Get Excited: Exploring Spatiotemporal Hawkes Processes
A Case Study of Earthquakes in Western North America

AI Statement: Generative AI tools were not used in this work or the preparation of this submission.

Abstract

Hawkes (1971) introduced the concept of the Hawkes process, a point process that captures self-exciting behavior along with a background rate of events. The Hawkes process has since been extended to the spatiotemporal realm, allowing the prediction of event rates as a function of space, time, and event marks. One of the primary areas of study for these processes has been seismic phenomena. Earthquake occurrences exhibit strong clustering in both space and time, motivating the use of Hawkes models. In this study, we analyze United States Geological Survey data to investigate whether a spatiotemporal Hawkes process adequately captures observed seismic patterns. We fit several spatiotemporal Hawkes process models using multiple estimation approaches. Of these models and methods, we find that the Gaussian Epidemic Type Aftershock Sequence (ETAS) model is the best performer. Diagnostic analyses reveal that the fitted Gaussian ETAS model captures key features of earthquake triggering dynamics, and also display limitations of the model. These findings illustrate the effectiveness of Hawkes-type models for representing earthquake occurrences and provide a basis for evaluating spatiotemporal self-exciting behavior in other event-driven systems.

Contents

1. Introduction

Analyzing and understanding data over a spatiotemporal domain is often useful due to its relevance in connection to phenomena such as global temperatures, animal or human movements, and climate patterns. However, one may be interested in phenomena where the spatiotemporal location of events is the data itself. Take, for instance, the spatiotemporal locations of volcanic eruptions, bird spottings or earthquakes. Spatiotemporal point processes help to characterize and model these types of occurrences as a natural generalization of Poisson processes.

Many phenomena cannot be modeled with a homogenous point process due to self-dependence. Certain spatiotemporal processes depend upon the history of the process itself. Examples of point processes that include self-dependence are earthquakes, disease cases, and crime (Park et al., 2021; Reinhart, 2018). The occurrence of an earthquake would intuitively increase the likelihood of other seismic events (aftershocks) in a nearby area for a short while. A case of an infectious illness is easily spread to others nearby, creating exponential growth. In instances of low-level offenses, such as speeding, a person charged with that offense is less likely to commit it again in the near future (Laub et al., 2022).

The latter example is a "self-regulating" process, wherein the instance of an event measured by the point process depresses the probability of the event occurring soon after. In contrast, the former pair can both be described as "self-exciting". These processes are more likely to occur in rapid succession. This paper is primarily concerned with self-exciting point processes, which can be described by a Hawkes process.

Alan Hawkes, a British statistician, developed what is now know as the Hawkes process in the early 1970s (Hawkes, 1971). Hawkes sought to create a better model of point processes, as there was little literature concerning self-dependent processes at the time. Hawkes initially approached the problem with a motivation he described as purely mathematical, but the possibility of application was always understood, especially in the seismic and epidemic realms (Hawkes & Chen, 2021). Today, Hawkes' process is widely used in various fields such as in criminology (Park et. al. 2019) and natural science (Xu et al., 2023). Spatiotemporal Hawkes process can even be used for generating predictions of a spike in drug overdoses in a specific geographical area (Chiang, 2022).

For this paper's part, in order to explore spatiotemporal Hawkes processes, we work by way of an example analysis of earthquake data from the United States Geological Survey. We begin by exploring the requisite mathematics and statistics to develop an understanding of spatiotemporal Hawkes process before exploring visualization techniques with the data. Then, we explore how spatiotemporal Hawkes process models can be specified and estimated through theory and practice, before discussing model diagnostics relevant to this class of point processes. Finally, we apply these methods to the earthquake data and examine the results

2. Point Processes and Hawkes Processes

To understand spatiotemporal Hawkes processes, we first consider univariate point processes before defining a univariate Hawkes process. These simpler processes generalize neatly to the spatiotemporal domain and further to marked spatiotemporal Hawkes processes. We begin with the following definition.

Definition 2.1 (Univariate Point Process, (Laub et al., 2022)): Let $\{\boldsymbol{T}\} = \{T_1, T_2, ...\}$ be a sequence of random variables such that each T_n has a sample space of $[0, \infty)$. If $P(0 \leq T_1 \leq T_2 \leq ...) = 1$, and the probability of observing infinite events in a finite period of time is zero, we call $\{\boldsymbol{T}\}$ a point process.

Aa univariate point process is recording the times of occurrence of some event with each successive event time necessarily not occurring before the previous event. The observed values of $\{T_1, T_2, \ldots\}$ simply are the times which the viewed event occurred. Alternatively, a point process can be viewed as a counting process, $N(u)$, which outputs the number of events that have occurred prior to time u. As one may suspect, the counting process is simply a step function any univariate point process can be viewed as a univariate counting process and vice versa. We visualize the connection of a point process and counting process in the following example.

Example (Point Process): Consider observing the arrival times of patrons to a coffee shop over the time interval $[0, U]$. Then the counting process, $N(u)$, describes the number of visitors before time u and the point process $\{\boldsymbol{T}\}$ records the arrival times, $\{T_1, T_2, \ldots\}$, of the visitors. Let customers arrive at times $T_1, T_2, \ldots, T_5$. We visualize this in Figure 1.

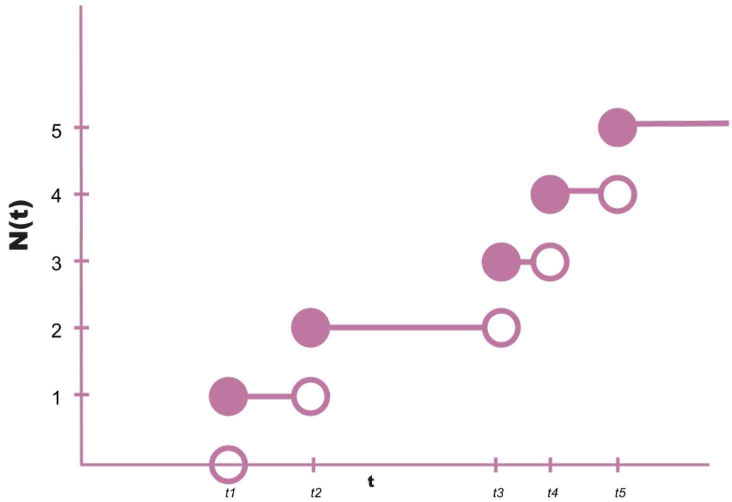


Figure 1: Visualizing a point process.

We also define the history of a point process (or counting process) up to a certain time t, a concept crucial to the Hawkes process and expanding the applicability of these processes. The history records all of the information contained within a point process up until a certain time, as explained in the following definition.

Definition 2.2 (History of Point Process, (Bernabeu et al., 2024)): Let $\{\boldsymbol{T}\} = \{T_1, T_2, \ldots\}$ be a univariate point process with values in the interval $[0, \infty)$. Given some $t \in [0, \infty)$ we define $\mathcal{H}_t$ as the history of events to time t. That is $\mathcal{H}_t = \{T_k : T_k \in [0, t)\}$.

Point processes can also be thought of as characterizing the time until the next event arrival based upon the history of the process itself. This idea naturally extends itself to the notion of "self-exciting" and "self-regulating" processes as the occurrence of an event could increase or decrease the chance the next arrival time occurs in some interval.

The cumulative distribution for the next arrival time, T_{k+1} in a point process, $\boldsymbol{T}$ with history $\mathcal{H}_{t_k}$ can then be written as

$$F\left(t|\mathcal{H}_{t_k}\right) = \int_{t_k}^{t} \mathbb{P}\left(T_{k+1} \in [s, s+ds]|\mathcal{H}_{t_k}\right) ds.$$

Unfortunately, this conditional distribution and associated joint probability density function are often very difficult to work with. Therefore, we turn to the conditional intensity function, which we now define.

Definition 2.3 (Univariate Conditional Intensity Function, (Bernabeu et al., 2024)): Let $N(t)$ be a counting process and $\mathcal{H}(t)$ its associated histories. If there exists a well-defined function $\lambda(t)$ such that

$$\lambda(t) = \lim_{h \to 0} \frac{\mathbb{E}[N(t+h) - N(t) \mid \mathcal{H}(t)]}{h},$$

we say that $\lambda(t)$ is the conditional intensity function of N.

One can conceptualize the conditional intensity function as asking, "given the history of the process, at what rate are arrivals expected to occur in the next infinitesimal interval?". This corresponds to the cumulative distribution for the next arrival time, but in practice is much easier to work with. One can also see how the definition appears similar to that of a derivative, which illuminates how $\lambda(t)$ captures information about event arrival rates. In practice, the conditional intensity function must be defined such that infinite events cannot occur. The conditional intensity function nicely characterizes different types of point processes as seen in the examples below.

Example (Self-Exciting Point Process): Figure 2 showcases what a conditional intensity function may look like for a self-exciting point process. A specific instance that may fit this $\lambda(t)$ is the occurrences of some crimes. In some cases, the occurrence of a crime can cause the chance for retaliatory crimes to increase in the following time period, which is followed by a decline to baseline levels of likelihood. (Zhuang & Mateu, 2018).

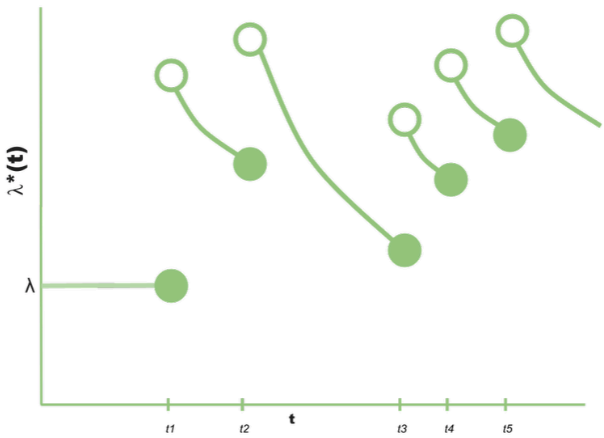


Figure 2: Conditional Intensity Function for a self-exciting point process.
Note: λ represents the background rate while $\lambda^*(t)$ is the conditional intensity function at time t.

Example (Self-Regulating Point Process): Consider observing an occurrence of a student turning in an assignment late. After seeing their professor's disappointment or the associated grade penalty, the student would be less likely to submit an assignment late for some period of time. However, as the time since the last late assignment lengthens, the student may be more likely to submit an assignment late again. This corresponds to the conditional intensity function visualized in Figure 3.

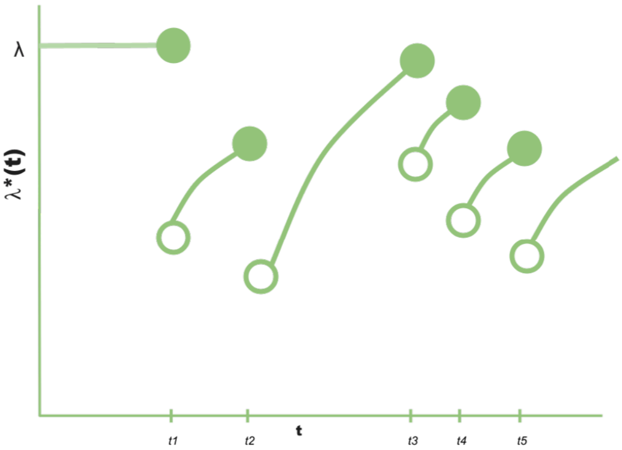


Figure 3: A self-regulating point process.
Note: λ represents the background rate while $\lambda^*(t)$ is the conditional intensity function at time t.

We formally define a self-regulating point process as one in which an event arrival causes the conditional intensity function to decrease, and a self-exciting point process as one in which an event causes an increase to the conditional intensity function. In Figures 2 and 3, the conditional intensity functions decay back to some base rate λ after the occurrence of an event. This can be intuitive if one considers how events further in the past often have less bearing on the future than more recent events.

Univariate Hawkes processes are specific instances of self-exciting point processes that can be defined completely by their conditional intensity function. We now specify the formal definition of a univariate Hawkes process.

Definition 2.4 (Univariate Hawkes process, (Bernabeu et al., 2024)): Let N be a univariate point process. If the process can be completely defined by its conditional intensity function $\lambda(t)$, we say that N is a Hawkes process.

For a Hawkes process we can write a general conditional intensity function as

$$\lambda(t) = \mu(t) + \sum_{i:t_i<t} v(t - t_i).$$

This can naturally be thought of as some background rate denoted above as $\mu(t)$, along with a function $v(t)$ that controls the excitation process. This excitement function $v(t)$ is alternatively called the triggering function or triggering kernel. When we take the summation of $v(t - t_i)$ over all previous events, we are adding up all of the contributions to the conditional intensity function from past events. For earthquake modeling, this is analogous to understanding how each prior earthquake increase the chance of seeing another seismic event in a given period of time. Temporal Hawkes processes have often been used for high-frequency economic dynamics as well as immigration-birth models (Bacry & Muzy, 2013; Laub et al., 2025).

The univariate Hawkes process also must meet some criteria in order to be valid. Specifically, $\lambda(t)$ must be non-negative and the integral of $v(t)$ over the interval $[0, \infty]$ must be less than 1 to avoid infinitely many events occurring. Moreover, this method of defining a Hawkes process assumes for the sake of simplicity that two events can not occur simultaneously at the same time t. Some definitions of the Hawkes process address this by allowing a very small chance of observing multiple events in a short time interval, where the chance becomes 0 as the length of interval approaches zero. We now explore an example of what a Hawkes conditional intensity function may look like algebraically (Bernabeu et al., 2024; Laub et al., 2022).

Example (Hawkes Conditional Intensity Function): Consider $\lambda(t) = \mu(t) + \sum_{i:t_i<t} v(t - t_i)$, the conditional intensity function for some univariate Hawkes process. We can let $\mu(t) = c$ for some $c \in \mathbb{R}^+$ and $v(t) = \alpha e^{-\beta}$ for some $\alpha, \beta \in \mathbb{R}^+$ to see

$$\lambda(t) = c + \sum_{i:t_i<t} \alpha e^{-\beta(t-t_i)}.$$

The above example showcases a Hawkes conditional intensity function that has a constant background rate c which governs the chance of an event absent of any impact from the past, and an exponentially decaying excitation function $v(t)$. A realization of this process could look like the crime example discussed previously and visualized in Figure 2. We discuss more possible conditional intensity functions, especially those pertinent to earthquake models and spatiotemporal processes in Section 4.

The Hawkes process can be extended from the temporal case to the spatiotemporal domain. The spatiotemporal case is analogous to the temporal one with each observation also including information about its location in space.

Definition 2.5 (Spatiotemporal Hawkes Process, (Bernabeu et al., 2024)): Let *K* be a spatiotemporal Hawkes process. K has a history $\mathcal{H}$ made up of observations at locations $s_i = \{s_1 = (x_n, y_1)....s_n = (x_n, y_n)\}$ and times $t_i = \{t_1, t_2...t_n\}$. *K* has a conditional intensity function

$$\lambda(s, t|\mathcal{H}_t) = \mu(s, t) + \sum_{\{i:t_i<t\}} g(s - s_i, t - t_i).$$

In the above equation, the triggering function g is commonly assumed to be separable by time and space, such that $g(s, t) = f(s)h(t)$ as a simplifying assumption.

