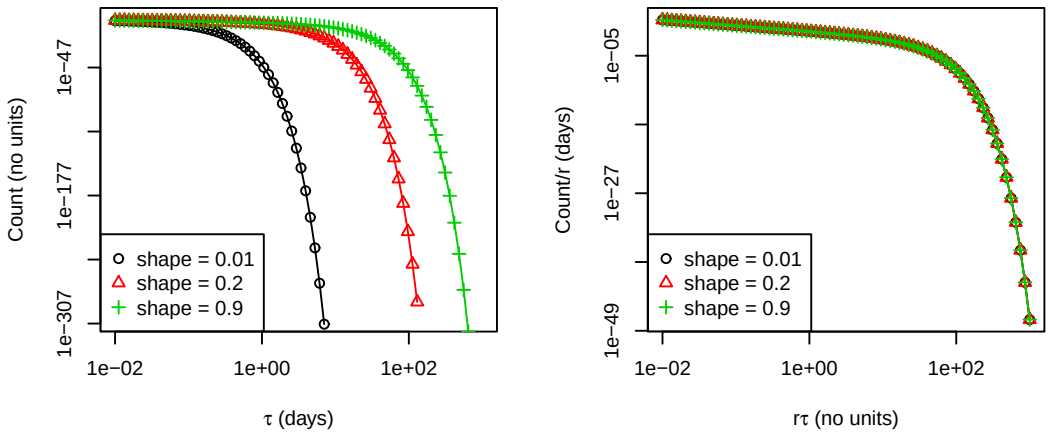


Figure 7.3: Gamma pdfs with rate parameter equal to 0.1 and different shape parameters as indicated in the legend. In the pdf plot (left), the slope of the power-law segment is constant while the point of exponential decay varies. Under rescaling by the mean rate r (right), data collapse is observed.

account, e.g. by formal Information Criteria (e.g. Burnham and Anderson, 2002). This filtering is also quite arbitrary: aftershock sequences are often considered as a Poisson process with a time-varying rate; if it were practical to do so, one could take very small segments that could each be approximated as a stationary Poisson process, rescale their inter-event time distributions by the rate, and obtain trivial data collapse onto an exponential distribution. Corral's procedure resembles this, but uses finite-sized segments that inevitably are not perfectly stationary, giving an approximate gamma distribution instead. Such a procedure yields no insight into the dynamics underlying longer-term inter-event time distributions. It would work regardless of the true nature of the distribution which explains how this data collapse has been claimed even in volcanic settings.

Bak et al. (2002)'s rescaling is a little different; it doesn't require selection of stationary periods. The rescaling of the x-axis is by a term representing the average event rate, similarly to Corral's, but the rescaling of the y-axis is done with a term involving the inter-event times themselves raised to some power. It is not clear what the meaning of this would be as they don't explain how they arrived at it, but the data collapse is visually less clear and convincing than Corral's; it is possible to see bimodality for example in some of their data sets (see figure 1.5 in the Introduction).

I believe that the bimodality in the distribution was overlooked in literature proposing universality mainly because this literature was focussed on earthquakes as an SOC process, and most SOC systems do not have the temporal clustering of events that earthquakes exhibit, which gives those other systems more simple unimodal distributions of intervals between the events. The earthquake process is really two processes, a spontaneous process produced by slow driving of tectonic plates (analogous to avalanches in the SOC sandpile paradigm) and an aftershock process following Omori's law. Bak et al. (2002) did recognise these different processes in some way and interpreted the form of the curve in terms of them: they recognised that the exponential decay at long intervals represented uncorrelated values, and conjectured that the "kink" in the curve marked a transition from a correlated to an uncorrelated regime. Their "kink" is equivalent to my second peak, the peak in the underlying uncorrelated distribution, so this is confirmed by my work except that it is not a sharp transition but instead a gradual crossover. However they also stated that the data collapse, and the dependence of the kink's position on the length scale and magnitude range considered, indicate that "there is no separate relaxation mechanism for aftershocks" and that the distinction between main events and aftershocks is arbitrary and has no absolute meaning. These statements seem to confuse events with intervals (i.e. with event pairs) and to negate the very notion of considering intervals as correlated or uncorrelated. It seems the motivation was to argue for a unification of the two processes. But my work has made it clear that the statistics of these two processes do not scale in the same way with region size and other parameters, and so it follows that they can not be treated as a single process obeying a unified scaling law.

In summary, gamma distributions with similar rate parameters will exhibit a data collapse under rescaling by the reciprocal of the mean value, which is a mathematical property and has nothing to do with criticality. The requirement of approximately stationary periods in earthquake data effectively produced exactly this situation for inter-event time distributions. The bimodality in the distribution was previously overlooked in the quest to unify the various aspects of the earthquake process, including the two processes of spontaneous event generation and aftershocks, whose timescales I have shown cannot be unified.

7.2 Aftershocks as a second critical process

There are some interesting parallels between the criticality discussed in the Introduction and the aftershock process. Critical systems generally involve neighbour-to-neighbour coupling, producing events as cascades of perturbations, triggered almost immediately after the initial onset of the cascade. The critical coupling level—that of each perturbation triggering on average one further perturbation—gives a power-law distribution of event (i.e. cascade) sizes. Aftershocks could perhaps be thought of as a higher-order process of a similar nature. They are triggered at a greater distance and time; however, there is a similar "critical" point at a branching ratio of 1, as explored by Sornette and Helmstetter (2002) and others and reviewed in the Introduction in section 1.3.1. It is critical in the sense that, approached from below, it is the value of n at which seismicity first has the potential to propagate indefinitely. At this point the global Omori law is also a single power law, as opposed to the subcritical case where the exponent switches to a higher (negative) value at a characteristic time (see section 1.3.1). Measured branching ratios are often significantly less than 1; however, the true effective branching ratio may well be higher than it seems due to undetected small earthquakes (Sornette and Werner, 2005b). The aftershock process may therefore be a critical or only slightly subcritical process. Although aftershocks are in a sense a separate process from the basic events, they happen in the same crust, so it stands to reason that they would show similar signs of self-organised criticality.

It is interesting to note that in the Ising model for magnetic Barkhausen noise, interactions occur between the cascades: cascades are arrested by the spontaneous flipping of a neighbouring dipole due to the temperature term, and so the spontaneous rate effectively moderates the coupling rather than being independent of it. This is why the temperature needs to be tuned to a specific value to see crackling noise. In SOC models such as the sandpile, it is usually assumed that the rate of spontaneous events (grains being added) is much smaller than the rate at which avalanches propagate, so that the cascades do not overlap in time and space and the coupling is instead moderated by the continual heterogeneity in the sand surface. In the case of earthquakes, the existence of aftershocks means that such spatio-temporal overlapping can occur, both in terms of the basic events (while the aftershock rate is high in the wake of a large event) and also in terms of aftershock sequences themselves being spatially and temporally extended processes that can potentially overlap in both domains at once. In the ETAS model, each separate aftershock sequence is viewed as independent and does not interact with the others; in reality perhaps there may occasionally be some interaction.

What we know empirically about aftershocks is that their count as a function of distance, time (the Omori law), and parent size (the productivity law) obeys power laws. Transferring the criticality concept into the aftershock process, the analogue for events would be entire aftershock sequences, so we may wonder about the size distribution of whole sequences; is it a power law comparable to the Gutenberg–Richter relation for individual events? This is calculated both in terms of the numbers of events and in terms of the total magnitude, as follows.

From the productivity law, the cumulative distribution of the number of events N in a sequence would be given by:

$$F(N) \quad = \quad p(e^{\alpha m} \leq N)$$

$$
\begin{aligned}
&= p(\alpha m \leq \ln N) \\
&= p(m \leq \frac{\ln N}{\alpha}) \\
&= \int_0^{\frac{\ln N}{\alpha}} \beta e^{-\beta m} dm \\
&= -e^{-(\frac{\beta}{\alpha} \ln N)} + e^0 \\
&= 1 - N^{-\frac{\beta}{\alpha}} \qquad (7.3)
\end{aligned}
$$

The probability density function of N is then obtained by differentiating with respect to N and is a power law:

$$f(N) = \frac{\beta}{\alpha} N^{-\frac{\beta}{\alpha}-1} \qquad (7.4)$$

For calculating the sequence size in terms of the total magnitude, we can start with the Gutenberg–Richter law for the sequence:

$$N_m = e^{A-\beta m} \qquad (7.5)$$

The parameter A can be considered to be determined by the total number of aftershocks, $N = e^{\alpha m_p}$ (where m_p is the parent magnitude). A is obtained by setting N equal to the integral of N_m evaluated from 0 to ∞. This gives:

$$
\begin{aligned}
e^{\alpha m_p} &= \int_0^{\infty} e^{A-\beta m} dm \\
&= e^A \left[\frac{e^{-\beta m}}{-\beta} \right]_0^{\infty} \\
&= \frac{e^A}{\beta} \\
\Rightarrow A &= \alpha m_p \ln \beta \qquad (7.6)
\end{aligned}
$$

Substituting this into (7.5) gives:

$$N_m = \beta^{\alpha m_p} e^{-\beta m} \qquad (7.7)$$

The total magnitude of the sequence is obtained as $m_{seq} = \int_0^{\infty} N_m m dm = \beta^{\alpha m_p - 2}$. Having now obtained the quantity we want (m_{seq}) as a function of m_p, we can express the cumulative distribution of m_{seq} as

$$
\begin{aligned}
F(m_{seq}) &= p(\beta^{\alpha m_p - 2} \leq m_{seq}) \\
&= p\left(m_p \leq \frac{1}{\alpha} \left(\frac{\ln m_{seq}}{\ln \beta} + 2 \right) \right) \qquad (7.8)
\end{aligned}
$$

and knowing the distribution of m_p (the simple Gutenberg–Richter pdf $f(m_p) = \beta e^{-\beta m_p}$), we can calculate this:

$$
\begin{aligned}
F(m_{seq}) &= \int_0^{\frac{1}{\alpha}\left(\frac{\ln m_{seq}}{\ln \beta}+2\right)} \beta e^{-\beta m_p} dm_p \\
&= 1 - m_{seq}^{-\frac{\beta}{\alpha \ln \beta}} e^{-2\frac{\beta}{\alpha}} \qquad (7.9)
\end{aligned}
$$

Differentiating with respect to m_{seq} gives the pdf of m_{seq}:

$$
f(m_{seq}) = \frac{\beta}{\alpha \ln \beta} e^{-2\frac{\beta}{\alpha}} m_{seq}^{-\frac{\beta}{\alpha \ln \beta}-1} \qquad (7.10)
$$

Based on the Gutenberg–Richter and the productivity law, then, we find that the distribution of cascade sizes—with size considered either in terms of numbers of events or total magnitude of the events—is an inverse power law, with an exponent that depends on the relative values of the productivity exponent α and the Gutenberg–Richter exponent β. This can be considered a whole-sequence version of the Gutenberg–Richter law and represents another parallel with the criticality idea. Of course, the self-similarity in the aftershock statistics that this calculation is based on, such as the productivity law, may also be a consequence of the criticality and only valid at or near $n = 1$, but this condition is likely to hold, at least approximately.

The BASS model (Turcotte et al., 2007) is a "fully self-similar" statistical model in which there is self-similarity in the aftershock process not just in terms of the numbers of events produced by parents of different magnitude, but also in the (average) magnitude relationships between the parent and offspring. Constraining the average number of aftershocks produced by an event of a given size using Gutenberg–Richter and Båth's law, rather than a productivity law, achieves this. It proposes essentially that $\alpha = \beta$ (Holliday et al., 2008b). This does make intuitive sense, because if $\alpha < \beta$ the smallest parent events will trigger a disproportionately large number of events comparable to their own size or larger as compared to the larger parent events (Holliday et al., 2008a); one would instead expect the magnitude difference between the parent and its largest aftershock (on average) to be constant with respect to parent magnitude. However, the issue is that this parameterisation in ETAS results in an infinite branching ratio (assuming no finite maximum magnitude). The papers that present the BASS model note this fact, and give only examples of simulating an aftershock sequence from a single seeding event with BASS, which perhaps suggests that it wouldn't be possible in the BASS model to simulate a catalogue from a Poissonian seeding rate and achieve a stable event rate.

It is easy to verify that the branching ratio is indeed infinite when there is no upper magnitude cutoff. The number of (direct) offspring from a parent event of magnitude $m_p = m + m_{min}$ is $N_d = e^{\beta(m-\Delta m^*)}$, where Δm^* is the constant from Båth's law (Turcotte et al., 2009). Thus the branching ratio is $n = \int_0^\infty e^{\beta(m-\Delta m^*)} \beta e^{-\beta m} dm$. The terms involving m cancel leaving a constant term, which yields infinity when integrated over the limits 0 to ∞.

There is another fully self-similar model by Vere-Jones (Vere-Jones, 2005) which has $\alpha = \beta$ and avoids this instability issue by constraining the magnitudes of the aftershocks directly by the magnitude of the parent, an idea which is not empirically-based (Sornette and Werner, 2005a).

As a way out of this quandary, we can consider the fact that the Gutenberg–Richter power law cannot extend indefinitely and must be truncated, for example it may be a gamma distribution; a finite maximum or corner magnitude is needed to keep the total energy release

rate from being unphysically infinite (Al-Kindy and Main, 2003). If we want to specify a finite effective maximum magnitude, the expression for the branching ratio in the ETAS model when $\alpha = \beta$ is given by:

$$n = \frac{Ac}{p-1}\frac{\beta(m_{max} - m_0)}{1 - e^{-\beta(m_{max}-m_0)}} \tag{7.11}$$

It is therefore (analytically) possible to get a branching ratio of around 1 if the range of triggering magnitudes is constrained. If the latter is deemed to be of the order of 10, for example, with typical order-of-magnitude estimates for c and p, $n \sim 1$ would require A to be of the order of 0.01 (or K of the order of 10^{-3}), which is perhaps plausible based on the literature survey shown in figure 1.2 of the Introduction—although given the parameter inversion uncertainties highlighted in chapter 5, it is difficult to be sure.

There is another issue in that the Omori parameter p has to be greater than 1. If this is not the case, integration of the Omori pdf over all time becomes unbounded, meaning that there is no value of the factor K (or A) for which we can have a finite-length aftershock sequence. In an ETAS context with a constant seeding rate, the overall event rate would become unbounded also. Measured values of p (see figure 1.2 in the Introduction) are often less than 1. This may partly be explained by the presence of secondary (and higher-order) aftershocks, but I am not aware of any physical reason why p would have to be greater than 1. For $p \leq 1$ there would need to be a temporal limit to the sequence in order to have a stable event rate. This is certainly possible and some studies point towards it (e.g. Narteau et al., 2002) although due to insufficient data and overlapping in the tails of aftershock sequences, it is difficult to prove.

The earthquake process is bounded in space due to the finite size of the Earth's crust or individual seismic zones. If the earthquake and aftershock process is bounded in the domains of magnitude and time also, it may be possible that it is a doubly (near-)critical and self-similar process while also being stable and convergent.

7.3 Applications and consequences of my results

7.3.1 Convergence

It was shown in chapter 3 that a sequential sample of inter-event times does not resemble a random sample from the underlying distribution, and hence the mean value for a finite sequential sample has a significantly larger error than for a random sample. The error varies between catalogues, however. We are now in a position to explain this in terms of temporal overlapping of aftershock sequences. When sequences do not overlap much, the average values within sequential samples can vary quite widely as aftershock sequences progress without the interruption of a new spontaneous event. Sequential intervals within an aftershock sequence tend to be similarly-valued, and so are correlated *with each other* as well as arising from correlated event pairs. At higher spontaneous rates, therefore, the convergence is faster because the temporal correlations in the data are weakened by the overlapping, and the variability in the event rate over time is smaller (see figure 5.6 in chapter 5).

As has been shown in chapter 5, the spontaneous rate is connected with region size. Indeed, the paper in chapter 3 demonstrated that the catalogue for the whole globe converges faster than the regional catalogues for New Zealand and Southern California; this is exactly what

would be expected for a high spontaneous rate. The effect of the spontaneous rate μ in the ETAS model on the convergence curves was not studied in the paper, but I show this here for completeness. Figure 7.4 shows convergence curves comparable to figure 3.14, for synthetic ETAS catalogues with three different values of μ. The pattern, as expected, is similar to that for the real catalogues spanning different-sized areas, with higher spontaneous rates producing faster convergence.

