16TH EUROPEAN CONFERENCE ON EARTHQUAKE ENGINEERING THESSALONIKI 18 - 21 JUNE 2018

PARAMETER ESTIMATION METHODS FOR MODELING OF TIME AND SPACE INTERACTIONS OF EARTHQUAKE RUPTURE

Luis CEFERINO[1], Anne KIREMIDJIAN[2], Greg G. DEIERLEIN[3]

ABSTRACT

Earthquake rupture occurrence modeling is the basis of seismic risk and Performance-based Earthquake Engineering (PBEE). This paper summarizes a probabilistic formulation for modeling time and space interactions of rupture occurrence of earthquake mainshocks in a tectonic region. The formulation represents the elastic-rebound behavior in the tectonic plates and models the stress interaction of neighboring areas of the faults through spatial correlation. The paper also describes two methods for the estimation of the model parameters. The first method is a simple approach that estimates the parameters in pairs and then calibrates the spatial correlation parameter. The second one uses a Bayesian updating to estimate all parameters simultaneously. Both approaches are applied to model the rupture occurrence of large interface earthquakes on the subduction zone along the Coast of Lima, Peru. The Bayesian updating is demonstrated to be a more reliable estimation technique of the two approaches as it predicts a hazard rate that is closer to the data. The simple approach overpredicts the hazard rate by more than 25% for areas where data are sparse.

Keywords: elastic-rebound theory; earthquake rupture; time-dependent hazard.

1. INTRODUCTION

A salient component earthquake risk and PBEE analyses is the probabilistic modeling of earthquake occurrences. In this paper, a probabilistic formulation for modeling time and space rupture interactions for large earthquake mainshocks is summarized. This formulation has an underlying physical interpretation and is an alternative to existing models that are data-driven but lack such an interpretation (e.g., Helmstetter and Werner, 2014) and others that lack model consistency (i.e., mismatch between the assumed and simulated distribution of rupture interarrival times in the fault) (e.g, Field et al., 2015). The model discretizes a tectonic fault into small sections and uses multiple Brownian Passage Time (BPT) distributions to model rupture occurrence at each of the sections. The BPT distributions represent a Brownian Relaxation Oscillator (BRO) acting at each section (Matthews et al., 2002). The BROs can be interpreted as representing a stress accumulation in each section comprised of two additive components: one stress component increasing at a constant rate and one random component behaving as a Brownian motion. Whenever there is a rupture, the stress resets to initial conditions and starts the cycle again. The BRO aims to represent the elastic-rebound behavior, the canonical theory of macroscopic earthquake tectonic behavior (Reid, 1911). Moreover, the model introduces a spatial correlation formulation that represents the stress interactions among neighboring zones in the tectonic region. A 1-D version of the model was originally proposed by Ceferino et al. (2017), and the 2-D version of the model and an in-depth study of its properties was presented in Ceferino et al. (2018a).

[1]PhD Candidate, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305, USA, ceferino@stanford.edu
[2]Professor, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305, USA, ask@stanford.edu
[3]Professor, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305, USA, ggd@stanford.edu

Additionally, this paper describes two approaches for estimation of the model parameters. The first one is a simple method, called hereafter 'simple approach', where the parameters of the BPT distributions are estimated in each section separately. The correlation parameter is then calibrated to match key seismic properties of the tectonic region. The second one is a Bayesian updating that estimates all parameters of the model simultaneously. This approach uses a Metropolis-Hastings (MH) Monte Carlo Markov Chain (MCMC) to explore the parameter space and sample from the joint posterior distribution of the parameters. The Bayesian updating performs a joint estimation of all the model parameters, therefore, it can be more suitable than the simple approach for parameter estimation for scarce data (e.g., in earthquake catalogs of large ruptures). An application to model large interface earthquake ruptures in the subduction zone along the Coast of Lima, Peru, is also presented to compare both approaches. The comparison reveals the differences in the estimates of the parameters using both approaches and also reveals how these differences propagate to the seismic hazard in the region.

The paper is organized as follows. First, a brief overview of the model is presented. Second, the two approaches for parameter estimation are described. Third, the application and the comparison of both approaches are detailed. Finally, key conclusions and further steps are provided.