As one may suspect, since we were able to generalize univariate Hawkes processes to higher dimensions, it is possible to extend them even further. This is most commonly done by 'marking' the spatiotemporal point process. Marking can be considered as taking an observed event with time *t* and location s and adding another dimension - the mark m. Now, an observed point is represented as $x = (t, s, m)$.

A relevant example of a marked Hawkes process is the study of earthquakes. Each seismic event possesses a spatial location *s* (its origin) and a time of occurrence *t*. But, in addition, all earthquakes have a marker *m*, the magnitude of the earthquake. The Epidemic-Type Aftershock Sequence (ETAS), which has been applied to earthquakes and other fields, is perhaps the consummate example of such a process, and is due to be explored in greater depth later in this paper (Ogata, 1998).

Examples are not limited to earthquakes. Marks can consist of the size of an animal that has been spotted or the value of a financial trade. Marks can also be categorical - the crime committed, the strain of illness suffered, or the result of a shot on goal. We conclude this section by defining a general marked spatiotemporal Hawkes process.

Definition 2.6 (Marked Spatiotemporal Hawkes Process): Let L be a marked spatiotemporal Hawkes process. L has a history $\mathcal{H}$ with a set of observations $\{x_i = (t_i, s_i, m_i) \in \mathcal{H} : t_i < t\}$. L has a conditional intensity function

$$\lambda(s, t, m|\mathcal{H}_t) = \mu(s, t, m) + \sum_{\{i:t_i<t\}} g(s - s_i, t - t_i, m_i).$$

3. Data Exploration and Visualization

For our example analysis, we collect 17,438 observations of earthquakes recorded by the United States Geological Survey (USGS). Each observation contains the spatial coordinates of the earthquake in latitude and longitude as well as the time (UTC) of the event. Additionally, information about the depth of the earthquake and its associated magnitude are recorded. The data range temporally from January 1st, 2020 through December 31st, 2025 and spatially between 24.6° and 50° latitude and between $-125.0°$ and $-65.0°$ longitude. Similar to the USGS, we call this rectangular region the conterminous United States. The minimum magnitude threshold for recording an earthquake within this data set was 2.5 on the Moment Magnitude scale (U.S. Geological Survey, 2026).

The USGS also tracks if the recorded seismic events were caused by non-geological sources, such as mining explosions, ensuring confidence that these observations are actually earthquakes or associated aftershocks. It should be noted that, according to the USGS, earthquakes of different magnitudes are best measured via different methods. Each earthquake notes the method of measurement used, but all events are on the same scale.

We chose seismic data in part because it is a large reason why many of the models we explore have been created and popularized, and in part because seismic data is important to study and understand. The ability to predict and understand earthquakes and aftershocks is especially important. It allows us to properly construct and regulate housing in earthquake-prone areas, and can help spark evacuations in situations where we may predict extreme aftershocks. The Seismological Society of America extols the study's virtues even in areas of communication - finding where fiber-optic cables can be placed such that they remain safe from fault lines (Seismological Society of America, 2025).

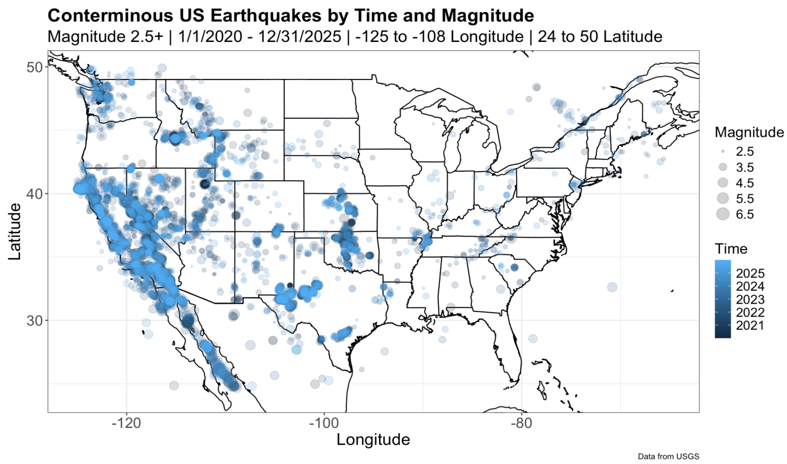


Figure 4: Spatiotemporal locations of all events in the data.

Figure 4 showcases the distribution of earthquakes and their respective magnitudes over time and space across the entire spatiotemporal domain of the data. Each dot corresponds to a seismic event plotted by its recorded longitude and latitude, with color representing the year of the event and dot size displaying magnitude, where larger magnitudes correspond to larger dots.

Based on the spatial distribution of the earthquakes in Figure 4, which displays clustering in the Western part of the continent, we elect now to explore in greater depth earthquakes in the Western region of North America. Specifically, we restrict our analysis to seismic events occurring west of $-108°$ longitude. Moreover, we opt to focus our study on events between June 1st 2021 and June 1st 2025 and on earthquakes greater than, or equal to, 3.5 in magnitude. These filters will aid in reducing computational intensity while also cutting out large swaths of the continent where earthquakes are uncommon. We now focus our exploratory data analysis on this reduced spatiotemporal region.

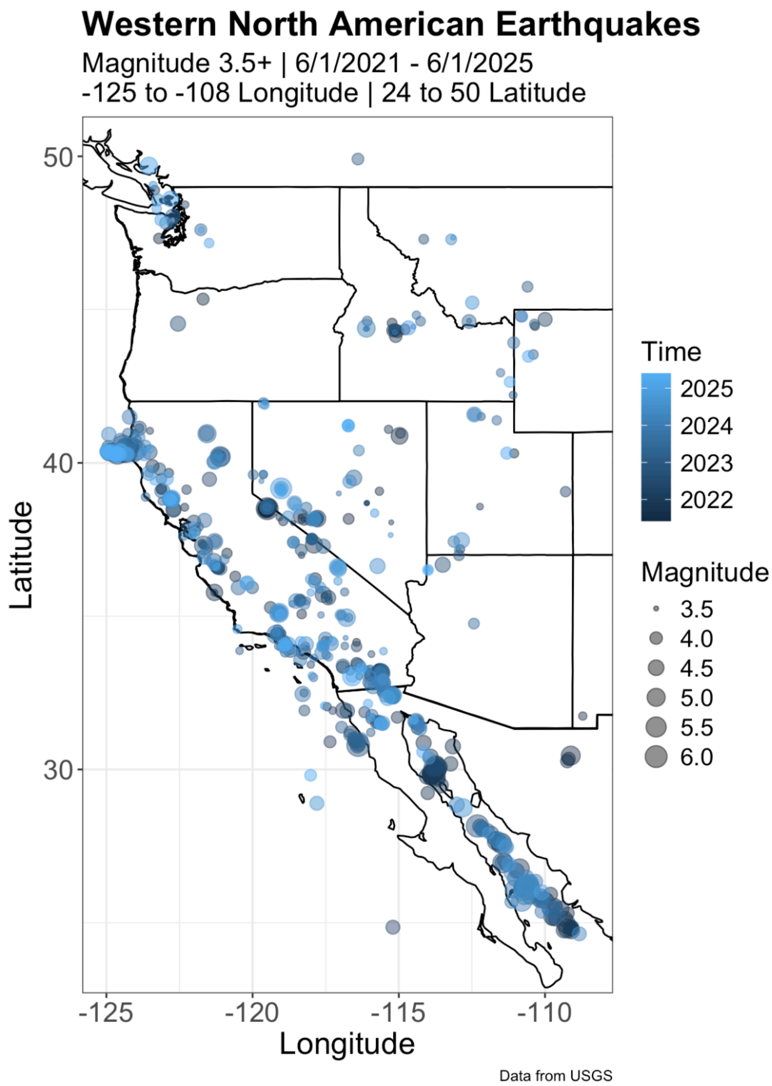


Figure 5: Spatiotemporal locations of the sub-setted data.

Figure 5 displays the spatial distribution of the sub-setted data. The plot clearly exhibits spatial patterns, with the majority of events clustering primarily in California, Nevada, and the Pacific Northwest. Driven by the San Andreas fault and other active faults, these regions are highly tectonically active (U.S. Geological Survey, n.d.).

The figure also shows the persistence of seismic events over space, providing a view of the spatial variability of earthquake occurrences and motivating the inclusion of a spatially varying background intensity in Hawkes process models. The presence of clustered events further suggests that earthquakes are not independent in space, providing qualitative support for modeling seismic activity as a spatiotemporal self-exciting process. We also visualize the spatial density of the events via contour lines in Section 8.1.

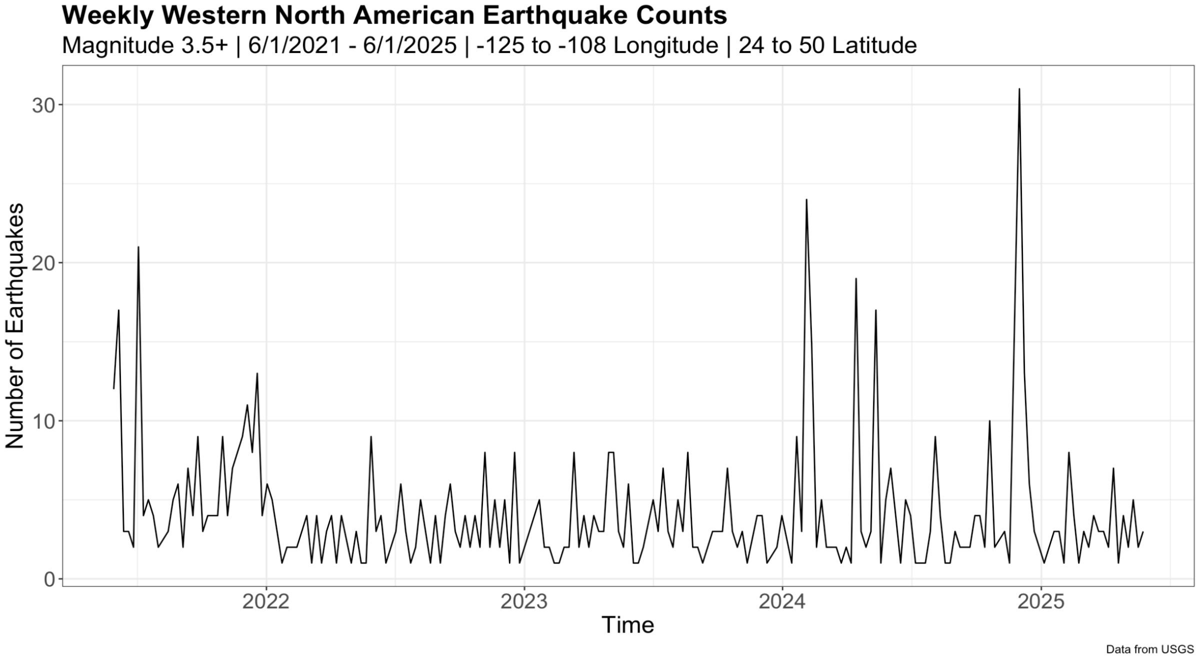


Figure 6: Earthquake counts by week in sub-setted data from June 1st, 2021 to June 1st, 2025.

Beyond spatial visualization, the temporal structure of earthquake occurrences provides strong evidence that the underlying process is self-exciting rather than memoryless. Figure 6 illustrates weekly earthquake counts from June 1st, 2021 to June 1st, 2025, revealing pronounced bursts of activity in which elevated event counts persist for short periods before gradually decaying to baseline. Such clustering behavior is inconsistent with a homogeneous Poisson process, which would produce relatively stable counts over time, but is characteristic of self-exciting dynamics in which events temporarily increase the likelihood of future events.

The USGS data also comes equipped with metrics displaying error of recorded depth of the seismic event and location variables. Error values appear to be relatively small and, for the purposes of this study, we assume there to be no error in the spatiotemporal event coordinates and the magnitude. Plots of error metrics and descriptions can be found in Section 8.2.

4. Methods

4.1. Specification and Estimation

In a Hawkes process modeling, one can apply parametric and non-parametric approaches to the specification and estimation step. The parametric approaches are often less computationally expensive, but force a model form to be selected. A non-parametric approach, therefore, is sometimes preferable as it does not require initial model form specification.

If we were to choose the parametric method, we must initially specify our model. This specification, as is typical, is not final, and can and should be revisited upon estimation. We must specify the background rate μ and the triggering function v. The background rate, in the original configuration of the Hawkes process, is assumed to be constant in time (Hawkes, 1971). There is a typical rate of occurrences, and the triggering function is the only piece that lends itself to heterogeneity in the process. In the earthquake case, Ogata recommends using a constant background rate when the number of events in the given time span is especially small (Ogata, 1999).

However, in many modern applications, the background rate is heterogeneous and time-dependent, or separated by time and space. In such a case, $\mu(s, t) = \mu_s(s)\mu_t(t)$ (Zhuang et al., 2006). In one case, Omi, Hirata, and Aihara discuss the background rate of trading patterns, and display that there is typically a bimodal "U" shaped pattern in background trades, with trade volume peaking at the beginning and end of a trading session (Omi et al., 2017). In this case, we would need to utilize a function to represent our background rate μ, rather than a constant.

The triggering function v also presents multiple options. As mentioned in Section 2, the triggering function is also separable by space and time. The "original" temporal triggering function was exponential decay, initially deployed by Hawkes as it simplified his theoretical derivations (Hawkes, 1971; Laub et al., 2022). This triggering kernel takes the form $v(t) = \alpha e^{-\beta(t)}$ where $\alpha, \beta > 0$. The parameter α can be understood as the instantaneous increase in intensity caused by an arrival to the system, while β is the rate at which the arrival's influence decays exponentially.

A commonly used alternative temporal function in seismic modeling is the inverse power-law kernel. The inverse power-law kernel is represented as $v(t) = \frac{k}{(c+t)^p}$. Here, k is analogous to α of exponential decay, controlling the magnitude of self-excitation. c prevents the kernel from blowing up at $t = 0$, t is of course time, and p can be interpreted similarly to β from before, determining the rate of decay in the conditional intensity function (Batra, 2025). In essence, a power-law kernel has a "long memory", with influence of events decaying more slowly, while exponential kernels see influence decay exponentially.

The spatial domain can also be modeled as possessing an exponential triggering function, which can be expressed as $f(s) = \beta e^{-\beta \, \|s - s_i\|}$. The spatial power-law can be expressed as $f(s) = \left(1 + \frac{\|s-s_i\|}{\gamma}\right)^{-q}$. Alternatively, the Gaussian function $f(s = x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{(x-x_i)^2}{2\sigma_x^2} - \frac{(y-y_i)^2}{2\sigma_y^2}\right)$ can be utilized, which assigns strong influence to nearby events and a squared exponential decrease in influence as distance increases (Bernabeu et al., 2024).

Batra compares power-law and exponential kernels with regard to modeling financial trades temporally. Here, he comes to the conclusion that the power law is superior in that context, and often provides additional flexibility in modeling real-world data (Batra, 2025). Additionally, the power-law kernel is more commonly used in seismic modeling, but that does not mean that it is always superior. Batra notes that an exponential kernel provides more simplicity and tractability in many cases. When modeling Hawkes processes, one should typically assess the performance of different types of kernels for best fit.