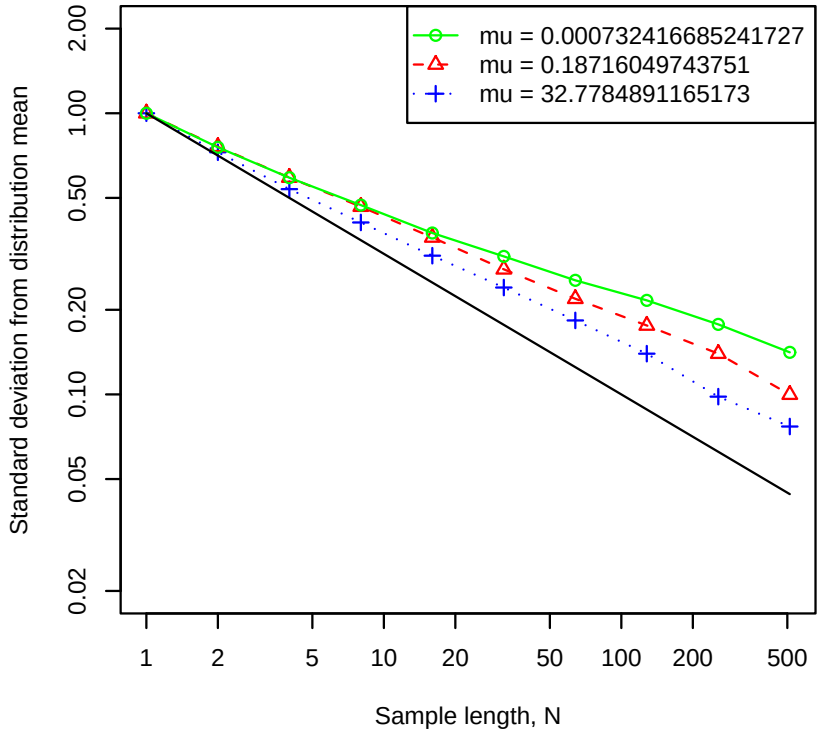


Figure 7.4: Convergence of the mean inter-event time with increasing sample size N for ETAS simulations with three different values of μ. The solid black line is $1/\sqrt{N}$; the lower μ is, the slower the decay of the error. Other ETAS parameters used in each simulation were: $A = 10$, $\alpha = 1$, $c = 0.01$, $p = 1.2$.

7.3.2 Exponential decay in the distribution

Hainzl et al. (2006) tested the analytical theory of Molchan (2005) which specified that the effective parameter of exponential decay ($1/\theta$ in equation (7.1)) in the distribution, after rescaling by the mean event rate, would be equal to the fraction of background (or spontaneous) events in the catalogue. Confirmation of this equality also supports non-universality as it shows that there is not a data collapse under this rescaling. This technique has been utilised in subsequent studies to estimate the background fraction (e.g. Traversa and Grasso, 2010).

The fraction of events that are background depends only on the extent of aftershock triggering and not on the background rate itself, and is equal to $1 - n$ where n is the branching ratio (Helmstetter and Sornette, 2003d). Thus the theory essentially says that the branching ratio

can be determined from the exponential decay of the distribution at long intervals. However we note that in figure 4.8 of chapter 4—reproduced here as figure 7.5—the rate of this decay in the rescaled distribution plot can be seen to depend also on μ, the background rate. This could perhaps explain some of the scatter in Hainzl et al. (2006)'s values around the correct ones for the ETAS simulations.

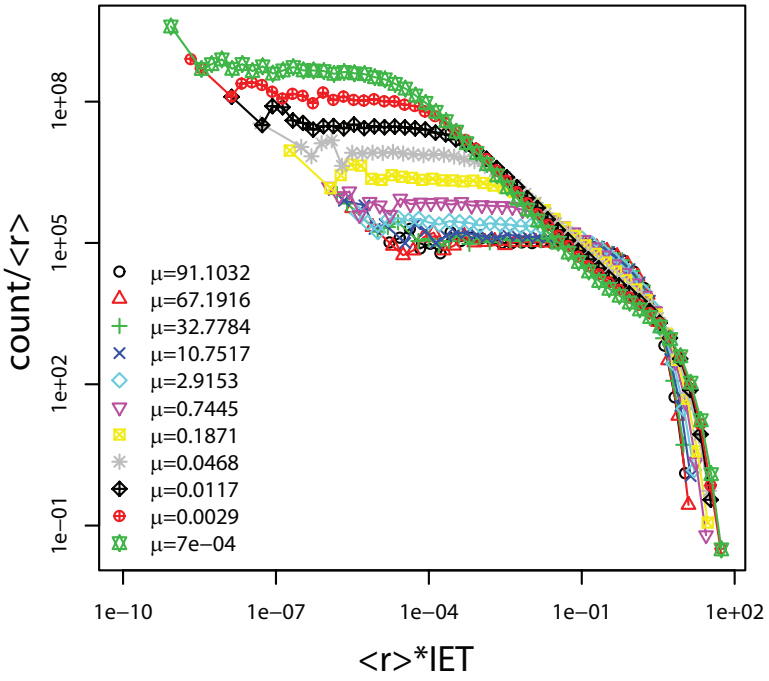


Figure 7.5: Inter-event time (IET) histograms for 11 ETAS simulations with μ values as indicated, plotted with normalisation by the bin widths, and rescaled by the mean event rate r. The other ETAS parameters used were $A = 10$, $\alpha = 1$, $c = 0.01$ and $p = 1.2$ in all cases. The rate of exponential decay at long intervals can be seen to vary with μ.

Assuming the inter-event time distribution approximates a gamma form (as is implied in the procedure), under rescaling by the mean event rate, equation (7.2) indicates that the rate of exponential decay becomes k, the rate parameter—that is, the exponent for the (approximate) power-law segment of the curve. The theory of Molchan then indicates a dependence of this exponent on the branching ratio. This can be framed in terms of my thesis results by considering that the effective power-law exponent is determined by the relative heights of the two peaks (for correlated and uncorrelated intervals), whose crossover can be approximated as power-law. In the case of no aftershocks, the distribution is exponential and the "power-law exponent" reduces to zero, which corresponds to $k = 1$, i.e. a background fraction of 1 according to Molchan's theory—which is obviously correct. As the branching ratio grows, that is, the background fraction decreases, a second peak at short intervals emerges and grows larger (see figure 4.7 in chapter 4). This makes the exponent of the approximate power-law segment increasingly negative, which corresponds to a decreasing value of k. Thus the relationship is consistent with my results, but additionally we recognise that the effective value of k depends on the background rate as well as the branching ratio.

7.3.3 Subsequent work by others in connection with my results

The results of the paper in chapter 4 of the thesis have been cited in several subsequent papers.

In earlier work, Bottiglieri et al. (2009b,a) utilised the universality model for the earthquake inter-event time distribution. In response to my results, they presented a new definition of universality in earthquake inter-event times (Bottiglieri et al., 2010), better taking into account the fact that multiple processes contribute to the form of the distribution. The authors claim four temporal parameters control this form. However, this is a large number of renormalising constants, and this even more broad notion of universality is very far from both the concept of universality classes applied in critical point systems, and the more general concept of data collapse invoked by Bak et al. (2002). Even with such a definition, it is clear in light of my results that universal data collapse onto a single curve for a wide variety of geographical regions remains out of the question in the general case. When μ is altered drastically, the inter-event times are fundamentally different ones and can never be rescaled to fit onto each other. The paradigm of overlapping sequences and masking of triggering effects, I believe, yields more insight than the loosely-defined and evolving notion of universality.

Traversa and Grasso (2010) study earthquake inter-event time distributions at Etna and Vesuvius, and find that the data from inter-eruptive phases collapses on a common gamma distribution whereas the data from dyke intrusion periods is more exponential. They correctly infer a higher background or seeding rate in the latter case, citing my result that the inter-event time distribution becomes more exponential as the seeding rate increases. They also use ETAS simulations to verify the explanation for the different shape. However in inferring similarities in the underlying processes between the inter-eruptive phases and purely tectonic datasets from elsewhere, they miss the point that the apparent gamma distribution is easily obtained by data selection, particularly the size of geographical area considered which largely controls this shape. Here I have shown in Chapter 6 that it is equally possible to obtain clear bimodal distributions, and that variations in the bimodality reflect spatial variability in volcano-tectonic setting.

Wu et al. (2010) applied my model to the intervals between messages in human communication systems. The distribution of such intervals is also bimodal, arising from Poissonian initiations of conversation coupled with mutual replying at a much faster rate that perhaps decays as an Omori-like power law.

Zhao et al. (2010) propose the metric of Local Variation (Lv) for quantifying the "burstiness" of the events in a series, similar to Coefficient of Variation (Cv) but with the non-stationary variation removed. In connection with my result on the non-universality of the distribution, they say that "using Lv may capture the difference more vividly". They divide the Earth into spatial cells and compute Lv for each, and find it to decrease with increasing event rate. The scatter also decreases, perhaps because a larger number of events are contained in the samples when the rate is higher. The decreasing burstiness with increasing rate is taken as suggesting that fewer events are correlated in high-rate areas, which is the wrong interpretation, confusing events with intervals. The temporal overlapping of aftershock sequences in high-rate situations simply means the correlations are hidden or masked. Their analysis of different tectonic settings is problematic in my view because the seeding rate—controlled by the spatial bounds of the dataset—has a much bigger impact than anything else, and that has not been accounted for. They find that Lv has a negative relationship with the largest magnitude in the catalogue, and speculate that the burstiness is "weakened by the occurrence of mega earthquakes." I would

argue instead that the biggest events occur in high-rate areas due to the Gutenberg–Richter statistics, and so Lv is lower as a consequence of the high overall rate rather than the large event.

Krishna Mohan and Revathi (2011) present a way of determining the correlations in a series of earthquakes from their spatial and magnitude relationships. They acknowledge my result on the bimodality of the inter-event time distribution, and claim to have extracted the correlated component through their method, which appears to have a double power-law distribution. While the shortest intervals are due to aftershocks, the steeper power law at longer distances is attributed to longer-term correlations caused by the healing and re-rupture of faults, termed "pure spatial recurrences". The crossover point is proposed as a criterion for selecting or identifying aftershocks.

On the other hand, there have been a number of studies subsequent to my work that have continued to model the earthquake inter-event time distribution in terms of gamma, exponential, Weibull, and power laws (e.g. Garavaglia and Pavani, 2011; Erisoglu et al., 2011; Hasumi et al., 2010; Akimoto et al., 2010). Some discuss correlations in the time series in a similar manner to Bak et al. (2002), using this to explain the different regimes in the distribution, but without the clarity of distinguishing between categories of intervals as I have done (e.g. Talbi and Yamazaki, 2010). The data collapse under rescaling by the event rate is still being pursued (e.g. Niccolini et al., 2010, 2011; Mendes et al., 2010; Timar and Kun, 2011; Bottiglieri et al., 2009a). Surprisingly Nekrasova et al. (2011) use the Unified Scaling Law for Earthquakes of Bak et al. (2002) despite this having been superseded multiple times.

7.4 Inferring changes in seismicity

The key assumptions in ETAS, of a time-independent externally driven background source plus cascades of triggered events obeying time-independent parameters, form a robust null hypothesis for seismicity. Any proposed variations in seismicity due to some other separate causal mechanism must be shown to reject this null hypothesis. Often this is not the case. For example, figure 7.6 (c) is reproduced from Scholz et al. (1973) in which it is claimed to be an observation of a precursory drop in event rate prior to a large aftershock, motivated by the dilatancy-hardening theory. However it is easy to show that such an observation can arise by chance from even a stationary Poisson process of events, if data is selected. Figure 7.6 (a) shows a simulation of a Poisson process (basically an ETAS simulation with zero branching ratio) with a mean rate of 50 events per day, simulated for a year. In figure 7.6 (b), a six-day subset of this simulation has been hand-picked in which a decrease in the event rate very similar to that in figure 7.6 (c) can be seen. Obviously a full ETAS model with aftershocks would be an even harder null hypothesis to reject.

Of course true temporal variations must sometimes occur in natural seismicity due to fluid or magma-driven triggering. Similar variations will occur with induced anthropogenic seismicity as we adopt new energy technologies such as geothermal heat mining and geological storage of CO2. We would like to be able to infer such changes in the underlying properties in order to improve hazard estimations but we must take care to quantify the uncertainty. Essentially we are trying to infer the properties of a marked point process from data that is 'unmarked', at least in terms of the causal connections between events—a task that poses significant challenges. This is an important research area and an avenue in which the insights gained through my PhD

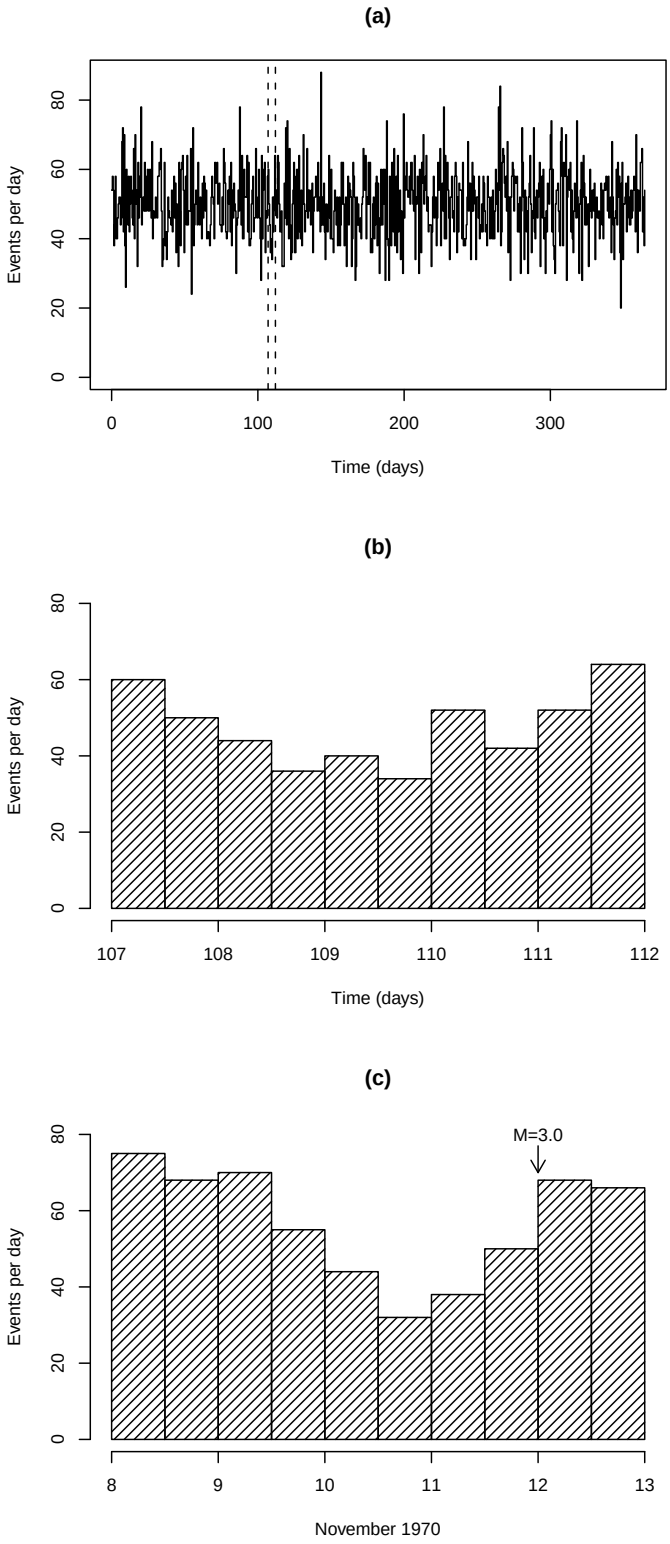


Figure 7.6: (a) A simulated Poisson process of mean 50 events per day, simulated for one year. (b) A six-day subset of the simulation (between dashed vertical lines in (a)) in which a drop in event rate similar to that in (c) can be seen. (c) Reproduced from Figure 7 of Scholz et al. (1973): histogram showing a drop in event rate prior to a large M3.0 aftershock (marked in the figure) of a 1970 earthquake in Fairbanks, Alaska.

work on earthquake time series could provide an important contribution.

In chapter 5 I questioned the procedure of inferring a time-varying background rate through a moving-window ETAS inversion, and showed that an over-estimation of μ can occur when there is strong overlapping of aftershock sequences, making it difficult to infer or quantify an elevated background rate. Some other studies (Lombardi et al., 2006, 2010) carry out similar moving-window inversion analysis to infer time-varying parameters, while others (Ogata, 1992, 1999; Zhuang, 2000; Ma and Zhuang, 2001) look for specific points in time where a change in the parameters occurs, using Akaike Information Criterion (AIC; discussed in the next section) to select between models. Other studies look for long-term clustering not captured by ETAS (Lombardi and Marzocchi, 2007; Faenza et al., 2004), or cyclic additions to the background term (Faenza et al., 2007; Matsu'ura and Karakama, 2005). Much of this work is guided by the identification of change points in residual analysis.

There are two types of residual analysis. Firstly, the residual point process, obtained by transforming the event times t_i to

$$\tau_i = \int_0^{t_i} \lambda(s|H_s)ds \tag{7.12}$$

where λ is the ETAS conditional intensity function, should be a Poisson process with rate 1 if the model is a good fit to the data (Ogata, 1988). Secondly, another type of residual analysis is obtained by thinning the catalogue based on the stochastic identification of aftershocks (Schoenberg, 2003; Zhuang et al., 2005). The thinned process represents the background and can be plotted as a cumulative function of time. This technique comes from the stochastic declustering of Zhuang et al. (2002) in which, after ETAS parameters for the catalogue have been determined, each event may be assigned a probability of being triggered by any of the previous events and (conversely) a probability p_{bg} of being background:

$$\begin{aligned} p_{bg} &= \frac{\mu}{\lambda(t|H_t)} \\ &= \frac{\mu}{\mu + AS} \end{aligned} \tag{7.13}$$

where AS refers to the aftershocks term. Deviations from linearity in the cumulative background probability as a function of time are then taken to indicate temporal variations in the background rate.