2. MODEL SUMMARY

In this model, the tectonic fault is discretized into N sections. For example, Figure 1 shows the discretization of a tectonic fault idealized as a line. This line can represent an idealized long strike-slip fault such as the San Andreas Fault or a subduction zone idealized along its strike direction. The vector $X_t = \{X_t(1), \ldots, X_t(N)\}$ is defined as the rupture vector at year t. Each element $X_t(j)$ corresponds to each section of the discretized fault. If there is a rupture at year t, $X_t(j)$ equals 1 and 0 otherwise. Figure 1 shows an earthquake rupture in red at year t, then all $X_t(j)$ corresponding to the sections in red will equal 1 and the others will equal 0 at year t. Additionally, the model assumes that the rupture occurrence depends on the vector $T_t = \{T_t(1), \ldots, T_t(N)\}$, that contains the time since the previous rupture at each section of the fault. Each year, $T_t(j)$ is either increased by one if there is no rupture (i.e., $X_t(j) = 0$) or reset if there is one (i.e., $X_t(j) = 1$). Equation (1) shows the relationship between $T_t(j)$ and $X_t(j)$.

$$T_{t+1}(j) = T_t(j)\{1 - X_t(j)\} + 1 \tag{1}$$

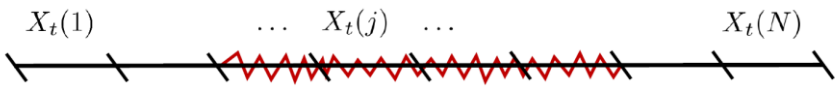


Figure 1. Model discretization of a 1-D representation of a tectonic fault. Extracted from Ceferino et al. (2017).

The formulation uses a correlated Multivariate Bernoulli distribution to model the earthquake rupture occurrence at year t as represented in Equation (2). The model is time-dependent since the rupture is dependent on the vector T_t, which contains the years since the last rupture in each section. The vector $p_t = \{T_t(1), \ldots, T_t(N)\}$ contains the marginal probabilities of rupture occurrence at each section at year t and is function of T_t as described below. Additionally, the ruptures of sections interact over the space since the model includes spatial correlation, which makes large rupture more likely. The spatial correlation in the model will be further explained below.

$$X_t | T_t \sim Multivariate\ Bernoulli(p_t) \tag{2}$$

The model assumes a Brownian Relaxation Oscillator (BRO) occurring at each section. Intuitively, the BRO represents a stress accumulation in the tectonic plates comprised of 2 additive components: one stress that increases over time with a constant rate, and one random stress resulting from a Brownian motion. As soon as the BRO reaches certain stress threshold (i.e., failure state), the BRO resets to initial stress conditions. Alternative interpretations of the BRO can be found in Matthews et al. (2002). The

rupture interarrival time (time between failures) is distributed as a Brownian Passage Time (BPT) distribution. In this model, each fault section has a corresponding BPT distribution, which is defined by two parameters: the mean interarrival time μ_j and the coefficient of variation (CoV) α_j. Earthquake catalogs only contain earthquake rupture interarrival times rather than the progressive stress accumulation through earthquake cycles. Therefore, only the parameters of the BPT (μ_j and α_j) can be estimated from such catalogs. Matthews et al. (2002) developed relationships among the parameters of the BRO (e.g., initial stress conditions, the constant-rate component, the Brownian motion component and the stress threshold) and the parameters of the BPT that can be used to infer possible associated BRO behaviors in a fault.

Using the Bayes rule, $p_t(j)$ can be found as a function of μ_j, α_j, and $T_t(j)$ as shown in Equation (3a), 3b and 3c (see Ceferino et al. (2018)), where $\Phi[.]$ is the standard normal cumulative distribution function.

$$p_t(j) = \frac{(\Phi[u_1(T_t(j))] - \Phi[u_1(T_t(j)-1)]) + e^{\frac{-2}{\alpha_j}}(\Phi[-u_2(T_t(j))] - \Phi[-u_2(T_t(j)-1)])}{1 - \left(\Phi[u_1(T_t(j)-1)] + e^{\frac{-2}{\alpha_j}}\Phi[-u_2(T_t(j)-1)]\right)} \quad (3a)$$

$$u_1(T) = \alpha_j^{-1}\left[T^{1/2}\mu_j^{-1/2} - T^{-1/2}\mu_j^{1/2}\right] \quad (3b)$$

$$u_2(T) = \alpha_j^{-1}\left[T^{1/2}\mu_j^{-1/2} + T^{-1/2}\mu_j^{1/2}\right] \quad (3c)$$