4.1.1. Likelihood

We now move to describing the methods we employ in estimating a Hawkes process for our data. We will begin with the most common method for self-exciting point processes, maximum likelihood. Maximum likelihood is theoretically performed in the same way as in any other scenario, by working to the log-likelihood and maximizing

$$\log(L(\theta)) = \sum_{i=1}^{n} \log \lambda(s_i, t_i \mid \mathcal{H}_t i)) - \int_T \int_W \lambda(s_i, t_i \mid \mathcal{H}_{ti}) \; ds \; dt \text{ (Bernabeu et al., 2024).}$$

For small datasets where the likelihood is computationally tractable, numerical evaluation of this log-likelihood can be utilized. This, however, can be challenging. Numerical evaluation is at times incredibly computationally expensive, and the log-likelihood can be nearly flat for swaths of the parameter space (Reinhart, 2018).

Given this, partial likelihood emerges as an alternative method. It is still difficult to compute analytically, but is much more convenient for computation. Here, we condition on locations s_i and t_i, and examine the log-likelihood for the observed time-ordered events, as shown below.

$$p_i = \frac{\lambda(s_i, t_i \mid \mathcal{H}_{ti})}{\int_W \lambda(s_i, t_i \mid \mathcal{H}_{ti}) ds}$$

$$L_{p(\theta)} = \sum_{i=1}^{n} \log p_i \text{ (Bernabeu et al., 2024)}$$

This partial likelihood maximization shares the general asymptotic properties of maximum likelihood estimators. However, it loses efficiency when compared to a full maximum likelihood estimator, and often provides wider confidence intervals. While certainly being more *feasible* than a maximum likelihood estimator, it is not always fully effective.

We implement this form of estimation to our data via the use of the 'stelfi' package (Jones-Todd & van Helsdingen, 2024). This package allows us to fit a model with a conditional intensity function defined for an event at location s and time t, with process history $\mathcal{H}_t$) as:

$$\lambda(s, t \mid \mathcal{H}_t) = \mu + \alpha \sum_{i:t_i<t} \exp(-\beta * (t - t_i)) f_X(s - s_i),$$

where α, β are unknown parameters, and f_X is the probability density function of $X \sim N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$ (Jones-Todd & van Helsdingen, 2024). The 'stelfi' framework is built on Hawkes process theory and is formulated as a Poisson process with a self-exciting triggering mechanism driven by the history of observed events. The conditional intensity at location (x, y) and time t depends on a constant background rate μ and a triggering component formed by summing the contributions of all previous earthquakes. Temporal triggering is modeled using an exponentially decaying kernel determined by β, a decay parameter controlling how quickly the influence of past events diminishes over time. The strength of the triggering is scaled by α, the productivity parameter, which controls the degree of clustering in the process. Spatial triggering is modeled using a bivariate Gaussian kernel, parameterized by spatial scale parameters that control the dispersion of triggered events around the initial event.

Due to the constant background of this model, we henceforth call this model the constant background model. One potential issue with this model for our data is the fact that the background rate is almost assuredly not constant across space (see Figure 5). Although there is a lack of flexibility in what this model can account for, it at least provides a starting point for trying to understand the complex dynamics of our seismic data.

4.1.2. Stochastic Declustering

As mentioned before, it is often preferred to fit some models non-parametrically. In other cases, a semi-parametric approach would be best. One way to do either is by stochastic declustering. There are many methods for stochastic declustering, but we shall explore the model-based version in depth. In this method, which was first brought forth by Zhuang, Ogata, and Vere-Jones in 2002 and elucidated by Renihart in 2018, the triggering function g is assumed to have a parametric form while the background rate μ is estimated non-parametrically, by determining background events from the data.

In this method, we begin by initializing the background rate with a crude assumption of the background rate $\hat{\mu}(s)$ being constant. We then continue by fitting the parameters of the conditional intensity function λ by using maximum likelihood, depending on need. Each event i is given a latent quantity u_i that indicates whether an event is part of the background or is triggered by another event. From there, we calculate the probability $\Pr(u_i \neq 0)$that each event i was triggered by another. These probabilities are then used to update the background intensity through kernel smoothing, with the bandwidth also dependent upon some chosen number of nearest-neighbors. That is, we form a new $\hat{\mu}^*$ with the estimator $\hat{\mu}^* = \frac{1}{T'} \sum_{i=1}^{n} 1 - \Pr(u_i \neq 0) k(s - s_i)$ where T' is the length of our observation period and k is the triggering kernel. Then, we compare our last estimate $\hat{\mu}$ and $\hat{\mu}^*$. If $\max_s |\hat{\mu} - \hat{\mu}^*| < \varepsilon$ for a tolerance ε that we choose we end the process. If not, we repeat the above steps until the maximum difference is below the tolerance threshold.

Finally, we thin the data. For each event, we classify it as a background event with probability $1 - \Pr(u_i \neq 0)$; otherwise we eliminate it as triggered by another event. After this step, only the background events remain, which allows us to estimate the background rate directly. The models fit via this method are described in the following subsection.

4.1.3. Epidemic Type Aftershock Sequence Modeling

We implement stochastic declustering via the 'ETAS' package (Jalilian, 2019). This package allows us to create a catalogue from our data set that includes all earthquakes with magnitude magnitude 3.5 and above. The events in the catalogue before the study began, and those outside the spatial domain, are called complementary events. The complementary events are included within the history of the model which help overcome edge issues as without outside events, it is difficult to discern which events at the edge of the spatiotemporal domain are background events or triggered events (Jalilian, 2019). However, these events are not directly considering when finding coefficient estimates in the stochastic declustering process. The catalogue specifically contains all events from the original data that were at least 3.5 in magnitude and occurred before June 1st, 2025. In Figure 7, we showcase the catalogue visualization provided by the 'ETAS' package. The events within the study period and region are in blue, while the complementary events are in gray.

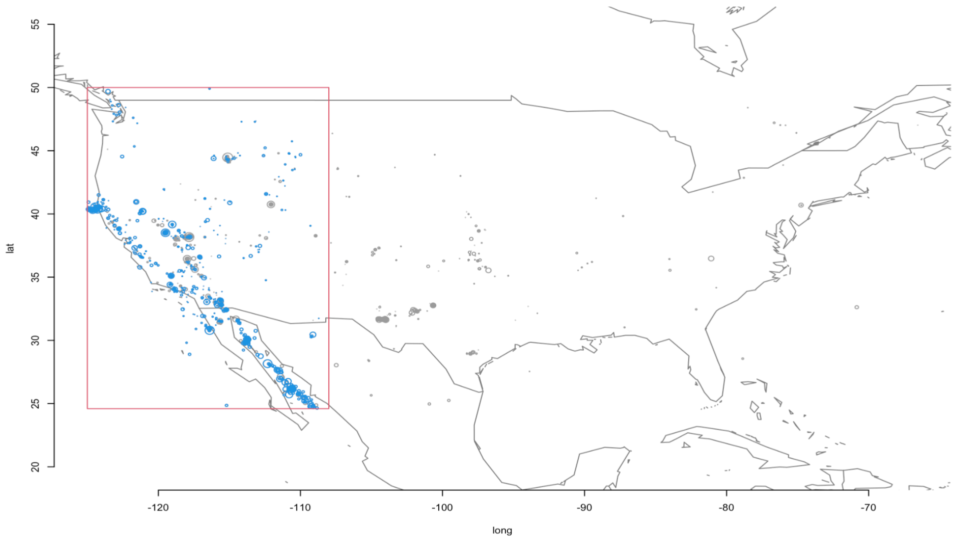


Figure 7: Plot of the created earthquake catalogue via the 'ETAS' package. The spatial region of interested is within the red box.

The 'ETAS' package is equipped with the ability to fit two different Epidemic Type Aftershock Sequence models. Emerged in the 1980s, these models were developed through the work of Yosihiko Ogata, and began as a statistical framework to describe the clustering of earthquakes without needing to manually differentiate aftershocks from mainshocks. The ETAS framework extended previous work on self-exciting point processes, including the Omori–Utsu law for aftershock decay and the Gutenberg–Richter magnitude–frequency relationship. The ETAS model, however, introduced a key idea that every earthquake can trigger subsequent earthquakes, which treated seismic activity as a self-exciting point process. Over time, ETAS became a cornerstone of earthquake forecasting and seismic analysis. Throughout the 1990s, the model was refined with improved estimation methods and extended from a purely temporal formulations to a full spatiotemporal framework that incorporates both spatial clustering and magnitude dependence. (Ogata, 2011).

Importantly, the conditional intensity function for an ETAS model incorporates magnitude, allowing the model to capture the intuition that larger earthquakes should have more aftershocks and a wider range of impact.

The conditional intensity function of these models is the product of a magnitude frequency function and the sum of a spatially varying background rate and a triggering function. It requires a minimum magnitude threshold m_0, which is 3.5 in our case. That is, for an event at location (x, y), time t and magnitude m with process history $\mathcal{H}_t$,

$$\lambda(t, x, y, m \mid \mathcal{H}_t) = \nu(m)(\mu u(x, y) + \sum_{i:t<t_i} k(m_i)g(t - t_i)f(x - x_i, y - y_i, m_i),$$

where $\mu > 0$ is the constant overall background rate, $u(x, y)$ is the varying background intensity that integrates to 1 over the spatial domain, $\nu(m) = \beta \exp(-\beta(m - m_0))$, is the probability density function of an event's magnitude, $k(m) = A \exp(\alpha(m - m_0))$ is the expected number of events triggered from an event of magnitude m, $g(t) = \frac{p-1}{c}\left(1 + \frac{t}{c}\right)^{-p}$ is the probability density function of the occurrence times of the triggered events, and $f(x, y, m)$ is the probability function of the locations of the triggered events, with unknown parameters $A, \alpha, c, \beta > 0$, and $p > 1$.

The estimated parameters each have clear interpretations within the ETAS framework. Parameter A controls the overall magnitude of the aftershock production, scaling the number of triggered events produced by past earthquakes. The magnitude sensitivity component α; it determines how strongly the magnitude of an earthquake influences its ability to produce aftershocks, with larger earthquakes triggering greater number of aftershocks. Temporal behavior is governed by the Omori-Utsu law parameters c and p. With c being the short-term time shift that stabilizes the decay immediately after an event, and the temporal decay parameter p shaping the long-term decay of triggered seismicity over time, with a larger p corresponding to a faster decay of the aftershock rate.

The Inverse Power ETAS model defines

$$f(x, y, m) = \frac{q-1}{\pi\sigma(m)}\left(1 + \frac{x^2 + y^2}{\sigma(m)}\right)^{-q}$$

where $\sigma(m) = D \exp(\gamma(m - m_0))$, x and y are the spatial offsets of the triggering event and $D, \gamma > 0$, and $q > 1$ are unknown parameters. That is, the locations of triggered events are assumed to follow a long-tailed Inverse Power law distribution centered at the parent earthquake. Based on the spatial kernel we dub this the Inverse Power model.

The parameters in the $f(x, y, m)$ can also be defined with q controlling the rate of spatial decay, with larger values corresponding to faster decay and tighter clustering, and D controlling the range of an earthquake's impact and γ fine tuning the spatial influence allowing larger magnitude earthquakes to have larger impact ranges. Larger values indicate increased influence of higher-magnitude earthquakes.

The second ETAS model we call the Gaussian model and it contains a slightly different spatial decay function where it assumes that the triggered events follow a light-tailed Gaussian distribution around the parent earthquake. It has the form:

$$f(x, y \mid m) = \frac{1}{2\pi\sigma(m)} \exp\left(-\frac{x^2 + y^2}{2\sigma(m)}\right).$$

The spatial influence decays exponentially with squared distance from the triggering event. The Gaussian kernel produces smoother and more localized clustering.

For both the Inverse Power and the Gaussian ETAS models, we fit two versions with differing degrees of smoothness dictated by the number of nearest neighbors used in bandwidth calculations. Specifically we consider 5 nearest neighbors as base smoothness, and 7 as the smoother model. In order to fit the models, we first supply an initial guess of parameter values to begin the stochastic delustering process. Each respective model converges to similar parameter estimates from a variety of parameter starting points.

There exist numerous alternative methods for estimation that we do not employ in this paper. Bayesian approaches, for example, utilize Markov Chains to estimate the conditional intensity. In some settings, machine learning has been used to model the process and has demonstrated competitive performance to other methods. There are also other approaches to stochastic declustering, including a model-independent form where the entirety of the conditional intensity function is estimated non-parametrically (Bernabeu et al., 2024; Reinhart, 2018; Zhuang et al., 2002).

4.2. Model Diagnostics

To assess the adequacy of our fitted models, we employ several diagnostic techniques. These diagnostics are designed to evaluate whether the estimated conditional intensity function successfully captures the spatiotemporal clustering present in the earthquake data, and whether any systematic structure remains unexplained after model fitting. If there exists structure, whether temporally or spatially, it is a sign that the model is inadequate.

The first technique we utilize is called thinning. Thinning provides a way to identify inconsistencies between the observed data and the fitted model by comparing observed event occurrences to those expected under the estimated conditional intensity. The goal of this process is to utilize the conditional intensity function and random number generation to be left with just the estimated background events in the data. If the model has a good fit, the leftover process should be a homogenous Poisson process (Bernabeu et al., 2024).

The procedure begins by computing the fitted conditional intensity function at each observed event. Events are then probabilistically retained with chance equal to $\frac{p}{\lambda(s,t)}$ where p is the minimum of the conditional intensity function over all events in the data, and $\lambda(s,t)$ is the estimated conditional intensity for the event. Events occurring in regions of high modeled intensity are hence more likely to be thinned. For the constant background model, Inverse Power ETAS model, and Gaussian ETAS model, we are able to utilize the estimated conditional intensity values for each event.

Under a correctly specified Hawkes process model, the thinned process should behave as a homogeneous Poisson process. This expectation provides a concrete diagnostic criterion. In particular, temporal residual plots should exhibit no systematic trends over time, instead fluctuating randomly around a constant level. Long runs of positive or negative residuals indicate that the model fails to adequately account for excitation or decay dynamics in the original process.

Similarly, inter-event times of the residual process should be exponentially distributed. (Bittelli et al., 2022). Persistent departures from these patterns indicate a lack of the ability for the model to account for the temporally exciting nature of the data.

Spatially, there should not be any clustering left within the thinned process. This can be examined visually but also via Ripley's K-function, as suggested by Reinhart. This function is able to analyze how many events occur near another event in a given range (Ripley, 2018). Applied to our spatial domain, we should expect the function to behave similarly to how it behaves for homogenous Poisson processes. We implement this via the 'Kest' function in the 'spatstat.explore' package (Baddeley & Turner, 2005).

Thinning, however, can be uninformative when the minimum of the conditional intensity function is small as very few events can remain in the thinned process. One can turn to a process known as super-thinning which involves simulating new data points as well, though we opt to focus on other diagnostic methods (Reinhart, 2018).

The next diagnostic method we employ utilizes Voronoi tessellations to evaluate the spatial adequacy of the fitted model. Voronoi tessellations partition the spatial domain into regions based on proximity to observed event locations, ensuring that each cell contains exactly one observed event. Each Voronoi cell contains the location of one event and all points in the spatial domain that are the closest to that event, compared to any other. Therefore, the observed count in each cell is 1. We visualize the Voronoi tesselation of our spatial domain in Figure 8.

