Final Technical Report
Bayesian ETAS: Towards Improved Operational Aftershock Forecasting: Collaborative Research with UW and USGS

PI: Peter Guttorp
Professor Emeritus of Statistics
Box 354322
University of Washington
Seattle, WA 98195-4322
Fax: +1-206-685-7419
guttorp@uw.edu

Supported Graduate Student: Max Schneider
PhD Candidate in Statistics
Box 354322
University of Washington
Seattle, WA 98195-4322
maxs15@uw.edu

Award Number: G20AP00061
Term covered by the award: June 15, 2020 - June 15, 2021

Acknowledgement of Support and Disclaimer: This material is based upon work supported by the U.S. Geological Survey under Grant No. G20AP00061. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Geological Survey.

Publications under this Award

Max Schneider and Peter Guttorp. *Bayesian ETAS: Towards Improved Earthquake Rate Models in the Pacific Northwest.* AGU Fall Meeting, December 2020, virtual.

Max Schneider and Peter Guttorp. *Bayesian ETAS: Towards Improved Earthquake Rate Models in the Pacific Northwest.* Seismological Society of America Annual Meeting, April 2021, virtual.

Max Schneider and Peter Guttorp. *Bayesian ETAS: Towards Improved Earthquake Rate Models in the Pacific Northwest.* Joint Statistical Meetings, August 2021, virtual.

Abstract

The Pacific Northwest (PNW) of North America has substantial earthquake risk, both due to the Cascadia megathrust fault but also crustal and intraplate events that occur under the region's population centers. Aftershock forecasts for the PNW require statistical modelling of a catalog of its past earthquakes; however, the PNW's limited catalog contains multiple tectonic regimes, as well as earthquake swarms, which complicate statistical seismicity modelling. The Epidemic-Type Aftershock Sequence model (ETAS) is a top-performing spatiotemporal point process model which parameterizes the rates of earthquakes and aftershocks within a seismic region. Typically, maximum likelihood estimation is used to fit ETAS to an earthquake catalog; however, the ETAS likelihood suffers from flatness near its optima, parameter correlation and numerical instability, making likelihood-based estimates less reliable. We present a Bayesian procedure to estimate ETAS parameters, such that parameter estimates and uncertainty can be reliably quantified, even for small and complex catalogs like the PNW. The procedure is conditional on knowing which earthquakes triggered which aftershocks; this latent structure and the ETAS parameters are estimated stepwise. The procedure uses a Gibbs sampler to conditionally estimate the posterior distributions of each part of the model. We simulate several synthetic catalogs and test the modelling procedure, showing posterior distributions that are well-centered on true values and follow previously reported patterns. We also demonstrate the procedure on a new catalog for the continental PNW. This catalog is merged from three existing catalogs with automated procedures for duplicate detection and identification of earthquake swarms. We compare parameter estimates between catalogs without swarms and classified into the region's tectonic regimes. More detailed information about PNW aftershocks can be estimated using Bayesian ETAS than using simpler seismicity models.

1 Introduction

1.1 Motivation

Since 2018, the USGS has released automated aftershock forecasts following earthquakes with magnitude above 5. These forecasts are based on statistical seismicity models of both global and local catalogs for the region in which the earthquake took place. At present, the USGS relies on the Reasenberg-Jones (RJ) model [Reasenberg and Jones, 1989], which is a simplified model based on several seismological laws. A more realistic representation of seismicity is the Epidemic-Type Aftershock Sequence (ETAS) model [Ogata, 1998], which has performed well in multiple evaluation experiments of earthquake forecasts [Nanjo et al., 2012, Schorlemmer et al., 2018]. The ETAS model is advantageous over the RJ model because it incorporates the spatial dimension of aftershock occurrence and allows for aftershocks to trigger their own sequences, accounting for cascading sequences, as found in nature.

The ETAS model is a point process model that describes the interdepence between earthquakes in a catalog, considering each as a point in some space (either time or space-time), equipped with a magnitude M. The spatiotemporal ETAS model describes the total seismicity rate $\lambda(t, x, y|\mathcal{H}_t)$, at a given time t and location (x, y), based on the catalog of events up to that time $\mathcal{H}_t$ as:

$$\lambda(t, x, y|\mathcal{H}_t) = \mu_{ST} + \sum_{j}^{t_j<t} k(M_j)h(t - t_j)g(x - x_j, y - y_j), \tag{1}$$

where

$$\begin{aligned} k(M_j) &= K \exp(\alpha(M_j - M_0)) \\ h(t - t_j) &= (t - t_j + c)^{-p} \\ g(x - x_j, y - y_j) &= \frac{q-1}{\pi d^{1-q}}((x - x_j)^2 + (y - y_j)^2 + d)^{-q}, \end{aligned}$$

and μ_{ST} is a spatiotemporal background seismicity rate that is assumed constant for the entire region; $k(M_j)$ is the Utsu productivity law, with global productivity parameter K and magnitude scaling parameter α; $h(t - t_j)$ is the (modified) Omori temporal decay function, with offset parameter c and decay parameter p; and $g(x - x_j, y - y_j)$ is a normalized Utsu-Seki spatial decay function, with offset parameter d and decay parameter q.

1.2 Statistical Issues with the ETAS Likelihood

The model described by Equation 1 assumes, among other simplifications, that background seismicity is stationary with respect to time and space. The latter assumption can be relaxed by letting μ vary spatially ($\mu_{ST} = \mu_{ST}(x, y)$), but stationarity in time is fundamental in most statistical studies of time-limited catalogs.

Another central assumption of the ETAS model is that a catalog can be split into sets of background earthquakes and aftershock sequences. Some authors use rules based on windows in space and time to identify aftershock sequences in a catalog and apply the ETAS model individually to each sequence [Zhang et al., 2020, Guo and Ogata, 1997]. However, such deterministic splits of the catalog ignore the fact that there is never certainty about which events belong to a given sequence. Methods exists for a probabilistic assignment of every earthquake into the background set or an aftershock sequence [Zhuang et al., 2002]. These better reflect the inherent uncertainty in the classification of aftershocks in a catalog.

Given a split of the catalog between background and aftershock events, ETAS parameters $\theta = \{\mu, K, \alpha, c, p, d, q\}$ may be estimated. The parameter estimates (and their accompanying uncertanties) are key not just in describing a region's aftershock patterns, but in forecasting aftershocks after large earthquakes, including those released by the USGS [Michael, 2018, Llenos and Michael, 2019, Hardebeck et al., 2019, Michael et al., 2020]. Parameters are typically estimated by maximizing the ETAS log-likelihood:

$$\begin{aligned} l(\theta|\mathcal{H}_T) &= \sum_{i=1}^{n} \log(\lambda(t_i, x_i, y_i|\mathcal{H}_{t_i})) - \int_0^T \int\int_S \lambda(t, x, y|\mathcal{H}_t) dx dy dt \\ &= SUM + INT. \end{aligned}$$

Several problems have been documented with using this likelihood to perform inference. Simply evaluating the likelihood function is non-trivial, particularly the integral term, for which no closed-form expression exists, forcing researchers to use either numerical solutions [Ogata, 1998, Lippiello et al., 2014] or other approximations [Schoenberg, 2013]. Optimizing the likelihood function similarly requires iterative numerical procedures, with final maximum-likelihood estimates (MLEs) being found when the likelihood iteratively converges to its maximum. But the likelihood is known to be flat near the optima for most ETAS parameters [Veen and Schoenberg, 2008]. Furthermore, parameter estimates are correlated between several parameter pairs [Chu et al., 2011, Guo and Ogata, 1997], leading to a likelihood surface that is multimodal [Lombardi, 2015]. All these issues make ETAS MLEs unstable, potentially biased [Seif et al., 2017, Harte, 2013] and often dependent on the initial values used to start the numerical procedure.

Furthermore, quantifying uncertainty around these parameter estimates is usually based on the inverse Hessian of the likelihood function. This is rooted in large-sample statistical reasoning, which time-limited earthquake catalogs rarely satisfy. It also requires further numerical procedures and approximations, which are again problematic even for large catalogs. As such, uncertainty quantification about ETAS parameter estimates based on the traditional likelihood is unstable. It may even be unattainable for small catalogs with complex seismicity [Harte, 2018], as the numerical procedures involved may fail to converge. Despite these limitations, the vast majority of published ETAS estimates and uncertainties rely on this likelihood function (e.g, Zhuang [2011], Zhang et al. [2020]).

These statistical challenges with the ETAS likelihood have been reported for decades and researchers have made different attempts to ameliorate them. In Veen and Schoenberg [2008], the authors utilize the (unknown) *branching structure* of the catalog, $B = \{B_i\}$, for all catalog earthquakes i, where

$$B_i = \begin{cases} 0 \text{ if earthquake } i \text{ was produced by background process} \\ j \text{ if earthquake } i \text{ was triggered by previous earthquake } j. \end{cases}$$

Conditional on values for B_i, they derive a new likelihood function, which we term the *branching likelihood*. ETAS parameters can then be estimated in a iterative procedure: first, fix some B based on initial guesses for the ETAS parameters θ. Then, iteratively maximize the branching likelihood to obtain ETAS parameter estimates $\hat{\theta}$ and then update B based on the new estimates $\hat{\theta}$. Iterate the previous step until estimates $\hat{\theta}$ converge to a final solution. The authors show that MLEs based on this estimation prcoedure are less biased and more robust to choice of initial values than traditional MLEs; however, maximizing even this simpler likelihood still requires a numerical procedure that may not converge for smaller catalogs. Other estimation techniques have proposed using different numerical procedures for the traditional ETAS likelihood [Lombardi, 2015, Kasahara et al., 2016], but they do not address the fundamental challenges inherent in maximizing this function.

1.3 Bayesian Inference for the ETAS Model

Bayesian inference is a statistical paradigm for parameter estimation that eschews maximizing a model's likelihood in order to find parameter estimates. Rather, it considers the parameter itself as a random variable that possesses a probability distribution. Sampling from this *posterior* distribution requires specifying the prior knowledge on this parameter (also in the form of a distribution) and then scaling it by the model likelihood. Bayesian inference is available for spatiotemporal point process models and is often applied to studies of natural events [Guttorp and Thorarinsdottir, 2012]. This framework has gotten some use in previous research with the ETAS model. Several studies have investigated how ETAS aftershock forecasts can be updated using the Bayesian approach during an ongoing sequence [Ebrahimian and Jalayer, 2017, Omi et al., 2015], but these require the initial computation of parameters MLEs.

Recent work has tackled the Bayesian estimation of ETAS parameters from catalogs. Ross [2017] and Kolev and Ross [2019] work with the temporal-only ETAS model and, similar to Veen and Schoenberg [2008], derive a branching likelihood that does not have many of the statistical issues in the traditional likelihood. They use a Monte Carlo Markov Chain routine to sample from the parameter posterior distributions to more rigorously quantify uncertainty in the parameter estimates. Ross [2021] extends this to the spatiotemporal ETAS model.

We expand on this foundation by showing the utility of the Bayesian inference approach for the spatiotemporal ETAS model. We then apply it to a new catalog for the Pacific Northwest region, which has specific issues that require a customized modelling approach.

1.4 Pacific Northwest Seismicity

The Pacific Northwest of North America has high seismic risk due to a diversity of fault systems, some of which lie under population centers. Statistical seismicity modelling for the Pacific Northwest is complicated by the complex tectonics of the region. The Juan de Fuca plate subducts under the North American continental plate and the resulting fault zone is capable of generating massive megathrust events. The subducting slab itself produces deep, intraslab earthquakes which are hypothesized to produce few aftershocks. There are also numerous shallow faults located in the crust of the North American plate that produce crustal earthquakes that are usually (though not always) smaller in magnitude. These earthquakes are also thought to have more productive aftershocks than intraslab earthquakes.

Furthermore, the region is known to produce earthquake swarms, which are spatiotemporal clusters of earthquakes that impedes splitting a catalog into background and triggered seismicity. Swarm events are not triggered by a previous earthquake as in aftershock sequences; rather, they occur because of aseismic sources, such as from underground fluids (volcanic magma or groundwater) or anthropogenic sources (such as wastewater injection). Swarms manifest in a catalog as brief surges in the seismicity rate for a subregion, beyond what is typical for its background seismicity. The many volcanoes of the PNW contribute to the presence of swarms in this catalog [Bostock et al., 2019].

Even though the region's seismicity is complex, the PNW does not have a large set of contemporary instrumentally-measured earthquakes [Malone, 2019]. Due to sizable seismicity at the US/Canada border, any catalog for the PNW region must consider earthquakes recorded by both US and Canadian seismic networks. Combined international PNW catalogs have been quantitatively studied (e.g., Bostock et al. [2019]), though almost no previous studies have been made of aftershock patterns across the PNW. The one recent exception found that intraslab events do indeed appear to have lower productivities than crustal mainshocks [Gomberg and Bodin, 2021]. This study was based on the Reasenberg-Jones model, with sequence-specific parameters fit using MLE. Given the importance of aftershock forecasts to emergency managers and other potential users [Gomberg and Jakobitz, 2013], the spatiotemporal ETAS model may better characterize and thus forecast aftershock patterns in the PNW. But the heterogenuous and limited catalog problematizes any traditional MLE-based ETAS model for the region. Our methodology for Bayesian inference for the ETAS model will thus be instrumental for estimating aftershock parameters together with their uncertainties.

In this technical report, we first explain in detail the Bayesian ETAS methodology and outline our steps for simulating catalogs on which to test it. We then describe a novel international PNW catalog with merged data from three data sources. We provide parameter estimation results that show the efficacy of our methodology under simulated catalogs, and then show results for the PNW catalog.

2 Methods

This section describes the spatiotemporal ETAS model, its branching likelihood and the Bayesian inference procedure we developed to estimate parameters and quantify their uncertainty. It also explains our procedures to simulate catalogs from an ETAS model and to elicit priors from experts.

2.1 ETAS Model and Bayesian Formulation

2.1.1 Spatiotemporal ETAS Likelihood

Consider $l(Y|\theta)$, the traditional log-likelihood function for the spatiotemporal ETAS model given in Ogata [1998]:

$$l(\theta|\mathcal{H}_T) = \sum_{i=1}^{n} \log(\lambda(t_i, x_i, y_i|\mathcal{H}_{t_i})) - \int_0^T \int\int_S \lambda(t, x, y|\mathcal{H}_t) dxdydt \tag{2}$$

As mentioned in Section 1.2, this likelihood is challenging, if not impossible, to optimize due to its multimodality, flatness near parameter optima, and parameter correlation.

The integral term of equation 2 can be rewritten as:

$$\int_0^T \int\int_S \lambda(t, x, y|\mathcal{H}_t) dxdydt = \mu T|S| + \int_0^T \int\int_S \int_{M_0}^{\infty} N(ds, d\xi, d\eta, dM) \cdot \int_0^{T-s} \int\int_{S-(\xi,\eta)} a(t, x, y, M) dtdxdy$$
$$INT = INT_b + INT_{trig},$$

where $ds, d\xi, d\eta, dM$ refers to an infinitesimal interval in time-2dspace-magnitude, the triggering function $a(t, x, y, M) = K \exp(\alpha(M_j - M_0))(t - t_j + c)^{-p} \frac{q-1}{\pi d^{1-q}}((x - x_j)^2 + (y - y_j)^2 + d)^{-q}$, T is the catalog period, S is the two-dimensional catalog spatial zone and $|S|$ is its area. In this model, since we assume a constant background rate in space, $\mu_{ST}(x, y) = \mu \frac{1}{|S|}$.