The model also includes a spatial correlation function through a spherical correlogram as shown in Equations (4). Other types of correlogram were also tested in Ceferino et al. (2018a). The correlogram outputs the correlation between rupture occurrence at two sections i and j during a given year, and it is function of the distance $dist(i,j)$ between these sections and the parameter γ. The correlation of rupture occurrence decays with distance, and γ defines the decaying rate. Intuitively, the correlogram introduces an implicit seismic interaction in the BROs (i.e., stress paths) in adjacent sections.

$$\rho_{i,j} = \exp\left(-\left(\frac{dist(i,j)}{\gamma}\right)^2\right) \quad (4)$$

For example, a section with a significant probability of rupture during year t (i.e., interpreted as large accumulated stress) can trigger a rupture in adjacent faults through the spatial correlation. This behavior resembles suggested theories for rupture propagation from the earthquake nucleation point to all the extent of the rupture area (e.g., Ellsworth and Beroza, 1995). In this model, it is assumed that the rupture of multiple adjacent sections during one year t are generated in one single large earthquake event, whereas rupture of non-adjacent sections are generated by corresponding multiple earthquake events during the year.

Finally, it is noteworthy that the Probability Distribution Function (PDF) of the correlated Multivariate Bernoulli distribution presented in Equation (2) lacks closed-form solution as a function of the vector p_t and the spatial correlation $\rho_{i,j}$. PDF evaluation is key for Bayesian parameter estimation as it will be shown below. In order to overcome this issue, an approximation to Equation (2) is used through the "Copulas" method. The method has been shown to be accurate in different applications (e.g., Jin et al., 2015). Equation (5) shows this approximation. First, a vector Z_t with N elements is defined. Z_t is distributed as a multivariate normal distribution with a mean equal to a 0-valued vector and a covariance matrix equal to the covariance of $X_t|T_t$ (i.e., whose elements come from the correlogram function $\rho_{i,j}$). $\Phi[.]$ is the cumulative normal distribution and $\mathbf{1}\{.\}$ is the indicator function.

$$X_t(j) = \mathbf{1}\{\Phi[Z_t(j)] < p_t(j)\} \quad (5)$$

3. PARAMETER ESTIMATION

In this section, two parameter estimation techniques are described: a first procedure, denominated "simple approach", that estimates μ_j and α_j at each section separately and then calibrates γ to match tectonic properties of the fault, and a second procedure based on Bayesian updating that estimates μ_j and α_j at all sections and γ simultaneously.

3.1 Simple approach

This procedure was proposed in Ceferino et al. (2018a). Here, the mean interarrival rupture time μ_j and the CoV α_j at each section are estimated separately using Maximum Likelihood Estimation (MLE). Tweedie (1957) developed expressions for estimating the mean $\hat{\mu}_j$ and the variance $\hat{\sigma}_j{}^2$ of a BPT distribution from n interarrival time samples $t_{j1}, t_{j2}, \ldots, t_{jn}$ at section j. These expressions are shown in Equations (6a) and (6b). Using these expressions, the CoV $\hat{\alpha}_j$ can be found as $\hat{\sigma}_j/\hat{\mu}_j$.

$$\hat{\mu}_j = \frac{1}{n}\sum_{k=1}^{n} t_{jk} \tag{6a}$$

$$\hat{\sigma}_j{}^2 = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{\hat{\mu}_j{}^3}{t_{jk}} - \hat{\mu}_j{}^2\right) \tag{6b}$$

After estimating μ_j and α_j at each section, the correlogram parameter γ is calibrated to match (1) the annual exceedance rates of magnitudes in the tectonic region, and (2) the annual seismic moment release in different sections of the fault.