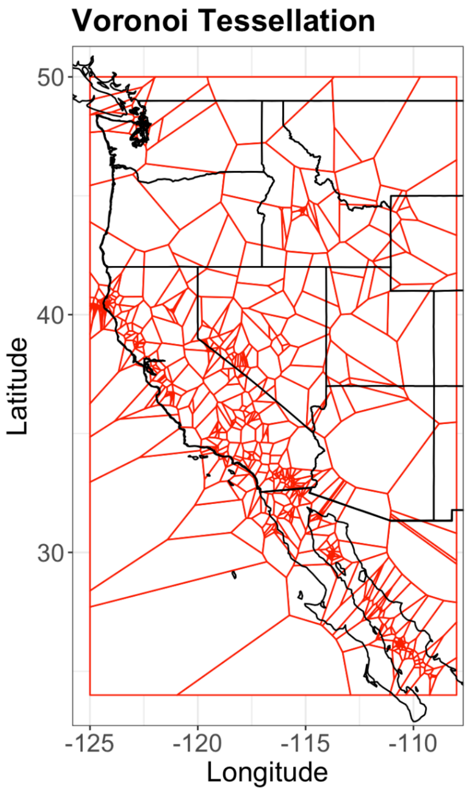


Figure 8: The constructed Voronoi tesselation of the data partitioning the spatial domain. Each cell contains exactly one event and all points in the spatial domain closest to it compared to any other event location.

Within each Voronoi cell, the expected number of events under the fitted model can be computed by integrating the conditional intensity over the cell over the temporal domain. This integration must be computed numerically due to the complex nature of the conditional intensity function. In our case we utilize the generation of random spatiotemporal points within a cell in order to gain a grasp on what its average conditional intensity is. Specifically, we generate 100 points within each cell uniformly over space and time and compute the average conditional intensity over those points. To obtain a numeric approximation of the expected number of events inside the cell, We then multiply this average by the area of the cell and by the temporal range of the data. This provides an approximation to the integral of the conditional intensity over the spatiotemporal region.

While calculating the conditional intensity for the random points for the Constant Background model with the 'stelfi' package is straightforward, the ETAS models require additional steps due to the non-parametric background rate. To approximate the background rate at a random point, we first construct a large grid over the study region so that we can observe the estimated background rate in each cell via the 'ETAS' package. For any randomly generated point, we then identify the grid cell in which it falls and assign to it the corresponding estimated background rate from that cell.

To account for the magnitude of the earthquake, we generate a random magnitude for each random point adhering to the logarithmic scale of moment magnitude. Residuals are then formed by subtracting the expected value from the observed count of 1. Under a well-specified model, Voronoi residuals should be centered near zero and display no coherent spatial structure. Systematic spatial patterns, such as clusters of positive or negative residuals, suggest over- or under-estimation of intensity in specific regions and may indicate deficiencies in the assumed background rate or spatial triggering kernel (Reinhart, 2018).

As one may suspect, this process can be carried out using other geometric partitions of the spatial domain. For example, one could do this using some sort of grid though the size of the grid cells. In particular, one may instead impose a regular grid over the study region and compute analogous residuals within each grid cell. In this case, rather than the pre-set 1, the observed count is the number of events falling in a given cell. While grid-based diagnostics are conceptually simpler, their interpretability highly depends on the choice of grid resolution. Small grids mean low expected number of events per cell and skewed distribution of residuals; large grids will cancel out over -and under-prediction within single cell (Reinhart, 2018). In contrast, the Voronoi tessellation avoid this arbitrariness by adapting cell size to the local event density, which is especially advantageous for ETAS models where spatial heterogeneity arises both from background variation and self-excitation.

Additionally, another diagnostic in the 'ETAS' package we consider is based on transformed event times constructed from the fitted conditional intensity. Each event time is mapped to a transformed time obtained by integrating the estimated temporal intensity up to that event. As per the time-rescaling theorem, under a correctly specified model, the collection of transformed times should follow a unit-rate Poisson process (Brown et al., 2002). As a result, plots of the transformed times $\tau_1, \tau_2, ...\tau_n$ against the event indices should exhibit an approximately linear trend $y = x$, and a Q-Q plot of $U_j = 1 - \exp\left(-\left(\tau_j - \tau_{j-1}\right)\right)$ should show agreement with the theoretical quantiles of a $U(0, 1)$ distribution (Jalilian, 2019). With these estimation and diagnostic methods in hand, we move to the results of our analysis.

5. Results

In this section we explore the results and validity of our best performing model - the Gaussian Epidemic Type Aftershock Sequence Model. The results of the constant background model, the inverse power 5-nearest neighbors ETAS model, and the inverse power 7-nearest neighbors ETAS model can be found in Section 8.3, Section 8.4, and Section 8.5, respectively.

Recall that the Gaussian ETAS Model conditional intensity function also takes the form explained in Section 4.1.3, with a Gaussian spatial kernel. The estimated coefficients for the base smoothness, and slightly smoother fit models are summarized with their standard deviations in Table 1 and Table 2 respectively.

Parameter	β	μ	A	c	α	p	D	γ
Estimate	1.91	0.72	1.36	0.0012	0.82	1.02	0.012	1.20
Standard Error	0.0044	0.013	0.23	0.096	0.032	0.0045	0.019	0.016

Table 1: Gaussian base model coefficient estimates (5 nearest-neighbors).

Parameter	β	μ	A	c	α	p	D	γ
Estimate	1.91	0.72	4.32	0.0011	0.83	1.01	0.012	1.20
Standard Error	0.0044	0.014	0.83	0.096	0.031	0.0044	0.018	0.015

Table 2: Gaussian smoother model coefficient estimates (7 nearest-neighbors).

For brevity, and on the basis of an AIC comparison, we focus on the base smoothness Gaussian ETAS Model, using the 5 nearest-neighbors. Visuals and diagnostics for the smoother Gaussian ETAS Model are contained in Section 8.7. An example reading of this table: The coefficients of A and α can be input into K to determine that the expected number of aftershocks for a magnitude 4.5 earthquake is around 3.09.

We can also visualize the non-parametrically estimated background rate over the spatial domain of the model as well as the clustering coefficients which describe the proportion of earthquakes that stem from other seismic events. Specifically, these are equal to $1 - \frac{\hat{u}(x,y)}{\hat{\Lambda}(x,y)}$ where $\hat{u}(x, y)$ is the estimated background rate at (x, y) and $\hat{\Lambda}(x, y)$ is the estimated total spatial intensity function at the same location. These coefficients capture the average intensity relative to the background rate in a given area, helping to showcase regions where seismic activity is mostly made up of aftershocks (Jalilian, 2019).

The background rate and clustering coefficients for the model are depicted in Figure 9. We see that while the background rate remains similar to the Inverse Power ETAS model, the clustering coefficients appear different over space. Notably, the areas of high coefficient values seem to be more tightly clustered with more areas of very low values throughout the spatial domain. Additionally, the area east of the Gulf of California seems to have much smaller clustering values than in the inverse power models. This is a result of the differing probability density functions for the spatial locations of events triggered by another earthquake. Many conclusions can be made from these plots as well. For example, it seems as if most background events in Idaho occur in one very specific region with wide-spread aftershock potential. This could be helpful information, for instance, if one wanted to notify citizens of when potential aftershocks may occur as the parent events are could be closely monitored.

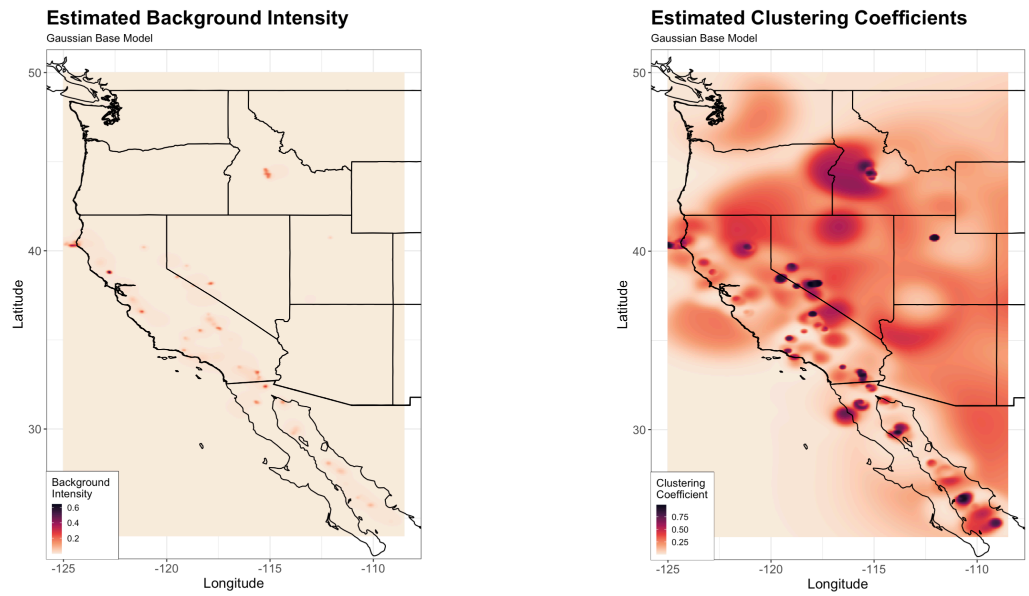


Figure 9: (L) The non-parametrically estimated background intensity for the Gaussian base model. (R) The spatial clustering coefficients for the Gaussian base model

Due to a low minimum conditional intensity function across the events, thinning is uninformative in this case as only around 10 events survive the process and we turn to the rest of our diagnostic methods.

The different spatial kernel does improve the spatial residuals compared to the Inverse Power ETAS model as evidenced by the raw spatial residuals and Voronoi cells depicted in Figure 10. The raw spatial residuals have less pattern but still remain mostly all negative. The very low values in Nevada and Idaho remain as well, but the values in California seem improved. The Voronoi cells seem to have some residual patterns as well. For instance, there appears to be many light colored cells at the bottom of the spatial region and darker colored ones in the top right. The Voronoi residuals have median and mean values of 0.51 and 0.050, respectively. There are certainly still some very low residual values, but the mean is an improvement compared to the residuals of other models. Also, the temporal residuals appear as desired (see Section 8.6) and the model seems to be accurately capturing the time-dependent patterns in the data.

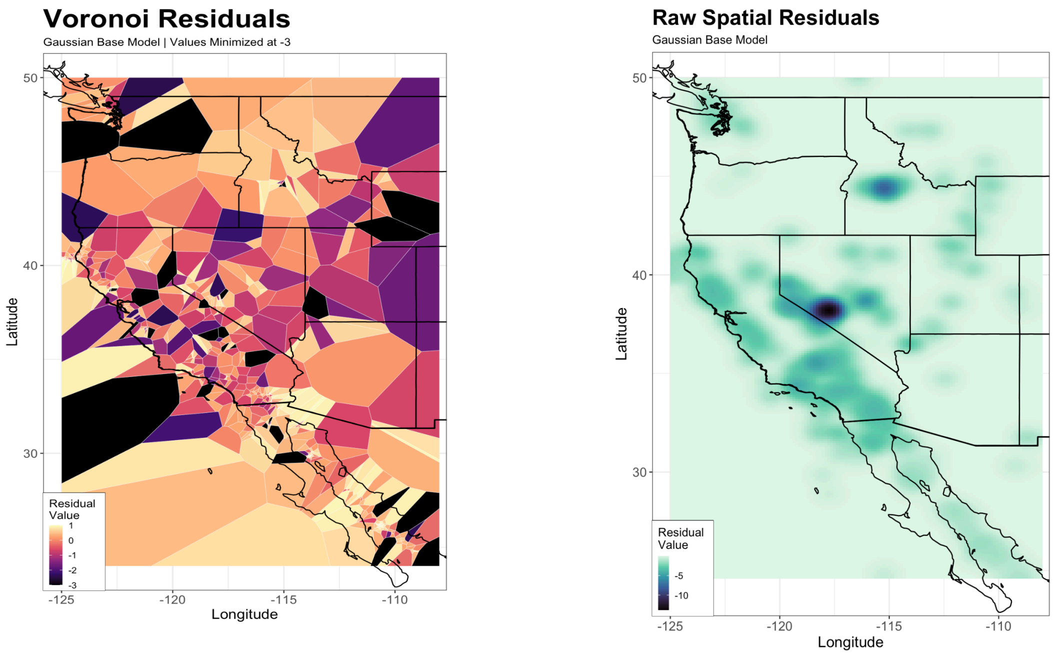


Figure 10: (L) The Voronoi residuals for the Gaussian base model. (R) The raw smoothed spatial residuals for the model have almost only non-positive values, but appear less problematic then the Inverse Power model spatial residuals.

Due to the improved, though not perfect, spatial residuals for this model (and due to it having a lower AIC compared to the slightly smoother Gaussian ETAS model), we determine that the base smoothness Gaussian ETAS model is the best of the tested options for modeling the Western North American Earthquake data.

The probabilities of each event either being a background or trigged event that result from stochastic delustering can also be visualized. Assuming that the model fit is adequate, we can visualize the most likely background events and the most likely triggering events as we do in Figure 11. This, in conjunction with Figure 9 helps us to conclude that where parent earthquakes occur and where their aftershocks do.

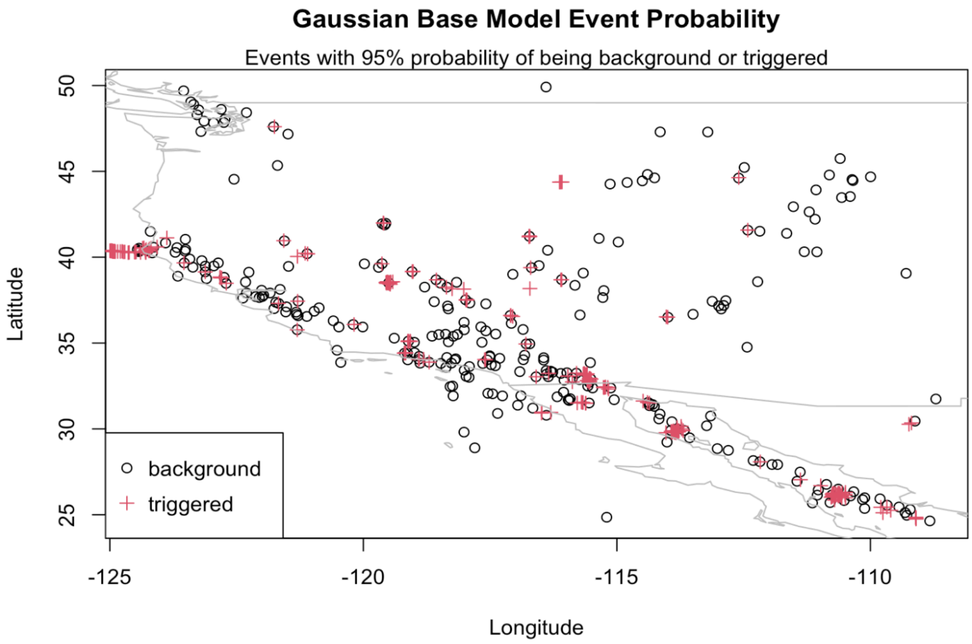


Figure 11: The spatial locations of events predicted to either be a background or triggered event with 95% confidence.