For example, Zhuang et al. (2005) plot the inferred cumulative background rate for a subset of Taiwan earthquake data and point out quiescence prior to a large event. This is not backed up with statistical tests and is based only on around 40 inferred background events over 100 years. Lombardi et al. (2010) use both types of residual analysis to assess the suitability of a stationary ETAS model for the 1997–98 Colfiorito sequence in the Umbria–Marche region of Italy and find that under a single maximum-likelihood parameterisation for the sequence, there are change points. When ETAS is fitted to the region's data over a longer period containing this sequence, from 1981–2002, a Kolmogorov-Smirnov test does not reject Poisson residuals. However, a separate inversion for the 1997–98 sequence alone indicates a higher background rate for this period, and ETAS—even when fitted just to the sequence in this way—under-predicts the event rate during the peaks in event rate by around 50%; not surprisingly then

the Poisson process is rejected for the sequence's residual plots by the Kolmogorov-Smirnov test. A time-varying inverted background with a 10-day window is preferred by AIC over the single parameterisation for the sequence. The question this raises for me is, how reliable is this inversion given that the background rate appears to be high (and thus sequences of triggered events are likely to be temporally overlapping)? Is it reliable enough to warrant so many parameters in its description? Could the under-prediction of peak rates be partly due to an insufficient triggering term—a likely consequence of masking by overlapping?

It is possible that the inclusion of spatial data for the events, for example in the space–time ETAS inversion algorithm of Zhuang et al. (2002), may assist the inversion of temporal parameters even when aftershock sequences are temporally overlapping. However, inverting the spatial component of the model is not straightforward: the spatial distribution of aftershocks around a parent event is less well-established empirically and there are several versions in use; the matter is complicated by the complex spatial structure of events and hence the inadequacy of a point approximation particularly for larger events (Hainzl et al., 2008). Additionally, the spatial positioning of events may not help in declustering if the aftershock sequences are overlapping in space as well as in time, which is likely to be the case in a fluid-driven sequence or swarm. Figure 7.7 shows the Colfiorito data analysed by Lombardi et al. (2010), with the 1981–2002 data in the top two plots and the 1997–98 sequence isolated in the bottom two plots. The inter-event time distribution and the space–time nearest-neighbour plot (see Appendix section A.14) both show a bimodal separation of same-sequence and inter-sequence relationships between events for the region over the whole 20 years. For the sequence alone, however, there is no clear separation of these two types of event pairs in either plot. This implies that the spatial data would not always help much in separating aftershocks from spontaneous events.

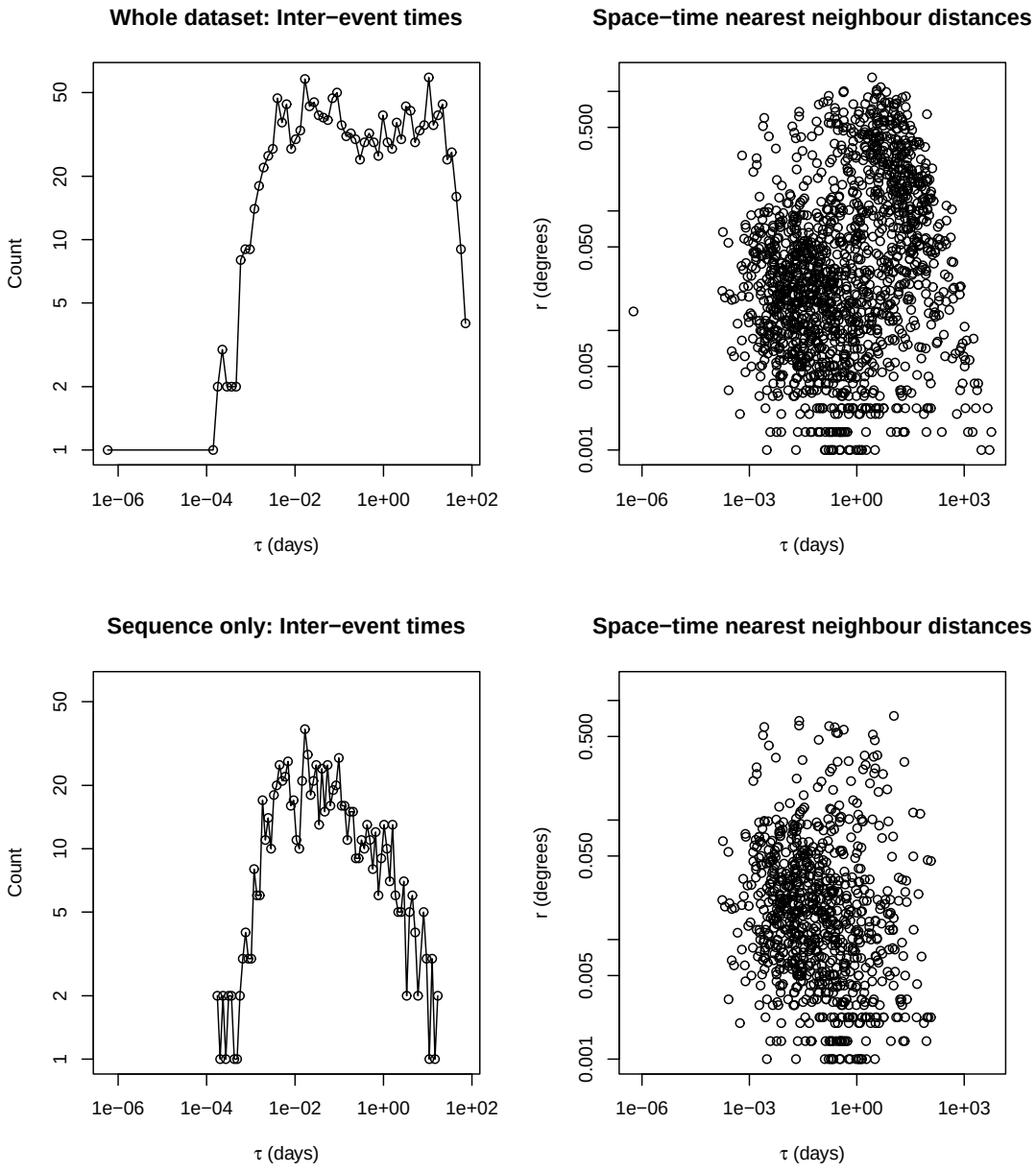


Figure 7.7: Colfiorito data from Lombardi et al. (2010) with events of $M \leq 2.5$ removed. The top plots show data from 1981–2002; the bottom plots show only 3rd May 1997–17th August 1998 data which is the Colfiorito sequence. Inter-event time histograms and nearest-neighbour space–time distance plots are shown on the left and right respectively. Clear bimodality, indicating that aftershock patterns are not masked, is only present for the long-term data.

7.5 Ideas for future work

First of all, I am interested in exploring the reliability of AIC in selecting the correct model. The basic definition of AIC is:

$$AIC = -2 \times \max \text{ LL} + 2 \times \text{no. of parameters} \tag{7.14}$$

where LL refers to log-likelihood (equation 1.7 of the Introduction). (In section 5.2 I used a corrected version of this formula for small sample sizes, but with more than a few events the difference is negligible.) For a combined AIC in the case of two model fits either side of a change point, the AIC is:

$$AIC = AIC_1 + AIC_2 + 2k(N) \tag{7.15}$$

where the subscripts refer to before (1) and after (2) the change point, and $k(N)$ represents the contribution of the change point as an additional parameter which is a function of the total data length N (Ogata, 1992). Its value is around 3. The additional penalty for using two ETAS parameterisations instead of a single one, then, is twice the difference in the number of ETAS parameters used ($2 \times 5 = 10$) plus $2k(N) \simeq 6$; in total a penalty of around 16. The difference in log-likelihood between the double and single parameterisation must be larger than this if the double model is to be preferred. The results of my inversions on synthetic catalogues showed that there can be a large uncertainty in both the inverted parameter values and the standard error, as well as a bias when sequences overlap to a large extent. I did not study the uncertainty or variability in the maximum log-likelihood values, but in light of my results, I suspect the penalty for extra parameters may sometimes be dwarfed by the larger uncertainty of the log-likelihood, especially given that the log-likelihood values themselves are typically orders of magnitude larger than the penalty.

Secondly, I think it would be interesting to test under what circumstances these residual plots show apparent change points and what can be inferred from them. Wang et al. (2010) show that the different tests for stationarity give different results and that the stationarity of the inferred background process strongly depends on the declustering algorithm used, with the Zhuang et al. (2005) method most often leaving a stationary background. The problem is made even worse by the inhomogeneity of some earthquake catalogues due to changes in the magnitude determining procedures, for example; Abhey Bhansal (in unpublished work) discovered a change of slope that could be explained by exactly this for Sumatran data prior to the 2004 Boxing Day mega-earthquake there. It is useful to produce the residual plots for synthetic ETAS catalogues, which have the advantage that the model perfectly matches the data and the parameter values are known, so that we can explore the effect of various parameters and also explore what happens when the estimated parameters used to produce the plot are significantly different from the true values. In the remainder of this section I focus on the inferred cumulative background probability plot as an example of what exploration may be carried out.

Figure 7.8 shows an example of this plot (see Appendix section A.16 for how this is done). The solid black curve is the background rate inferred using parameter values equal to the true (known) parameters; i.e., assuming that the parameter inversion had produced the correct result. The red and blue curves are produced using wrong values of μ (too small and too large,

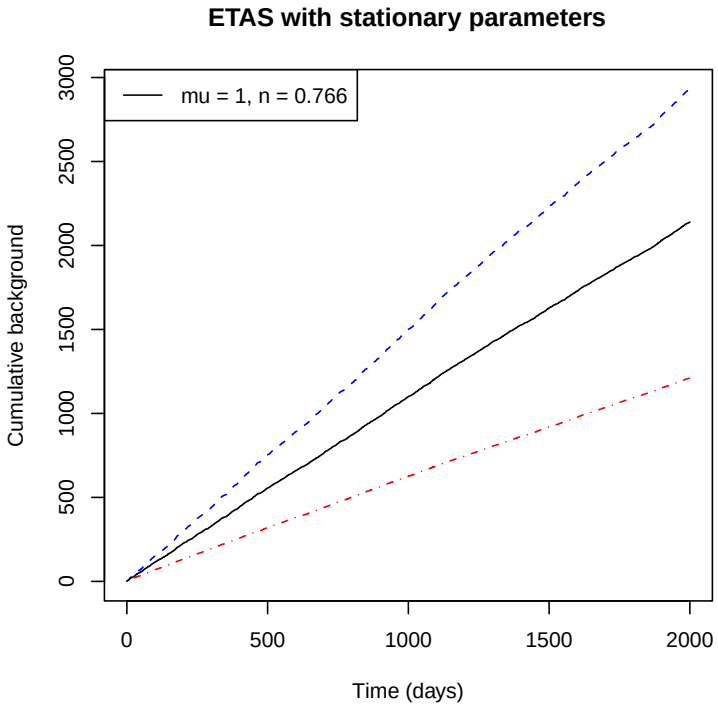


Figure 7.8: Inferred cumulative background from a stationary ETAS simulation, using ETAS conditional intensity to calculate the background probability with correct parameters (solid black curve), and with μ 50% above (blue) and 50% below (red) the correct value. In all cases, a constant background rate is inferred.

respectively), with branching ratio n adjusted to keep the overall event rate approximately accurate. Straight lines are produced in all cases. This is also found to be true even without adjusting n (not shown), and is not dependent on the value of μ used in the simulation. So the first observation is that when the stationary ETAS model is a good description of the data, the background will look uniform in time, regardless of whether the parameters have been inverted correctly.

The next step is to perform the analysis on synthetic catalogues in which there is a period of elevated background rate, as this is often inferred to be the case in real seismicity during swarms or other anomalously high activity. I produced ETAS simulations of 2000 days in which the background rate is increased temporarily, from 0.1 to 1, during the time period 1000–1100 days (see Appendix section A.15 for how this is done). Figure 7.9 shows the true cumulative number of background events against time for such a simulation, to illustrate this effect.

The inferred background for this situation is shown in figure 7.10, for 100 simulations of each of two values of the branching ratio (with low and high n shown in the top and bottom plots, respectively). As before, the solid black curves in both plots show the result of inferring the background assuming the parameter inversion produced "correct" results—using a constant μ value of 0.1 (which is the correct value outside of the elevated period) and having all other parameters at their correct values. A temporary increase in slope is observed during the period 1000–1100 days, indicating the raised background rate. But comparing the two plots, we note that the increase in slope depends strongly on the true branching ratio: the rate change is only detected strongly if n is low.

A further effect to note is that when the value of μ used to compute the probabilities is higher than the true value (the blue curves in figure 7.10), the change in slope is more noticeable. This is interesting to know because my results in chapter 5 showed that under some circumstances, the inverted value of μ is likely to be significantly larger than the true effective value. When the inverted branching ratio is high, the Zhuang et al. (2005) calculation accommodates most of the increased event rate in the aftershocks term. The calculated probability of an event being a background event (equation (7.13)) becomes much smaller due to the expectation of many aftershocks, which to some extent cancels the tendency for it to increase through summing many such terms over the time period. In the opposite case of low inverted n and high inverted μ, the relative reduction in the aftershocks term (n being lower than the true value) means that it can't accommodate so much of the increase in the event rate.

Finally, we consider simulated catalogues where there is a period of elevated n, to discover whether this can be distinguished from the elevated μ case using this type of plot. I use similar ETAS simulations of 2000 days in which n is increased temporarily from $\sim$0.15 to $\sim$0.77 during the time period 1000–1500 days. Regardless of the true μ value, the slope of the inferred background curve changes during the period of elevated triggering (figure 7.11). An elevated background would be the wrong interpretation in this case although importantly, in terms of the plot, it is indistinguishable from the case where the background rate is genuinely raised. Again, the higher the value of μ used in the calculations, the more dramatic the change in slope.

To summarise, when the temporally-homogeneous ETAS model is a good match to the data, the inferred background using the Zhuang et al. (2002) stochastic declustering will tend to look uniform in time regardless of whether the parameters have been inverted correctly. When the effective branching ratio is temporarily raised, this will (wrongly) appear as a raised background

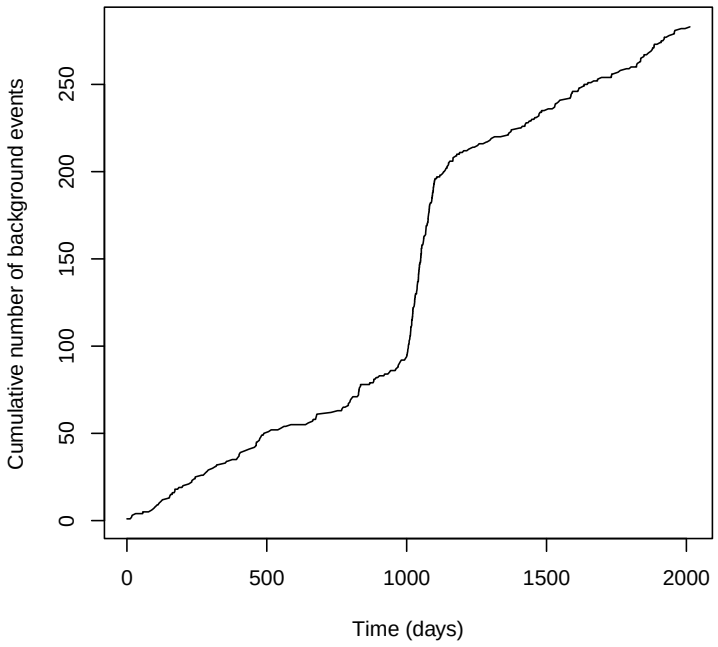
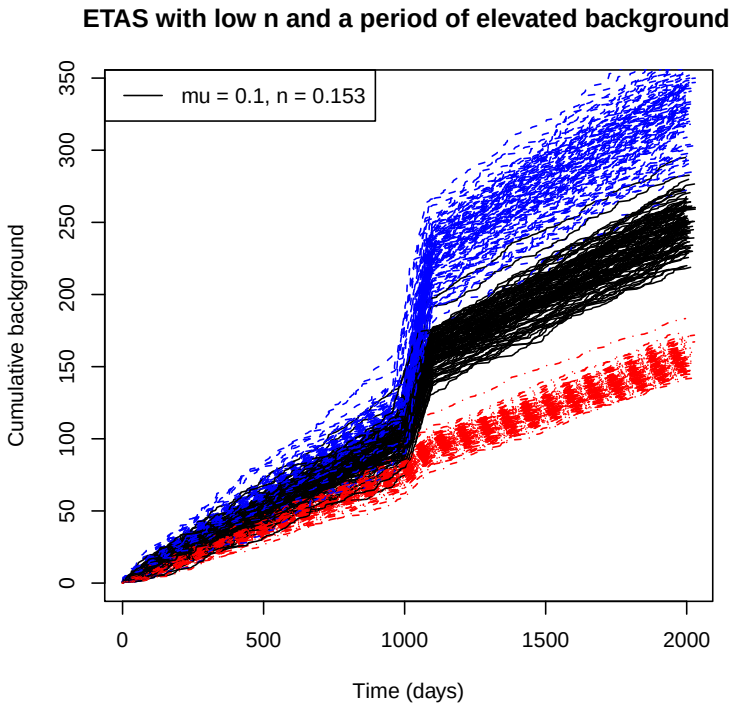


Figure 7.9: Cumulative (true) number of background events from an ETAS simulation with a period of elevated background rate, from 0.1 to 1 events per day, during the time period 1000–1100 days.