3.2 Bayesian updating

This procedure was proposed by Ceferino et al. (2018b). The details of the mathematical formulation of the procedure can be found in that study. Unlike the previous procedure, the Bayesian approach estimates the joint posterior of all the parameters of the model at once. Equation (7) shows the formula for estimating the posterior of the parameters of the model. α and μ are vectors that contain the CoVs $\{\alpha_1, \ldots, \alpha_N\}$ and the means $\{\mu_1, \ldots, \mu_N\}$ of the interarrival time at each section, respectively. The posterior equation is a fraction whose numerator is the product of the probability of observing a rupture history $P(X|\alpha, \mu, \gamma)$ and the prior of the parameters $P(\alpha, \mu, \gamma)$. The denominator of the posterior equation is the integral of this product over all the parameter space.

$$P(\alpha, \mu, \gamma|X) = \frac{P(X|\alpha, \mu, \gamma)P(\alpha, \mu, \gamma)}{\int\int\int P(X|\alpha, \mu, \gamma)P(\alpha, \mu, \gamma)d\alpha d\mu d\gamma} \tag{7}$$

$P(X|\alpha, \mu, \gamma)$ and $P(\alpha, \mu, \gamma)$ can be computed quickly. The prior $P(\alpha, \mu, \gamma)$ was taken as a multivariate lognormal distribution. All the parameters were considered mutually independent in the prior. Yet, this does not imply that the parameters will be independent in the posterior. Equation (8) shows how to evaluate $P(\alpha, \mu, \gamma)$, where each element of the product ($P(\gamma)$, and all $P(\mu_j)$, $P(\alpha_j)$) is a lognormal PDF. The application in the next section shows the procedure for prior parameter selection.

$$P(\alpha, \mu, \gamma) = P(\gamma)\prod_{j=1}^{N} P(\mu_j)P(\alpha_j) \tag{8}$$

The probability of observing the data $P(X|\alpha, \mu, \gamma)$ can be assessed as shown in Equations (9a) and (9b). These expressions were derived in Ceferino et al. (2018b). Equation (9a) shows that $P(X|\alpha, \mu, \gamma)$ can be assessed as a product of rupture occurrence during each of the H years of the catalog. Equation (9a) shows how to evaluate the probability of rupture occurrence $P(X_t|T_t)$ during each year t. This is estimated as the probability of being in the intersection of the regions defined by A_j, which is either

$Z_t(j) \leq \Phi^{-1}[p_t(j)]$ if there is a rupture at year t in section j, or $Z_t(j) > \Phi^{-1}[p_t(j)]$ otherwise.

$$P(X|\alpha,\mu,\gamma) = \prod_{t=1}^{H} P(X_t|T_t) \tag{9a}$$

$$P(X_t|T_t) = P\left(\cap_{j=1}^{N} A_j\right), \text{ where } \begin{Bmatrix} A_j = \{Z_t(j) \leq \Phi^{-1}[p_t(j)]\}\ if\ X_t(j) = 1, \\ or\ \ A_j = \{Z_t(j) > \Phi^{-1}[p_t(j)]\}\ otherwise \end{Bmatrix} \tag{9b}$$