6. Conclusions

This study demonstrates that the Gaussian ETAS model, and Hawkes processes more broadly, are well-suited for modeling earthquake data. Although the specific models we employed do not fully capture the spatial variability observed in our dataset, they adequately describe the temporal clustering patterns. Different triggering kernels and conditional intensity specifications appear to improve spatial residual behavior, and with sufficient refinement, these models should be capable of fully capturing the spatial structure. Our selected model helps explain where parent earthquakes and their aftershocks occur, how many aftershocks are expected from an earthquake of a given magnitude, and how one event increases the likelihood of future nearby events.

More broadly, the usefulness of these models is self-evident. Understanding and predicting earthquake and aftershock behavior is particularly important for the Western U.S. region examined in this study, as well as for many communities worldwide that live near active faults. Well-estimated models can inform evacuation planning, building codes, and aftershock preparedness. Beyond seismology, Hawkes processes have important applications in epidemics and crime, where modeling the spread or clustering of events is crucial. Recent research has also expanded into finance, a rapidly evolving field that demands continual methodological adaptation. These innovative applications demonstrate the model's strong potential for further growth (Bacry & Muzy, 2013; Chiang, 2022; Xu et al., 2023).

Despite these strengths, our study has several limitations. Although Hawkes processes were introduced five decades ago, the field remained relatively dormant for much of the 20th century and only recently experienced a surge in interest. As such, available software for estimation and simulation is often rudimentary and lacks flexibility in kernel specification. Jalilian and Zhuang's ETAS package in R proved helpful, but it has relatively few contemporary alternatives (Jalilian, 2019). Additionally, we restricted our analysis to earthquakes of magnitude 3.5 and above, meaning that smaller-scale seismic activity was not incorporated into the model.

If we were to deepen our study of Hawkes processes, we would begin by refining the Gaussian ETAS specification to better account for spatial variability in the data. A natural next step would involve simulation. We would explore several simulation approaches, including acceptance–rejection algorithms, Ogata's method, and the parent–offspring approach (Bernabeu et al., 2024; Reinhart, 2018). Simulation methods and software are widely available, and deeper familiarity with these techniques would strengthen both model validation and diagnostic procedures, including methods such as super-thinning (Reinhart, 2018). Expanding the spatiotemporal domain of the model would also provide a valuable avenue for future research.

7. Bibliography

Bacry, E., & Muzy, J. (2013). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, *14*. https://doi.org/10.1080/14697688.2014.897000

Baddeley, A., & Turner, R. (2005). spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software*, *12*(6), 1–42. https://doi.org/10.18637/jss.v012.i06

Batra, N. (2025). *Modelling High-Frequency Data with Bivariate Hawkes Processes: Power-Law vs. Exponential Kernels*. https://doi.org/10.48550/arXiv.2503.14814

Bernabeu, A., Zhuang, J., & Mateu, J. (2024). Spatio-Temporal Hawkes Point Processes: A Review. *Journal of Agricultural, Biological and Environmental Statistics*, *30*, 89–119. https://doi.org/https://doi.org/10.1007/s13253-024-00653-7

Bittelli, M., Olmi, R., & Rosa, R. (2022). Poisson Processes. In *Random Process Analysis With R: Random Process Analysis With R*. Oxford University Press. https://doi.org/10.1093/oso/9780198862512.003.0004

Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., & Frank, L. M. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, *14*(2), 325–346. https://doi.org/0.1162/08997660252741149

Chiang, W.-H. (2022). *New Spatio-Temporal Hawkes Process Models for Social Good* (p. 118) [Doctoral dissertation]. https://www.proquest.com/dissertations-theses/new-spatio-temporal-hawkes-process-models-social/docview/2838332381/se-2

Hawkes, A. G. (1971). Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, *58*(1), 83–90. https://doi.org/https://doi.org/10.2307/2334319

Hawkes, A., & Chen, J. (2021). A personal history of Hawkes process. *Proceedings of the Institute of Statistical Mathematic*, *69*(2), 123–143. https://orca.cardiff.ac.uk/id/eprint/141988

Jalilian, A. (2019). ETAS: An R Package for Fitting the Space-Time ETAS Model to Earthquake Data. *Journal of Statistical Software, Code Snippets*, *88*(1), 1–39. https://doi.org/10.18637/jss.v088.c01

Jones-Todd, C. M., & van Helsdingen, A. B. M. (2024). stelfi: An R package for fitting Hawkes and log-Gaussian Cox point process models. *Ecology and Evolution*, *14*, e11005. https://doi.org/10.1002/ece3.11005

Laub, P., Lee, Y., & Taimre, T. (2022). *The Elements of Hawkes Processes*. Springer Cham. https://doi.org/https://doi.org/10.1007/978-3-030-84639-8

Laub, P., Lee, Y., Pollett, P., & Taimre, T. (2025). Hawkes Models and Their Applications. *Annual Review Statistics and Its Application*, *12*, 233–258. https://doi.org/10.48550/arXiv.2405.10527

Ogata, Y. (1998). Space-Time Point-Process Models for Earthquake Occurrences. *Annals of the Institute of Statistical Mathematics*, *50*, 379–402. https://doi.org/10.1023/A:1003403601725

Ogata, Y. (1999). Seismicity Analysis through Point-process Modeling: A Review. *Pure and Applied Geophysics*, *155*, 471–507. https://doi.org/10.1007/s000240050275

Ogata, Y. (2011). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth, Planets and Space*, *63*, 217–229. https://doi.org/10.5047/eps.2010.09.001

Omi, T., Hirata, Y., & Aihara, K. (2017). Hawkes process model with a time-dependent background rate and its application to high-frequency financial data. *Physical Review E*, *96*. https://doi.org/10.1103/PhysRevE.96.012303

Park, J., Schoenberg, F. P., Bertozzi, A. L., & Brantingham, P. J. (2021). Investigating Clustering and Violence Interruption in Gang-Related Violent Crime Data Using Spatial–Temporal Point Processes With Covariates. *Journal of the American Statistical Association*, *116*(536), 1674–1687. https://doi.org/10.1080/01621459.2021.1898408

Reinhart, A. (2018). A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications. *Statistical Science*, *33*(3), 299–318. https://www.jstor.org/stable/26770999

Ripley, B. D. (2018). Modelling Spatial Patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(2), 172–192. https://doi.org/10.1111/j.2517-6161.1977.tb01615.x

Seismological Society of America. (2025, June). *Statement on the Importance of Seismic Research and Monitoring for National Defense*. https://www.seismosoc.org/us-government-relations/importance-of-seismic-research-and-monitoring-for-national-defense/#:~:text=Seismic%20research%2Fmonitoring%20provides%20information,the%202022%20Nord%20Stream%20pipeline

U.S. Geological Survey. (n.d.). *Faults — Earthquake Hazards Program*. Retrieved February 19, 2026, from https://www.usgs.gov/programs/earthquake-hazards/faults#data

U.S. Geological Survey. (2026, January). *Search Earthquake Catalog*. https://earthquake.usgs.gov/earthquakes/search/

Xu, C., Xie, Y., Vazquez, D. A. Z., Yao, R., & Qiu, F. (2023,). *Spatio-Temporal Wildfire Prediction using Multi-Modal Data*. https://arxiv.org/abs/2207.13250

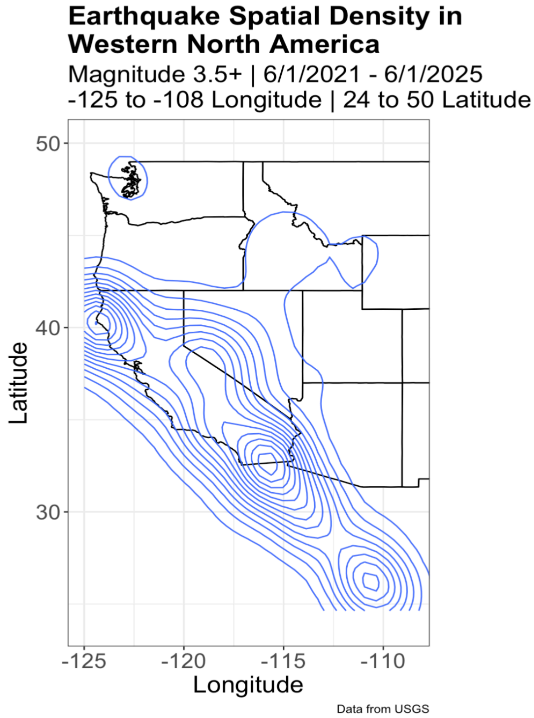
Zhuang, J., & Mateu, J. (2018). A Semiparametric Spatiotemporal Hawkes-Type Point Process Model with Periodic Background for Crime Data. *Journal of the Royal Statistical Society: Series a (Statistics in Society)*, *182*. https://doi.org/10.1111/rssa.12429

Zhuang, J., Ogata, Y., & Vere-Jones, D. (2002). Stochastic Declustering of Space-Time Earthquake Occurrences. *Journal of the American Statistical Association*, *97*, 369–380. https://doi.org/10.1198/016214502760046925

Zhuang, J., Ogata, Y., & Vere-Jones, D. (2006). Diagnostic Analysis of Space-Time Branching Processes for Earthquakes. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, & D. Stoyan (Eds.), *Case Studies in Spatial Point Process Modeling* (pp. 275–292). Springer New York. https://doi.org/10.1007/0-387-31144-0_15

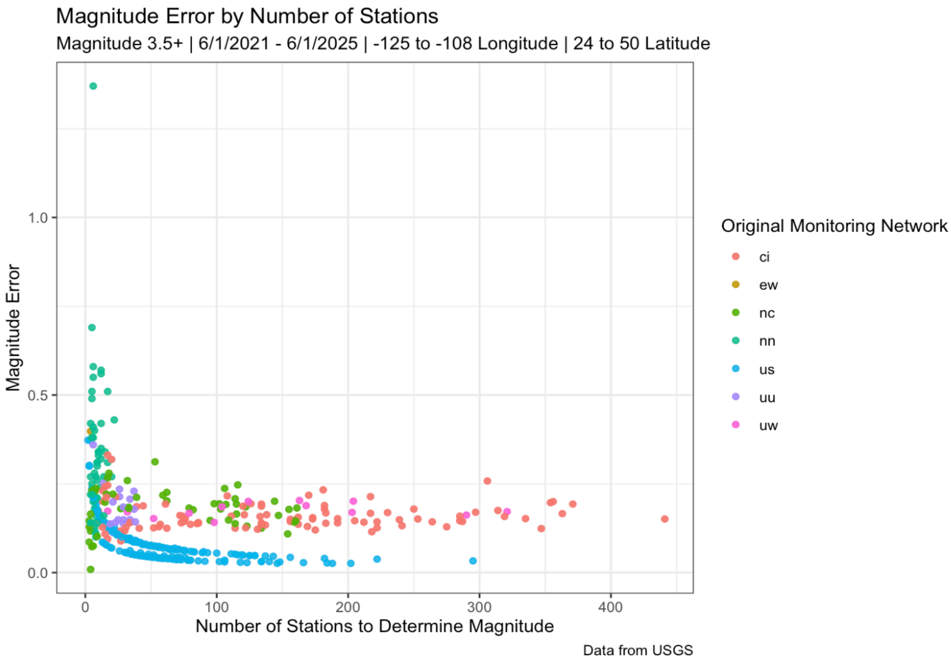
8. Appendix

8.1. Spatial Density

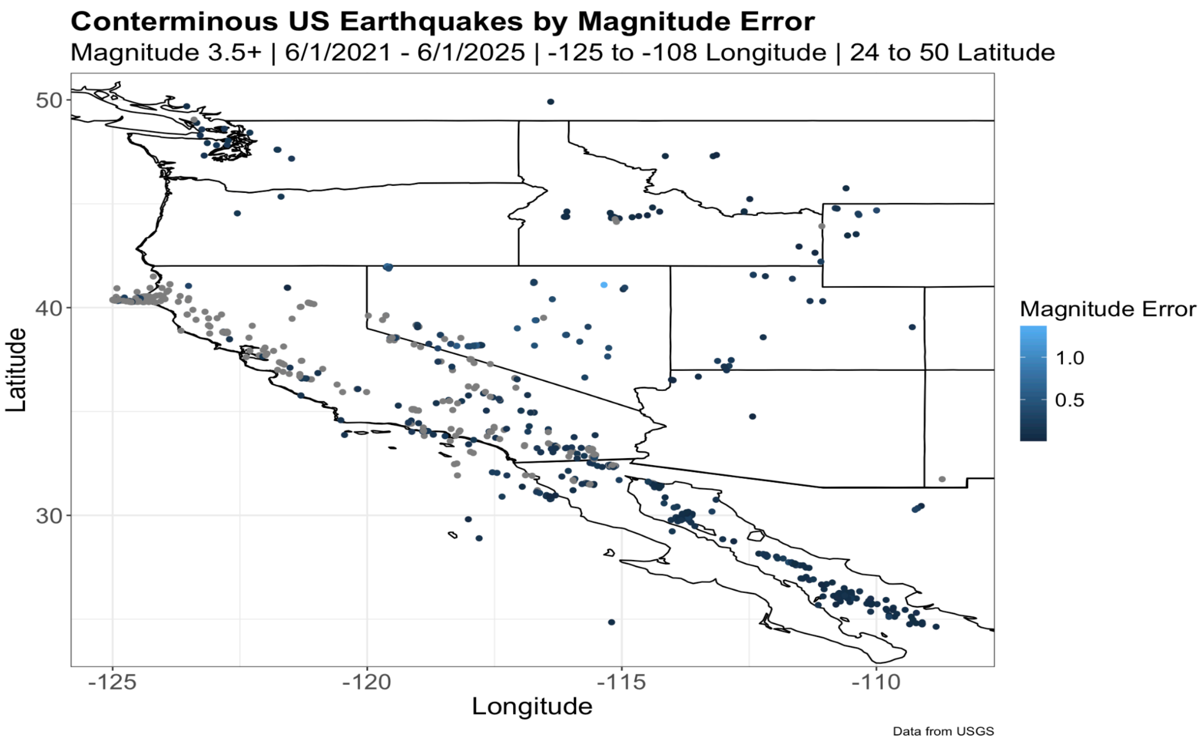


The above plot represents the kernel density of earthquake locations in the sub-setted data. In contrast to individually plotting each event, this plot provides a summary of the spatial concentration of seismic activity for the sub-setted data. The visualization uses darker colors to represent lower density and lighter colors to represent higher density. Notably, the highest density occurs at the center of each hotspot, with density gradually decreasing in the surrounding areas. The presence and shape of these hotspots further emphasize the self-exciting structure that earthquake activity exhibits. Similar to Figure 5, clear seismic hotspots are visible in the Western region of the United States, with relatively sparse activity in the central regions.