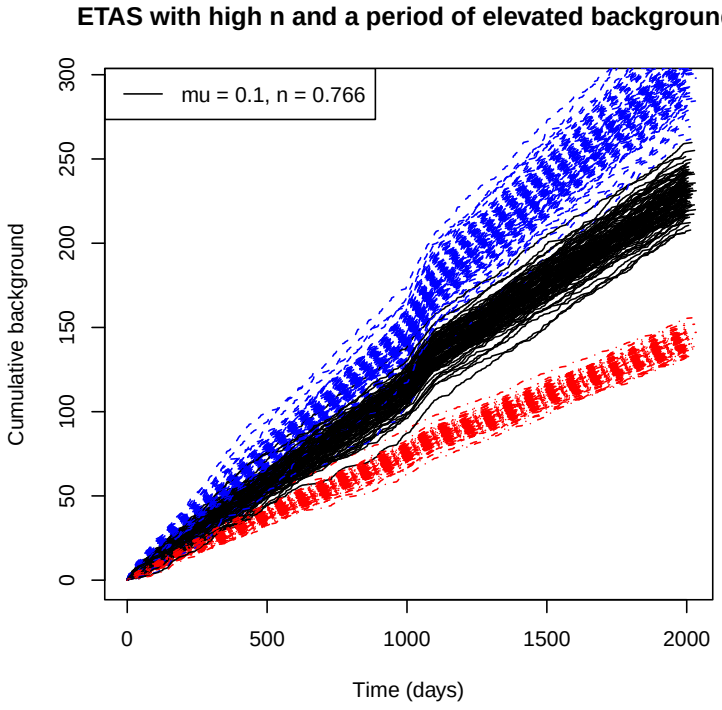


Figure 7.10: Top: inferred cumulative background from 100 ETAS simulations with a period of elevated background rate as in the previous figure. The background probability is calculated with correct parameters (solid black curve), and with μ 50% above (blue) and 50% below (red) the correct value (where "correct" refers to the values used in the simulation excluding the elevated background period). The period of elevated background rate is visible, although diminished when the values of μ and n used to compute the probability are lower and higher than the true values, respectively. Bottom: similar, except that the simulations were run with a higher value of n. The change in slope is much less clear in this case.

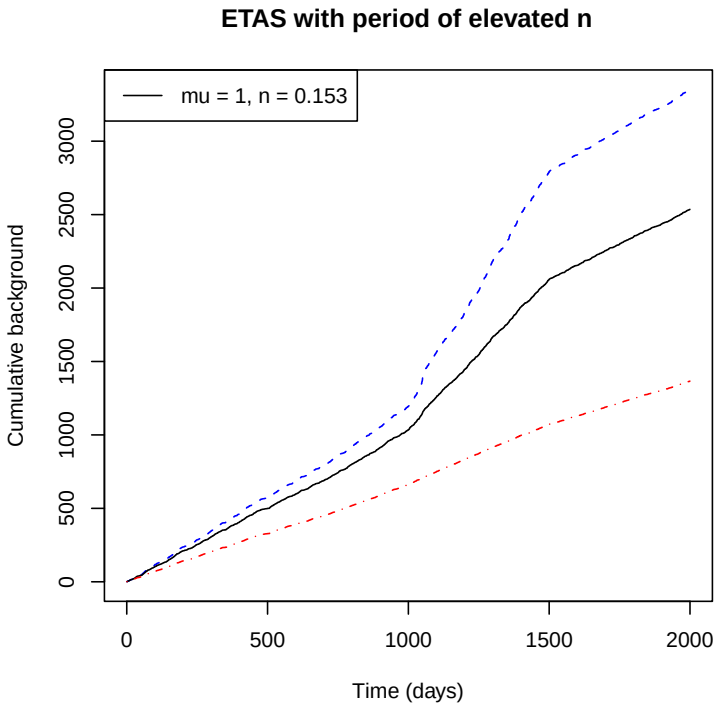


Figure 7.11: Inferred cumulative background from an ETAS simulation with a period of elevated branching ratio n, from 0.15 to 0.77, during the time period 1000–1500 days. The background probability is calculated with correct parameters (solid black curve), and with μ 50% above (blue) and 50% below (red) the correct value (where "correct" refers to the values used in the simulation excluding the elevated triggering period). In all cases, the period of elevated n shows up as an elevated background rate, although diminished when the values of μ and n used are lower and higher than the true values, respectively.

rate using this analysis. When the effective background rate is temporarily raised, this will also appear as a raised background rate, but more strongly so the lower the effective n is. It will show up more strongly if the inverted μ value is too high and the inverted n is too low.

In short, this type of plot can give an indication of a temporal change in seismicity, but it does not necessarily reflect a change in the background rate, nor does any change in slope give a true reflection of the strength of the change. My results lead me to expect that greater temporal overlapping of sequences would make it more difficult to detect spontaneous rate changes, and in any case, quantifying the uncertainty within these techniques is a worthwhile goal for the future.

Chapter 8

Conclusions

While exhibiting many of the key properties of self-organised critical (SOC) systems, natural seismicity has some significant differences to other such systems, notably due to the presence of aftershocks. This adds a secondary near-critical process, with a causal spatio-temporal organisation that is fundamentally different to that of the random spontaneous process, albeit occurring in the same crust on the same fractal fault network. Critically, the spontaneous event rate is a function of data selection criteria such as the chosen spatial bounds of the catalogue. In contrast the aftershock rate does not vary so much with the area of the spatial sample, as aftershocks mostly occur close in space to the parent event. As a result the earthquake process cannot be unified into a single scaling law for all events.

Recent attempts to achieve universal scaling laws for earthquake spatio-temporal processes, for example in the inter-event time distribution, ignored this fundamental difference, and were facilitated by subjective data selection (notably for visually 'stationary' time series); by the mathematical properties of the gamma distribution used to fit the data; and by the graphical methods of visualisation used. Instead the time intervals between successive earthquakes have a fundamentally bimodal distribution in the general case, reflecting the two different timescales for spontaneous and triggered events in a straightforward way. This property can be captured very well by the statistical Epidemic-Type Aftershock Sequence (ETAS) model with its explicit spontaneous and triggered components. The model also quantitatively explains the slow convergence of simple metrics such as mean inter-event time or mean event rate to the central limit in real earthquake catalogues.

The spontaneous rate (or, as a rough proxy, region size) strongly influences the extent to which aftershock sequences overlap in time. This in turn influences the proportion of inter-event times that result from successive events that are not causally related. While the ETAS triggering parameters also have an effect, the spontaneous rate is more tunable since it depends on the spatial bounds of the data set, so is more practically relevant. As the spontaneous event rate increases, aftershock sequences increasingly overlap. This has a number of interesting and important consequences:

- The inter-event time distribution becomes exponential.
- The time series of events resembles a stationary Poisson process.
- The mean inter-event time converges more rapidly with respect to the number N of

recorded events, approaching the rate $1/\sqrt{N}$ of the Central Limit Theorem, consistent with a comparative absence of correlations between successive events.

- The maximum-likelihood inversion of ETAS parameters becomes less accurate, with standard errors that typically do not accurately reflect the true error, and with a systematic bias towards a larger spontaneous rate and smaller branching ratio than the true values.

The last point is a particularly good illustration of how aftershock characteristics are 'masked' when they overlap, making the underlying parameters difficult to determine; at a fundamental level the properties of a marked point process such as earthquake magnitude time series cannot be determined uniquely from unmarked data. This is important because ETAS parameter inversions are routinely carried out to assist in the geophysical interpretation of past sequences of events and in recent attempts at real-time operational earthquake hazard forecasting. Further work is needed to develop pragmatic ways of minimising the impact of this source of bias.

Bimodality in the inter-event time distribution is also observed in volcanic earthquake data, but do not necessarily conform to the ETAS model, in contrast to all the examples of natural seismicity considered here. In such settings the spontaneous process may not be stationary (e.g. due to magma injection), and additional timescales may be involved (e.g. due to stress relaxation and/or time-dependent weakening of the volcanic edifice). As a consequence bimodality may result instead from a superposition of transient (aftershock-like) and accelerating processes, consistent with data from the caldera at Kilauea, Hawaii.

Appendix A

Additional R Codes

This Appendix contains further R codes used in the research: sections A.1 to A.9 present self-contained functions referred to from the Methods chapter 2, and section A.10 onwards concern more chapter-specific methods which are referred to from the individual chapters.

A.1 SSLib function to simulate the temporal ETAS model

The following is the SSLib version of the simulation function for creating synthetic ETAS catalogues, referred to in Methods section 2.3.

```
pp.sim.default
function (data, params, cif, TT, output = FALSE, seed = 5, magn.sim = 1,
    stopping.condition = NULL, max.rate = NA)
{
    if (!is.null(data)) {
        use <- (data[, "time"] < TT[1])
        if (sum(use) == 0)
            data <- NULL
        else data <- data[use, c("time", "magnitude")]
    }
    set.seed(seed)
    if (output)
        cat("\nSimulation Begins\nTime              Magnitude\n")
    ti <- TT[1]
    repeat {
        if (is.null(attr(cif, "rate")))
            stop("rate attribute is not specified on cif")
        else if (attr(cif, "rate") == "decreasing") {
            Rmax <- cif(data = data, eval.pts = ti, params = params,
                t.plus = TRUE)
            tau <- rexp(1, rate = Rmax)
            rate <- cif(data = data, eval.pts = ti + tau, params = params)
            if (rate > Rmax)
```

```
24                  stop("cif is not decreasing")
25          }
26          else if (attr(cif, "rate") == "bounded") {
27              Rmax <- max.rate
28              tau <- rexp(1, rate = Rmax)
29              rate <- cif(data = data, eval.pts = ti + tau, params = params)
30              if (rate > Rmax)
31                  stop("cif is not bounded by max.rate")
32          }
33          else if (attr(cif, "rate") == "increasing") {
34              Rmax <- 0
35              rate <- cif(data = data, eval.pts = ti, params = params,
36                  t.plus = TRUE)
37              tmax <- ti
38              while (rate > Rmax) {
39                  ti <- tmax
40                  tmax <- ti + qexp(0.7, rate = rate)
41                  Rmax <- cif(data = data, eval.pts = tmax, params = params)
42                  tau <- rexp(1, rate = Rmax)
43                  rate <- cif(data = data, eval.pts = ti + tau,
44                    params = params)
45              }
46          }
47          else stop("unknown value of rate attribute on cif")
48          ti <- ti + tau
49          if (ti > TT[2])
50              break
51          if (runif(1, 0, 1) <= rate/Rmax) {
52              if (is.vector(magn.sim)) {
53                  magnitude <- rexp(1, magn.sim[1] * log(10))
54                  if (length(magn.sim) == 2)
55                    magnitude <- min(magnitude, magn.sim[2])
56              }
57              else if (is.matrix(magn.sim))
58                  magnitude <- as.numeric(dimnames(magn.sim)[[1]][seq(1,
59                    nrow(magn.sim))[magn.sim > runif(1)][1]])
60              if (output)
61                  cat(paste(ti, "    ", magnitude, "\n"))
62              data <- rbind(data, c(time = ti, magnitude = magnitude))
63              if (!is.null(stopping.condition))
64                  if (stopping.condition(data))
65                    break
66          }
67      }
68      return(as.data.frame(data))
```

```
}
```

A.2 Function to simulate the temporal ETAS model

The following code is my own ETAS simulation function, based on the SSLib version in the previous section, referred to in Methods section 2.3. It calls the functions `create.aftershocks` presented in the next section (A.3), and `calc.n` (section A.4).

```
etas.sim <- function (params, max.events, m.min=0, seed = 5, bvalue = 1)
{
    set.seed(seed)

    n <- calc.n(params, bvalue)
    mu <- params[1]
    if(max.events < 100000 & n > 0) working.max.events <- 100000
    else working.max.events <- max.events


    # work out the temporal length we should run each aftershock sequence
    fraction <- 0.05
    event.rate <- mu * (1 + n*(1-fraction)/(1-n))
    sim.length <- working.max.events/event.rate
    cc <- params[4]
    theta <- params[5] - 1
    sequence.length <- max((cc * fraction^(-1/theta)), sim.length)

    # shorten sequence.length if it is going to result in too many wraps
    num.spont <- sim.length * mu
    max.poss.wraps <- sequence.length/sim.length
    if(max.poss.wraps * num.spont > 1000)
    {
        max.poss.wraps <- 1000
        sequence.length <- max.poss.wraps * sim.length
        fraction <- (cc/sequence.length)^theta
    }


    # create background sequence to pass into simulation: fill the whole
    # time period (plus an extra event) OR if max.events is specified,
    # estimate the required number of background events conservatively
    bg.times <- c()
    bg.magnitudes <- c()
    bg.gas.indices <- c()
    index <- 1
    ti <- 0
```

```
38      min.bg.events <- working.max.events * 10^(bvalue*m.min) / (1+n)
39      repeat
40      {
41          # increment ti
42          tau <- rexp(1, rate = mu)
43          ti <- ti + tau
44
45          # magnitude
46          magnitude <- rexp(1, bvalue * log(10))
47
48          # add the event to bg.events
49          bg.times <- c(bg.times,ti)
50          bg.magnitudes <- c(bg.magnitudes,magnitude)
51          bg.gas.indices <- c(bg.gas.indices,as.character(index))
52
53          index <- index + 1
54
55          # check if we have (more than) enough independent events
56          if(length(bg.times) >= min.bg.events) break
57      }
58
59
60      # set the background rate (mu) to 0
61      params[1] <- 0
62
63
64      # create the aftershocks
65      all.events <- create.aftershocks(times=bg.times,
66          magnitudes=bg.magnitudes, gas.indices=bg.gas.indices,
67          params=params, m.min=m.min, bvalue=bvalue,
68          sequence.length=sequence.length, start.time=0,
69          max.events=working.max.events)
70
71
72      # extract the data
73      times <- all.events$times
74      magnitudes <- all.events$magnitudes
75      gas.indices <- all.events$gas.indices
76      remove(all.events)
77
78
79      # order the events chronologically
80      ii <- order(times)
81      times <- times[ii]
82      magnitudes <- magnitudes[ii]
```

```
83      gas.indices <- gas.indices[ii]
84
85
86      # truncate data to required length
87      max.time <- Inf
88      if(length(times) > max.events) max.time <- times[max.events]
89      use <- times <= max.time
90      times <- times[use]
91      magnitudes <- magnitudes[use]
92      gas.indices <- gas.indices[use]
93
94      return.object <- list(time=times, magnitude=magnitudes,
95          gas.indices=gas.indices)
96      sim <- as.data.frame(return.object)
97      return(sim)
98  }
```