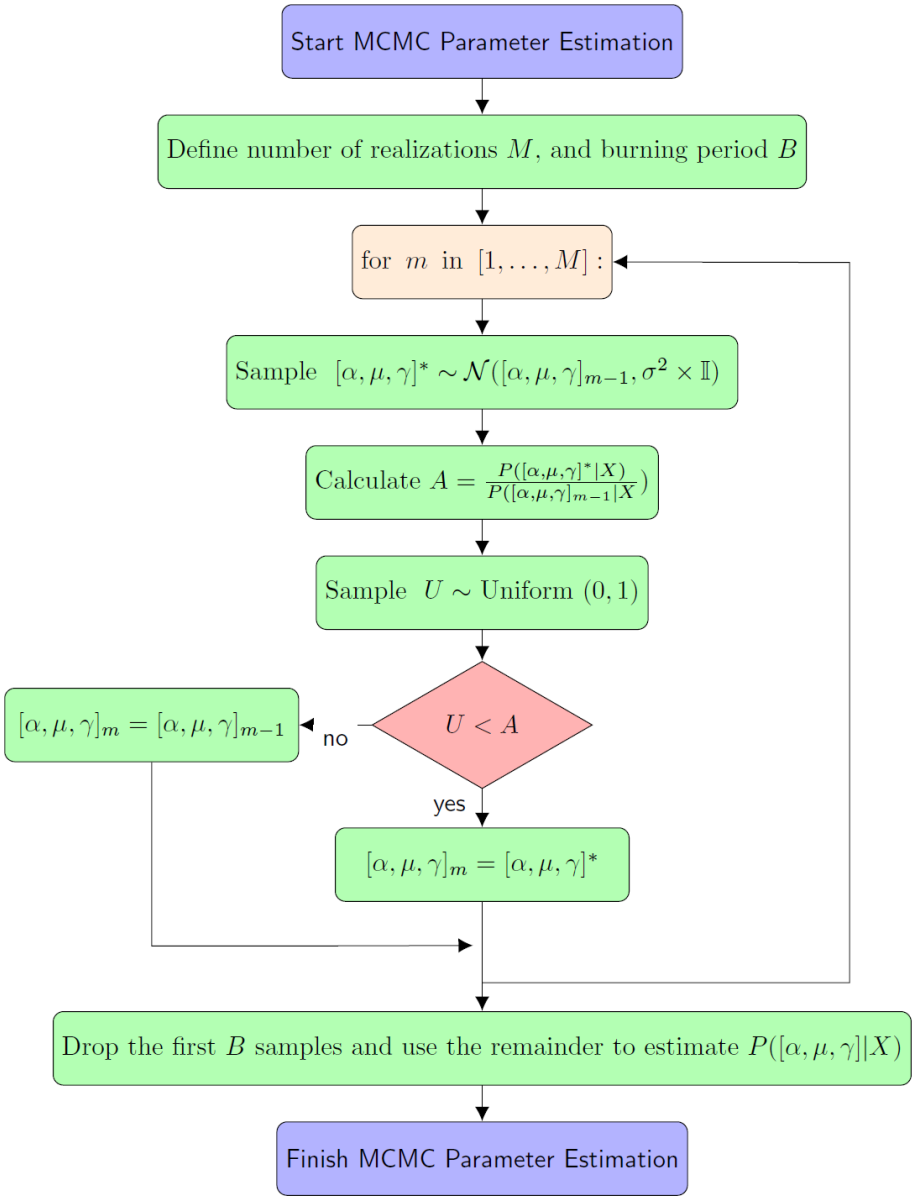


Figure 2. Algorithm for Bayesian parameter estimation.

As mentioned earlier, $P(X|\alpha,\mu,\gamma)$ and $P(\alpha,\mu,\gamma)$, whose product is the numerator of the posterior in Equation (7), can be quickly evaluated. However, the integral in the denominator of the posterior cannot be evaluated since the parameter space has high dimensionality and the probabilistic models are complex. Because direct integration is computationally unfeasible, a Markov Chain Monte Carlo (MCMC) was used to sample from the posterior. MCMC allows to sample from a PDF that can be only partially evaluated. Note that the denominator of the posterior is constant and is not a function of the parameters α,μ,γ since the integral is done through all the parameter space. This characteristic is key in order for the MCMC to work.

A Metropolis-Hasting (MH) algorithm is used to solve for the MCMC. In short, the parameter space is explored through a random walk of *M* steps. The random walk takes steps that are normally distributed, and the size of the step is empirically calibrated to have a good sample "mixing" (i.e., be large enough to explore all the parameter space, but not that large to get "lost" in low-probability regions). At each step, the relative probability gain of the new step is compared to the previous one. The new step is accepted with probability equal to the relative probability gain, or rejected otherwise. After certain number of steps, the MCMC will reach equilibrium and start sampling from the posterior distribution. The initial steps of the MCMC are usually discarded since they are considered to be the burning period of the method. Figure 2 shows the workflow of this algorithm, and the next section details its use with an application.