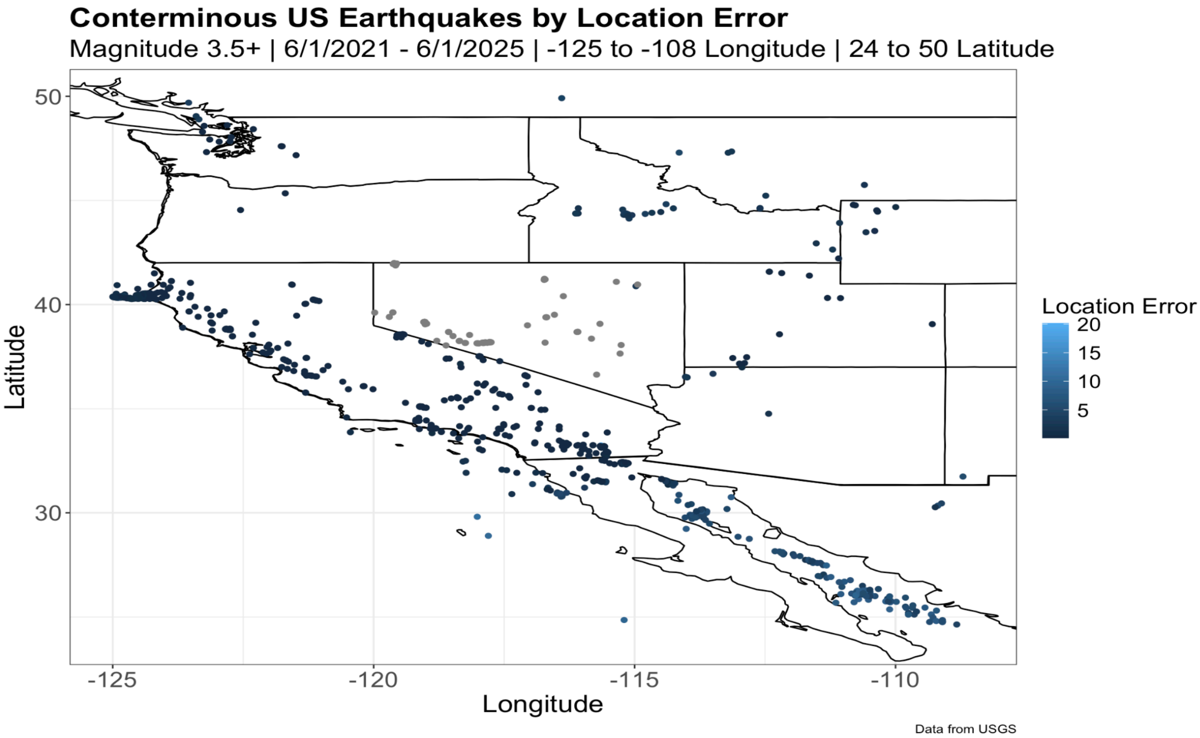
8.2. Error EDA



The relationship between the number of stations used to estimate an earthquake's magnitude and the associated magnitude error is visualized above with each point representing one earthquake, colored by the original reporting network. The plot shows a clear inverse relationship between the variables for one network, and almost no relationship for another. Overall, as the number of stations used to calculate the magnitude increases, the magnitude error rapidly decreases and then stabilizes.



The magnitude error appears to have some higher values in Nevada and many missing values along the coast of California. For the most part, however, these values seem small and relatively homogenous.



The location error values also appear similar to the magnitude error values spatially, except with missing data in Nevada and higher error values around the Gulf of California.

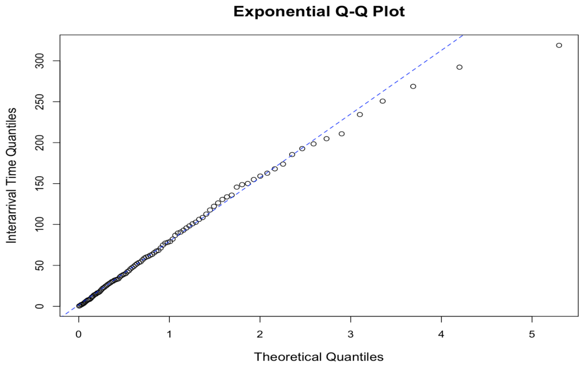
8.3. Constant Background Model Results

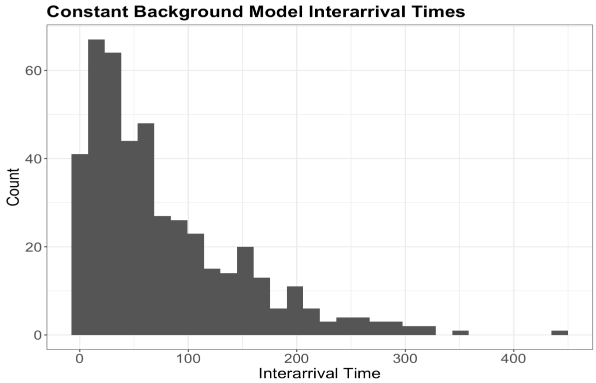
The constant background model estimated coefficients are described in the table below. The imply a small, constant background rate across the entire spatiotemporal domain.

The diagnostics for the constant background model show the parameter estimations to be relatively stable with small standard errors, and the thinning indicates that the model does not capture the underlying seismic process. The thinned process contains a large number of events that do not follow a homogeneous Poisson process spatially, which is not consistent with the background rate for Western North America.

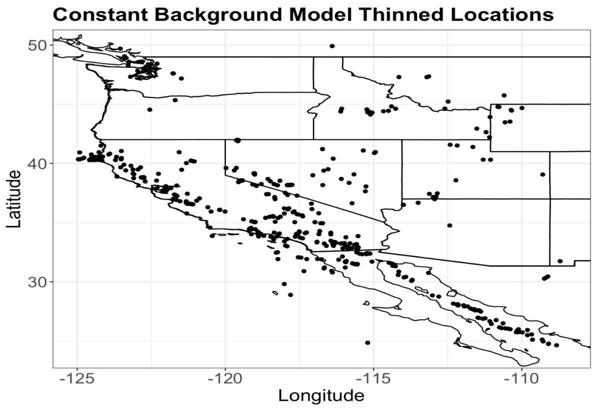
Parameter	μ	α	β	σ_x	σ_y	ρ
Estimate	0.000033	0.0046	0.020	0.28	0.23	0.38
Standard Error	0.0000017	0.00070	0.0029	0.025	0.039	0.14

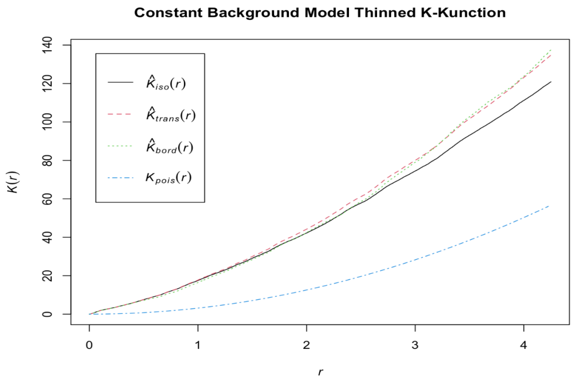
The thinned process contains a sizable 438 events. The interarrival times appear to somewhat follow an exponential distribution as evidenced by the plots below. This implies that the model is close to adequately describing the temporal trends in the data.



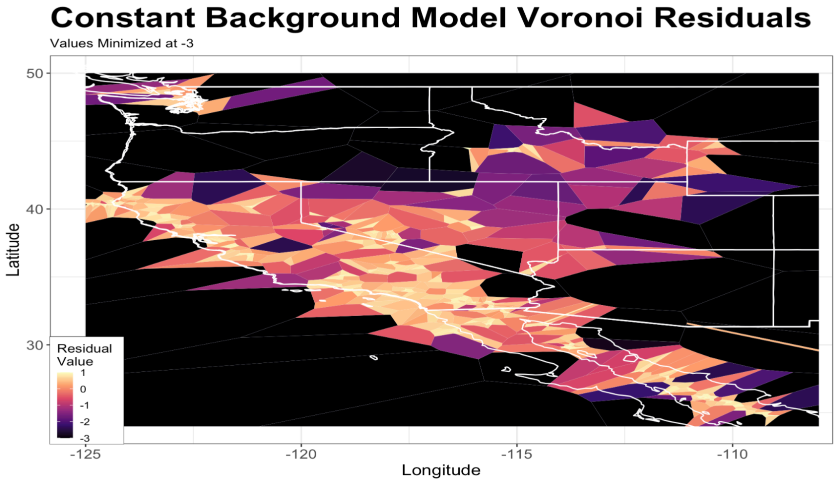


Spatially, though, the thinned process is certainly not a homogenous Poisson process. We see in the plots below that the spatial locations of the earthquakes in the thinned process still contain clustering. Moreover, regardless of which edge effect correction is chosen, the thinned process still lands far from a Poisson process in terms of its K-function values. The temporal diagnostics show that the model follows an exponential distribution, showing that the model is capturing the temporal behavior. Both the K-function results and the Voronoi residual plots show the model's deviation from homogeneity along with strong clustering in California and along Baja California, with low residual values in less dense regions. Overall, the diagnostics display that the constant background model is a useful baseline but does not adequately capture the spatial structure of the data.





Additionally, the Voronoi cells for this model, even coloring cells with residual values less than -3 as the darkest possible color on the scale, have major patterns. For instance, there are high residuals within the clustered portion of the data and very negative ones around the edges. This adheres to our expectations of the true background rate being non-constant as we are over-predicting the number of earthquakes in areas where seismic events are sparse in the data, due to the assumption of it being constant in this model

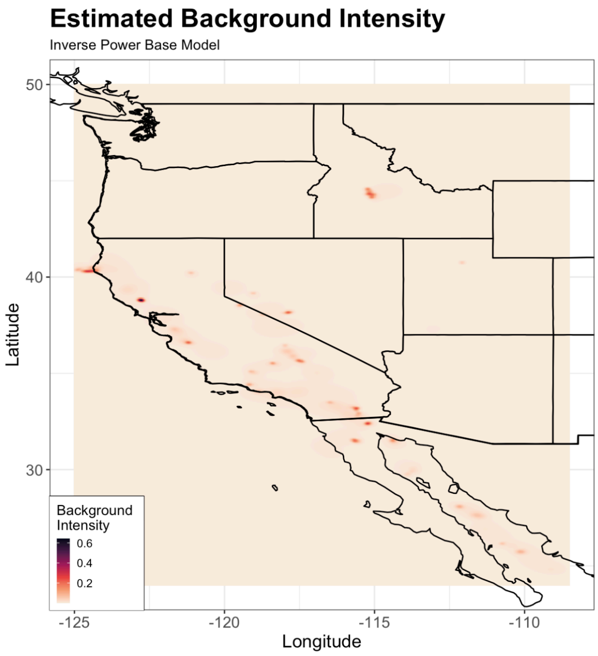


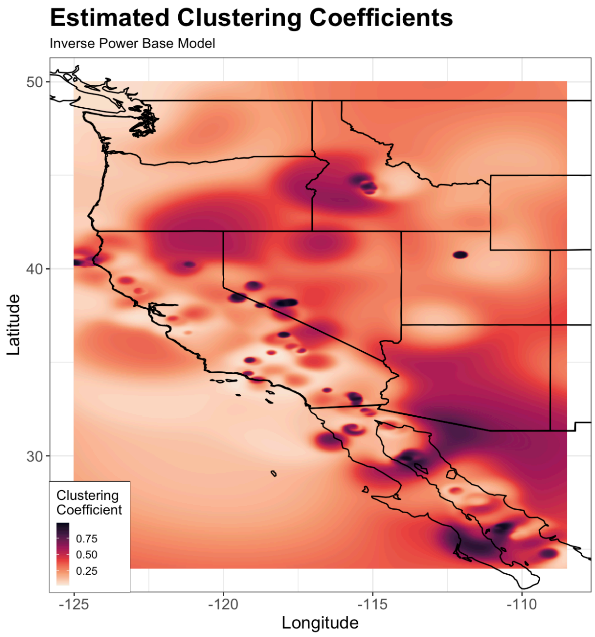
8.4. Inverse Power Base ETAS Model Results

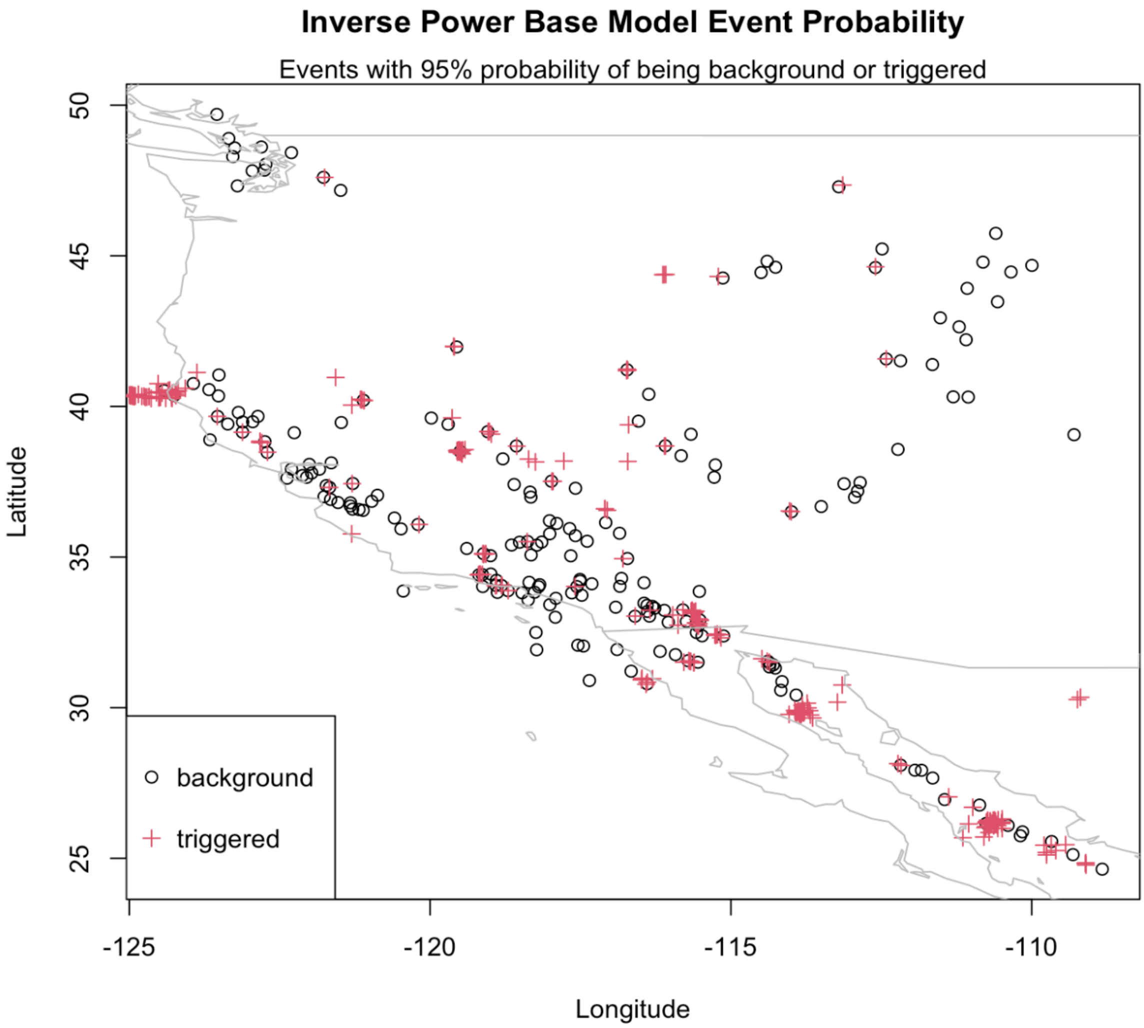
The parameter estimates for the inverse power ETAS model fit with 5 nearest-neighbor bandwidths are described in the table below.