We thus decompose INT into the background component (b) and triggered component ($trig$) (the latter two integral terms). In INT_{trig}, the first integral term can be interpreted as "how many triggered earthquakes occur in this infinitesimal interval, when integrated over all of time, space and magnitude" and the second integral term as "what is the contribution to the total triggered intensity from these earthquakes". The first integral in INT_{trig} thus translates to a sum over all earthquakes in the catalog leading to

$$INT_{trig} = \sum_{j=1}^{n} \int_0^{T-t_j} \int\int_{S^{(j)}} a(t_j, x_j, y_j, M_j) dtdxdy,$$

with $S^{(j)} = \{(x - x_j, y - y_j) : (x, y) \in S\}$. This can be split into its temporal and spatial components:

$$INT_{trig} = \sum_{j=1}^{n} \int_0^{T-t_j} \frac{K \exp(\alpha(M_j - M_0))}{(t + c)^p} dt \int\int_{S^{(j)}} \frac{q-1}{\pi d^{1-q}} \frac{1}{(x^2 + y^2 + d)^q} dxdy.$$

2.1.2 Evaluating the Spatial Integral

A particular challenge of the likelihood function is in evaluating the spatial integral in INT_{trig}. Ogata [1998] originally suggested to evaluate this intergral by radially approximating it across the catalog zone. Section 6.3 in the Appendix provides a comprehensive explanation of this radial approximation and several other approaches to handling this integral. The radial approximation relies on setting several arbitrary choices on how to split the region and is computationally expensive, so we investigated other approaches. It should be noted that the most common approach in the literature is simply to set this integral equal to unity, allowing to ignore it [Schoenberg, 2013]. While this is more computationally efficient and does not require arbitrary choices, it does not work well for seismic regions with seismicity near its borders, which is the case for the PNW (see Figure 3). We required a different approach to evaluate the spatial integral.

We developed a way to solve for the spatial integral in one dimension analytically (say, x). This yields a function that is an integral over another dimension (y). This integral does not have a closed form solution but can be numerically evaluated. Such an analytical+numerical approach has been used in previous ETAS studies [Werner et al., 2011]. Specifically, solving the (indefinite) spatial integral, G, in one dimension yields:

$$\begin{aligned} G &= \int (x^2 + y^2 + d)^{-q} dx \\ &= x(d + x^2 + y^2)^{-q}(1 + \frac{x^2}{d + y^2})^q \cdot {}_2F_1(\frac{1}{2}, q; \frac{3}{2}; \frac{-x^2}{y^2 + d}), \end{aligned}$$

where ${}_2F_1$ is the hyper-geometric function that results when integrating a function with a negative power.

We can now rewrite the two-dimensional spatial integral as:

$$\int_{S_y^{(j)}} \int_{S_x^{(j)}} (x^2 + y^2 + d)^{-q} dx dy$$
$$= \int_{S_y^{(j)}} x(d + x^2 + y^2)^{-q} (1 + \frac{x^2}{d + y^2})^q \cdot {}_2F_1(\frac{1}{2}, q; \frac{3}{2}; \frac{-x^2}{y^2 + d}) dy,$$

where $S_x^{(j)}$ and $S_y^{(j)}$ are the range of the spatial zone in each direction.

This expression can be evaluated by one-dimensional numerical integration, which is faster and less sensitive to initial values than two-dimensional numerical integration. The standard approach is to integrate via some quadrature rule with the most common one being Gaussian quadrature. We use Gaussian-Kronrod quadrature (in R's function "integrate") with a convergence tolerance of 10^{-6} for each earthquake.

Computing an updated G value for each earthquake for each proposed d/q combination would be computationally costly, especially within an iterative MCMC procedure (see following section). Our solution is to first compute G for a somewhat fine grid of d/q values. This is done for every earthquake in the catalog, since the product in the INT_{trig} term is evaluated earthquake-wise. The grids used in the synthetic catalog analysis are presented in Table 16 in Appendix Section 6.3.4.

We then interpolate from the grid of G values for a specific proposed d/q combination. We use bilinear interpolation because of observed smoothness in the G surfaces for several synthetic catalogs. Figure 14 in Appendix Section shows G surfaces for two earthquakes, near the spatial boundary (left plot, lower G values) and in the middle of the region (right plot, higher G values, as there is more space to integrate over).

2.1.3 Branching Likelihood for ETAS

Now assume that, for each catalog event i, we know B as given in Section 1.2. Conditional on this B, a branching likelihood $L(Y|\theta, B)$ can be derived (we provide this derivation in Section 6.4). For the spatiotemporal ETAS model, this branching likelihood is

$$L(Y|\theta, B) = \exp(-\mu_{ST} T) \prod_{j=1}^{n} \mu_{ST} \prod_{j=1}^{n} \Big(\exp(-k(M_j) H(T - t_j) G((x - x_j, y - y_j)) k(M_j)^{|A_j|} \prod_{t_i \in A_j} h(t_i - t_j) g(x_i - x_j, y_i - y_j) \Big),$$

where

$$H(T - t_j) = \int_0^{T - t_j} h(t) dt,$$
$$G(x - x_j, y - y_j) = \int \int_{S^{(j)}} g(x - x_j, y - y_j) dx dy,$$

A_j is the set of all earthquake triggered by earthquake j and A_0 is the set of all earthquake not triggered by a previous earthquake (background events). In the branching likelihood, the first two terms are the contribution of the background rate to the likelihood and the terms inside the $\prod_{j=1}^{n}$ are the aftershock contribution to the likelihood: integral part and sum part, respectively.

We will work with the log-likelihood $l(Y|\theta, B) = \log(L(Y|\theta, B))$:

$$l(Y|\theta, B) = |A_0| \log(\mu_{ST}) + \sum_{j=1}^{n} |A_j| \log(k(M_j)) + \sum_{t_i \in A_j} \log(h(t_i - t_j)) + \log(g(x_i - x_j, y_i - y_j)) \quad (3)$$

$$+ (-\mu_{ST} T |S|) + \sum_{j=1}^{n} -k(M_j) H(T - t_j) G(x - x_j, y - y_j)$$

Figure 1 shows a surface for $l(Y|\theta, B)$ with varying K and α near their true values for Catalog 2A (see next section). Figure 2 shows a surface for $l(Y|\theta, B)$ with varying d and q near their true values for Catalog 3D. Other plots for these two parameter combos for different synthetics looked similar.

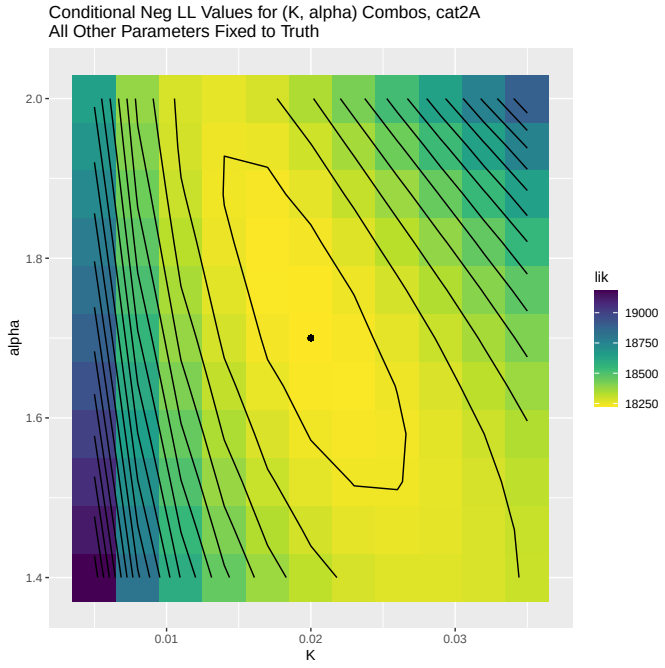

Figure 1: Branching likelihood surface for Catalog 2A, varying K and α near their true values and keeping other parameters fixed to their true values. Note that this shows the negative log-likelihood values (lowest values are the maximal values of the log-likelihood values). Contours for every 50 values of the likelihood. The point marks the true K and α for Catalog 2A.

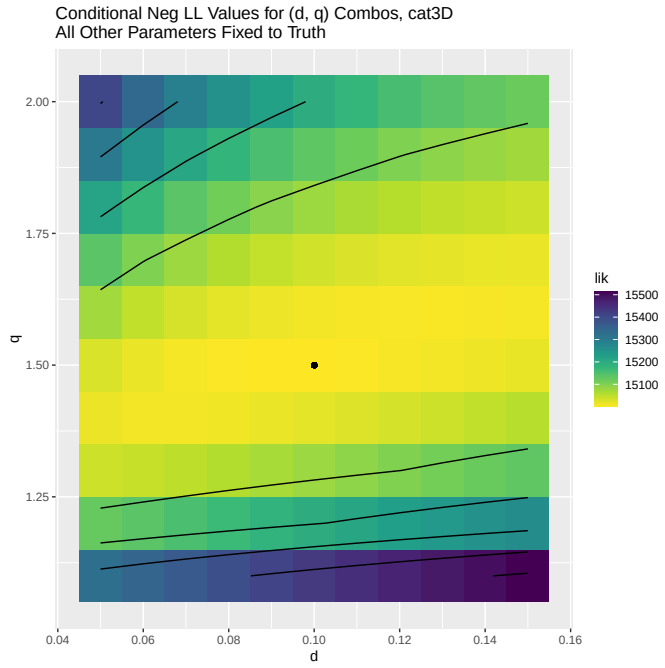

Figure 2: Branching likelihood surface for Catalog 3D, varying d and q near their true values and keeping other parameters fixed to their true values. Contours for every 100 values of the likelihood.

Parameter	Default Prior
B_i	Uniform on previous events
μ_{ST}	Gamma(0.1, 0.1)
K	Uniform(0, 10)
α	Uniform(0, 10)
c	Uniform(0, 8)
p	Uniform(1, 8)
d	Uniform(0, 8)
q	Uniform(1, 8)

Table 1: Priors for spatiotemporal ETAS parameters taken from Ross [2021]. No difference was found in the posteriors for K between the default and alternative Gamma priors.

2.1.4 Bayesian Inference for ETAS Parameters

Bayesian inference is based on drawing samples from the posterior distribution for a parameter of interest. The parameter's posterior is proportional to its prior distribution (our initial beliefs about the values the parameter can take) scaled by the likelihood function (which expresses how likely any given parameter value is, given the data we have observed and our model). That is, the data enters the model through the likelihood and only adjusts the prior to the extent that the data contains information about the parameter. Priors are described in Table 1.

We use the branching likelihood $L(Y|\theta, B)$ with the spatial integral term as described in Section 2.1.2. So every posterior distribution $p(\theta)$ for parameter θ will be expressed proportional to the prior, $\pi(\theta)$, multiplied by $L(Y|\theta, B)$. Furthermore, we will work with the log-posteriors, so

$$\begin{aligned} p(\theta) &\propto \pi(\theta) L(Y|\theta, B) \\ \log(p(\theta)) &\propto \log(\pi(\theta)) + l(Y|\theta, B). \end{aligned}$$

We can re-express $l(Y|\theta, B)$ from equation 3 as:

$$\begin{aligned} l(Y|\theta, B) =& |A_0| \log(\mu_{ST}) - \mu_{ST} T|S| \\ &+ \sum_{j=1}^{n} -k(M_j) H(T - t_j) G(x - x_j, y - y_j) \\ &+ \sum_{j=1}^{n} |A_j| \log(k(M_j)) + \sum_{t_i \in A_j} \log(h(t_i - t_j)) + \log(g(x_i - x_j, y_i - y_j)) \\ =& \, SUM_{bkgd} + INT_{bkgd} + INT_{trig} + SUM_{trig}. \end{aligned}$$

Unlike in the traditional likelihood function, the only term here that depends on all three components of the triggering model is INT_{trig}, which equals $\sum_{j=1}^{n} -k(M_j) H(T - t_j) G(x - x_j, y - y_j)$.

We will pair the parameters belonging to each model component and sample from their conditional posteriors, conditional on fixed values for the other ETAS parameters. The log-posterior density functions for each parameter pair are derived in Appendix Section 6.5. We use a Gibbs sampler procedure [Ross, 2017, 2021] to sample from each conditional posterior distribution, fixing all other parameters to their current values. We take the common approach of sampling posterior values directly from each conditional posterior distribution with the Metropolis-Hastings sampler [Robert and Casella, 1999], a type of Markov Chain Monte Carlo sampler.

We use the following procedure to sample values from each parameter's posterior distribution:

1. Set initial values for ETAS parameters $B_i, \theta = \{\mu, K, \alpha, c, p, d, q\}$. Again, we assume a spatially constant background rate so $\mu_{ST}(x, y) = \mu \frac{1}{|S|}$.
2. Take a uniform prior distribution on B_i (each previous earthquake has equal probability of triggering earthquake i, and it is as probable to be a background event). Sample from the posterior of B_i by

randomly drawing an integer from 0 to j ($j < i$) with the following probability weights:

$$P(B_i = j|Y,\theta) = \begin{cases} \frac{\mu_{ST}}{\mu_{ST}+\sum_{j=1}^{i-1} k(M_j)h(t_j)g(x-x_j,y-y_j)} & \text{if j = 0} \\ \frac{k(M_j)h(t_j)g(x-x_j,y-y_j)}{\mu_{ST}+\sum_{j=1}^{i-1} k(M_j)h(t_j)g(x-x_j,y-y_j)} & \text{if } j \in 1,2,\ldots,i-1 \end{cases}$$

This is equivalent to sampling from a discrete Uniform distribution with probabilities given by $P(B_i = j|Y,\theta)$.

3. Using the B_i information, sample from the posteriors of ETAS parameters:
   - For μ, use a Gamma prior, which means the posterior is also Gamma and can be directly sampled.
   - For (K,α), use improper priors and a random walk Metropolis-Hastings algorithm to sample from the posterior.
   - For (c,p), use Uniform priors over a broad interval and a random walk Metropolis-Hastings algorithm to sample from the posterior.
   - For (d,q), use Uniform priors over a broad interval and a random walk Metropolis-Hastings algorithm to sample from the posterior.
4. Repeat steps 2 and 3, updating the B_i with new posterior draws for θ and then the estimation of θ using the updated B_i, for a fixed number of iterations (usually 10,000).

We implemented this computationally as a set of C++ functions that are used within the R statistical software.

2.2 Simulation of Spatiotemporal ETAS Catalog

In order to evaluate the utility of the Bayesian ETAS procedure and our computational implementation, we estimate parameters for catalogs simulated from the spatiotemporal ETAS model. We used several fixed parameter sets and spatiotemporal square region $[0,T]\times[0,S]^2$. To account for know boundary issues [Harte, 2013], we simulate catalogs for a larger auxiliary space-time window and estimate parameters for a smaller space-time window $[0,T_{target}]\times[0,S_{target}]^2$, where $T = T_{target} + T_{margin}$ and $S = S_{target} + S_{margin}$.