4. EFFECT ON SEISMIC HAZARD

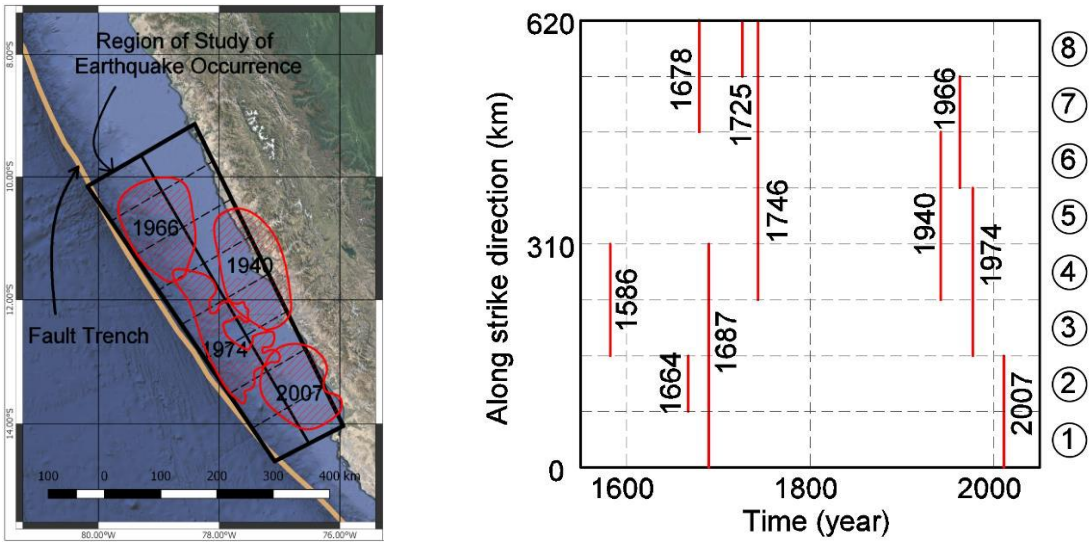
Both procedures for parameter estimation were applied here. A description of the discrepancies of these parameter estimates and their implication for earthquake hazard computation are also presented.

4.1 Earthquake data

The data from a 450-year earthquake catalog of large earthquakes in the subduction zone along the Coast of Lima, Peru, were used. Only interface events larger than Mw 7.5 occurring between the Nazca Ridge and the Mendaña Fracture were included in the catalog.

Figure 3a shows this tectonic region in a black polygon. It also shows in red four previous earthquake rupture areas occurring in this region. The subduction zone was idealized as a line along the strike direction. A section length of 77.5 km was used, which roughly corresponds to a Mw of 7.5 (Wells & Coppersmith, 1994). Therefore, the rupture of one section would be equivalent to the minimum Mw in the catalog. Then, the fault was further discretized into 8 sections (i.e., the total length of the fault is 620 km).

Figure 3b shows the 450-year earthquake catalog. The red lines represent earthquakes that rupture different fault sections (shown in the Y-axis) occurring at different years (X-axis). The ruptures were idealized to match the sections. Only ruptures larger than half the section along the strike were considered.



a) Tectonic region of analysis b) Historic catalog of earthquake ruptures

Figure 3. Earthquake data.

4.2. Simple approach

The parameters μ_j and α_j were estimated separately for each of the eight sections of the tectonic region

using Equations (6a) and 6b. The correlation parameter γ was calibrated such that the probabilistic model matches both the annual rate of magnitude exceedance in the whole tectonic region and the annual seismic moment release at each of the sections. Ceferino et al. (2018a) shows the procedure and details the result of performing such a calibration on a 450-year earthquake catalog. Table 1 displays the estimates of μ, α, and γ resulting from this approach in columns 2, 3, and 4, respectively. The 450-year catalog is scarce data since earthquake cycles take several centuries, and as a result, the parameters were estimated with just a few data points. For example, there is only one rupture interarrival of 320 years in Section 1 (in the South) since only two events occurred there in the catalog. Then, the parameter α_1 could not be estimated. Lack of data also caused few unreliable estimates of the mean interarrival time. For example, the mean μ_8 was calculated to be 34 years from two data points: 47 and 21. However, these two data points came from three ruptures occurring during the 450 years in Section 8, averaging one event each 150 years. This large discrepancy in interarrival time estimations (34 vs 150) is due to (1) the scarcity of data, and (2) that this procedure does not include the information that the next interarrival time will be longer than 271 years (because the last rupture was in 1746).

Most α values were in the range between 0.4 and 1.0. Only α_2 and α_6 were larger than 1.0, whereas α_1 was not calculated due to the insufficient data. Previous studies have shown that α (the CoV or aperiodicity of the rupture interarrival times) can take a wide set of values (Sykes and Menke, 2006). Though the α values estimated with this approach might seem high relative to existing empirical studies, such values fall within the range found in other tectonic regions.