Parameter	β	μ	A	c	α	p	D	q	γ
Estimate	1.92	0.71	1.11	0.0013	1.06	1.02	0.000022	1.37	1.54
Standard Error	0.0044	0.014	0.20	0.091	0.021	0.0044	0.067	0.0069	0.034

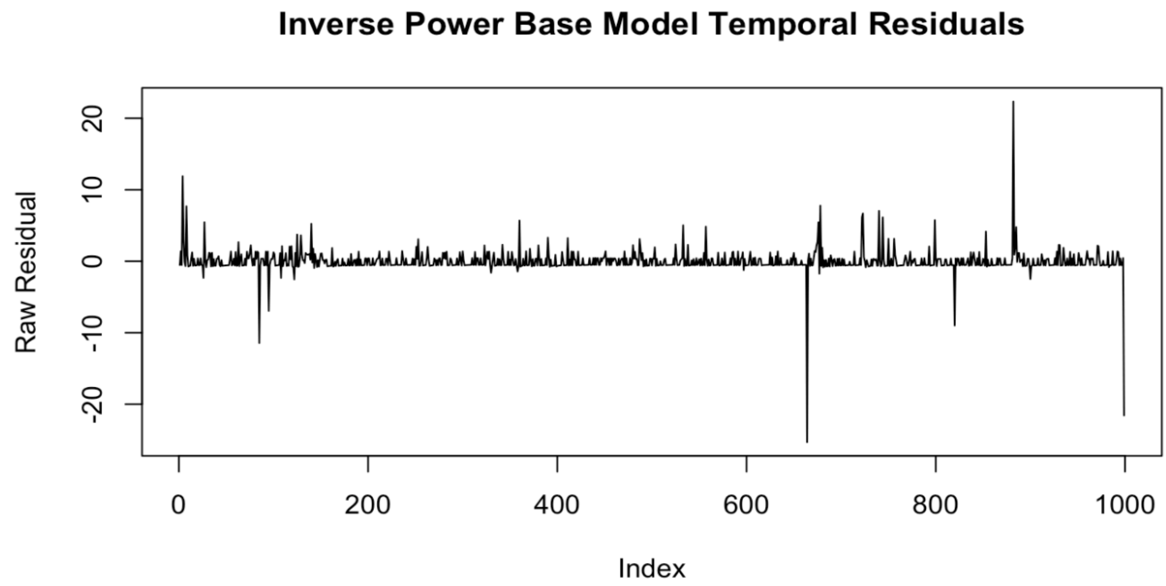
We visualize the background intensity and clustering coefficients for this model, as well as the most probable background and triggered events in the three plots below.

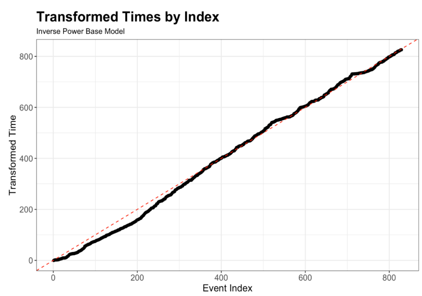


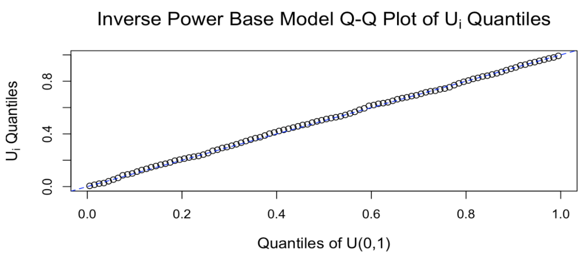




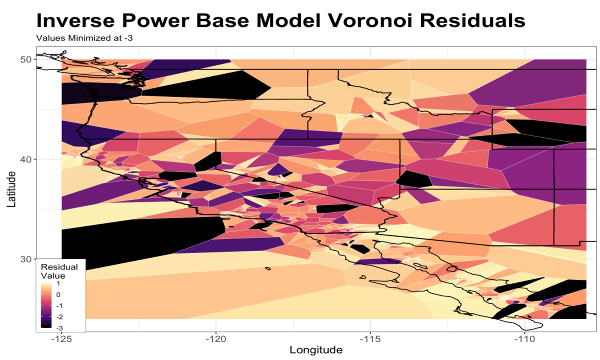
The temporal residuals have a mean of 0.00072 and appear to be independent from one another suggesting a good temporal fit for the model. Moreover, the transformed times appear to follow a unit rate Poisson process as evidenced by the plots below.

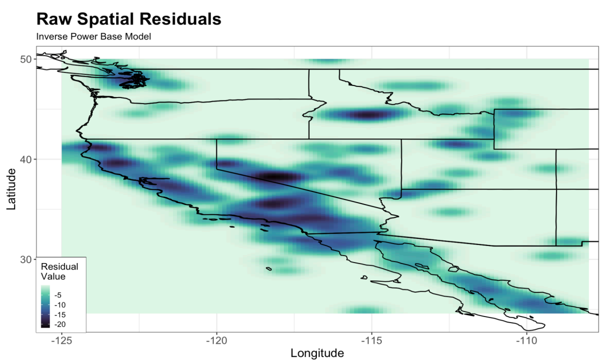






The raw spatial residuals and the Voronoi residuals depicted below showcase that there are still spatial patterns left accounted for. As an example, consider the clusters of low spatial residuals in Washington and Idaho, as well as along the coast of California and in the Gulf of California. Additionally, the Voronoi residuals contain some residue patterns including clusters of high residual values in the southeast corner of the spatial domain and lower residual values in central California. The median and mean of the Voronoi residual values are 0.57 and 0.12, respectively, also indicating a degree of spatial inadequacy for this model.

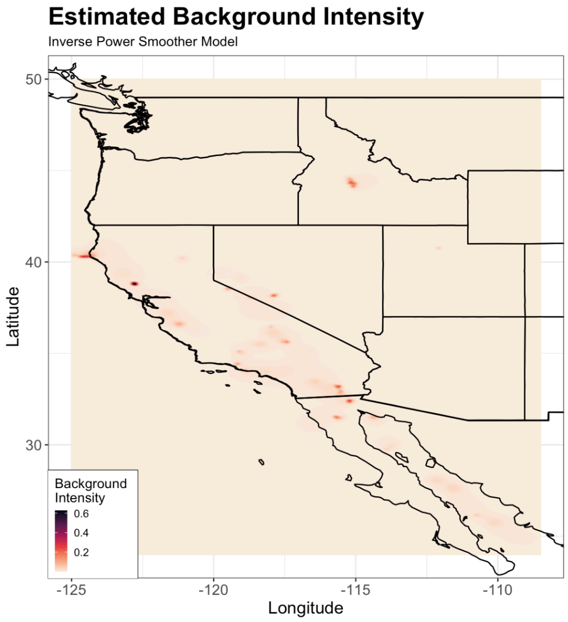


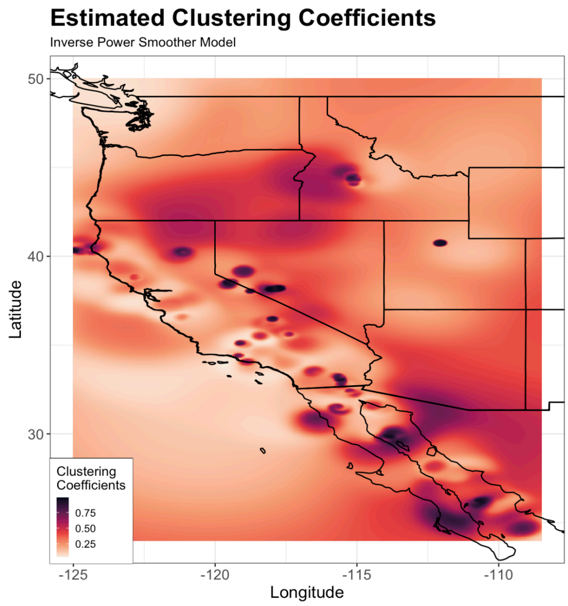


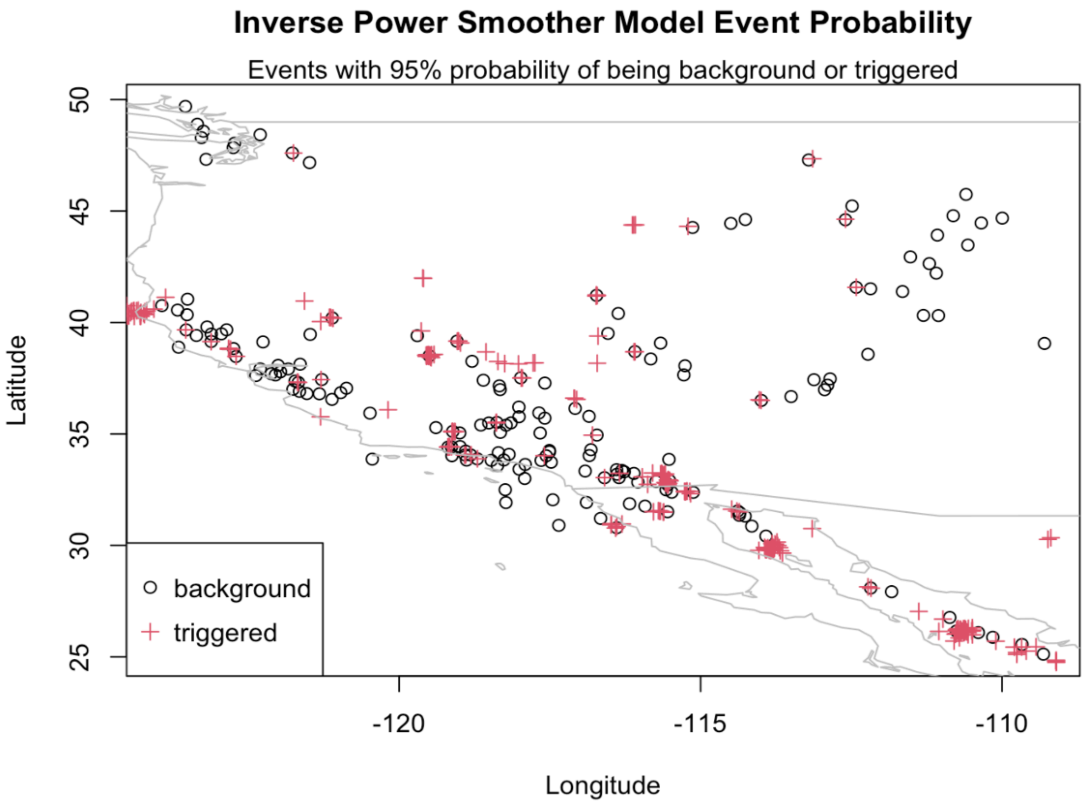
8.5. Inverse Power Smoother ETAS Model Results

The parameter estimates for the inverse power ETAS model fit with 7 nearest-neighbor bandwidths are described in the table below.

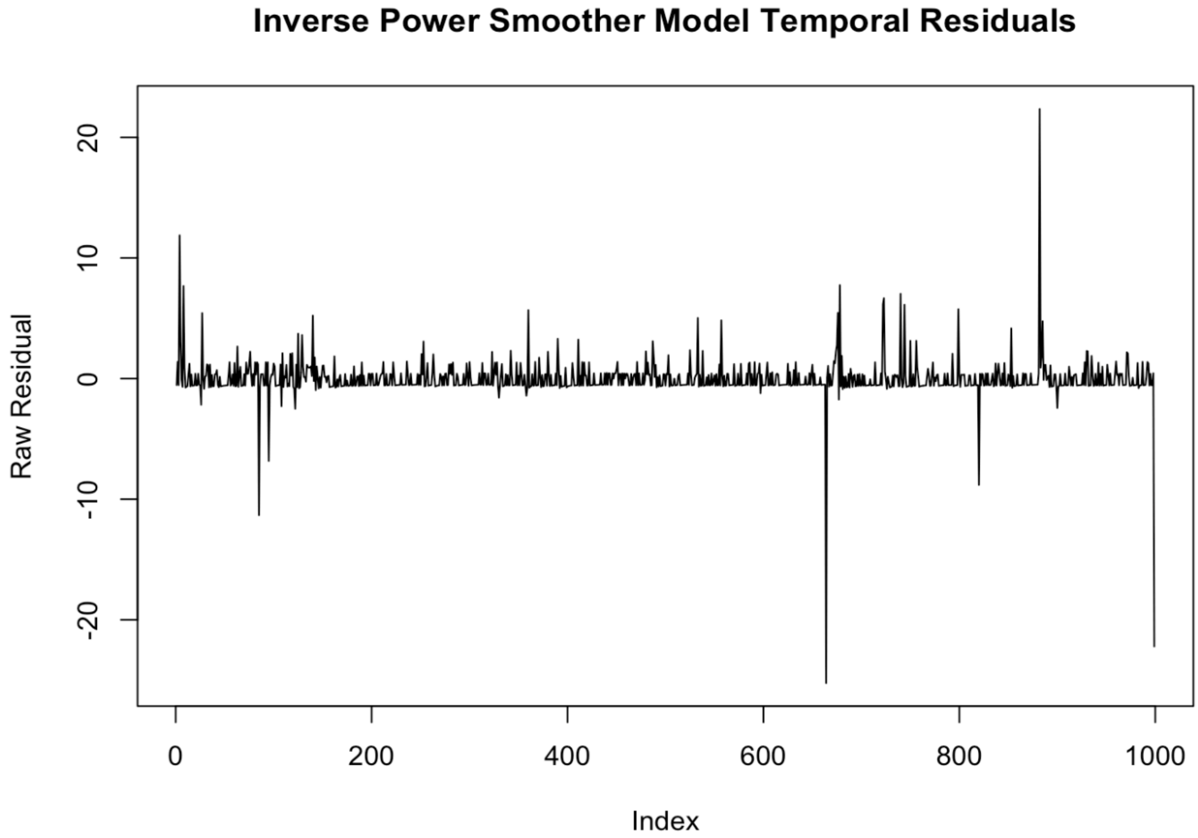
Parameter	β	μ	A	c	α	p	D	q	γ
Estimate	1.92	0.71	2.56	0.0012	1.06	1.0081	0.000023	1.37	1.54
Standard Error	0.0044	0.015	0.50	0.093	0.021	0.0043	0.066	0.0069	0.033

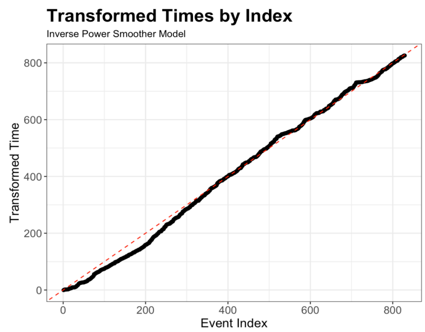


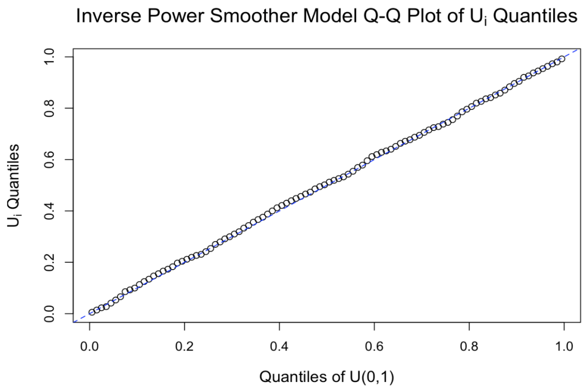




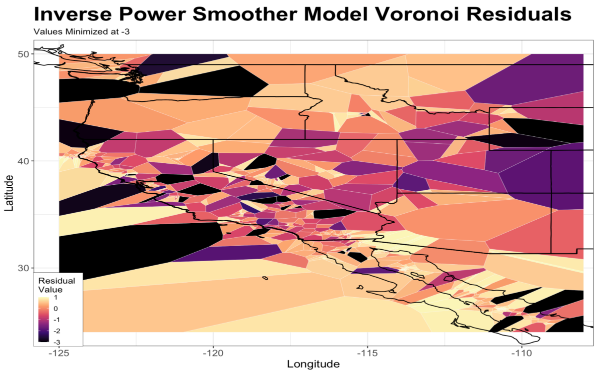
The above plots appear really similar to the plots showcased in Section 8.4. The temporal residuals and transformed times also appear to be as desired, indicating that the temporal variation is accurately accounted for by this model.