A.3 Function to create aftershocks in the temporal ETAS model

This function is called by the ETAS simulation algorithm in the previous section. It calls the function `conditional.intensity` shown in section A.5.

```
1   create.aftershocks <- function(times, magnitudes, gas.indices, params,
2       m.min=0, bvalue, sequence.length, start.time, max.events)
3   {
4       # objects to eventually hold all events (those passed in to the
5       # function, and aftershocks created for them)
6       new.times <- c()
7       new.magnitudes <- c()
8       new.gas.indices <- c()
9
10      for(i in 1:length(times))
11      {
12          # select an event
13          parent.time <- times[i]
14          parent.mag <- magnitudes[i]
15          parent.index <- gas.indices[i]
16
17          # make its time the starting time
18          ti <- parent.time
19
20          # if this is a b/g event (no dot in index), check if the number of
21          # events collected so far is enough yet
22          if(i > 1 & length(grep("[0-9]+[.][0-9].*",parent.index))==0)
```

```
23          {
24              num.aftershocks <- length(new.times[new.magnitudes >= m.min])
25              num.mainshocks <-
26                  length(times[magnitudes >= m.min & times < ti])
27              if((num.aftershocks + num.mainshocks) > max.events)
28              {
29                  # we have enough events, so we can finish here.
30                  end.time <- times[i-1]
31
32                  # throw away unused events
33                  use <- magnitudes >= m.min & times < ti
34                  times <- times[use]
35                  magnitudes <- magnitudes[use]
36                  gas.indices <- gas.indices[use]
37                  use <- new.magnitudes >= m.min
38                  new.times <- new.times[use]
39                  new.magnitudes <- new.magnitudes[use]
40                  new.gas.indices <- new.gas.indices[use]
41
42
43                  # wrap aftershocks occurring after the end time around to
44                  # the beginning, and reset the first-generation part of
45                  # their index to break the connection with the later ones
46                  event.indices <- which(new.times > end.time)
47                  if(length(event.indices) > 0)
48                  {
49                      diffs <- new.times[event.indices] - end.time
50                      sim.length <- end.time - start.time
51                      number.of.wraps <- ceiling(diffs/sim.length)
52                      to.add <- number.of.wraps *
53                          as.numeric(gas.indices[num.mainshocks])
54                      all.parent.indices <- as.numeric(sub("[.].*", "",
55                          new.gas.indices[event.indices]))
56                      new.parent.indices <- all.parent.indices + to.add
57                      for(j in 1:length(new.parent.indices))
58                      {
59                          new.gas.indices[event.indices[j]] <- sub("^[0-9]+",
60                              new.parent.indices[j],
61                              new.gas.indices[event.indices[j]])
62                      }
63                      new.times[event.indices] <- new.times[event.indices] -
64                          (number.of.wraps * sim.length)
65                  }
66
67                  # include mainshocks in returned events
```

```
68                  new.times <- c(times, new.times)
69                  new.magnitudes <- c(magnitudes, new.magnitudes)
70                  new.gas.indices <- c(gas.indices, new.gas.indices)
71
72                  final.events <- list(times=new.times,
73                      magnitudes=new.magnitudes,
74                      gas.indices=new.gas.indices)
75                  return(final.events)
76              }
77          }
78          # objects to hold its aftershocks
79          aftershock.times <- c()
80          aftershock.magnitudes <- c()
81          aftershock.gas.indices <- c()
82
83          # calculate the initial rate, at the time of this event
84          Rmax <- conditional.intensity(data.mag=parent.mag,
85              data.time=parent.time, eval.time=ti, params=params)
86          repeat
87          {
88              # increment the time by an appropriate amount, tau
89              if(Rmax > 0) tau <- rexp(1, rate = Rmax)
90              else tau <- Inf
91              ti <- ti + tau
92
93              # check if this sequence is long enough yet
94              if ((ti-parent.time) > sequence.length) break
95
96              # calculate the new rate
97              rate <- conditional.intensity(data.mag=parent.mag,
98                  data.time=parent.time, eval.time=ti, params=params)
99
100             # decide whether to create an aftershock at this time
101             # (thinning method)
102             if (runif(1, 0, 1) <= rate/Rmax)
103             {
104                 # select a magnitude
105                 new.mag <- rexp(1, bvalue * log(10))
106
107                 # make the index the same as the parent's but with ".i"
108                 # appended, where i is the aftershock number within this
109                 # sequence
110                 new.index <- paste(parent.index, ".",
111                     length(aftershock.times), sep="")
112
```

```
113                 # add the aftershock to the data objects
114                 aftershock.times <- c(aftershock.times, ti)
115                 aftershock.magnitudes <- c(aftershock.magnitudes, new.mag)
116                 aftershock.gas.indices <-
117                     c(aftershock.gas.indices, new.index)
118             }
119
120             # re-set the initial rate for the next iteration
121             Rmax <- rate
122         }
123
124         if(length(aftershock.times) > 0)
125         {
126             # now create aftershock sequences for each of the aftershocks
127             aftershocks <- create.aftershocks(times=aftershock.times,
128                 magnitudes=aftershock.magnitudes,
129                 gas.indices=aftershock.gas.indices, params=params,
130                 m.min=m.min, bvalue=bvalue,
131                 sequence.length=sequence.length, start.time=start.time,
132                 max.events=max.events)
133             remove(aftershock.times)
134             remove(aftershock.magnitudes)
135             remove(aftershock.gas.indices)
136
137             # add the new generations to our collection of events
138             new.times <- c(new.times, aftershocks$time)
139             new.magnitudes <- c(new.magnitudes, aftershocks$magnitude)
140             new.gas.indices <- c(new.gas.indices, aftershocks$gas.indices)
141         }
142     }
143
144     # include parent events in returned events
145     new.times <- c(times, new.times)
146     new.magnitudes <- c(magnitudes, new.magnitudes)
147     new.gas.indices <- c(gas.indices, new.gas.indices)
148
149     events <- list(times=new.times, magnitudes=new.magnitudes,
150         gas.indices=new.gas.indices)
151     return(events)
152 }
```

A.4 Function to calculate the branching ratio

A simple function to calculate the branching ratio for a set of ETAS parameter values.

```
calc.n <- function(params, bvalue, m.max=NULL, m0=NULL)
{
    beta <- bvalue*log(10)
    A <- params[2]
    alpha <- params[3]
    c <- params[4]
    p <- params[5]

    n <- A*c/(p-1) * beta/(beta-alpha)
    if(!is.null(m.max))
    {
        n <- n *
            (1-exp(-(beta-alpha)*(m.max-m0)))/(1-exp(-beta*(m.max-m0)))
    }

    return(n)
}
```

A.5 Function to calculate the ETAS conditional intensity

A simple function to calculate the conditional intensity for an ETAS model with specified parameters and event history.

```
conditional.intensity <- function(data.mag, data.time, eval.time, params)
{
    mu <- params[1]
    A <- params[2]
    alpha <- params[3]
    CC <- params[4]
    P <- params[5]

    if (length(data.time) > 0)
    {
        triggering.ci <- A * sum(exp(params[3] * data.mag) *
            (1 + (eval.time - data.time)/CC)^(-P))
    }
    else triggering.ci <- 0
    ci <- mu + triggering.ci
    return(ci)
}

```

A.6 Function to simulate the temporal ETAS model with a single seeding event

A version of the ETAS simulation function that creates a single spontaneous event at time 0. Can be used to simulate the average global aftershock sequence as described in Methods section 2.3.

```
etas.sim.agas <- function (params, seed = 5, bvalue = 1,
    sequence.length=NULL)
{
    set.seed(seed)


    if(is.null(sequence.length))
    {
        # work out the temporal length we should run the aftershock
        # sequence to
        fraction <- 0.05
        cc <- params[4]
        theta <- params[5] - 1
        sequence.length <- cc * fraction^(-1/theta)
    }


    # create seeding event first
    magnitude <- rexp(1, bvalue * log(10))


    # set the background rate (mu) to 0
    params[1] <- 0


    # create the aftershocks
    all.events <- create.aftershocks(times=0, magnitudes=magnitude,
        gas.indices=1, params=params, bvalue=bvalue,
        sequence.length=sequence.length, start.time=0, max.events=Inf)


    # extract the data
    times <- all.events$times
    magnitudes <- all.events$magnitudes
    gas.indices <- all.events$gas.indices
    remove(all.events)


```

```
39      # order the events chronologically
40      ii <- order(times)
41      times <- times[ii]
42      magnitudes <- magnitudes[ii]
43      gas.indices <- gas.indices[ii]
44
45
46      return.object <- list(time=times, magnitude=magnitudes,
47          gas.indices=gas.indices)
48      sim <- as.data.frame(return.object)
49      return(sim)
50  }
```

A.7 Function to create an inter-event time histogram

The following code creates a logarithmic histogram of inter-event times with 80 bins (by default), as described in Methods section 2.5.

```
1   create.iet.hist <- function(iets, breaks=NULL, normalise=FAlSE)
2   {
3       # create logarithmic bins
4       if(is.null(breaks))
5           breaks <- seq(from=min(log10(iets)), to=max(log10(iets)),
6               length.out=80)
7
8       # histogram the logarithm of the IETs
9       histogram <- hist(log10(iets), breaks=breaks, plot=FALSE)
10      counts <- histogram$counts
11      breaks <- 10^histogram$breaks
12
13      if(normalise)
14      {
15          bin.widths <-
16              breaks[2:length(breaks)] - breaks[1:(length(breaks)-1)]
17          counts <- counts/bin.widths
18      }
19
20      x <- breaks[1:(length(breaks)-1)]
21      y <- counts
22
23      return.obj <- structure(list(x=x, y=y), .Names = c("x", "y"))
24      return(return.obj)
25  }
```

A.8 Function to simulate the spatial ETAS model

The following code is my spatial ETAS simulation function, based on the temporal version in section A.2, and referred to in Methods section 2.4. It calls a function `create.aftershocks.spatial` presented in the next section A.9.

```
etas.sim.spatial <- function (params, coords, max.events, seed = 5,
    bvalue = 1)
{
    set.seed(seed)


    # work out the temporal length we should run each aftershock sequence
    fraction <- 0.05 ## this is the fraction of aftershocks we want
                     ## each sequence to be short by, on average
    cc <- params[4]
    theta <- params[5] - 1
    sequence.length <- cc * fraction^(-1/theta)


    # work out the spatial length each sequence should be on average
    fraction <- 0.05 ## this is the fraction of aftershocks we want
                     ## each sequence to be short by, on average
    dd <- params[6]
    qq <- params[7]
    sequence.spatial.length <- dd * sqrt(fraction^(-1/(qq-1)) - 1)


    # create background sequence to pass into simulation, estimating the
    # required number of background events conservatively
    bg.times <- c()
    bg.magnitudes <- c()
    bg.xs <- c()
    bg.ys <- c()
    bg.gas.indices <- c()
    mu <- params[1]
    if(length(coords)==2)
    {
        centre <- coords[1]
        radius <- coords[2]
    }
    else
    {
        x1 <- coords[1]
        x2 <- coords[2]
        y1 <- coords[3]
```

```
41          y2 <- coords[4]
42      }
43      index <- 1
44      ti <- 0
45      n <- calc.n(params,bvalue)
46      repeat
47      {
48          # increment ti
49          tau <- rexp(1, rate = mu)
50          ti <- ti + tau
51
52          # magnitude
53          magnitude <- rexp(1, bvalue * log(10))
54
55          # coords
56          reject <- FALSE
57          if(length(coords)==2)
58          {
59              x <- runif(1, (centre-radius-2*sequence.spatial.length),
60                  (centre+radius+2*sequence.spatial.length))
61              y <- runif(1, (centre-radius-2*sequence.spatial.length),
62                  (centre+radius+2*sequence.spatial.length))
63              if(sqrt((x-centre)^2 + (y-centre)^2) >
64                  (radius + 2*sequence.spatial.length)) reject <- TRUE
65          }
66          else
67          {
68              x <- runif(1, (x1-2*sequence.spatial.length),
69                  (x2+2*sequence.spatial.length))
70              y <- runif(1, (y1-2*sequence.spatial.length),
71                  (y2+2*sequence.spatial.length))
72          }
73
74          if(!reject)
75          {
76              # add the event to bg.events
77              bg.times <- c(bg.times,ti)
78              bg.magnitudes <- c(bg.magnitudes,magnitude)
79              bg.xs <- c(bg.xs,x)
80              bg.ys <- c(bg.ys,y)
81              bg.gas.indices <- c(bg.gas.indices,as.character(index))
82
83              index <- index + 1
84
85              # check if we have (more than) enough independent events
```

```
            if(length(bg.times) >= max.events/(1+n)) break
        }
    }


    # set the background rate (mu) to 0
    params[1] <- 0


    # create the aftershocks
    all.events <- create.aftershocks.spatial(times=bg.times,
        magnitudes=bg.magnitudes, xs=bg.xs, ys=bg.ys,
        gas.indices=bg.gas.indices, params=params, coords=coords,
        bvalue=bvalue, sequence.length=sequence.length, start.time=0,
        max.events=max.events)


    # extract the data
    times <- all.events$times
    magnitudes <- all.events$magnitudes
    xs <- all.events$xs
    ys <- all.events$ys
    gas.indices <- all.events$gas.indices
    remove(all.events)


    # order the events chronologically
    ii <- order(times)
    times <- times[ii]
    magnitudes <- magnitudes[ii]
    xs <- xs[ii]
    ys <- ys[ii]
    gas.indices <- gas.indices[ii]


    # truncate data to required length
    max.time <- Inf
    if(length(times) > max.events) max.time <- times[max.events]
    use <- times <= max.time
    times <- times[use]
    magnitudes <- magnitudes[use]
    xs <- xs[use]
    ys <- ys[use]
    gas.indices <- gas.indices[use]

```

```
131
132     return.object <- list(time=times, magnitude=magnitudes, x=xs, y=ys,
133         gas.indices=gas.indices)
134     sim <- as.data.frame(return.object)
135     return(sim)
136 }
137
```

A.9 Function to create aftershocks in the spatial ETAS model

This function is called by the spatial ETAS simulation algorithm in the previous section.

```
1  create.aftershocks.spatial <- function(times, magnitudes, xs, ys,
2      gas.indices, params, coords, bvalue, sequence.length, start.time,
3      max.events)
4  {
5      if(length(coords)==2)
6      {
7          centre <- coords[1]
8          radius <- coords[2]
9      }
10     else
11     {
12         x1 <- coords[1]
13         x2 <- coords[2]
14         y1 <- coords[3]
15         y2 <- coords[4]
16     }
17
18     # objects to eventually hold all events (those passed in to the
19     # function, and aftershocks created for them)
20     new.times <- c()
21     new.magnitudes <- c()
22     new.xs <- c()
23     new.ys <- c()
24     new.gas.indices <- c()
25
26     num.mainshocks <- 0
27     for(i in 1:length(times))
28     {
29         # select an event
30         parent.time <- times[i]
31         parent.mag <- magnitudes[i]
32         parent.x <- xs[i]
33         parent.y <- ys[i]
```

```
34          parent.index <- gas.indices[i]
35
36          # make its time the starting time
37          ti <- parent.time
38
39          # if this is a b/g event (no dot in index), check if the number of
40          # events collected so far is enough yet
41          if(i > 1 & length(grep("[0-9]+[.][0-9].*",parent.index))==0)
42          {
43              if(length(coords)==2)
44              {
45                  new.rs <- sqrt((new.xs-centre)^2 + (new.ys-centre)^2)
46                  use <- new.rs <= radius
47              }
48              else use <-
49                  new.xs < x2 & new.xs >= x1 & new.ys < y2 & new.ys >= y1
50              num.aftershocks <- length(new.times[use])
51              if((num.aftershocks + num.mainshocks) > max.events)
52              {
53                  # we have enough events, so we can finish here.
54                  end.time <- times[i-1]
55
56                  # throw away unused mainshocks
57                  times <- times[1:(i-1)]
58                  magnitudes <- magnitudes[1:(i-1)]
59                  xs <- xs[1:(i-1)]
60                  ys <- ys[1:(i-1)]
61                  gas.indices <- gas.indices[1:(i-1)]
62
63                  # remove events outside boundaries
64                  new.times <- new.times[use]
65                  new.magnitudes <- new.magnitudes[use]
66                  new.xs <- new.xs[use]
67                  new.ys <- new.ys[use]
68                  new.gas.indices <- new.gas.indices[use]
69                  if(length(coords)==2)
70                  {
71                      rs <- sqrt((xs-centre)^2 + (ys-centre)^2)
72                      use <- rs <= radius
73                  }
74                  else use <- xs < x2 & xs >= x1 & ys < y2 & ys >= y1
75                  times <- times[use]
76                  magnitudes <- magnitudes[use]
77                  xs <- xs[use]
78                  ys <- ys[use]
```

```
79                  gas.indices <- gas.indices[use]
80
81                  # wrap aftershocks occurring after the end time around to
82                  # the beginning, and reset the first-generation part of
83                  # their index to break the connection with the later ones
84                  event.indices <- which(new.times > end.time)
85                  diffs <- new.times[event.indices] - end.time
86                  sim.length <- end.time - start.time
87                  wraps <- floor(diffs/sim.length)
88                  all.parent.indices <- as.numeric(sub("[.].*", "",
89                      new.gas.indices[event.indices]))
90                  to.add <- wraps * max(all.parent.indices)
91                  new.parent.indices <- all.parent.indices + to.add
92                  for(j in 1:length(new.parent.indices))
93                  {
94                      new.gas.indices[event.indices[j]] <- sub("^[0-9]+",
95                          new.parent.indices[j],
96                          new.gas.indices[event.indices[j]])
97                  }
98                  new.times[event.indices] <- new.times[event.indices] -
99                      (wraps * sim.length)
100
101                 # include mainshocks in returned events
102                 new.times <- c(times, new.times)
103                 new.magnitudes <- c(magnitudes, new.magnitudes)
104                 new.xs <- c(xs, new.xs)
105                 new.ys <- c(ys, new.ys)
106                 new.gas.indices <- c(gas.indices, new.gas.indices)
107
108                 final.events <- list(times=new.times,
109                     magnitudes=new.magnitudes, xs=new.xs, ys=new.ys,
110                     gas.indices=new.gas.indices)
111                 return(final.events)
112             }
113         }
114
115         # increment the number of mainshocks if this mainshock is within
116         # the spatial boundaries
117         if(length(coords)==2)
118             if(sqrt((parent.x-centre)^2 + (parent.y-centre)^2) <= radius)
119                 num.mainshocks <- num.mainshocks + 1
120         else if(length(coords)==4)
121             if(parent.x < x2 & parent.x >= x1 & parent.y < y2 &
122                 parent.y >= y1)
123                 num.mainshocks <- num.mainshocks + 1
```

```
124
125         # objects to hold its aftershocks
126         aftershock.times <- c()
127         aftershock.magnitudes <- c()
128         aftershock.xs <- c()
129         aftershock.ys <- c()
130         aftershock.gas.indices <- c()
131
132         # calculate the initial rate, at the time of this event
133         Rmax <- conditional.intensity(data.mag=parent.mag,
134             data.time=parent.time, eval.time=ti, params=params)
135         repeat
136         {
137             # increment the time by an appropriate amount, tau
138             if(Rmax > 0) tau <- rexp(1, rate = Rmax)
139             else tau <- Inf
140             ti <- ti + tau
141
142             # check if this sequence is long enough yet
143             if ((ti-parent.time) > sequence.length) break
144
145             # calculate the new rate
146             rate <- conditional.intensity(data.mag=parent.mag,
147                 data.time=parent.time, eval.time=ti, params=params)
148
149             # decide whether to create an aftershock at this time
150             # (thinning method)
151             if (runif(1, 0, 1) <= rate/Rmax)
152             {
153                 # select a magnitude
154                 new.mag <- rexp(1, bvalue * log(10))
155
156                 # select a distance
157                 dd <- params[6]
158                 qq <- params[7]
159                 u <- runif(1, 0, 1)
160                 root <- sqrt((1 - u)^(-1/(qq-1)) - 1)
161                 distance <- dd * root
162
163                 # select an orientation and work out the event coordinates
164                 angle <- runif(1, 0, 2*pi)
165                 new.x <- parent.x + distance*sin(angle)
166                 new.y <- parent.y + distance*cos(angle)
167
168                 # make the index the same as the parent's but with ".i"
```

```
169                 # appended, where i is the aftershock number within this
170                 # sequence
171                 new.index <- paste(parent.index, ".",
172                     length(aftershock.times), sep="")
173
174                 # add the aftershock to the data objects
175                 aftershock.times <- c(aftershock.times, ti)
176                 aftershock.magnitudes <- c(aftershock.magnitudes, new.mag)
177                 aftershock.xs <- c(aftershock.xs, new.x)
178                 aftershock.ys <- c(aftershock.ys, new.y)
179                 aftershock.gas.indices <-
180                     c(aftershock.gas.indices, new.index)
181             }
182
183             # re-set the initial rate for the next iteration
184             Rmax <- rate
185         }
186
187         if(length(aftershock.times) > 0)
188         {
189             # create aftershock sequences for each of these aftershocks
190             aftershocks <-
191                 create.aftershocks.spatial(times=aftershock.times,
192                 magnitudes=aftershock.magnitudes, xs=aftershock.xs,
193                 ys=aftershock.ys, gas.indices=aftershock.gas.indices,
194                 params=params, coords=coords, bvalue=bvalue,
195                 sequence.length=sequence.length, start.time=start.time,
196                 max.events=max.events)
197             remove(aftershock.times)
198             remove(aftershock.magnitudes)
199             remove(aftershock.xs)
200             remove(aftershock.ys)
201             remove(aftershock.gas.indices)
202
203             # add the created aftershocks to our collection of events
204             new.times <- c(new.times, aftershocks$time)
205             new.magnitudes <- c(new.magnitudes, aftershocks$magnitude)
206             new.xs <- c(new.xs, aftershocks$xs)
207             new.ys <- c(new.ys, aftershocks$ys)
208             new.gas.indices <- c(new.gas.indices, aftershocks$gas.indices)
209         }
210     }
211
212     # include parent events in returned events
213     new.times <- c(times, new.times)
```

```
214       new.magnitudes <- c(magnitudes, new.magnitudes)
215       new.xs <- c(xs, new.xs)
216       new.ys <- c(ys, new.ys)
217       new.gas.indices <- c(gas.indices, new.gas.indices)
218
219       events <- list(times=new.times, magnitudes=new.magnitudes, xs=new.xs,
220           ys=new.ys, gas.indices=new.gas.indices)
221       return(events)
222   }
```