Our simulation, based on Zhuang and Touati [2015], follows the following steps:

1. Draw a set of n_{bkgd} background earthquakes, where $n_{bkgd} \sim Pois(\mu \cdot T)$, where T is the catalog duration (eg, 2000 or 5000 days).
   (a) For each background event i, simulate a random time $t_i \sim Unif(0,T)$, a random location $x_i \sim Unif(0,S), y_i \sim Unif(0,S)$, and random magnitude $M_i \sim GR(\beta)$, the Gutenberg-Richter law with slope β.
2. Set $T_{as} = 2 \cdot T$ as the maximum aftershock duration. For each background event (t_i, M_i), simulate a random number of triggered events $n_{trig,i} \sim Pois(\lambda_{trig,i})$, where

$$\lambda_{trig,i}(t_i,M_i) = K\exp(\alpha\cdot(M_i - M_0))\cdot((T_{as}+c)^{1-p} - c^{1-p})\cdot\frac{1}{1-p}.$$

   This is the integral of the unnormalized Omori temporal density from $t = 0$ to $t = T_{as}$, scaled by the time-independent productivity law. This represents the expected number of triggered events within the finite time window [0, T_{as}]. Thus, the number of triggered events is a function solely of the mainshock's magnitude.
3. Each of these triggered events ($j = 1,\ldots,n_{trig,i}$) needs a random time, location, magnitude and number of triggered events.
   (a) Draw a random time using the inverse transform sampling technique. The inverse transform for the modified Omori law is derived in Section 6.1. Simulate $r \sim Unif(0,1)$ and draw a random time difference Δt_j from [0, T_{as}]:

$$\Delta t_j = (r\cdot((T_{as}+c)^{1-p} - c^{1-p}) + c^{1-p})^{1/1-p} - c.$$

   Set the triggered event's time $t_j = t_i + \Delta t_j$. Remove any times that fall outside T.
   (b) Draw a random location using inverse transform sampling from the Utsu-Seki law. The inverse transform for the modified Utsu-Seki law is derived in Section 6.1. This has two steps:
      i. Simulate a random distance by drawing $r \sim Unif(0,1)$ and drawing a random distance Δs_j from [0, S]:

$$\Delta s_j = \sqrt{[(r-1)(-d^{1-q})]^{1/1-q} - d}.$$

ii. Draw a random angle in the circle, uniformly from $(0, 2\pi)$. The aftershock is located at the polar coordinate determined by this random distance and angle.

(c) Draw a random magnitude $\sim GR(\beta)$.

(d) Draw a random number of triggered events again using $n_{trig,j} \sim Pois(\lambda_{trig,j})$, where

$$\lambda_{trig,j}(t_j, M_j) = K \exp(\alpha \cdot (M_j - M_0)) \cdot ((T_{as} + c)^{1-p} - c^{1-p}) \cdot \frac{1}{1-p}.$$

4. Repeat the above step for all generations of triggered events until there are no more. The number of triggered events is guaranteed to converge to 0 for subcritical catalogs [van der Elst, 2017].

	Description	μ	K	α	c	p	β	T	n_{aux}	n_{tar}	$n_{M>4}$	br. ra
Catalog 2	Low α, high K	0.1	0.02	1.7	0.05	1.08	ln(10)	20000	8121	1590	45	0.67
Catalog 3	High α	0.1	0.006	ln(10)	0.05	1.08	ln(10)	20000	5866	1384	46	0.63
Units		$\frac{\text{events}}{(\text{days}\cdot\text{km}^2)}$	events	$\frac{1}{\text{M}}$	days	-	-	days	events	events	events	-

Table 2: Synthetic test catalogs (M_{min}=2.5 for all). Catalogs were simulated for an auxiliary spatial window of (0,1000; $S = 1000km^2$) and values for the last four columns are for the target window of (250,750) only ($S_{target} = 500km^2$). To account for the temporal boundary issues, we used $T = 20000$ days and $T_{target} = 15000$ days.

	d	q
A	1	2
B	1	1.5
C	0.1	2
D	0.1	1.5
Units	km^2	-

Table 3: Simulation values for spatial parameters. Catalogs 2 and 3 had all four variations simulated.

The above procedure was performed for the "A" version of the catalog (d=1, q=2, see Table 3). We then used this catalog to reassigned earthquakes to different locations based on the other d and q combinations. Thus, the total number of earthquakes is kept fixed and aftershocks are simply relocated based on different spatial parameters. Taking as input the auxiliary catalog (in our case, S=1000) and a d/q combination, we use the following procedure:

1. Identify all mainshocks as earthquakes that were not triggered by a previous earthquake. Reassign all mainshock locations uniformly, so $x_i \sim Unif(0, S), y_i \sim Unif(0, S)$ for all mainshocks i.
2. For each mainshock, find its triggered events. Assign to each one a random distance Δs_j from the mainshock and random angle as in Step 3b above. This provides the (x, y) of the relocated aftershock j.

3 Pacific Northwest Catalog

We collected a new complete catalog of contemporary seismicity for the international PNW region. We consider all earthquakes from 1970-2018 instrumentally observed by the authoritative seismic network for a rectangular box around the states of Washington and Oregon (*target* region) and a larger *auxiliary* region, which comprises of the target zone and a margin zone around it; see Figure 3. The authoritative networks are the Pacific Northwest Seismic Network (PNSN) for all earthquakes in Washington and Oregon, the Geological Survey of Canada (GSC) for all earthquakes in British Columbia, and other regional networks for the US states bordering Washington and Oregon, as included in the Advanced National Seismic System's ComCat catalog.

We consider all earthquakes above magnitude 2.0, as this is believed to be the magnitude of completeness [Malone, 2020] for the PNSN, for the vast majority of this time period and region. However, according to Malone [2019], "in the early decades of the 1970s and 1980s, southern Oregon was not monitored at all by the PNSN ...', and so we follow their example of only considering the region above 42.6°as complete. We build models on the catalog both down to 42°and 42.6°. Table 4 has information about the three data sources going into the catalog to be modelled.

Catalog	Latitude	Longitude	Magnitude	Time Window
Target (PNSN)	42°× 49°	-125°× -116.5°	M2.0+	1970-01-01 to 2019-01-01
Target (GSC)	42°× 49°	-125°× -116.5°	M2.0+	1985-01-01 to 2019-01-01
Auxiliary (PNSN)	41°× 50°	-126°× -115.5°	M2.0+	1970-01-01 to 2019-01-01
Auxiliary (GSC)	41°× 50°	-126°× -115.5°	M2.0+	1985-01-01 to 2019-01-01

Table 4: Target catalog parameters and number of events for raw catalogs. The latitude and longitude correspond to a rectangular box around the US states of Washington and Oregon, as well as a portion of British Columbia. The time window begins in 1970, when the instrumental record of the PNSN began. For GSC, the catalog begins in 1985, as location and magnitude calculations are inconsistent before and after 1985.

Duplicate events between catalogs were algorithmically identified and confirmed by seismologists from the seismic networks; we used authoritative solutions when possible. More details on this procedure are in Schneider et al. [Forthcoming]. Our careful study of duplicates resulted in 280 events entering the (auxiliary) catalog from outside the authoritative source.

We excluded blasts and attempted to identify earthquake swarms, which are common in the PNW [Bostock et al., 2019] and do not comply with the time-stationarity assumed by the ETAS model of the background rate. We first removed any seismicity within a 10 km radius circular area of Mt. St. Helens, which produced thousands of volcanic (non-tectonic) earthquakes. We then use a window-based deterministic procedure (similar to those used by Jacobs et al. [2013]) to identify earthquake clusters in horizontal space, time and depth (within the target zone only). These were classified by PNSN seismologists on a six-point scale:

- **0: Neither a swarm nor an aftershock sequence.** These earthquakes do not appear to be related enough to be a cluster or possess neither the characteristics of swarms, nor sequences.
- **1: Definitely a swarm.** The cluster has the features of a swarm, as far as times, locations, magnitudes, depths.
- **2: Likely/maybe a swarm.** The cluster exhibits more features of a swarm than an aftershock sequence, but it is not definite.
- **3: Don't know / too hard to tell.** The cluster could either be a swarm or an aftershock sequence.
- **4: Likely/maybe a sequence.** The cluster exhibits more features of an aftershock sequence than a swarm, but it is not definite.
- **5: Definitely a sequence.** The cluster has the features of an aftershock sequence, as far as times, locations, magnitudes, depths.

The full details of this process are given in Schneider et al. [Forthcoming]. We consider models with and without events in definite or likely swarms (see Figure 4). There are slight differences in the magnitude-frequency distribution of the catalog with and without maybe/confirmed (MC) swarms (see Figure 6).

The presence of the subducting slab leads to different tectonic regimes in this region, as described in Section 1.4. We used the Slab 2.0 model [Hayes et al., 2018] to calculate the distance between each event's hypocenter and the estimated location of the slab interface. As in Gomberg and Bodin [2021], we considered an event in the crustal regime if it was at least 10 km above the estimated slab interface, due to the uncertainty in the location of the interface (n=3935, see Figure 5, left). All events that were below the estimated slab interface were considered intraplate events ($n = 546$, see Figure 5, right); further restricting to events that were at least 10 km below the estimated interface would reduce this to less than 150 events, so we did not do this.

We computed the spatial integral G for all earthquakes in the PNW target zone, using the d/q grid given in Table 16.

4 ETAS Model Results

4.1 Results on Simulated Catalogs

MLE Inference Table 5 has MLEs all spatiotemporal ETAS parameters using the branching likelihood function, for four end-member synthetic catalogs. In general, MLEs are close to true values, though there is a slight bias in overpredicting μ and underpredicting p.

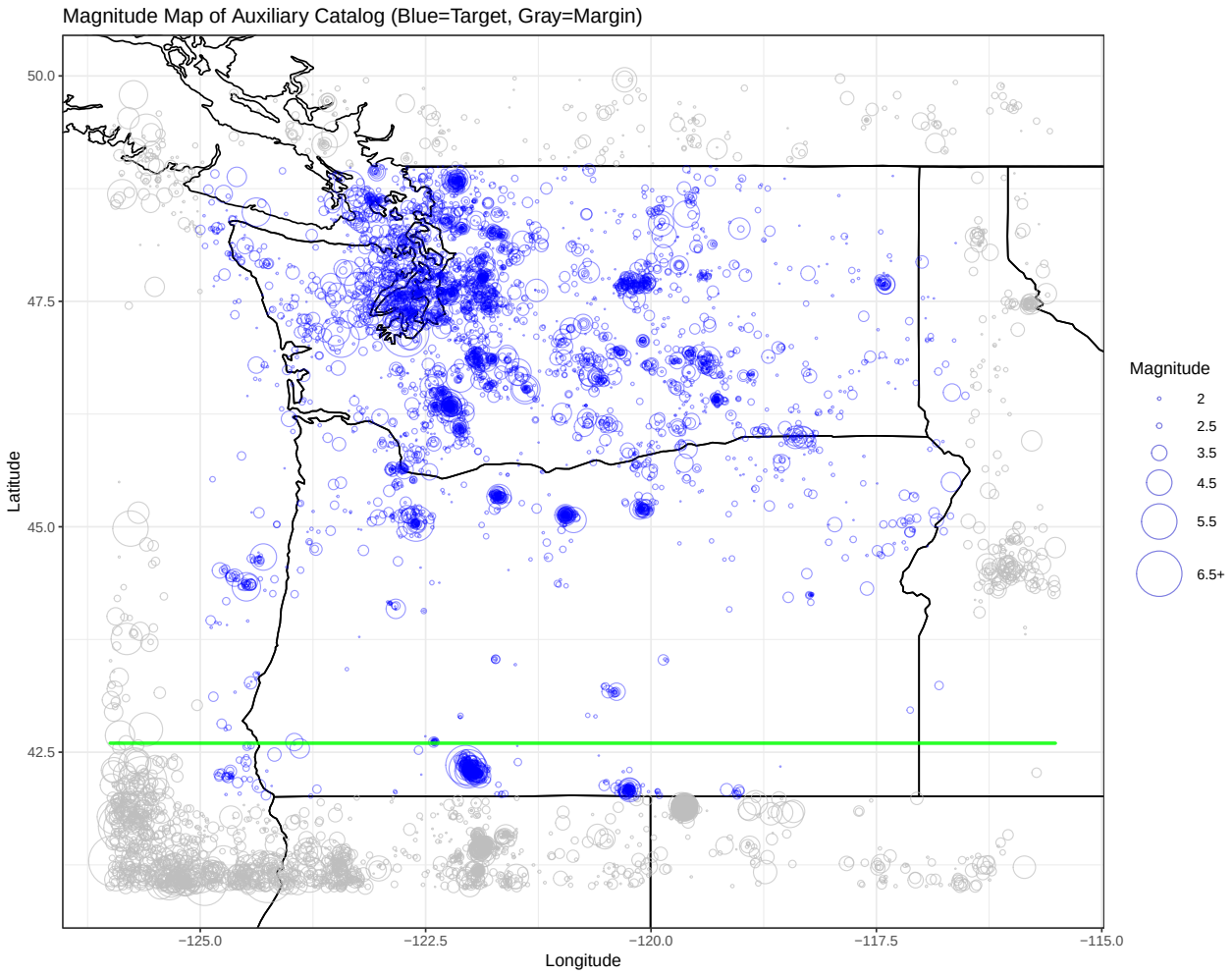

Figure 3: Instrumental catalog for the international PNW region. Target zone events are in blue and margin zone events in gray. Based on the arguments of Malone [2019], we consider only the zone north of 42.6°(above the green line) to be complete.

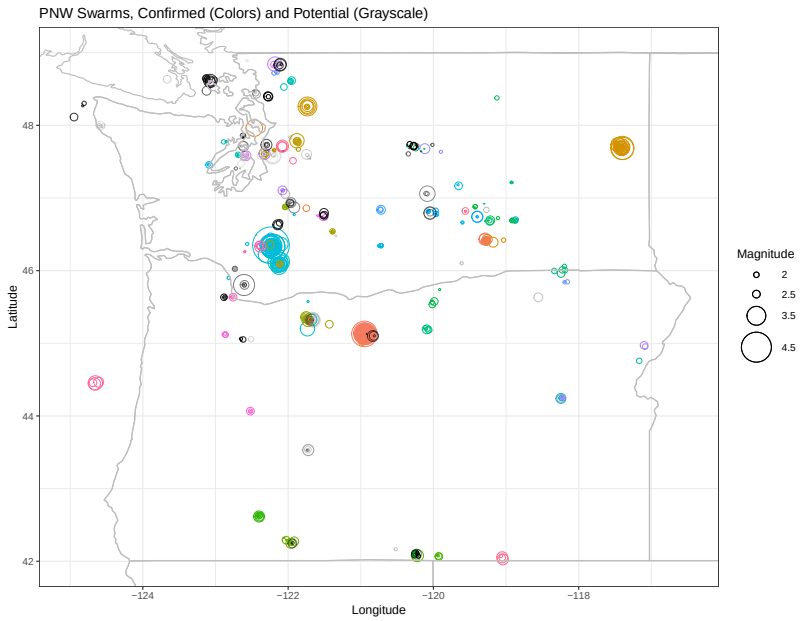

Figure 4: Earthquakes classified as definitely being in a swarm (in rainbow colors) and maybe being in a swarm (in grayscale colors).