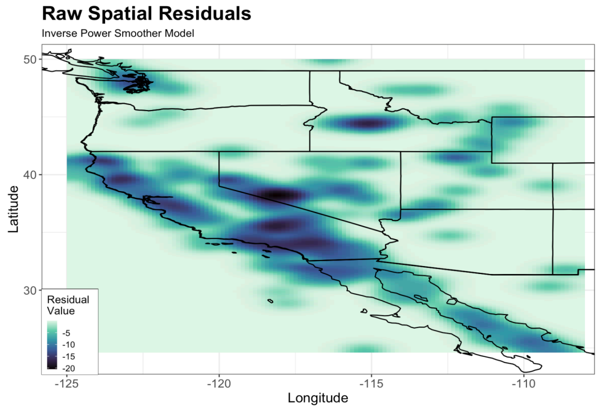




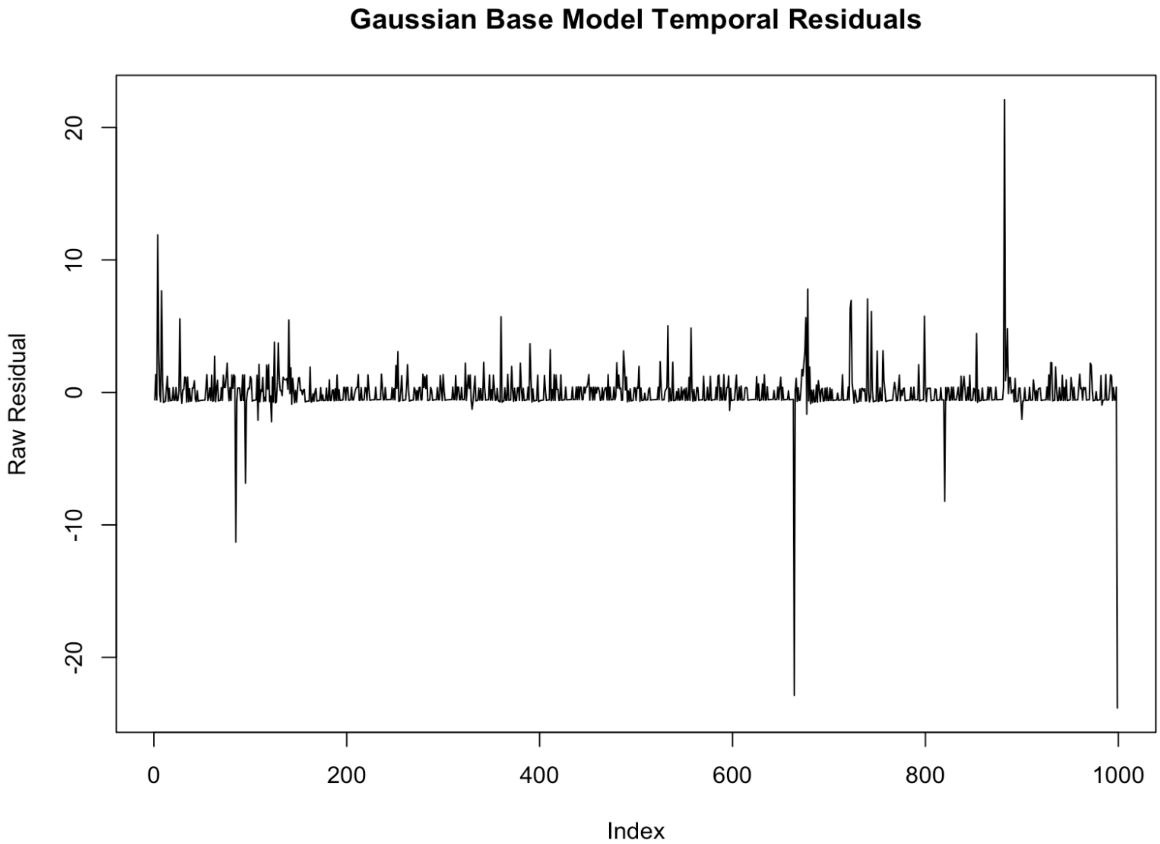


The spatial diagnostics below, however, indicate that this model contains issues with its spatial fit due to the leftover patterns in both the Voronoi residuals and the raw spatial residuals.

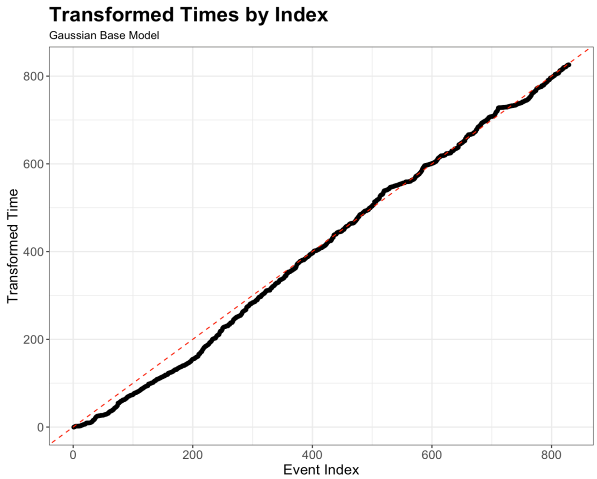


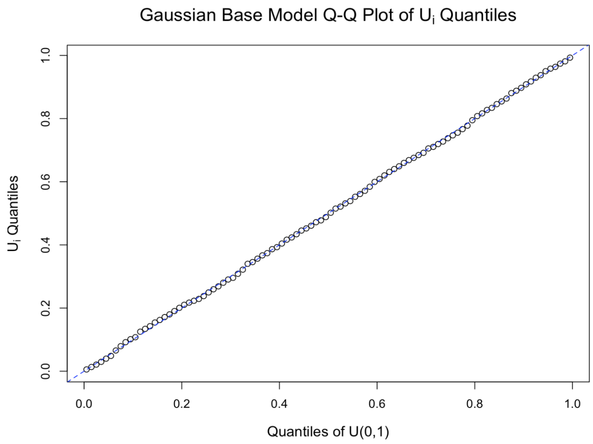


8.6. Gaussian Base Model Plots



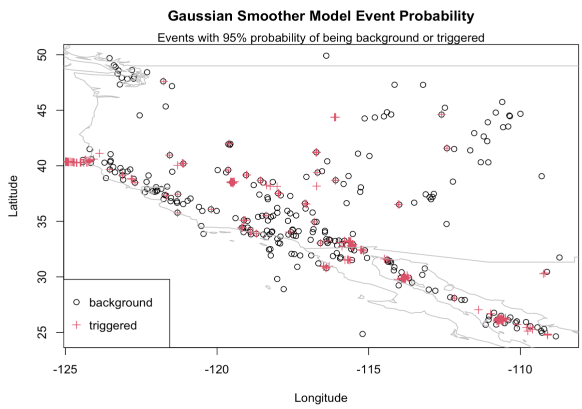
As seen in the image above, the raw temporal residuals for the model appear to be centered at zero and independent suggesting that the model is adequately accounting for temporal variation in the data. The transformed times and the associated U_i values also appear to follow the correct trends as depicted below.

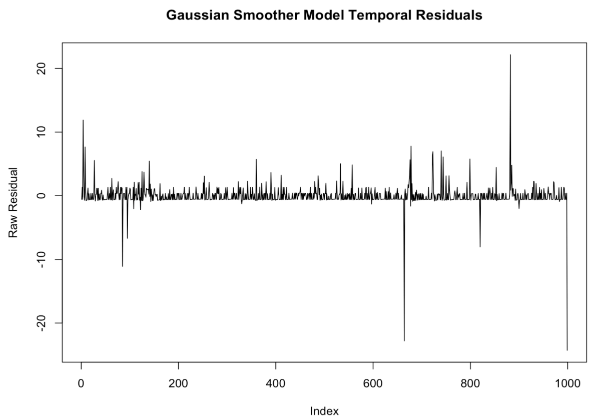


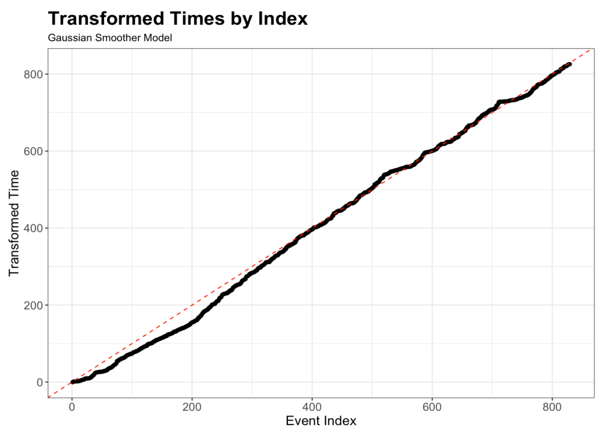


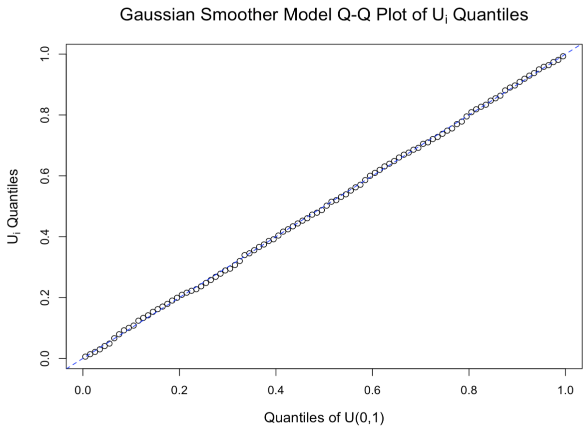
8.7. Gaussian Smoother Model Plots

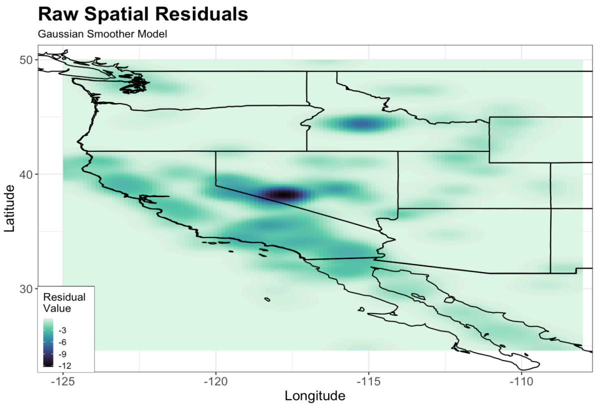
Below are the plots for the Gaussian ETAS model fit using 7 nearest-neighbors as opposed to 5. Overall, the plots appear very similar to those shown in Section 5.

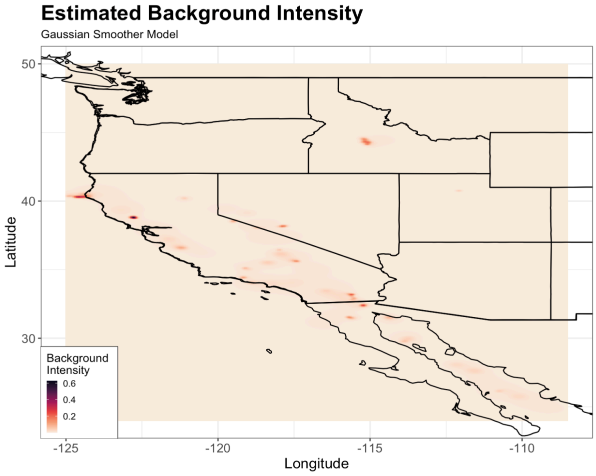


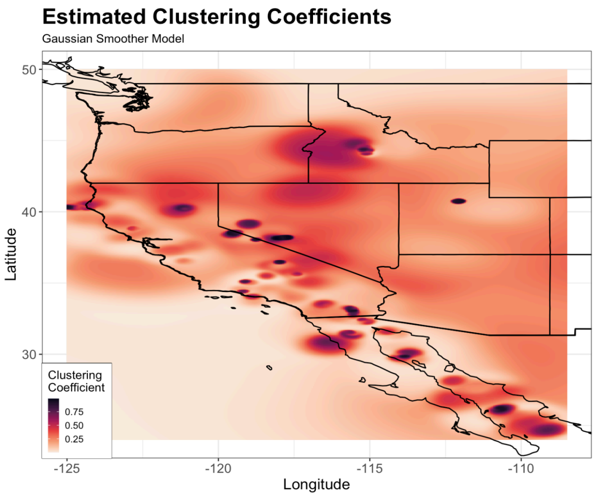


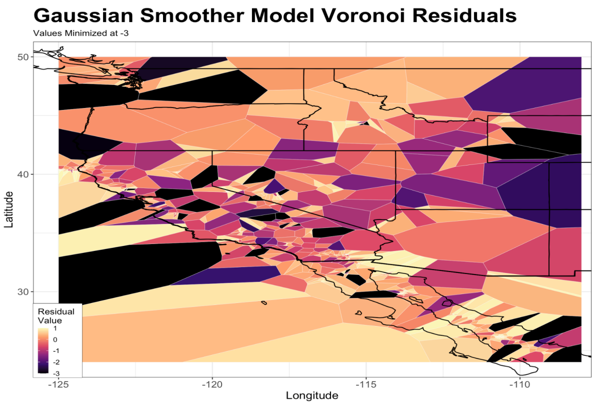





Raw Spatial Residuals
Gaussian Smoother Model
Latitude
Longitude
Residual Value
-3
-6
-9
-12
50
40
30
-125
-120
-115
-110


Estimated Background Intensity
Gaussian Smoother Model
Latitude
Longitude
Background Intensity
0.6
0.4
0.2
50
40
30
-125
-120
-115
-110


Estimated Clustering Coefficients
Gaussian Smoother Model
Latitude
Longitude
Clustering Coefficient
0.75
0.50
0.25
50
40
30
-125
-120
-115
-110


Gaussian Smoother Model Voronoi Residuals
Values Minimized at -3
Latitude
Longitude
Residual Value
1
0
-1
-2
-3
50
40
30
-125
-120
-115
-110