A.10 Analysing convergence of the mean inter-event time

The code for analysing the convergence of the mean inter-event time, as used in chapter 3, is included below. It was originally written by Mark Naylor and I used it (modified slightly) to analyse synthetic ETAS catalogues.

The input to the algorithm is a vector of inter-event times, `iets`. The procedure is to make samples from this vector and calculate the mean of each sample. The first part up to line 15 works out the number of samplings that can be made: each sampling will use a sample size of twice the previous sampling, starting at a sample size of 1, and with the constraint that we must be able to obtain at least 8 samples at each sampling. Lines 18–20 then discard any values that are unused at the largest sample size, so that we sample from the same set of values at every sample size.

Lines 26–43 perform the sequential sampling. For each sampling, an inner loop at lines 31–41 selects the inter-event times for each sample, calculates its mean, squares the deviation of this mean from the overall mean (for all inter-event times), and adds up these deviations as the sampling is done. After that, at line 42 the sum of deviations is divided by the number of samples and the square root is taken, to give a standard deviation.

Having done the samplings, the code then normalises the standard deviations by the first value at line 44, and then calculates the gradients in log space of the convergence curve at each sample size.

Lines 52–70 repeat the sampling but using random samples from the inter-event times rather than sequential samples. Line 59 is where this random sampling is done using the R function `sample`.

Finally the plotting is done in lines 75 to the end.

```
1   minSamples <- 8 # the minimum acceptable number of samples
2   numberOfIETs <- length(iets)
3
4   # find out how many samplings we can make given numberOfIETs
5   sampleLengths <- c()
6   sampleLength <- 1
7   numberOfSamplings <- 0
8   numberOfSamples <- numberOfIETs
9   while(numberOfSamples >= minSamples)
10  {
```

```
    numberOfSamplings <- numberOfSamplings+1
    sampleLengths <- c(sampleLengths, sampleLength)
    sampleLength <- 2*sampleLengths[length(sampleLengths)]
    numberOfSamples <- floor(numberOfIETs/sampleLength)
}

# hence how much data we should actually use
numberOfIETs <- sampleLengths[numberOfSamplings] *
    floor(numberOfIETs/sampleLengths[numberOfSamplings])
iets <- iets[1:numberOfIETs]

distributionMean <- mean(iets)


########## Sequential processing
std.dev <- vector(mode="numeric", length=numberOfSamplings)
for(k in 1:numberOfSamplings) ## for each sample size
{
    numberOfSamples <- floor(numberOfIETs/sampleLengths[k])
    sumOfDeviationsFromMean <- 0
    for( m in 1:numberOfSamples ) ## for each sample
    {
        sample.iet.indices <-
            ((m-1)*sampleLengths[k]+1):(m*sampleLengths[k])
        sample.iets <- iets[sample.iet.indices]

        # update sum of deviations from mean
        currentMean <- mean(sample.iets)
        sumOfDeviationsFromMean <-
            sumOfDeviationsFromMean + (currentMean-distributionMean)^2
    }
    std.dev[k] <- sqrt(sumOfDeviationsFromMean/numberOfSamples)
}
sd <- std.dev/std.dev[1]
gradient <-
    (log(sd[2:numberOfSamplings]) - log(sd[1:(numberOfSamplings-1)]))/
    (log(sampleLengths[2:numberOfSamplings]) -
    log(sampleLengths[1:(numberOfSamplings-1)]))


########## Random processing
random.std.dev <- vector(mode="numeric", length=numberOfSamplings)
for(k in 1:numberOfSamplings)
{
    numberOfSamples <- floor(numberOfIETs/sampleLengths[k])
```

```
56      sumOfDeviationsFromMean <- 0
57      for( m in 1:numberOfSamples )
58      {
59          sample.iets <- sample(iets, sampleLengths[k], replace=FALSE)
60          currentMean <- mean(sample.iets)
61          sumOfDeviationsFromMean <-
62              sumOfDeviationsFromMean + (currentMean-distributionMean)^2
63      }
64      random.std.dev[k] <- sqrt(sumOfDeviationsFromMean/numberOfSamples)
65  }
66  rsd <- random.std.dev/random.std.dev[1]
67  random.gradient <-
68      (log(rsd[2:numberOfSamplings]) - log(rsd[1:(numberOfSamplings-1)]))/
69      (log(sampleLengths[2:numberOfSamplings]) -
70      log(sampleLengths[1:(numberOfSamplings-1)]))
71
72
73  ########## Convergence Plots
74  # sequential
75  plot(sampleLengths, sd, type="o", pch=3, lty=3, col="blue", log="xy",
76      ylim=c(0.02,2), xlab="Sample length, N",
77      ylab="Standard deviation from distribution mean")
78
79  # random
80  points(sampleLengths, rsd, type="o", col="black", pch=3)
81
82  # 1 over root n line
83  y1 <- 1/sampleLengths^0.5
84  points(sampleLengths, y1, type="l")
85
86
87  ########## Gradient Plots
88  par(mfrow=c(1,2))
89
90  # sequential
91  plot(gradient, type="o", ylim=c(-1,0), ylab="Local gradient", col="blue")
92
93  # random
94  points(random.gradient, col="black", type="o")
95
96  # 1 over root n line
97  abline(h=-0.5)
```

A.11 Plotting a normalised histogram with Saichev and Sornette's analytic function

In chapter 5, I compared the straightforward histograms for ETAS simulations with normalised versions, also rescaled by the mean event rate, onto which I overplotted the analytic distribution of Saichev and Sornette (2007). The following code shows how this was done. The values are not normalised by the total count, only by the bin widths, and so Saichev and Sornette's function multiplies the result by the total count to match this. The vector `params` passed in on line 29 contains the ETAS parameters used.

```
1  Saichev.Sornette <- function(x, params, N)
2  {
3      n <- calc.n(params, 1)
4      lambda <- params[1]/(1-n)
5      epsilon <- lambda*params[4]
6      theta <- params[5] - 1
7      phi <- exp(-(1-n)*x - (n*epsilon^theta)*(x^(1-theta))/(1-theta))
8      f <- (1 - n + n * epsilon^theta * x^(-theta))^2 +
9          n * epsilon^theta * theta * x^(-1-theta)
10     return(N*f*phi)
11 }
12
13 times <- sim$time
14 times1 <- as.numeric(times[1:length(times)-1])
15 times2 <- as.numeric(times[2:length(times)])
16 iets <- times2-times1
17
18 mean.rate <- 1/mean(iets)
19 histogram <- create.iet.hist(iets, normalise=TRUE)
20 x <- histogram$x*mean.rate
21 y <- histogram$y/mean.rate
22
23 plot(x, y, log="xy", type="o", lty=1,
24     xlab=expression(paste("Rescaled inter-event time, ",tau,"*<r>")),
25     ylab="Normalized count/<r>")
26
27 SS <- function(x)
28 {
29     return(Saichev.Sornette(x, params, length(times)))
30 }
31 curve(SS, from=min(x), to=max(y), add=TRUE, col="blue", lwd=2)
32
33 legend(x="bottomleft", legend=c("Simulation","Analytic"), lty=c(1,1),
34     lwd=c(1,2), pch=c(1,NA), col=c("black","blue"))
35 abline(v=cc*mean.rate, lty=2)
```

A.12 Inter-event time histograms with error bars and fitted exponential

In chapter 5, I plotted histograms of inter-event times for spatial subsets of a global (PDE) catalogue of real events. I identified the peak count as being the peak of an underlying exponential contribution from uncorrelated events, and plotted this exponential on top of the histogram to demonstrate that there is a further contribution from correlated event pairs which grows larger as the spatial area is decreased. To further verify this, I added error bars to the histogram counts.

First, the inter-event times for the PDE subset are obtained and histogrammed as shown in the code below. In this example, a circle of radius 900km is selected. The PDE catalogue and the `subset.circle`, `julian` and `as.catalogue` functions are part of SSLib.

```
1   library(ssPDE)
2   data(PDE)
3   PDEsub <- subset.circle(PDE, centrelat=0, centrelong=130, minradius=0,
4       maxradius=900, minmag=5, maxday=julian(1,1,2005))
5   as.catalogue(PDEsub, "PDEsubcat")
6   times <- PDEsubcat$time
7   magnitudes <- PDEsubcat$magnitude
8   lat <- PDEsubcat$latitude
9   times <- times[!is.na(lat)]
10  magnitudes <- magnitudes[!is.na(lat)]
11  iets <- calc.iet(times)
12  iets <- iets[iets>0]
13  breaks <- seq(from=min(log10(iets)), to=max(log10(iets)), length.out=40)
14  histogram <- create.iet.hist(iets, breaks=breaks, normalise=FALSE)
```

The exponential distribution $f(x) = \lambda e^{-\lambda x}$ in a non-normalised histogram has a peak. The x-value at the peak is $\frac{1}{\lambda}$. To translate the distribution function into its non-normalised equivalent, we note that the rate of change of the bin sizes with respect to x is (Bonnet et al., 2001):

$$\frac{d(\ln x)}{dx} = \frac{1}{x} \tag{A.1}$$

so that the non-normalised frequency distribution becomes:

$$N(x) = f(x)\Delta x = f(x)x\Delta(\ln x) \tag{A.2}$$

where $\Delta(\ln x)$ are the bin widths in log space. (I use logarithm to base 10, so the bin width expression should really be $\ln 10\Delta(\log x)$, but the same result is obtained and it is simpler to derive in natural logs.)

Substituting $x = \frac{1}{\lambda}$ gives us the peak y value in terms of the bin widths:

$$\begin{aligned} y_{peak} &= \lambda e^{-\lambda x_{peak}} x_{peak} \Delta(\ln x) \\ &= e^{-1}\Delta(\ln x) \end{aligned} \tag{A.3}$$

Rearranging then gives the bin widths:

$$\Delta(\ln x) = \frac{y_{peak}}{e^{-1}} \tag{A.4}$$

So the frequency distribution function in terms of the peak y value is:

$$N(x) = \frac{f(x)xy_{peak}}{e^{-1}} \tag{A.5}$$

Identifying the peak of the inter-event time histogram as being coincident with the peak of the underlying exponential component for uncorrelated intervals, we can then fit the exponential by reading off the peak value as shown in the following code excerpt.

```
1  index <- which(histogram$y==max(histogram$y)) + 1
2  peak.x <- histogram$x[index]
3  peak.y <- histogram$y[index]
4  fitted.exponential <- function(x, lambda, peak.y)
5  {
6      return(lambda * exp(-lambda*x) * x * peak.y / exp(-1))
7  }
```

The error bars are plotted to show the 95% confidence limits of a binomial error distribution, which is a good approximation for histogram counting errors (Greenhough and Main, 2008). According to binomial theorem the probability of n successes out of N trials is:

$$\binom{N}{n} p^n (1-p)^{N-n} \tag{A.6}$$

where p is the probability of success. In terms of a histogram bin, n is the count within the bin and N is the total count. For 95% confidence limits, we want the values of p for which the probability of getting a count of at least n is 2.5% and 97.5%, respectively. The expression for this probability is:

$$\sum_{k=n}^{N} \binom{N}{k} p^k (1-p)^{N-k} \tag{A.7}$$

This happens to be equal to the cumulative beta distribution, $I_p(n, N-n+1)$ (Vermeesch, 2005). The beta quantile function (`qbeta` in R) is the inverse of this: it returns the value of p when given a value of the cumulative distribution and the two parameters.

The code to compute these p values is as follows:

```
1  n <- histogram$y
2  N <- sum(histogram$y)
3  p.plus <- c()
4  p.minus <- c()
5  for(j in 1:length(n))
6  {
7      p.plus <- c(p.plus, qbeta(0.975, n[j]+0.5, N-n[j]+0.5))
8      p.minus <- c(p.minus, qbeta(0.025, n[j]+0.5, N-n[j]+0.5))
9  }
```

The actual confidence intervals are then obtained by multiplying these by the total count N:

```
y.plus <- p.plus * N
y.minus <- p.minus * N
```

Finally the histogram with error bars and the fitted exponential may be plotted as follows:

```
errbar(x=histogram$x, y=histogram$y, yplus=y.plus, yminus=y.minus,
    log="xy", type="o", pch=1, lty=1,
    xlab=expression(paste("Inter-event time, ",tau," (days)")),
    ylab="Count")
curve(fitted.exponential(x, lambda=1/peak.x, peak.y=peak.y),
    from=min(iets), to=max(iets), col="red", add=TRUE)
```

A.13 Dividing a catalogue into spatial cells

In chapter 5 I divided the whole globe into spatial cells of equal size and concatenated the inter-event times from each cell, to use all events at each cell size and reduce counting errors. The code for this is below. The vector `cells` containing two values specifies the number of cells in the horizontal direction and the number in the vertical direction, respectively. In each case the number in the vertical direction was 2, so that the equator is the only line of constant latitude used to divide cells. Any other lines of latitude would result in unequally-sized cells.