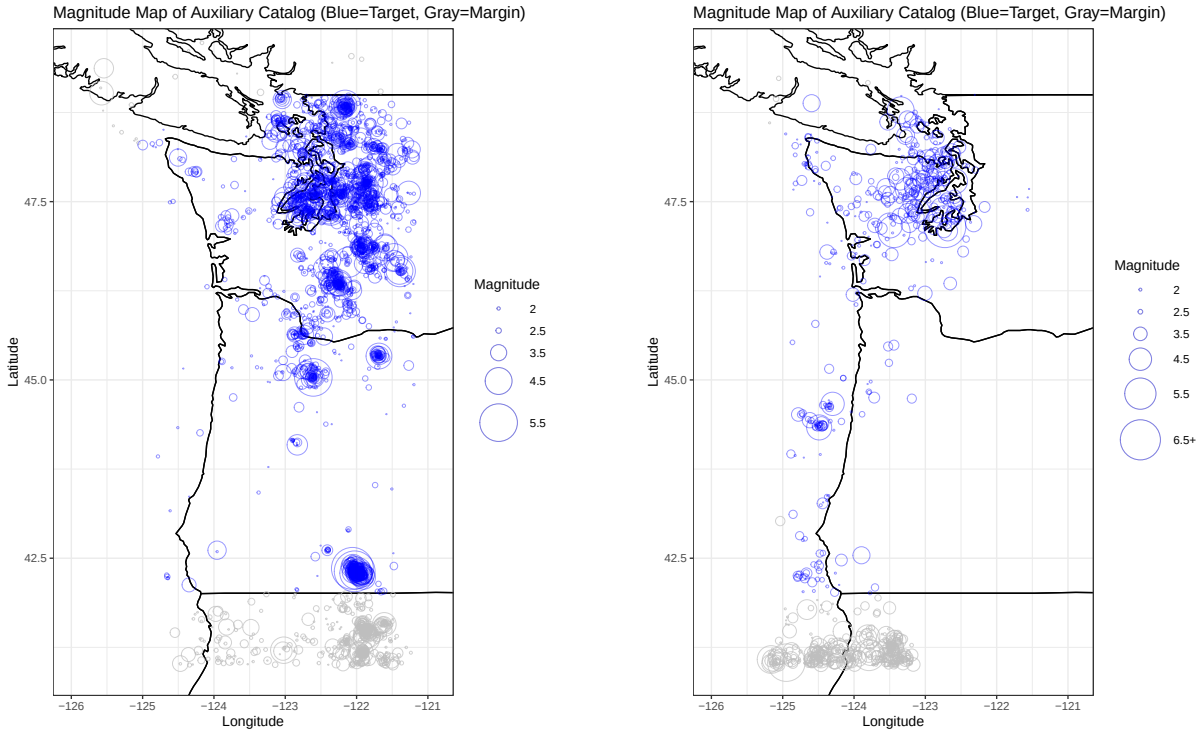

Figure 5: Earthquakes classified as below the slab (left) and earthquakes classified as above the slab (right), using the Slab 2.0 model [Hayes et al., 2018].

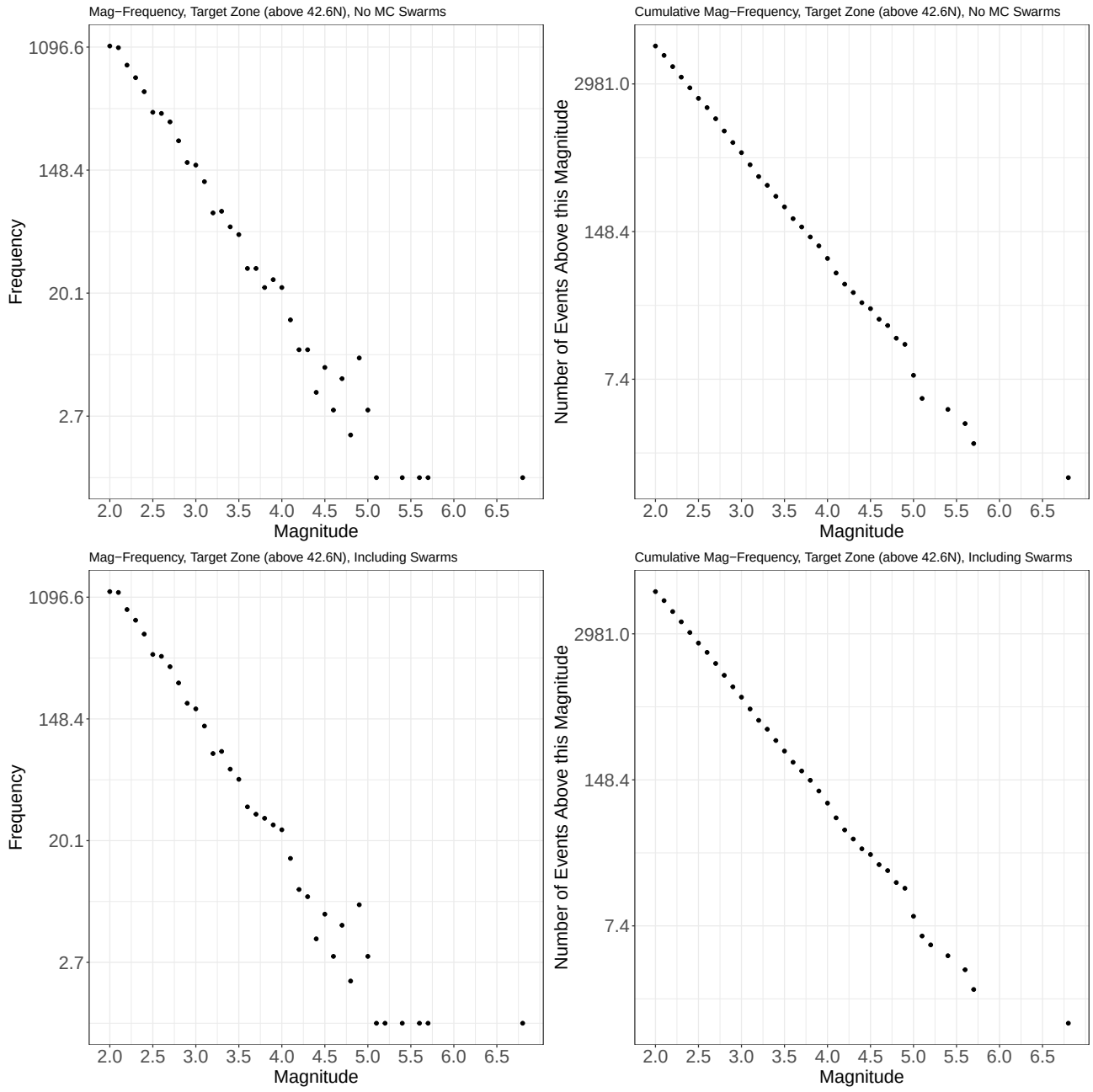

Figure 6: Magnitude frequency and cumulative plots for all earthquakes, and just those not in maybe/confirmed swarms, for the target region above 42.6°. Results look similar for the auxiliary region as well.

	Value	μ	K	α	c	p	d	q
Cat 2A	True	0.1	0.02	1.7	0.05	1.08	1	2
Cat 2A	MLEs	0.113 (0.005)	0.019 (0.001)	1.715 (0.034)	0.043 (0.008)	1.063 (0.012)	1.008 (0.124)	2.009 (0.0
Cat 2D	True	0.1	0.02	1.7	0.05	1.08	0.1	1.5
Cat 2D	MLEs	0.125 (0.005)	0.019 (0.001)	1.702 (0.019)	0.04 (0.005)	1.061 (0.009)	0.082 (0.007)	1.476 (0.01
Cat 3A	True	0.1	0.006	ln(10)	0.05	1.08	1	2
Cat 3A	MLEs	0.108 (0.005)	0.007 (0.001)	2.252 (0.034)	0.052 (0.011)	1.083 (0.014)	0.983 (0.121)	1.986 (0.07
Cat 3D	True	0.1	0.006	ln(10)	0.05	1.08	0.1	1.5
Cat 3D	MLEs	0.122 (0.005)	0.006 (0.001)	2.29 (0.036)	0.055 (0.012)	1.067 (0.015)	0.091 (0.011)	1.491 (0.02

Table 5: MLEs for spatiotemporal ETAS parameters in several synthetic catalogs, using the branching log-likelihood $l(Y|\theta, B)$ and Hessian-based standard errors in paranetheses. All runs have initial values close to true parameters, though results are insensitive to different initial values. Other catalogs not shown did equally well.

Bayesian Inference We evaluated the goodness of fit of the posterior distributions on simulated catalogs visually. For each parameter, we inspected whether the true value was contained within the posterior density curve. We also inspected whether the true value was systematically below or above the posterior mode to evaluate a bias in the estimation (towards over- or under-prediction of the parameter). We further examined trace plots for each parameter, showing the posterior sample by iteration. We considered whether the sampling was well-mixed or whether the sampler took excursions to a part of the parameter space. We finally examined scatterplot matrices of the posterior samples across all parameter pairs to evaluate the level of parameter correlation and trend.

Figure 7 shows the posterior results for the ETAS parameters θ for two simulated catalogs (Catalog 3A and 2D) using the following modelling settings: parameters initialized close to the true values, branching structure initialized close to the true branching structure, and prior distributions taken from Ross [2021]. Posteriors are usually well-centered around the true values (Figure 7) and trace plots show a well-mixed sampling of the posterior (Figure 8). Scatterplot matrices in Figure 9 show large and relatively linear correlations between parameters in the same model component (K and α; c and p; d and q) but also across model components (e.g, K and p). Our posterior estimates allow to more closely investigate parameter correlation that has been reported in past literature [Lombardi, 2015, Wang et al., 2010].

We ran the following sensitivity checks to check the robustness of our parameter estimation procedure and computational implementation.

Changes to model implementation:

- Initialized the ETAS parameters θ to values far from the true simulated values
- Initialized the branching structure B to all background events ($B_i = 0 \forall i$)
- Fixing the spatial integral for each earthquake $G_i = 1 \forall i$ rather than calculating it using our numerical+analytical+interpolation procedure.

Changes to synthetic catalogs:

- Re-simulated the synthetic catalogs with a longer time window
- Simulated 10 versions of each synthetic catalog

Posteriors estimated for each of these changes were similar to those under the initial model implementation, or with the initial catalogs. Furthermore, we observed qualitatively similar posterior results under all eight synthetic catalogs (Catalogs 2A-3D). These sensitivity checks suggested that not only was our computational implementation valid for modelling real catalogs, but that our posterior estimations are robust to different parameter initializations, which is a known problem for MLEs for the ETAS model (see Section 1.2).

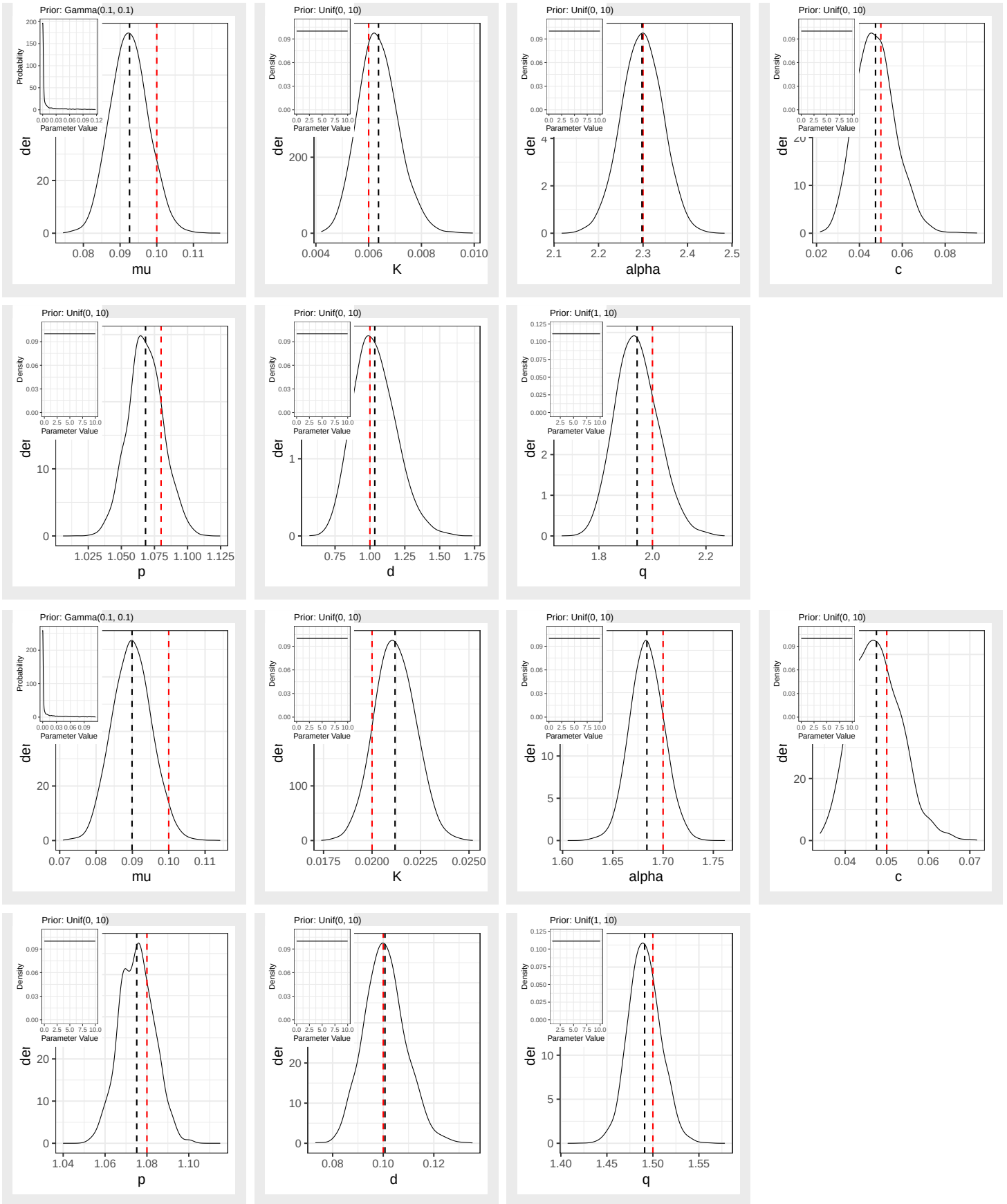

Figure 7: Sampled posteriors for ETAS parameters for synthetic Catalogs 3A (top) and 2D (bottom), with posterior means in black and true parameter values in red. Results were similar for the other synthetic catalogs.