The function `subset.rect` here also belongs to SSLib.

```
library(ssPDE)
data(PDE)
PDEsub <- subset.rect(PDE, minmag=5, maxday=julian(1,1,2005))
as.catalogue(PDEsub, "PDEsubcat")
times <- PDEsubcat$time
long <- PDEsubcat$longitude
lat <- PDEsubcat$latitude
times <- times[!is.na(lat)]
long <- long[!is.na(lat)]
lat <- lat[!is.na(lat)]

cells <- c(4,2)
width <- 360/cells[1]
height <- 360/cells[2]

iets <- c()
for(k in 1:cells[1])
{
    xmin <- 0 + (k-1)*width
    xmax <- 0 + k*width

    for(m in 1:cells[2])
```

```
23      {
24          ymin <- -180 + (m-1)*height
25          ymax <- -180 + m*height
26          this.times <-
27              times[long >= xmin & long <= xmax & lat >= ymin & lat <= ymax]
28          times1 <- as.numeric(this.times[1:length(this.times)-1])
29          times2 <- as.numeric(this.times[2:length(this.times)])
30          this.iets <- times2-times1
31          iets <- c(iets, this.iets)
32      }
33  }
34  iets <- iets[iets>0]
35  histogram <- create.iet.hist(iets, normalise=FALSE)
36  plot(histogram1$x, histogram1$y, log="xy", type="o",
37      xlab=expression(paste("Inter-event time, ",tau," (days)")),
38      ylab="Count")
```

A.14 Nearest-neighbour plots

My R functions for working out the space-time nearest neighbour of each event in a catalogue, for chapter 5, are included here. The following function `get.nn.indices` takes the times and spatial coordinates of the events as arguments, and then for each event, calculates the temporal and spatial intervals between it and each future event. It then multiplies these two intervals together, and records the index of the event for which this distance is the shortest—this is the event's nearest neighbour.

```
1   get.nn.indices <- function(times, xs, ys)
2   {
3       nn.indices <- c()
4       for(i in 1:(length(times)-1))
5       {
6           time.i <- times[i]
7           time.j <- times[(i+1):length(times)]
8           tau <- time.j - time.i
9
10          x.i <- xs[i]
11          x.j <- xs[(i+1):length(xs)]
12          x <- x.j - x.i
13
14          y.i <- ys[i]
15          y.j <- ys[(i+1):length(ys)]
16          y <- y.j - y.i
17
18          r <- sqrt(x^2 + y^2)
19
```

```
        product <- tau * r
        nn.index <- which(product == min(product)) + i
        nn.indices <- c(nn.indices, nn.index)
    }
    return(nn.indices)
}
```

The following functions `get.nn.tau` and `get.nn.r` take the collection of nearest-neighbour event indices and compute the list of temporal and spatial intervals to the nearest neighbours, respectively, for the catalogue.

```
get.nn.tau <- function(nn.indices, times)
{
    nn.tau <- vector(mode="numeric",length=(length(times)-1))
    nn.tau <- times[nn.indices] - times

    return(nn.tau)
}


get.nn.r <- function(nn.indices, xs, ys)
{
    nn.x <- vector(mode="numeric",length=(length(xs)-1))
    nn.x <- xs[nn.indices] - xs

    nn.y <- vector(mode="numeric",length=(length(xs)-1))
    nn.y <- ys[nn.indices] - ys

    nn.r <- sqrt(nn.x^2 + nn.y^2)

    return(nn.r)
}
```

The following code excerpt shows the usage of these functions for a spatial ETAS synthetic catalogue, `sim`. It plots the nearest neighbours as points in the time-space surface (using only the first 1000 events for the sake of clarity in the plot.)

```
times <- sim$time
xs <- sim$x
ys <- sim$y

nn.indices <- get.nn.indices(times, xs, ys)
nn.tau <- get.nn.tau(nn.indices, times)
nn.r <- get.nn.r(nn.indices, xs, ys)

nn.tau.short <- nn.tau[1:999]
nn.r.short <- nn.r[1:999]
```

```
11
12  plot(nn.tau.short, nn.r.short, log="xy", type="p",
13      xlab=expression(paste(tau," (days)")), ylab="r (km)")
```

A.15 ETAS simulations with time-varying parameters

In the Discussion (chapter 7), section 7.4, I present results from ETAS simulations with a period of raised background rate or raised branching ratio. The simulation algorithm in sections A.2 and A.3 was modified to do this. In both cases the usual procedure of wrapping aftershock sequences back to the start (see Methods section 2.3) is inappropriate because the simulation is not homogeneous in time. So instead of this, an arbitrary run-in period of 100000 days was added to the start of the simulation and then removed at the end.

For the raised background rate, the start and end times for the raised background period are passed in to the simulation function as a vector argument `mu.change.times`; another (`mu.vals`) specifies the values of μ to be used. During the loop for creating the background events (lines 39–57 in the function in section A.2), the following code ensures the correct value of μ is used at each point in time:

```
1  done <- FALSE
2  for(i in 1:length(mu.vals))
3  {
4      if(!done & ti < mu.change.times[(i+1)])
5      {
6          mu <- mu.vals[i]
7          done <- TRUE
8      }
9  }
```

For the raised branching ratio, similar arguments provide the start and end times for the period and the corresponding values of the parameter A, which are then passed into the `create.aftershocks` function and used to set A in the same way as for μ above. This is done during each sequence creation (lines 86–122 in the function in section A.3).

A.16 Inferred cumulative background probability

In the Discussion (chapter 7), section 7.4, I show plots of the inferred cumulative background from ETAS simulations based on a technique from Zhuang et al. (2005) related to stochastic declustering. The background is inferred using the following piece of code. Given `times` and `magnitudes` of events, ETAS parameters (`params`), and the ETAS `conditional.intensity` function, it records both the probability of an event (`event.prob`) and the background probability (`bg.prob`) at the time of each event, and at the end, divides the latter values by the former values and takes the cumulative sum:

```
1  bg.prob <- c()
2  event.prob <- c()
3  for(i in 2:length(times))
```

```
4   {
5       bg.prob <- c(bg.prob, params[1])
6       event.prob <- c(event.prob, conditional.intensity(magnitudes[1:(i-1)],
7           times[1:(i-1)], times[i], params))
8   }
9   prob.event.is.bg <- bg.prob/all.prob
10  cum.bg <- cumsum(prob.event.is.bg)
```

Bibliography

S. G. Abaimov, D. L. Turcotte, R. Shcherbakov, and J. B. Rundle. Recurrence and interoccurrence behavior of self-organized complex phenomena. *Nonlinear Processes in Geophysics*, 14 (4):455–464, 2007.

T. Akimoto, T. Hasumi, and Y. Aizawa. Characterization of intermittency in renewal processes: Application to earthquakes. *Phys. Rev. E*, 81(3):031133, 2010.

F. H. Al-Kindy and I. G. Main. Testing self-organized criticality in the crust using entropy: A regionalized study of the CMT global earthquake catalogue. *Journal of Geophysical Research-Solid Earth*, 108(B11), 2003.

E. G. Altmann and H. Kantz. Recurrence time analysis, long-term correlations, and extreme events. *Physical Review E*, 71(5), 2005.

M. Baiesi and M. Paczuski. Complex networks of earthquakes and aftershocks. *Nonlinear Processes in Geophysics*, 12(1):1–11, 2005.

P. Bak. *How Nature Works*. Springer, 1996.

P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality—an explanation of 1/f noise. *Physical Review Letters*, 59(4):381–384, 1987.

P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality. *Physical Review A*, 38(1): 364–374, 1988.

P. Bak, K. Christensen, L. Danon, and T. Scanlon. Unified scaling law for earthquakes. *Physical Review Letters*, 88(17), 2002.

E. Bonnet, O. Bour, N. E. Odling, P. Davy, I. Main, P. Cowie, and B. Berkowitz. Scaling of fracture systems in geological media. *Reviews of Geophysics*, 39(3):347–383, 2001.

M. Bottiglieri, C. Godano, and L. D'Auria. Distribution of volcanic earthquake recurrence intervals. *Journal of Geophysical Research-Solid Earth*, 114, 2009a.

M. Bottiglieri, E. Lippiello, C. Godano, and L. de Arcangelis. Identification and spatiotemporal organization of aftershocks. *Journal of Geophysical Research-Solid Earth*, 114(B13):B03303, 2009b.

M. Bottiglieri, L. de Arcangelis, C. Godano, and E. Lippiello. Multiple-time scaling and universal behavior of the earthquake interevent time distribution. *Physical Review Letters*, 104 (15), 2010.

J.-P. Bouchaud and M. Potters. *Theory of Financial Risks: From Statistical Physics to Risk Management*. Cambridge University Press, 2001.

K. P. Burnham and D. R. Anderson. Information and likelihood theory: A basis for model selection and inference. In *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, 2nd edition, 2002.

V. Carbone, L. Sorriso-Valvo, P. Harabaglia, and I. Guerra. Unified scaling law for waiting times between seismic events. *Europhysics Letters*, 71(6):1036–1042, 2005.

J. N. Carter, P. J. Ballester, Z. Tavassoli, and P. R. King. Our calibrated model has no predictive value: An example from the petroleum industry, 2004.

J. N. Carter, P. J. Ballester, Z. Tavassoli, and P. R. King. Our calibrated model has poor predictive value: An example from the petroleum industry. *Reliability Engineering & System Safety*, 91(10-11):1373–1381, 2006.

S. Castellaro and F. Mulargia. What criticality in cellular automata models of earthquakes? *Geophysical Journal International*, 150(2):483–493, 2002.

S. F. M. Chastin and I. G. Main. Statistical analysis of daily seismic event rate as a precursor to volcanic eruptions. *Geophysical Research Letters*, 30(13), 2003.

R. Console, M. Murru, and A. M. Lombardi. Refining earthquake clustering models. *Journal of Geophysical Research-Solid Earth*, 108(B10), 2003.

R. Console, M. Murru, F. Catalli, and G. Falcone. Real time forecasts through an earthquake clustering model constrained by the rate-and-state constitutive law: Comparison with a purely stochastic ETAS model. *Seismological Research Letters*, 78(1):49–56, 2007.

A. Corral. Local distributions and rate fluctuations in a unified scaling law for earthquakes. *Physical Review E*, 68(3), 2003.

A. Corral. Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. *Physical Review Letters*, 92(10), 2004.

A. Corral. Mixing of rescaled data and Bayesian inference for earthquake recurrence times. *Nonlinear Processes in Geophysics*, 12(1):89–100, 2005.

A. Corral. Universal earthquake-occurrence jumps, correlations with time, and anomalous diffusion. *Physical Review Letters*, 97(17), 2006a.

A. Corral. Statistical features of earthquake temporal occurrence. In P. Bhattacharyya and B. K. Chakrabarti, editors, *Modelling Critical and Catastrophic Phenomena in Geoscience: A Statistical Physics Approach*, volume 705 of *Lecture Notes in Physics*, pages 191–221, 2006b.

A. Corral. Structure of earthquake occurrence in space, time and magnitude. *Terra Nova*, 19: 337–343, 2007.

A. Corral. Point-occurrence self-similarity in crackling-noise systems and in other complex systems. *Journal of Statistical Mechanics-Theory and Experiment*, JAN 2009.

J. Davidsen and C. Goltz. Are seismic waiting time distributions universal? *Geophysical Research Letters*, 31(21), 2004.

J. Davidsen and M. Paczuski. Analysis of the spatial distribution between successive earthquakes. *Physical Review Letters*, 94(4), 2005.

J. Davidsen, S. Stanchits, and G. Dresen. Scaling and universality in rock fracture. *Physical Review Letters*, 98(12), 2007.

L. de Arcangelis, E. Lippiello, C. Godano, and M. Nicodemi. Statistical properties and universality in earthquake and solar flare occurrence. *European Physical Journal B*, 64(3-4): 551–555, 2008.

P. T. Delaney and R. P. Denlinger. Stabilization of volcanic flanks by dike intrusion: an example from Kilauea. *Bulletin of Volcanology*, 61(6):356–362, 1999.

P. T. Delaney, R. P. Denlinger, M. Lisowski, A. Miklius, P. G. Okubo, A. T. Okamura, and M. K. Sako. Volcanic spreading at Kilauea, 1976-1996. *Journal of Geophysical Research-Solid Earth*, 103(B8):18003–18023, 1998.

C. DeMets. Earthquake slip vectors and estimates of present-day plate motions. *Journal of Geophysical Research-Solid Earth*, 98(B4):6703–6714, 1993.

B. Enescu, K. Ito, and Z. R. Struzik. Wavelet-based multiscale resolution analysis of real and simulated time-series of earthquakes. *Geophysical Journal International*, 164(1):63–74, 2006.

M. Erisoglu, N. Calis, T. Servi, U. Erisoglu, and M. Topaksu. The mixture distribution models for interoccurence times of earthquakes. *Russian Geology and Geophysics*, 52(7):737–744, 2011.

L. Faenza, W. Marzocchi, A. M. Lombardi, and R. Console. Some insights into the time clustering of large earthquakes in Italy. *Annals of Geophysics*, 47(5):1635–1640, 2004.

L. Faenza, S. Hainzl, F. Scherbaum, and C. Beauval. Statistical analysis of time-dependent earthquake occurrence and its impact on hazard in the low seismicity region Lower Rhine Embayment. *Geophysical Journal International*, 171(2):797–806, 2007.

H. J. S. Feder and J. Feder. Self-organized criticality in a stick-slip process. *Physical Review Letters*, 66(20):2669–2672, 1991.

K. R. Felzer and E. E. Brodsky. Decay of aftershock density with distance indicates triggering by dynamic stress. *Nature*, 441(7094):735–738, 2006.

E. Garavaglia and R. Pavani. About earthquake forecasting by Markov renewal processes. *Methodology and Computing in Applied Probability*, 13(1):155–169, 2011.

V. V. Gnedenko and A. N. Kolmogorov. *Limit Distributions of Sums of Independent Random Variables*. Addison-Wesley, Boston, Mass., 1968.

J. Greenhough and I. G. Main. A Poisson model for earthquake frequency uncertainties in seismic hazard analysis. *Geophysical Research Letters*, 35(19), 2008.

J. Greenhough, A. F. Bell, and I. G. Main. Comment on "Relationship between accelerating seismicity and quiescence, two precursors to large earthquakes" by Arnaud Mignan and Rita Di Giovambattista. *Geophysical Research Letters*, 36, 2009.

P. Grossi and H. Kunreuther. *Catastrophe Modeling: A New Approach to Managing Risk*. Springer, New York, 2005.

Z. Q. Guo and Y. Ogata. Statistical relations between the parameters of aftershocks in time, space, and magnitude. *Journal of Geophysical Research-Solid Earth*, 102(B2):2857–2873, 1997.

S. Hainzl and Y. Ogata. Detecting fluid signals in seismicity data through statistical earthquake modeling. *Journal of Geophysical Research-Solid Earth*, 110(B5), 2005.

S. Hainzl, G. Zoller, and J. Kurths. Similar power laws for foreshock and aftershock sequences in a spring-block model for earthquakes. *Journal of Geophysical Research-Solid Earth*, 104 (B4):7243–7253, 1999.

S. Hainzl, G. Zoller, and F. Scherbaum. Earthquake clusters resulting from delayed rupture propagation in finite fault segments. *Journal of Geophysical Research-Solid Earth*, 108(B1), 2003.

S. Hainzl, F. Scherbaum, and C. Beauval. Estimating background activity based on interevent-time distribution. *Bulletin of the Seismological Society of America*, 96(1):313–320, 2006.

S. Hainzl, A. Christophersen, and B. Enescu. Impact of earthquake rupture extensions on parameter estimations of point-process models. *Bulletin of the Seismological Society of America*, 98(4):2066–2072, 2008.

J. L. Hardebeck, K. R. Felzer, and A. J. Michael. Improved tests reveal that the accelerating moment release hypothesis is statistically insignificant. *Journal of Geophysical Research-Solid Earth*, 113(B8), 2008.

D. Harte. Statistical Seismology Library (SSLib), 2007a.

D. Harte. *Users Guide for the Statistical Seismology Library*. Statistics Research Associates, P.O. Box 12 649 Thorndon, Wellington, New Zealand, May 2007b.

D. Harte and D. Vere-Jones. The entropy score and its uses in earthquake forecasting. *Pure and Applied Geophysics*, 162(6-7):1229–1253, 2005.

T. Hasumi. Interoccurrence time statistics in the two-dimensional Burridge-Knopoff earthquake model. *Physical Review E*, 76(2), 2007.

T. Hasumi, C.-C. Chen, T. Akimoto, and Y. Aizawa. The Weibull-log Weibull transition of interoccurrence time for synthetic and natural earthquakes. *Tectonophysics*, 485(1-4):9–16, 2010.

A. Helmstetter and D. Sornette. Diffusion of epicenters of earthquake aftershocks, Omori's law, and generalized continuous-time random walk models. *Physical Review E*, 66(6), 2002a.

A. Helmstetter and D. Sornette. Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research-Solid Earth*, 107(B10), 2002b.

A. Helmstetter and D. Sornette. Predictability in the epidemic-type aftershock sequence model of interacting triggered seismicity. *Journal of Geophysical Research-Solid Earth*, 108(B10), 2003a.