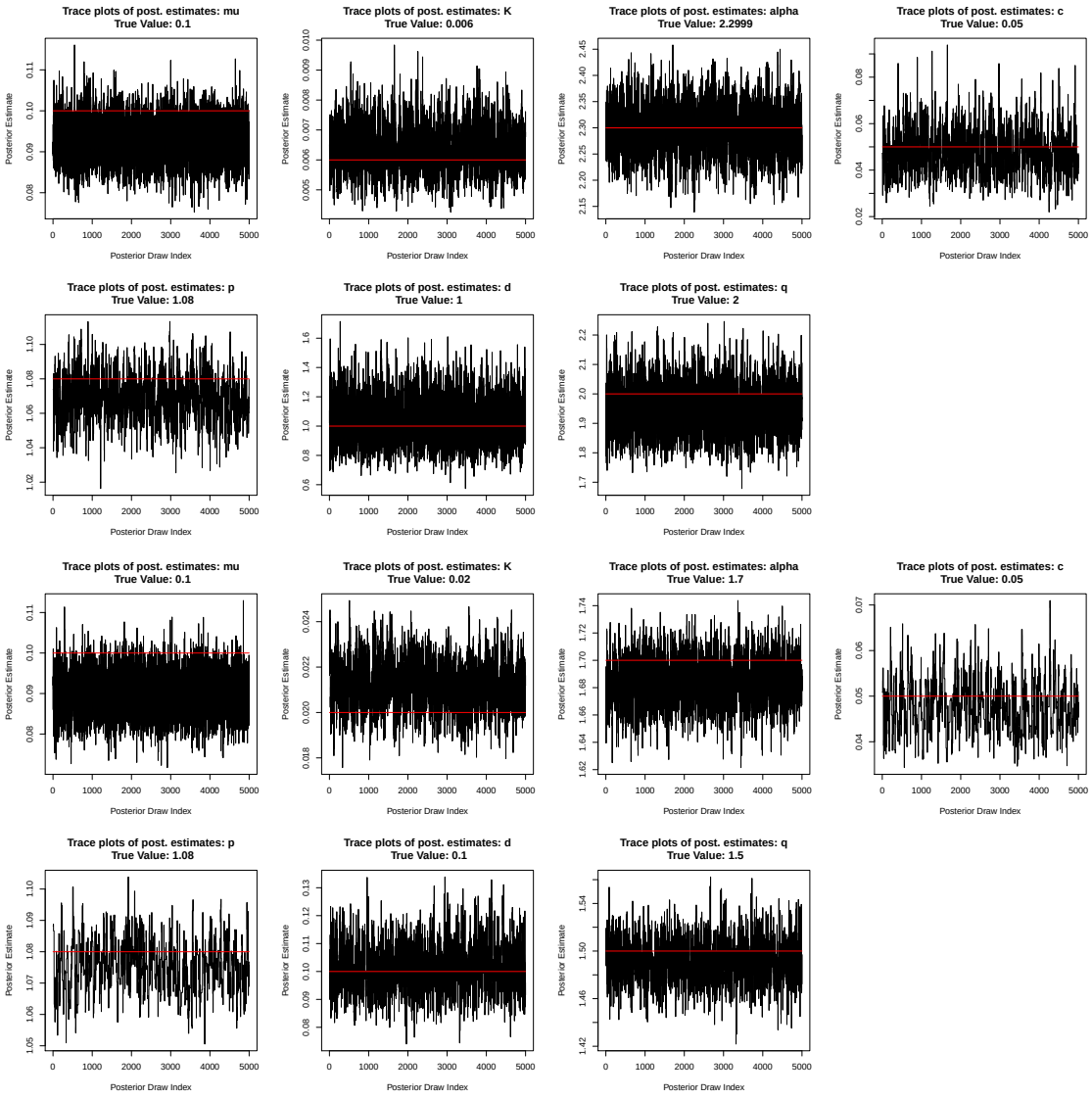

Figure 8: Trace plots for posterior samples for Catalogs 3A (top) and 2D (bottom), with true parameter values in red. Results were similar for the other synthetic catalogs.

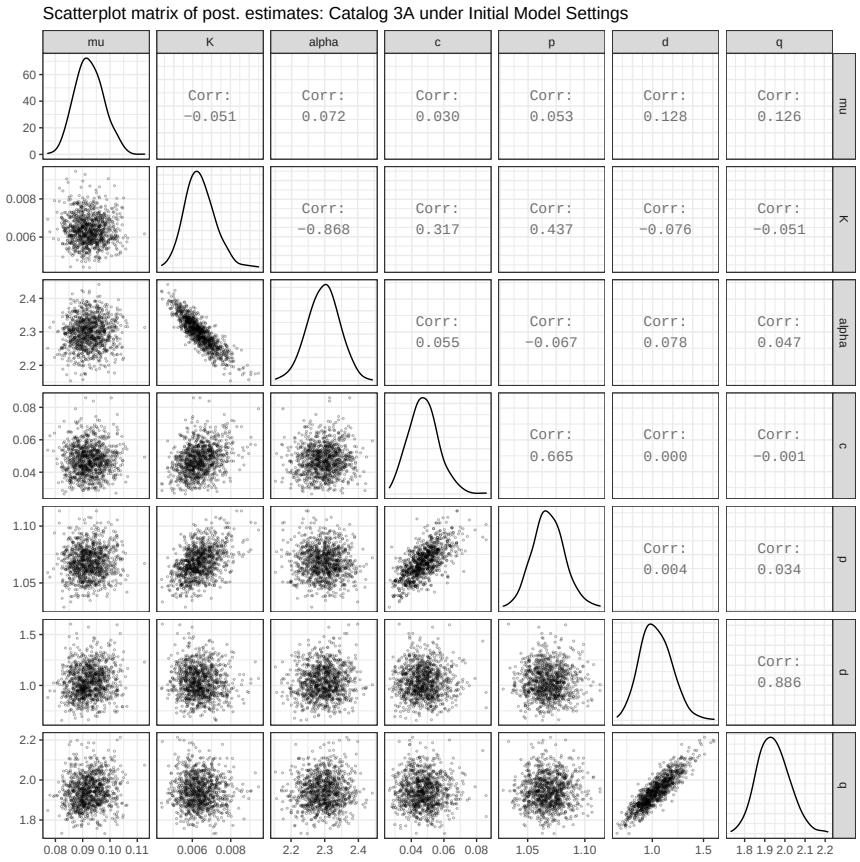

Figure 9: Scatterplot matrices for posterior samples for pairs of parameters for Catalog 3A (scatterplot matrix for Catalog 2D omitted). Results were similar for the other synthetic catalogs.

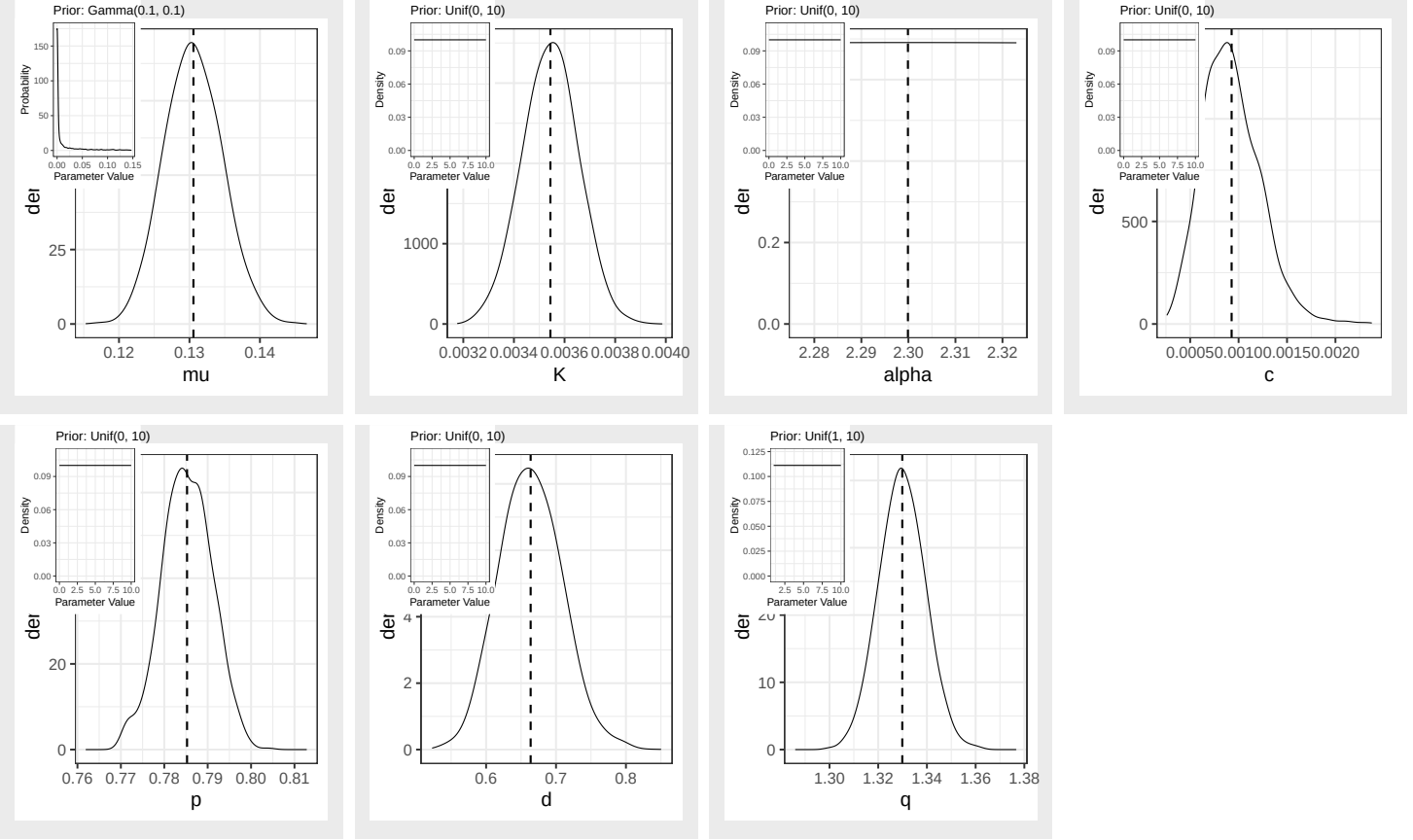

Figure 10: Sampled posteriors for ETAS parameters for PNW catalog, target zone (above 42.6°) without maybe/confirmed swarms, with α fixed to ln(10) and the spatial integral G fixed to 1. Posterior means marked with black lines.

4.2 Results on PNW Catalog

We sampled posterior distributions for the ETAS parameters for different subsets of the PNW catalog. For all results, we used non-specific prior distributions from the ETAS literature [Ross, 2021]. Figure 10 shows posterior densities for ETAS parameters for the catalog, for all earthquakes excluding maybe/confirmed swarms in the target zone north of 42.6°and covering all tectonic regimes (n=6441). These posterior functions use a fixed $\alpha = \ln(10)$, a spatial integral G_i fixed to 1 for all earthquakes i, and the Cat 2 initial values. Figures 11 and 12 show the trace plots and scatterplot matrices for these posterior values, respectively.

We ran a series of experiments to test how each of these modelling and catalog choices affect the posterior results.

Changes to model implementation:

- α parameter fixed to ln(10) (which assumes a self-similar catalog and a Gutenberg-Richter b value of 1) OR allowing α to be estimated by the procedure (see Table 6)
- spatial integral calculated as described in section 2.1.2 OR fixing it to 1, as advocated in Schoenberg [2013] and used by many authors (see Table 7)
- several sets of initial values (we omit this description)

Changes to PNW catalog:

- using the target zone OR the auxiliary zone (target+margin zones) (see Table 8)
- using all earthquakes OR without events that were in possible or definite swarms (with a value of 1 or 2 on the scale described in Section 3) (see Table 9)
- using all regimes OR just earthquakes classified as being in the crustal or deep regime (see Table 10)

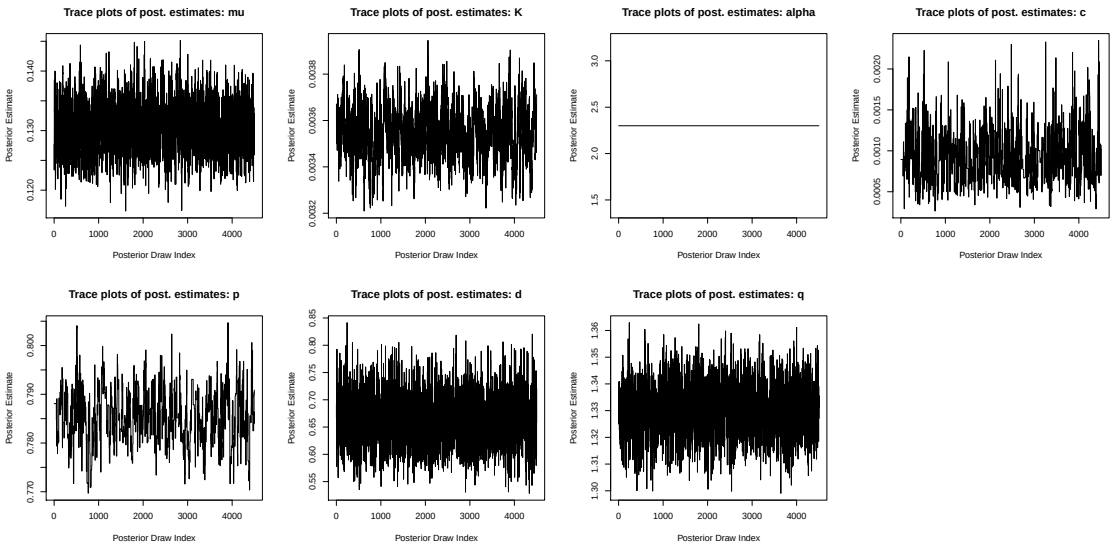

Figure 11: Trace plots for Bayesian inference for ETAS parameters for PNW catalog.

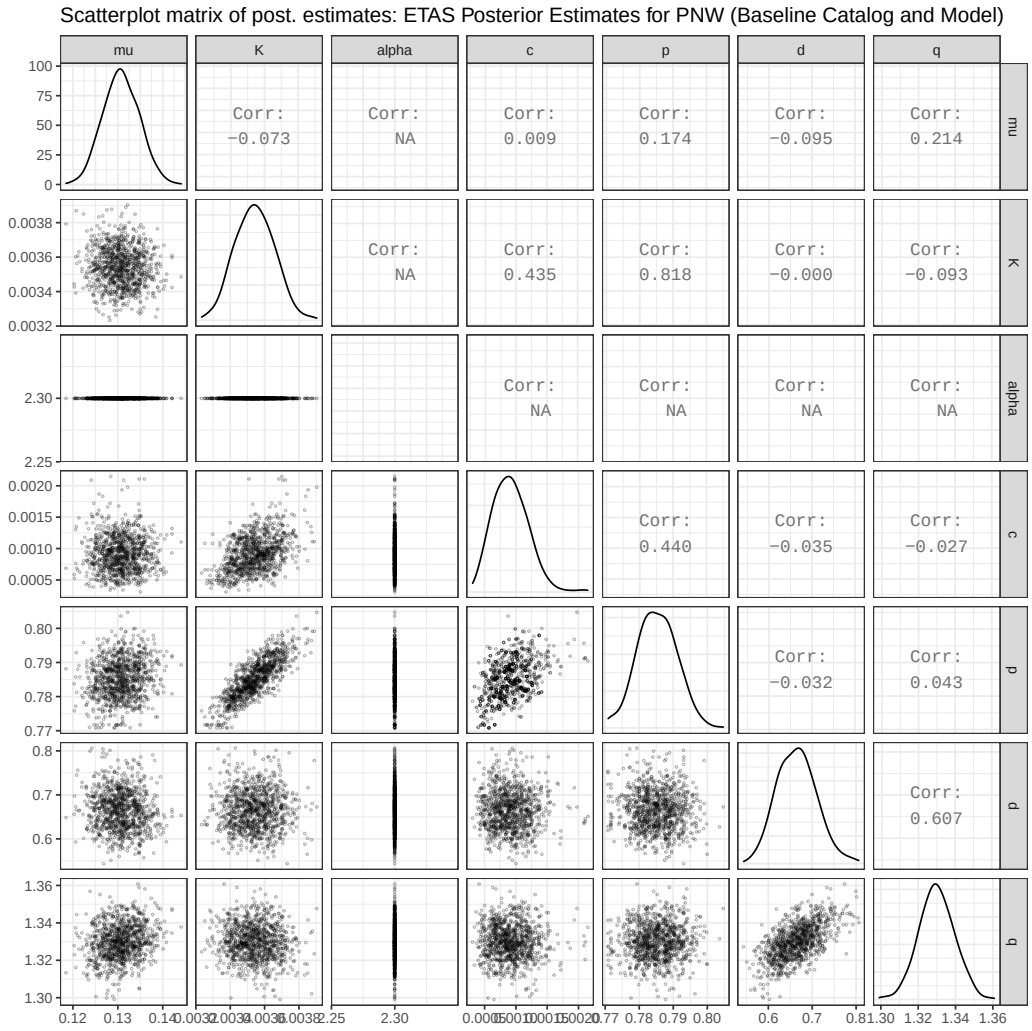

Figure 12: Scatterplot matrices for Bayesian inference for ETAS parameters for PNW catalog.

Experiment	α Fixed?	G Fixed?	Inits	Spatial Zone	Swarms?	Tectonics?	n
Baseline	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42.6°)	No MC swarms	All	6441
Free α	Free	Fixed to 1	Cat 2	Target (42.6°)	No MC swarms	All	6441

Table 6: Experiments with free α. Target (42.6°) means the target zone north of 42.6°. MC swarms are maybe/confirmed swarms. All tectonics means all eqks, regardless of their tectonic regime.

Experiment	α Fixed?	G Fixed?	Inits	Spatial Zone	Swarms?	Tectonics?	n
Baseline	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42.6°)	No MC swarms	All	6441
Calc, interpolate G	Fixed to ln(10)	Novel Solution	Cat 2	Target (42.6°)	No MC swarms	All	6441

Table 7: Experiments with calculated and interpolated G_i for all earthquakes i.

Experiment	α Fixed?	G Fixed?	Inits	Spatial Zone	Swarms?	Tectonics?	n
Baseline	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42.6°)	No MC swarms	All	6441
Target (all)	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42°)	No MC swarms	All	7119
Auxiliary (north)	Fixed to ln(10)	Fixed to 1	Cat 2	Auxiliary (42.6°)	No MC swarms	All	7866
Auxiliary (all)	Fixed to ln(10)	Fixed to 1	Cat 2	Auxiliary (42°)	No MC swarms	All	12457

Table 8: Experiments related to the spatial zone being modelled.

Experiment	α Fixed?	G Fixed?	Inits	Spatial Zone	Swarms?	Tectonics?	n
Baseline	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42.6°)	No MC swarms	All	6441
No confirmed swarms	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42.6°)	No C swarms	All	6574
Include swarms	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42.6°)	All events	All	7095

Table 9: Experiments related to which swarm events to exclude.

Experiment	α Fixed?	G Fixed?	Inits	Spatial Zone	Swarms?	Tectonics?	n
Baseline	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42.6°)	No MC swarms	All	6441
Crustal	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42.6°)	No MC swarms	Crustal (western only)	3935
Crustal+Eastern	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42.6°)	No MC swarms	Crustal (all)	5820
Deep	Fixed to ln(10)	Fixed to 1	Cat 2	Target (42.6°)	No MC swarms	Deep only	546

Table 10: Experiments related to tectonic regime.

Since posterior, trace and scatterplots looked similar between most runs, we omit these and just report summary statistics on the posterior values: their medians and middle 95% intervals (since posteriors tended to be symmetric). These are described in Tables 11 - 15.

The ETAS posteriors were insensitive to the following model choices: solving the spatial integral G or fixing it to 1 (Table 12); setting different initial values (table omitted). The results were also insensitive to the following catalog choices: using the target or auxiliary window (when using the 42.6°southern border, Table 13)); grouping Crustal+Eastern regimes or considering Crustal events only (except for μ, Table 14); excluding Maybe/Confirmed swarms v. Confirmed swarms only (Table 15).

The ETAS posteriors were sensitive to fixing $\alpha = \ln(10)$ or letting it be estimated (Table 11). They were also sensitive to the following catalog choices: setting the southern boundary at 42.6°or 42°(Table 13); deep v. non-deep earthquakes (Table 14); including (any kind of) swarms or not (Table 15).

5 Acknowledgements

We gratefully acknowledge the consistent and critical guidance and support of the Operational Aftershock Forecasting Team at USGS: Andrew Michael, Jeanne Hardebeck, Nicholas van der Elst, and Morgan Page.

	μ	K	α	c	p	d	q
Baseline	0.12 (0.11, 0.13)	0.0035 (0.0033, 0.0038)	ln(10)	9e-04 (4e-04, 0.0016)	0.79 (0.77, 0.8)	0.66 (0.58, 0.75)	1.33 (1.31, 1.35)
Free α	0.031 (0.026, 0.035)	0.028 (0.027, 0.03)	0.52 (0.45, 0.58)	0.001 (5e-04, 0.0017)	0.77 (0.76, 0.78)	0.93 (0.82, 1.06)	1.32 (1.31, 1.34)

Table 11: Results of experiments with free α (see Table 6).

	μ	K	α	c	p	d	q
Baseline	0.13 (0.13, 0.14)	0.0042 (0.004, 0.0044)	ln(10)	0.0018 (0.0011, 0.0028)	0.84 (0.83, 0.85)	0.71 (0.63, 0.81)	1.37 (1.35, 1.39
Solve G	0.13 (0.13, 0.14)	0.0043 (0.0041, 0.0045)	ln(10)	0.0019 (0.0012, 0.0028)	0.84 (0.83, 0.85)	0.72 (0.63, 0.81)	1.37 (1.34, 1.39)

Table 12: Results of experiments using our novel solution for spatial integral G (see Table 7).

	μ	K	α	c	p	d	q
Baseline	0.12 (0.11, 0.13)	0.0035 (0.0033, 0.0038)	ln(10)	9e-04 (4e-04, 0.0016)	0.79 (0.77, 0.8)	0.66 (0.58, 0.75)	1.33 (1.31, 1.35)
Target	0.14 (0.13, 0.14)	0.0053 (0.005, 0.0055)	ln(10)	0.0025 (0.0018, 0.0035)	0.87 (0.86, 0.88)	0.57 (0.51, 0.65)	1.38 (1.36, 1.4)
Aux N	0.13 (0.13, 0.14)	0.0034 (0.0032, 0.0036)	ln(10)	0.001 (6e-04, 0.0018)	0.79 (0.78, 0.8)	0.72 (0.63, 0.81)	1.32 (1.3, 1.33)
Aux	0.52 (0.51, 0.54)	0.0028 (0.0027, 0.0029)	ln(10)	0.0103 (0.0087, 0.0123)	1.03 (1.02, 1.04)	0.53 (0.48, 0.58)	1.52 (1.5, 1.54)

Table 13: Results of experiments for the spatial zone being modelled (see Table 8).

	μ	K	α	c	p	d	q
Baseline	0.12 (0.11, 0.13)	0.0035 (0.0033, 0.0038)	ln(10)	9e-04 (4e-04, 0.0016)	0.79 (0.77, 0.8)	0.66 (0.58, 0.75)	1.33 (1.31, 1.35)
No C Swarms	0.12 (0.11, 0.13)	0.0038 (0.0036, 0.004)	ln(10)	9e-04 (5e-04, 0.0016)	0.79 (0.78, 0.8)	0.63 (0.55, 0.72)	1.33 (1.31, 1.35)
Incl. Swarms	0.12 (0.12, 0.13)	0.0047 (0.0044, 0.0049)	ln(10)	0.0013 (8e-04, 0.0021)	0.83 (0.82, 0.84)	0.51 (0.45, 0.57)	1.34 (1.32, 1.36)

Table 14: Results of experiments for which swarm events to exclude (see Table 9).

	μ	K	α	c	p	d	q
Baseline	0.12 (0.11, 0.13)	0.0035 (0.0033, 0.0038)	ln(10)	9e-04 (4e-04, 0.0016)	0.79 (0.77, 0.8)	0.66 (0.58, 0.75)	1.33 (1.31, 1.35)
Crustal	0.022 (0.018, 0.025)	0.01 (0.0094, 0.0107)	ln(10)	9e-04 (4e-04, 0.0018)	0.77 (0.76, 0.78)	0.89 (0.76, 1.04)	1.34 (1.32, 1.36)
Crustal+ Eastern	0.044 (0.039, 0.049)	0.01 (0.0095, 0.0107)	ln(10)	9e-04 (4e-04, 0.0017)	0.77 (0.76, 0.78)	0.87 (0.76, 0.98)	1.31 (1.3, 1.33)
Deep	0.023 (0.02, 0.026)	2e-04 (1e-04, 3e-04)	ln(10)	0.0137 (0.0011, 0.0894)	0.8 (0.71, 0.9)	8.2 (4.89, 9.92)	1.41 (1.31, 1.57)

Table 15: Results of experiments related to tectonic regime (see Table 10).

We are also grateful for collaboration and assistance from Luke Blair, Patricia McCrory and Joan Gomberg at the USGS; Amy Wright, Stephen Malone, Paul Bodin, and Renate Hartog, at the Pacific Northwest Seismic Network; Camille Brillon, Taimi Mulder and John Cassidy at Geological Survey of Canada; Henry Flury and other colleagues at the University of Washington; and Sebastian Hainzl (German Research Center for the Geosciences) and Christian Molkenthin (University of Potsdam).

References

Michael G Bostock, Nikolas I Christensen, and Simon M Peacock. Seismicity in Cascadia. *Lithos*, 332:55–66, 2019.

Annie Chu, Frederic P Schoenberg, Peter Bird, David D Jackson, and Yan Y Kagan. Comparison of ETAS parameter estimates across different global tectonic zones. *Bulletin of the Seismological Society of America*, 101(5):2323–2339, 2011.

Hossein Ebrahimian and Fatemeh Jalayer. Robust seismicity forecasting based on Bayesian parameter estimation for epidemiological spatio-temporal aftershock clustering models. *Scientific reports*, 7(1):1–15, 2017.

Joan Gomberg and Paul Bodin. The productivity of Cascadia aftershock sequences. *Bulletin of the Seismological Society of America*, 111(3):1494–1507, 2021.

Joan Gomberg and Allen Jakobitz. A collaborative, user-producer, assessment of earthquake response products. *US Geological Survey Open-File Report*, 2013:1103, 2013.

Zhenqi Guo and Yosihiko Ogata. Statistical relations between the parameters of aftershocks in time, space, and magnitude. *Journal of Geophysical Research: Solid Earth*, 102(B2):2857–2873, 1997.

Peter Guttorp and Thordis L Thorarinsdottir. Bayesian inference for non-Markovian point processes. In *Advances and Challenges in Space-time Modelling of Natural Events*, pages 79–102. Springer, 2012.

Jeanne L Hardebeck, Andrea L Llenos, Andrew J Michael, Morgan T Page, and Nicholas Van Der Elst. Updated California aftershock parameters. *Seismological Research Letters*, 90(1):262–270, 2019.

DS Harte. Bias in fitting the ETAS model: a case study based on New Zealand seismicity. *Geophysical Journal International*, 192(1):390–412, 2013.

DS Harte. Effect of sample size on parameter estimates and earthquake forecasts. *Geophysical Journal International*, 214(2):759–772, 2018.

Gavin P Hayes, Ginevra L Moore, Daniel E Portner, Mike Hearne, Hanna Flamme, Maria Furtney, and Gregory M Smoczyk. Slab2, a comprehensive subduction zone geometry model. *Science*, 362(6410):58–61, 2018.

Katrina M Jacobs, Euan GC Smith, Martha K Savage, and Jiangcang Zhuang. Cumulative rate analysis (CURATE): A clustering algorithm for swarm dominated catalogs. *Journal of Geophysical Research: Solid Earth*, 118(2):553–569, 2013.

Amato Kasahara, Yuji Yagi, and Bogdan Enescu. etas_solve: A robust program to estimate the ETAS model parameters. *Seismological Research Letters*, 87(5):1143–1149, 2016.

Aleksandar A Kolev and Gordon J Ross. Inference for ETAS models with non-poissonian mainshock arrival times. *Statistics and Computing*, 29(5):915–931, 2019.

E Lippiello, F Giacco, L de Arcangelis, W Marzocchi, and C Godano. Parameter estimation in the ETAS model: Approximations and novel methods. *Bulletin of the Seismological Society of America*, 104(2): 985–994, 2014.

Andrea L Llenos and Andrew J Michael. Ensembles of ETAS models provide optimal operational earthquake forecasting during swarms: Insights from the 2015 San Ramon, California swarm. *Bulletin of the Seismological Society of America*, 109(6):2145–2158, 2019.

Anna Maria Lombardi. Estimation of the parameters of ETAS models by simulated annealing. *Scientific reports*, 5(1):1–11, 2015.

Stephen Malone. Private communication, 2020.

Stephen D Malone. An ominous (?) quiet in the Pacific Northwest, 2019.

Andrew J Michael, Sara K McBride, Jeanne L Hardebeck, Michael Barall, Eric Martinez, Morgan T Page, Nicholas van der Elst, Edward H Field, Kevin R Milner, and Anne M Wein. Statistical seismology and communication of the USGS operational aftershock forecasts for the 30 November 2018 Mw 7.1 Anchorage, Alaska, earthquake. *Seismological Research Letters*, 91(1):153–173, 2020.

Andrew Jay Michael. *On the potential duration of the aftershock sequence of the 2018 Anchorage earthquake*. US Department of the Interior, US Geological Survey, 2018.

KZ Nanjo, H Tsuruoka, S Yokoi, Y Ogata, G Falcone, N Hirata, Y Ishigaki, TH Jordan, K Kasahara, K Obara, et al. Predictability study on the aftershock sequence following the 2011 Tohoku-Oki, Japan, earthquake: first results. *Geophysical Journal International*, 191(2):653–658, 2012.

Yosihiko Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

Takahiro Omi, Yosihiko Ogata, Yoshito Hirata, and Kazuyuki Aihara. Intermediate-term forecasting of aftershocks from an early aftershock sequence: Bayesian and ensemble forecasting approaches. *Journal of Geophysical Research: Solid Earth*, 120(4):2561–2578, 2015.

Paul A Reasenberg and Lucile M Jones. Earthquake hazard after a mainshock in California. *Science*, 243 (4895):1173–1176, 1989.

Christian P Robert and George Casella. The Metropolis—Hastings algorithm. In *Monte Carlo Statistical Methods*, pages 231–283. Springer, 1999.

Gordon J. Ross. *bayesianETAS: Bayesian Estimation of the ETAS Model for Earthquake Occurrences*, 2017. URL `https://CRAN.R-project.org/package=bayesianETAS`. R package version 1.0.3.

Gordon J Ross. Bayesian estimation of the ETAS model for earthquake occurrences. *Bulletin of the Seismological Society of America*, 111(3):1473–1480, 2021.

Max Schneider, Henry Flury, and Peter Guttorp. Earthquake catalog processing and swarm identification for the Pacific Northwest. Forthcoming.

Frederic Paik Schoenberg. Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America*, 103(1):601–605, 2013.

Danijel Schorlemmer, Maximilian J Werner, Warner Marzocchi, Thomas H Jordan, Yosihiko Ogata, David D Jackson, Sum Mak, David A Rhoades, Matthew C Gerstenberger, Naoshi Hirata, et al. The Collaboratory for the Study of Earthquake Predictability: achievements and priorities. *Seismological Research Letters*, 89(4):1305–1313, 2018.

Stefanie Seif, Arnaud Mignan, Jeremy Douglas Zechar, Maximilian Jonas Werner, and Stefan Wiemer. Estimating ETAS: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, 122(1):449–469, 2017.

Nicholas J van der Elst. Accounting for orphaned aftershocks in the earthquake background rate. *Geophysical Journal International*, 211(2):1108–1118, 2017.

Alejandro Veen and Frederic 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.