A. Helmstetter and D. Sornette. Båth's law derived from the Gutenberg-Richter law and from aftershock properties. *Geophysical Research Letters*, 30(20), 2003b.

A. Helmstetter and D. Sornette. Foreshocks explained by cascades of triggered seismicity. *Journal of Geophysical Research-Solid Earth*, 108(B10), 2003c.

A. Helmstetter, S. Hergarten, and D. Sornette. Properties of foreshocks and aftershocks of the nonconservative self-organized critical Olami-Feder-Christensen model. *Physical Review E*, 70(4), 2004.

A. S. Helmstetter and D. Sornette. Importance of direct and indirect triggered seismicity in the ETAS model of seismicity. *Geophysical Research Letters*, 30(11), 2003d.

S. Hergarten and H. J. Neugebauer. Foreshocks and aftershocks in the Olami-Feder-Christensen model. *Physical Review Letters*, 88(23), 2002.

D. P. Heyman and M. J. Sobel. *Stochastic Models in Operations Research: Stochastic Processes and Operating Characteristics*, volume 1. Courier Dover Publications, 2004.

J. R. Holliday, D. L. Turcotte, and J. B. Rundle. A review of earthquake statistics: Fault and seismicity-based models, ETAS and BASS. *Pure and Applied Geophysics*, 165(6):1003–1024, 2008a.

J. R. Holliday, D. L. Turcotte, and J. B. Rundle. Self-similar branching of aftershock sequences. *Physica A: Statistical Mechanics and its Applications*, 387(4):933–943, 2008b.

M. Huc and I. G. Main. Anomalous stress diffusion in earthquake triggering: Correlation length, time dependence, and directionality. *Journal of Geophysical Research-Solid Earth*, 108(B7), 2003.

S. C. Jaume and L. R. Sykes. Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes. *Pure and Applied Geophysics*, 155(2-4):279–305, 1999.

K. Jonsdottir, M. Lindman, R. Roberts, B. Lund, and R. Bodvarsson. Modelling fundamental waiting time distributions for earthquake sequences. *Tectonophysics*, 424(3-4):195–208, 2006.

Y. Y. Kagan. Seismic moment distribution. *Geophysical Journal International*, 106(1):123–134, 1991.

Y. Y. Kagan. Observational evidence for earthquakes as a nonlinear dynamic process. *Physica D*, 77(1-3):160–192, 1994.

Y. Y. Kagan. Aftershock zone scaling. *Bulletin of the Seismological Society of America*, 92(2): 641–655, 2002.

Y. Y. Kagan and H. Houston. Relation between mainshock rupture process and Omori's law for aftershock moment release rate. *Geophysical Journal International*, 163(3):1039–1048, 2005.

F. W. Klein, R. Y. Koyanagi, J. S. Nakata, W. R. Tanigawa, R. W. Decker, T. L. Wright, and P. H. Stauffer. The seismicity of kilauea's magma system. In *Volcanism in Hawaii, edited*, pages 1019–1185. Hawaiian Volcano Observatory, 1987.

F. W. Klein, T. Wright, and J. Nakata. Aftershock decay, productivity, and stress rates in Hawaii: Indicators of temperature and stress from magma sources. *Journal of Geophysical Research-Solid Earth*, 111(B7), 2006.

T. R. Krishna Mohan and P. G. Revathi. Earthquake correlations and networks: A comparative study. *Phys. Rev. E*, 83(4):046109, 2011.

P. S. Laplace. *Theorie Analytique des Probabilities*. Courcier, Paris, 1812.

S. Lennartz, V. N. Livina, A. Bunde, and S. Havlin. Long-term memory in earthquakes and the distribution of interoccurrence times. *Europhysics Letters*, 81, 2008.

M. Lindman, K. Jonsdottir, R. Roberts, B. Lund, and R. Bodvarsson. Earthquakes descaled: On waiting time distributions and scaling laws. *Physical Review Letters*, 94(10), 2005.

M. Lindman, B. Lund, R. Roberts, and K. Jonsdottir. Physics of the Omori law: Inferences from interevent time distributions and pore pressure diffusion modeling. *Tectonophysics*, 424 (3-4):209–222, 2006.

E. Lippiello, L. de Arcangelis, and C. Godano. Memory in self-organized criticality. *Europhysics Letters*, 72(4):678–684, 2005.

E. Lippiello, L. de Arcangelis, and C. Godano. Influence of time and space correlations on earthquake magnitude. *Physical Review Letters*, 1(3), 2008.

W. B. Liu and L. Ma. A recent application of the ETAS model and a proposed method for prediction of strong aftershocks. *Pure and Applied Geophysics*, 163(11-12):2513–2528, 2006.

A. M. Lombardi and W. Marzocchi. Evidence of clustering and nonstationarity in the time distribution of large worldwide earthquakes. *Journal of Geophysical Research-Solid Earth*, 112(B2), 2007.

A. M. Lombardi, W. Marzocchi, and J. Selva. Exploring the evolution of a volcanic seismic swarm: The case of the 2000 Izu Islands swarm. *Geophysical Research Letters*, 33(7), 2006.

A. M. Lombardi, M. Cocco, and W. Marzocchi. On the increase of background seismicity rate during the 1997-1998 umbria-marche, central italy, sequence: Apparent variation or fluid-driven triggering? *Bulletin of the Seismological Society of America*, 100(3):1138–1152, 2010.

L. Ma and J. Zhuang. Relative quiescence within the Jiashi swarm in Xinjiang, China: an application of the ETAS point process model. *Journal of Applied Probability*, 38A:213–221, 2001.

I. Main. Statistical physics, seismogenesis, and seismic hazard. *Reviews of Geophysics*, 34(4): 433–462, 1996.

I. Main. Earthquake prediction debate—Concluding Remarks, http://www.nature.com/nature/debates/earthquake/equake_frameset.html, 1999.

I. G. Main, L. Li, J. McCloskey, and M. Naylor. Effect of the Sumatran mega-earthquake on the global magnitude cut-off and event rate. *Nature Geoscience*, 1(3):142, 2008.

R. S. Matsu'ura and I. Karakama. A point-process analysis of the Matsushiro earthquake swarm sequence: The effect of water on earthquake occurrence. *Pure and Applied Geophysics*, 162 (6-7):1319–1345, 2005.

R. S. Mendes, L. C. Malacarne, R. P. B. Santos, H. V. Ribeiro, and S. Picoli, Jr. Earthquake-like patterns of acoustic emission in crumpled plastic sheets. *EPL*, 92(2), 2010.

G Molchan. Interevent time distribution in seismicity: A theoretical approach. *Pure and Applied Geophysics*, 162(6-7):1135–1150, 2005.

G. Molchan and T. Kronrod. Seismic interevent time: A spatial scaling and multifractality. *Pure and Applied Geophysics*, 164(1):75–96, 2007.

J. Nakata. Hawaiian volcano observatory seismic data, January to December 2005., 2006.

C. Narteau, P. Shebalin, and M. Holschneider. Temporal limits of the power law aftershock decay rate. *Journal of Geophysical Research-Solid Earth*, 107(B12), 2002.

M. Naylor and H. D. Sinclair. Punctuated thrust deformation in the context of doubly vergent thrust wedges: Implications for the localization of uplift and exhumation. *Geology*, 35(6): 559–562, 2007.

M. Naylor, I. G. Main, and S. Touati. Quantifying uncertainty in mean earthquake interevent times for a finite sample. *Journal of Geophysical Research-Solid Earth*, 114(B01316), 2009.

A. Nekrasova, V. Kossobokov, A. Peresan, A. Aoudia, and G. F. Panza. A multiscale application of the unified scaling law for earthquakes in the Central Mediterranean Area and Alpine region. *Pure and Applied Geophysics*, 168(1-2, SI):297–327, 2011.

G. Niccolini, A. Schiavi, P. Tarizzo, A. Carpinteri, G. Lacidogna, and A. Manuello. Scaling in temporal occurrence of quasi-rigid-body vibration pulses due to macrofractures. *Physical Review E*, 82(4, Part 2), 2010.

G. Niccolini, A. Carpinteri, G. Lacidogna, and A. Manuello. Acoustic emission monitoring of the Syracuse Athena Temple: Scale invariance in the timing of ruptures. *Physical Review Letters*, 106(10), 2011.

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

Y. Ogata. Detection of precursory relative quiescence before great earthquakes through a statistical model. *Journal of Geophysical Research-Solid Earth*, 97(B13):19845–19871, 1992.

Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

Y. Ogata. Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics*, 155(2-4):471–507, 1999.

Y. Ogata and H. C. Zhuang. Space-time ETAS models and an improved extension. *Tectonophysics*, 413(1-2):13–23, 2006.

Y. Ogata, L. M. Jones, and S. Toda. When and where the aftershock activity was depressed: Contrasting decay patterns of the proximate large earthquakes in southern California. *Journal of Geophysical Research-Solid Earth*, 108(B6), 2003.

I. O. Ojala, I. G. Main, and B. T. Ngwenya. Strain rate and temperature dependence of Omori law scaling constants of AE data: Implications for earthquake foreshock-aftershock sequences. *Geophysical Research Letters*, 31(24), 2004.

Z. Olami, H. J. S. Feder, and K. Christensen. Self-organized criticality in a continuous, non-conservative cellular automaton modeling earthquakes. *Physical Review Letters*, 68(8):1244–1247, 1992.

M. Paczuski and S. Boettcher. Universality in sandpiles, interface depinning, and earthquake models. *Physical Review Letters*, 77(1):111–114, 1996.

S. D. Poisson. *Recherches sur la Probabilite des Jugements en Matiere Criminelle et en Matiere Civile, Precedees des Regles Generales du Calcul des Probabilities*. Bachelier, Paris, 1837.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

S. M. Ross. *Introduction to Probability Models*. Elsevier, New York, 8th edition, 2003.

A. Saichev and D. Sornette. "Universal" distribution of interearthquake times explained. *Physical Review Letters*, 97(7), 2006.

A. Saichev and D. Sornette. Theory of earthquake recurrence times. *Journal of Geophysical Research-Solid Earth*, 112(B4), 2007.

M. Sambridge and K. Mosegaard. Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40(3), 2002.

C. G. Sammis and D. Sornette. Positive feedback, memory, and the predictability of earthquakes. *Proceedings of the National Academy of Sciences of the United States of America*, 99:2501–2508, 2002.

F. P. Schoenberg. Multidimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98(464):789–795, 2003.

C. H. Scholz, L. R. Sykes, and Y. P. Aggarwal. Earthquake prediction: A physical basis. *Science*, 181:803–810, 1973.

J. P. Sethna, K. A. Dahmen, and C. R. Myers. Crackling noise. *Nature*, 410(6825):242–250, 2001.

R. Shcherbakov and D. L. Turcotte. A modified form of Båth's law. *Bulletin of the Seismological Society of America*, 94(5):1968–1975, 2004.

R. Shcherbakov, G. Yakovlev, D. L. Turcotte, and J. B. Rundle. Model for the distribution of aftershock interoccurrence times. *Physical Review Letters*, 95(21), 2005.

R. Shcherbakov, J. Van Aalsburg, J. B. Rundle, and D. L. Turcotte. Correlations in aftershock and seismicity patterns. *Tectonophysics*, 413(1-2):53–62, 2006.

D. Sornette and A. Helmstetter. Occurrence of finite-time singularities in epidemic models of rupture, earthquakes, and starquakes. *Physical Review Letters*, 89(15), 2002.

D. Sornette and M. J. Werner. Constraints on the size of the smallest triggering earthquake from the epidemic-type aftershock sequence model, Båth's law, and observed aftershock sequences. *Journal of Geophysical Research-Solid Earth*, 110(B8), 2005a.

D. Sornette and M. J. Werner. Apparent clustering and apparent background earthquakes biased by undetected seismicity. *Journal of Geophysical Research-Solid Earth*, 110(B9), 2005b.

D. Sornette, S. Utkin, and A. Saichev. Solution of the nonlinear theory and tests of earthquake recurrence times. *Physical Review E*, 77(6), 2008.

F. Tajima and H. Kanamori. Aftershock area expansion and mechanical heterogeneity of fault zone within subduction zones. *Geophysical Research Letters*, 12(6):345–348, 1985.

A. Talbi and F. Yamazaki. A mixed model for earthquake interevent times. *Journal of Seismology*, 14(2):289–307, 2010.

R. I. Tilling and J. J. Dvorak. Anatomy of a basaltic volcano. *Nature*, 363(6425):125–133, 1993.

G. Timar and F. Kun. Crackling noise in three-point bending of heterogeneous materials. *Physical Review E*, 83(4, Part 2), 2011.

S. Touati, M. Naylor, and I. G. Main. Origin and Nonuniversality of the Earthquake Interevent Time Distribution. *Physical Review Letters*, 102(16), 2009.

S. Touati, M. Naylor, I. G. Main, and M. Christie. Masking of earthquake triggering behavior by a high background rate and implications for epidemic-type aftershock sequence inversions. *Journal of Geophysical Research-Solid Earth*, 116, 2011.

P. Traversa and J.-R. Grasso. How is volcano seismicity different from tectonic seismicity? *Bulletin of the Seismological Society of America*, 100(4):1755–1769, 2010.

T. G. Trucano, L. P. Swiler, T. Igusa, W. L. Oberkampf, and M. Pilch. Calibration, validation, and sensitivity analysis: What's what. *Reliability Engineering & System Safety*, 91(10-11): 1331–1357, 2006.

D. L. Turcotte. Earthquake prediction. *Annual Review of Earth and Planetary Sciences*, 19: 263–281, 1991.

D. L. Turcotte. Self-organized criticality. *Reports on Progress in Physics*, 62(10):1377–1429, 1999.

D. L. Turcotte, J. R. Holliday, and J. B. Rundle. BASS, an alternative to ETAS. *Geophysical Research Letters*, 34(12), 2007.

D. L. Turcotte, S. G. Abaimov, I. Dobson, and J. B. Rundle. Implications of an inverse branching aftershock sequence model. *Phys. Rev. E*, 79(1):016101, 2009.

A. Veen and F. P. Schoenberg. Estimation of space-time branching process models in seismology using an EM-type algorithm. *Journal of the American Statistical Association*, 103(482):614–624, 2008.

D. Vere-Jones. A class of self-similar random measure. *Advances in Applied Probability*, 37(4): 908–914, 2005.

D. Vere-Jones. Foundations of statistical seismology. *Pure and Applied Geophysics*, 167(6-7): 645–653, 2010.

D. Vere-Jones, Y. Ben-Zion, and R. Zuniga. Statistical seismology. *Pure and Applied Geophysics*, 162(6-7):1023–1026, 2005.

P. Vermeesch. Statistical uncertainty associated with histograms in the Earth sciences. *Journal of Geophysical Research-Solid Earth*, 110(B2), 2005.

Q. Wang, D. D. Jackson, and J. Zhuang. Are spontaneous earthquakes stationary in California? *Journal of Geophysical Research-Solid Earth*, 115(B08310), 2010.

D. Weatherley. Recurrence interval statistics of cellular automaton seismicity models. *Pure and Applied Geophysics*, 163(9):1933–1947, 2006.

K. G. Wilson. Problems in physics with many scales of length. *Scientific American*, 241(2): 140–, 1979.

S. Wolfram. *A New Kind of Science*. Wolfram Media Inc, 2002.

Y. Wu, C. Zhou, J. Xiao, J. Kurths, and H. J. Schellnhuber. Evidence for a bimodal distribution in human communication. *Proceedings of the National Academy of Sciences*, 2010. doi: 10.1073/pnas.1013140107. URL `http://www.pnas.org/content/early/2010/10/14/1013140107.abstract`.

M. Wyss and D. C. Booth. The IASPEI procedure for the evaluation of earthquake precursors. *Geophysical Journal International*, 131(3):423–424, 1997.

M. Wyss, F. Klein, K. Nagamine, and S. Wiemer. Anomalously high b-values in the South Flank of Kilauea volcano, Hawaii: evidence for the distribution of magma below Kilauea's East rift zone. *Journal of Volcanology and Geothermal Research*, 106(1-2):23–37, 2001.

I. Zaliapin, A. Gabrielov, V. Keilis-Borok, and H. Wong. Clustering analysis of seismicity and aftershock identification. *Physical Review Letters*, 101(1), 2008.

X. Zhao, T. Omi, N. Matsuno, and S. Shinomoto. A non-universal aspect in the temporal occurrence of earthquakes. *New Journal of Physics*, 12(6):063010, 2010. URL `http://stacks.iop.org/1367-2630/12/i=6/a=063010`.

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

J. C. Zhuang. Statistical modelling of seismicity patterns before and after the 1990 Oct 5 Cape Palliser earthquake, New Zealand. *New Zealand Journal of Geology and Geophysics*, 43(3): 447–460, 2000.

J. C. Zhuang, C. P. Chang, Y. Ogata, and Y. I. Chen. A study on the background and clustering seismicity in the Taiwan region by using point process models. *Journal of Geophysical Research-Solid Earth*, 110(B5), 2005.

E. Zivot and J. Wang. *Modelling Financial Time Series With S-Plus*. Springer, New York, 2nd edition, 2006.