Qi Wang, Frederic Paik Schoenberg, and David D Jackson. Standard errors of parameter estimates in the ETAS model. *Bulletin of the Seismological Society of America*, 100(5A):1989–2001, 2010.

Maximilian J Werner, Agnès Helmstetter, David D Jackson, and Yan Y Kagan. High-resolution long-term and short-term earthquake forecasts for California. *Bulletin of the Seismological Society of America*, 101 (4):1630–1648, 2011.

Lizhong Zhang, Maximilian J Werner, and Katsuichiro Goda. Variability of ETAS parameters in global subduction zones and applications to mainshock–aftershock hazard assessment. *Bulletin of the Seismological Society of America*, 110(1):191–212, 2020.

J Zhuang and S Touati. Stochastic simulation of earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*, 29, 2015.

Jiancang Zhuang. Next-day earthquake forecasts for the Japan region generated by the ETAS model. *Earth, planets and space*, 63(3):207–216, 2011.

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

6 Appendix

6.1 Derivation of Inverse Transforms

In simulation step 3a, we need to draw random triggering times as specified by the ETAS triggering equation, normalized over the time interval $[0, T]$. To obtain a pdf for this distribution, we divide the triggering equation by its integral over the time interval $[0, T]$, using the result that $\int_0^T (t+c)^{-p} = ((T+c)^{1-p} - c^{1-p}) \cdot \frac{1}{1-p}$.

$$\begin{aligned} f(t) &= \frac{K \exp(\alpha \cdot (M_i - M_0)) \cdot (t+c)^{-p}}{\int_0^T K \exp(\alpha \cdot (M_i - M_0)) \cdot (t+c)^{-p} dt} \\ &= \frac{K \exp(\alpha \cdot (M_i - M_0)) \cdot (t+c)^{-p}}{K \exp(\alpha \cdot (M_i - M_0)) \cdot ((T+c)^{1-p} - c^{1-p}) \cdot \frac{1}{1-p}} \\ &= \frac{(t+c)^{-p}}{((T+c)^{1-p} - c^{1-p}) \cdot \frac{1}{1-p}} \end{aligned}$$

From this, we can use the cdf, $F(z)$, to derive the inverse transform function, from which we can sample triggering times using a random uniform number.

$$\begin{aligned} F(z) &= \int_0^z (1-p)(\frac{(t+c)^{-p}}{((T+c)^{1-p} - c^{1-p})}) dt \\ &= (1-p)(\frac{1}{((T+c)^{1-p} - c^{1-p})}) \int_0^z (t+c)^{-p} dt \\ &= (1-p)(\frac{1}{((T+c)^{1-p} - c^{1-p})})(\frac{1}{1-p}\Big((z+c)^{1-p} - c^{1-p}\Big)) \\ &= \frac{(z+c)^{1-p} - c^{1-p}}{(T+c)^{1-p} - c^{1-p}} \end{aligned}$$

We draw $u \sim Unif(0,1)$ and solve $F(F^{-1}(u)) = u$, to find $F^{-1}(u)$, the inverse transform function that will draw a sample from the desired pdf $f(t)$.

$$\begin{aligned} \frac{(F^{-1}(u)+c)^{1-p} - c^{1-p}}{(T+c)^{1-p} - c^{1-p}} &= u \\ (F^{-1}(u)+c)^{1-p} - c^{1-p} &= u((T+c)^{1-p} - c^{1-p}) \\ (F^{-1}(u)+c)^{1-p} &= u((T+c)^{1-p} - c^{1-p}) + c^{1-p} \\ F^{-1}(u)+c &= (u((T+c)^{1-p} - c^{1-p}) + c^{1-p})^{\frac{1}{1-p}} \\ F^{-1}(u) &= (u((T+c)^{1-p} - c^{1-p}) + c^{1-p})^{\frac{1}{1-p}} - c \end{aligned}$$

This is the uth percentile of the desired pdf.

I next derive the inverse transform for the Utsu-Seki spatial decay law (needed for simulation step 3b). First, we need to derive the normalization constant, Q_s, to get a spatial pdf:

$$\begin{aligned} & Q_s \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + y^2 + d)^{-q} dx dy \\ &= Q_s \int_0^{2\pi} \int_0^R (r^2 + d)^{-q} r dr d\theta \\ &= Q_s \int_0^{2\pi} \frac{1}{1-q} (\frac{(R+d)^{1-q}}{2} - \frac{d^{1-q}}{2}) d\theta \\ &= \frac{Q_s}{1-q} \pi ((R+d)^{1-q} - d^{1-q}) \end{aligned}$$

If we set $Q_s = \frac{q-1}{\pi d^{1-q}}$, then we would get

$$\begin{aligned}
&\frac{q-1}{\pi d^{1-q}} \frac{\pi}{1-q}((R+d)^{1-q} - d^{1-q}) \\
&= \frac{-1}{\pi d^{1-q}}((R+d)^{1-q} - d^{1-q}) \\
&= -\frac{-d^{1-q}}{d^{1-q}} \text{ as } R \to \infty \\
&= 1.
\end{aligned}$$

So the spatial pdf is $g(x,y) = \frac{q-1}{\pi d^{1-q}}(x^2+y^2+d)^{-q}$ as this integrates to 1 over the entire domain. We use the cdf, $G(z)$ to derive the inverse transform function.

$$\begin{aligned}
G(z) &= \int_{-z}^{z}\int_{-z}^{z} \frac{q-1}{\pi d^{1-q}}(x^2+y^2+d)^{-q} dxdy \\
&= \frac{q-1}{\pi d^{1-q}} \int_0^{2\pi}\int_0^{z} (r^2+d)^{-q} r dr d\theta \\
&= \frac{q-1}{\pi d^{1-q}} \frac{1}{2(1-q)} \int_0^{2\pi} (z^2+d)^{1-q} - d^{1-q} d\theta \\
&= \frac{q-1}{\pi d^{1-q}} \frac{2\pi}{2(1-q)}((z^2+d)^{1-q} - d^{1-q}) \\
&= \frac{(z^2+d)^{1-q}}{-d^{1-q}} - \frac{d^{1-q}}{-d^{1-q}} \\
&= \frac{(z^2+d)^{1-q}}{-d^{1-q}} + 1
\end{aligned}$$

We draw $u \sim Unif(0,1)$ and solve $G(G^{-1}(u)) = u$, to find $G^{-1}(u)$, the inverse transform function that will draw a sample from the desired pdf $g(x,y)$.

$$\begin{aligned}
\frac{(G^{-1}(u)^2+d)^{1-q}}{-d^{1-q}} + 1 &= u \\
(G^{-1}(u)^2+d)^{1-q} &= (u-1)(-d^{1-q}) \\
G^{-1}(u)^2 &= [(u-1)(-d^{1-q})]^{1/1-q} - d \\
G^{-1}(u) &= \sqrt{[(u-1)(-d^{1-q})]^{1/1-q} - d}
\end{aligned}$$

This is the uth percentile of the desired spatial pdf.

6.2 Derivation of branching likelihood

This can be derived directly from Ogata (1998)'s likelihood function:

$$\begin{aligned}
L(Y|\theta) &= \prod_{i=1}^{n} \lambda(t_i, x_i, y_i, M_i|\mathcal{H}_t, \theta) \exp(-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \lambda(t,x,y,M|\theta,\mathcal{H}_t) dtdxdy) \\
&= \prod_{i=1}^{n} \Big(\mu_{ST} + \sum_{j:t_j<t} k(M_j)h(t_j)g(x-x_j, y-y_j)\Big) \exp\Big(-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \mu_{ST} + \sum_{j:t_j<t} k(M_j)h(t_j)g(x-x_j,y-y_j) dtdxdy\Big)
\end{aligned}$$

Say we have $|S_0|$ background and $n - |S_0|$ triggered events in the catalog of size n. The background events are generated by a homogenuous Poisson process with intensity $\lambda(t|\mathcal{H}_t) = \mu_{ST}$. So,

$$
\begin{aligned}
L(Y|\theta, S_0) &= \prod_{i=1}^{|S_0|} \mu_{ST} \exp\Big(- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \mu_{ST} dt dx dy\Big) \\
&= \prod_{i=1}^{|S_0|} \mu_{ST} \exp(-\mu_{ST} T|A|) \\
&= \exp(-\mu_{ST} T|A|) \prod_{i=1}^{|S_0|} \mu_{ST} \\
&= \exp(-\mu_{ST} T|A|) \mu_{ST}^{|S_0|},
\end{aligned}
$$

where $|A|$ is the area of the spatial region being modelled.

Now each of the events triggered by earthquake j (those in set S_j) are generated by a time-inhomogenuous Poisson process with $\lambda(t|\mathcal{H}_t) = k(M_j)h(t-t_j)g(x-x_j, y-y_j)$. The contribution to the likelihood function for each aftershock sequence is:

$$
\begin{aligned}
L(Y|\theta, S_j) &= \prod_{i=1}^{|S_j|} k(M_j)h(t_i-t_j)g(x_i-x_j, y_i-y_j) \exp\Big(- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} k(M_j)h(t_j)g(x-x_j, y-y_j) dt dx dy\Big) \\
&= k(M_j)^{|S_j|} \prod_{i=1}^{|S_j|} h(t_i-t_j)g(x_i-x_j, y_i-y_j) \exp\Big(- k(M_j) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} h(t-t_j)g(x-x_j, y-y_j) dt dx dy\Big) \\
&= k(M_j)^{|S_j|} \prod_{i=1}^{|S_j|} h(t_i-t_j)g(x_i-x_j, y_i-y_j) \exp\Big(- k(M_j) \int_{0}^{\infty} h(t-t_j) dt \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x-x_j, y-y_j) dx dy\Big). \\
&= k(M_j)^{|S_j|} \prod_{i=1}^{|S_j|} h(t_i-t_j)g(x_i-x_j, y_i-y_j) \exp\Big(- k(M_j) H(T-t_j) G(x-x_j, y-y_j)\Big).
\end{aligned}
$$

As in the above likelihood term, the spatial integral $G(x-x_j, y-y_j)$ is evaluated using my analytical+numerical solution.

Now we combine the background and each aftershock contribution to the likelihood:

$$
\begin{aligned}
L(Y|\theta, B) &= L(Y|\theta, S_0) \cdot L(Y|\theta, S_1) \cdot L(Y|\theta, S_2) \cdot \ldots \cdot L(Y|\theta, S_n) \\
&= L(Y|\theta, S_0) \prod_{j=1}^{n} L(Y|\theta, S_j) \\
&= \exp(-\mu_{ST} T|A|) \mu_{ST}^{|S_0|} \prod_{j=1}^{n} \Big(\exp\Big((-k(M_j)H(T-t_j)G(x-x_j, y-y_j)\Big) k(M_j)^{|S_j|} \prod_{t_i \in S_j} h(t_i-t_j)g(x_i-x_j, y_i-y_j)\Big) \\
&= \Big[\mu_{ST}^{|S_0|} \prod_{j=1}^{n} k(M_j)^{|S_j|} \prod_{t_i \in S_j} h(t_i-t_j)g(x_i-x_j, y_i-y_j)\Big] \cdot \\
&\quad \Big[\exp(-\mu_{ST} T|A|) \prod_{j=1}^{n} \exp\Big((-k(M_j)H(T-t_j)G(x-x_j, y-y_j)\Big)\Big] \\
&= SUM \cdot INT.
\end{aligned}
$$

6.3 Additional material related to evaluating the spatial integral

6.3.1 Radially approximating the spatial integral

The spatial component of INT_{trig} (spatial integral) is particularly complicated and is not directly evaluable. Several authors (Ogata (1998), Jalilian (2019)) advocate to deal with the spatial integral by radially approximating it across the catalog zone.

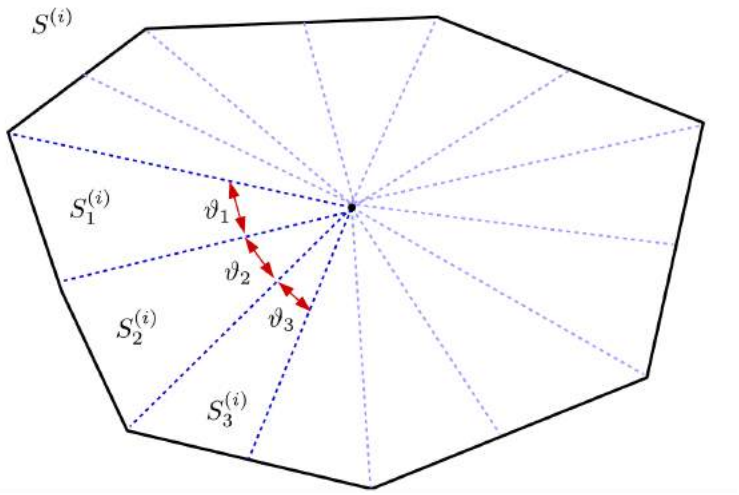

Figure 13: Jalilian (2019), Figure 1. Note that here, θ refers to the triangle's central angle but in my (and Ogata (1998)'s) notation, it is the azimuth angle.

For each mainshock j, partition the region $S^{(j)}$ into N subregions: $S_1^{(j)}, S_2^{(j)}, \ldots, S_N^{(j)}$. Do this by setting N knots along the boundary of $S^{(j)}$ and drawing a segment to connect the earthquake's location (the origin of $S^{(j)}$) to each knot. That is, we split $S^{(j)}$ into a set of N radial wedges (triangles) from the earthquake j in the center. See Figure 13 for a visual example (Jalilian 2019). Each of these triangles is defined radially by its radial segment r_k and azimuth angle θ_k $(k = 1, \ldots, N)$, and we order the N segments by θ, such that $0 < \theta_1 < \theta_2 < \ldots < \theta_N < 2\pi$.

Then, using Ogata (1998),

$$\int\int_{S^{(j)}} \frac{q-1}{\pi d^{1-q}} \frac{1}{(x^2+y^2+d)^q} dxdy \approx \sum_{k=1}^{N} G_k^{(i)}(x_i, y_i) \frac{\Delta_k}{2\pi},$$

where

$$G_k^{(i)} = \frac{\pi}{1-q}((r_k^2+d)^{1-q} - d^{1-q})$$
$$\Delta_k = \theta_{k+1} - \theta_k.$$

The first term in the sum evaluates the Utsu-Seki integral along the radial segment r_k and the second term scales it by a measure of the angular area of the wedge. This is the proportion of the circle (2π) corresponding to the difference between this central angle and the neighboring central angle. Ogata (1998) fails to specify the case of $k = N$, where $\Delta_N = \theta_{N+1} - \theta_N$ does not exist. We simply set $\Delta_N = \theta_N - \theta_{N-1}$.

The number of triangles N must be "sufficiently large for the accurate numerical approximation of the integral" (Ogata 1998). This is done by trial and error. Currently, we have set $N = 20$ but need to experiment with higher and lower Ns.

The above are standard practices for dealing with the spatiotemporal ETAS likelihood. We am adding one other piece to improve computation of this solution:

6.3.2 Direct evaluation of the spatial integral

The spatial integral $\int\int_{S^{(j)}} \frac{1}{(x^2+y^2+d)^q} dxdy$ may also be computed directly using numerical integral solvers in R. This would avoid some of the arbitrary choices that the radial approximation requires (eg, setting N, the number of knots).

There are many solutions in R for multidimensional integration. The R package cubature offers both deterministic and Monte Carlo (MC) direct integration. We examined two approaches:

- **hcubature**, which implements an h-adaptive cubature rule (McFee 1997). A cubature rule is a quadrature rule (integral evaluation procedure) applied to a multidimensional space. The h-adaptive version recursively partitions the integration area into smaller subareas, applying the same cubature rule to each, until convergence is achieved within each subarea. Eventually, the integral over the entire area converges. **Open question:** it is still unclear what the quadrature rule used by hcubature is. The documentation refers to Steven G. Johnson's C code, but the exact quadrature rule is unspecified.

- **vegas**, which implements a MC-based integration (Lepage 1980). The integral is computed at random points from a probability dist'n related to the integrand, drawn using importance sampling. The integral points are summed to estimate the full function integral.

I evaluated the spatial integral for all earthquakes in catalogs 3A and 3B under both hcubature and vegas; they yielded very similar values.

6.3.3 Speedier analytical+numerical evaluation of the spatial integral

We may also solve the spatial integral in one dimension analytically (say, x). This yields a function that is an integral over another dimension (y). This integral doesn't have a closed form solution but can be numerically evaluated as described above (1-dimensional numerical integration). Specifically, solving the (indefinite) spatial integral in one dimension yields

$$\begin{aligned}&\int (x^2+y^2+d)^{-q}dx\\&\quad= x(d+x^2+y^2)^{-q}(1+\frac{x^2}{d+y^2})^q\cdot {}_2F_1(\frac{1}{2},q;\frac{3}{2};\frac{-x^2}{y^2+d}),\end{aligned}$$

where ${}_2F_1$ is the hyper-geometric function that results when integrating a function with a negative power. Michael Barall evaluated this analytic solution in Mathematica and we confirmed it with https://www.wolframalpha.com/calculators/integral-calculator/.

We need the definite integral over our spatial dimensions, however:

$$\begin{aligned}&\int_{S_x^{(j)}} (x^2+y^2+d)^{-q}dx\\&\quad= \Big[x(d+x^2+y^2)^{-q}(1+\frac{x^2}{d+y^2})^q\cdot {}_2F_1(\frac{1}{2},q;\frac{3}{2};\frac{-x^2}{y^2+d})\Big]\Big|_{S_{xlower}^{(j)}}^{S_{xupper}^{(j)}},\end{aligned}$$

where $S_x^{(j)}$ is notation for the x-boundaries for box $S^{(j)}$ centered around mainshock j. In my simulated catalogs, the spatial regions S are all squares: $[250,750]^2$ (target) and $[0,1000]^2$ (full=target+auxiliary). So for target catalogs, the definite integral for mainshock location (x_j,y_j) is

$$\begin{aligned}&\Big[(750-x_j)(d+(750-x_j)^2+y^2)^{-q}(1+\frac{(750-x_j)^2}{d+y^2})^q\cdot {}_2F_1(\frac{1}{2},q;\frac{3}{2};\frac{-(750-x_j)^2}{y^2+d})\Big]-\\&\Big[(250-x_j)(d+(250-x_j)^2+y^2)^{-q}(1+\frac{(250-x_j)^2}{d+y^2})^q\cdot {}_2F_1(\frac{1}{2},q;\frac{3}{2};\frac{-(250-x_j)^2}{y^2+d})\Big]\end{aligned}$$

We can now rewrite the two-dimensional spatial integral as:

$$\begin{aligned}&\int_{S_y^{(j)}}\int_{S_x^{(j)}} (x^2+y^2+d)^{-q}dxdy\\&=\int_{S_y^{(j)}} x(d+x^2+y^2)^{-q}(1+\frac{x^2}{d+y^2})^q\cdot {}_2F_1(\frac{1}{2},q;\frac{3}{2};\frac{-x^2}{y^2+d})dy,\end{aligned}$$

and when applying to the target catalogs, this becomes

$$\begin{aligned}\int_{250-y_j}^{750-y_j}&\Big[(750-x_j)(d+(750-x_j)^2+y^2)^{-q}(1+\frac{(750-x_j)^2}{d+y^2})^q\cdot {}_2F_1(\frac{1}{2},q;\frac{3}{2};\frac{-(750-x_j)^2}{y^2+d})\Big]-\\&\Big[(250-x_j)(d+(250-x_j)^2+y^2)^{-q}(1+\frac{(250-x_j)^2}{d+y^2})^q\cdot {}_2F_1(\frac{1}{2},q;\frac{3}{2};\frac{-(250-x_j)^2}{y^2+d})\Big]dy.\end{aligned}$$

Catalog	true d	d sequence	true q	q sequence
A	1	from=0.3, to=1.75, by=0.05	2	from=1.25, to=2.75, by=0.05
B	1	from=0.3, to=1.75, by=0.05	1.5	from=1.05, to=2.25, by=0.05
C	0.1	from=0.025, to=0.175, by=0.005	2	from=1.25, to=2.75, by=0.05
D	0.1	from=0.025, to=0.175, by=0.005	1.5	from=1.05, to=2.25, by=0.05
PNW	-	from=0.3, to=2.0, by=0.08	-	from=1.2, to=1.6, by=0.04

Table 16: d/q grids for pre-computed G evaluations.

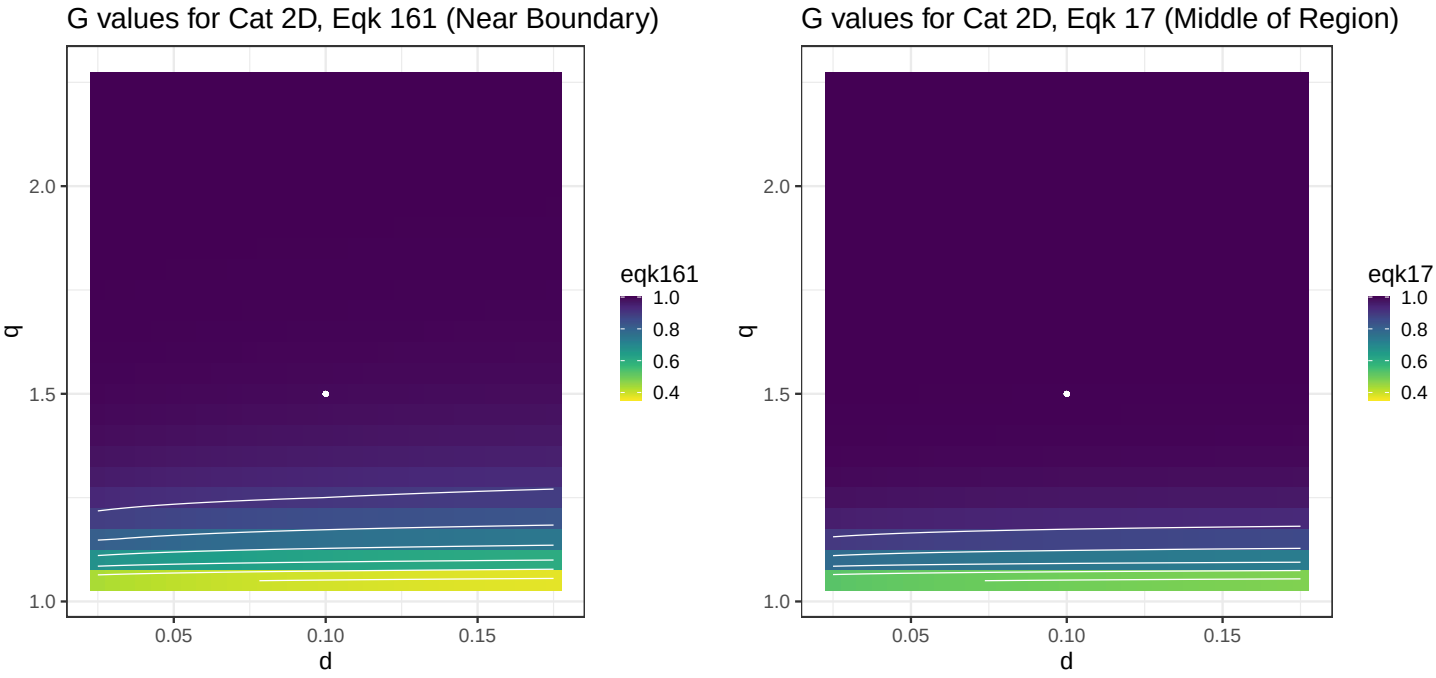

Figure 14: G surfaces for varying d and q around the true values for Catalog 2D. The integral values for an earthquake near the border (left) are smaller than the values for an earthquake near the center of the region (right). The values increase for increasing q, but the rate of increase gets smaller for q values above the truth. White contours show increase in 0.1 steps of integral value.

This expression can be evaluated by 1-dimensional numerical integration. The standard approach is to integrate via some quadrature rule with the most common one being Gaussian quadrature. We am using a version of this called Gaussian-Kronrod quadrature (in R's function "integrate") which has better accuracy for higher order functions. There are other options for the numerical procedure (ie, trapezodial quadrature, Clenshaw-Curtis quadrature) that we could experiment with, if needed.

I confirmed that the spatial integral values produced by this analytical+numerical solution were very close to those from the preceding 2D numerical approximation, for all earthquakes in several simulated catalogs.

6.3.4 Tables and figures for d/q grids for evaluating the spatial integral

In order to save computation in evaluating the ETAS branching likelihood as parameter posteriors are sampled, we pre-compute G, the spatial integral, for each earthquake, under a grid of d and q values given in Table 16. The surfaces of G values for two earthquakes in a synthetic catalog are given in Figure 14.

6.4 Derivation of branching likelihood

We can directly derived the branching likelihood of Ross [2021] from Ogata [1998]'s likelihood function:

$$L(Y|\theta) = \prod_{i=1}^{n} \lambda(t_i, x_i, y_i, M_i|\mathcal{H}_t, \theta) \exp(-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \lambda(t, x, y, M|\theta, \mathcal{H}_t) dt dx dy)$$
$$= \prod_{i=1}^{n} \Big(\mu_{ST} + \sum_{j:t_j<t} k(M_j)h(t_j)g(x-x_j, y-y_j)\Big) \exp\Big(-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \mu_{ST} + \sum_{j:t_j<t} k(M_j)h(t_j)g(x-x_j, y-y_j) dt dx dy\Big)$$

Say we have $|A_0|$ background and $n - |A_0|$ triggered events in the catalog of size n. The background events are generated by a homogenuous Poisson process with intensity $\lambda(t|\mathcal{H}_t) = \mu_{ST}$. So,

$$\begin{aligned} L(Y|\theta, A_0) &= \prod_{i=1}^{|A_0|} \mu_{ST} \exp\Big(-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \mu_{ST} dt dx dy\Big) \\ &= \prod_{i=1}^{|A_0|} \mu_{ST} \exp(-\mu_{ST} T|S|) \\ &= \exp(-\mu_{ST} T|S|) \prod_{i=1}^{|A_0|} \mu_{ST} \\ &= \exp(-\mu_{ST} T|S|) \mu_{ST}^{|A_0|}, \end{aligned}$$

where $|S|$ is the area of the spatial region being modelled.

Now each of the events triggered by earthquake j (those in set A_j) are generated by a time-inhomogenuous Poisson process with $\lambda(t|\mathcal{H}_t) = k(M_j)h(t-t_j)g(x-x_j, y-y_j)$. The contribution to the likelihood function for each aftershock sequence is:

$$\begin{aligned} L(Y|\theta, A_j) &= \prod_{i=1}^{|A_j|} k(M_j)h(t_i-t_j)g(x_i-x_j, y_i-y_j) \exp\Big(-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} k(M_j)h(t_j)g(x-x_j, y-y_j) dt dx dy\Big) \\ &= k(M_j)^{|A_j|} \prod_{i=1}^{|A_j|} h(t_i-t_j)g(x_i-x_j, y_i-y_j) \exp\Big(-k(M_j)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} h(t-t_j)g(x-x_j, y-y_j) dt dx dy\Big) \\ &= k(M_j)^{|A_j|} \prod_{i=1}^{|A_j|} h(t_i-t_j)g(x_i-x_j, y_i-y_j) \exp\Big(-k(M_j)\int_{0}^{\infty} h(t-t_j) dt \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x-x_j, y-y_j) dx dy\Big). \\ &= k(M_j)^{|A_j|} \prod_{i=1}^{|A_j|} h(t_i-t_j)g(x_i-x_j, y_i-y_j) \exp\Big(-k(M_j)H(T-t_j)G(x-x_j, y-y_j)\Big). \end{aligned}$$

As in the above likelihood term, the spatial integral $G(x-x_j, y-y_j)$ is evaluated using my analytical+numerical solution.

Now we combine the background and each aftershock contribution to the likelihood:

$$
\begin{aligned}
L(Y|\theta,B) &= L(Y|\theta,A_0)\cdot L(Y|\theta,A_1)\cdot L(Y|\theta,A_2)\cdot \ldots \cdot L(Y|\theta,A_n)\\
&= L(Y|\theta,A_0)\prod_{j=1}^{n} L(Y|\theta,A_j)\\
&= \exp(-\mu_{ST}T|S|)\mu_{ST}^{|A_0|}\prod_{j=1}^{n}\Big(\exp\Big((-k(M_j)H(T-t_j)G(x-x_j,y-y_j)\Big)k(M_j)^{|A_j|}\prod_{t_i\in A_j} h(t_i-t_j)g(x_i-x_j,y_i-y_j)\Big)\\
&= \Big[\mu_{ST}^{|A_0|}\prod_{j=1}^{n} k(M_j)^{|A_j|}\prod_{t_i\in A_j} h(t_i-t_j)g(x_i-x_j,y_i-y_j)\Big]\cdot\\
&\quad \Big[\exp(-\mu_{ST}T|S|)\prod_{j=1}^{n}\exp\Big((-k(M_j)H(T-t_j)G(x-x_j,y-y_j)\Big)\Big]\\
&= SUM\cdot INT.
\end{aligned}
$$

6.5 Derivations of log posterior density functions

Recall that posteriors for a given parameter pair are conditional on fixed values for the other ETAS parameters. In other words, for example. $p(K,\alpha) = p(K,\alpha|\mu,c,p,d,q)$. We remove this "conditional" part for notational ease.

$$
\begin{aligned}
\log(p(K,\alpha)) &\propto \log(\pi(K,\alpha)) + l(Y|K,\alpha,B)\\
&= 0 + (\sum_{j=1}^{n}|A_j|\log(k(M_j))) + INT_{trig}.\\
&= \Big(\sum_{j=1}^{n}|A_j|\log(K\exp(\alpha(M_j-M_0)))\Big) + INT_{trig}.
\end{aligned}
$$

$$
\begin{aligned}
\log(p(c,p)) &\propto \log(\pi(c,p)) + l(Y|c,p,B)\\
&= 0 + \sum_{j=1}^{n}\sum_{t_i\in A_j}\log(h(t_i-t_j)) + INT_{trig}\\
&= \Big(\sum_{j=1}^{n}\sum_{t_i\in A_j} -p(\log(t_i-t_j)+c)\Big) + INT_{trig}.
\end{aligned}
$$

$$
\begin{aligned}
\log(p(d,q)) &\propto \log(\pi(d,q)) + l(Y|d,q,B)\\
&= 0 + \sum_{j=1}^{n}\sum_{t_i\in A_j}\log(g(x_i-x_j,y_i-y_j)) + INT_{trig}\\
&= \Big(\sum_{j=1}^{n}\sum_{t_i\in A_j}\log(q-1) - \log(\pi) + (q-1)\log(d) + (-q(\log((x_i-x_j)^2+(y_i-y_j)^2+d))\Big) + INT_{trig}.
\end{aligned}
$$