4.3. Bayesian updating

The Bayesian updating was applied to 450-year catalog in order to estimate the posterior distribution of the parameters. The lognormal prior was set up as follows:

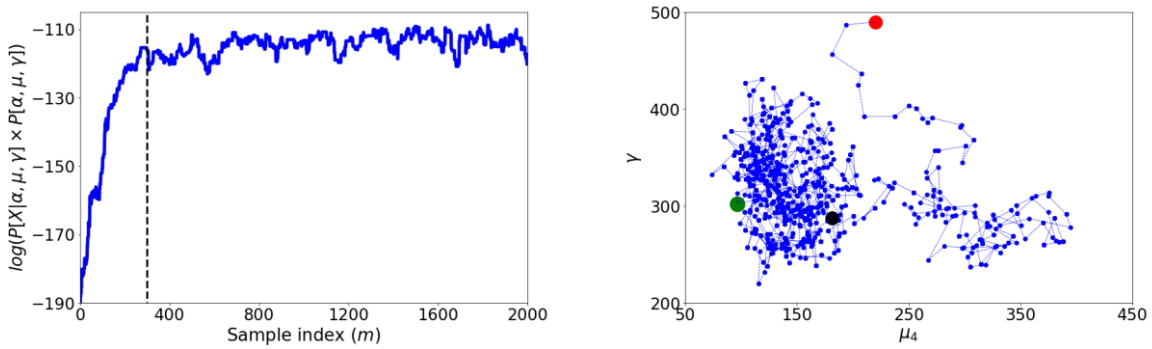
- The median of μ_j was taken as 175 years for all sections. This represents the average number of ruptured sections in the 450 years in the tectonic region. The logarithmic standard deviation of μ_j was taken as 0.3.
- The median of α_j was taken as 0.7. This value has been extensively used in other time-dependent hazard assessments with BPT distributions (e.g., Field et al., 2015). The logarithmic standard deviation of α_j was taken as 0.3.
- The median of γ was taken as 375 km, as suggested by the calibration described in the previous parameter estimation approach. The logarithmic standard deviation of γ was taken as 0.3.

Table 1. Estimated parameters using both the simple approach and the Bayesian updating.

Sections	**Simple Method**			**Bayesian Updating (MAP)**		
	μ	α	γ	μ	α	γ
1	--	--	375	148	0.93	356
2	172	1.73		140	0.77	
3	194	0.55		183	0.65	
4	97	0.7		132	0.8	
5	114	0.98		154	0.82	
6	110	1.18		114	0.69	
7	144	0.62		125	0.98	
8	34	0.41		156	0.41	

The MCMC followed the workflow outlined in Figure 2. 10,000 realizations of the posterior were

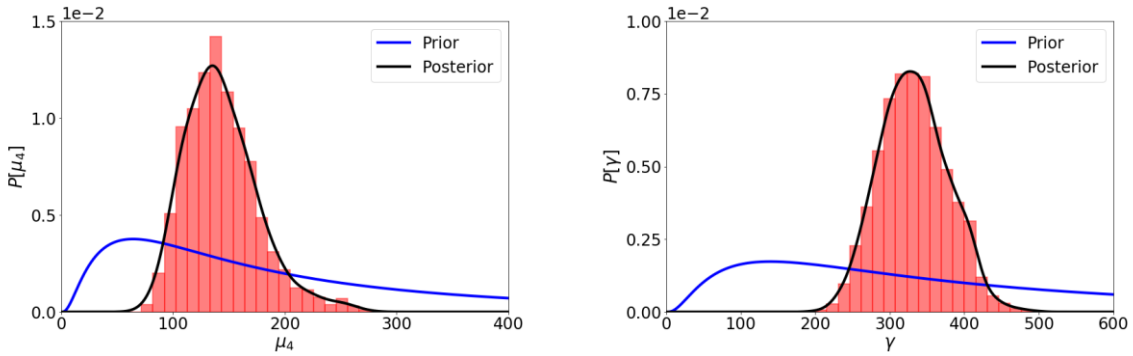
sampled with the random walk. Notice that the random walk explores a 17-dimensional parameters space (eight μ_j, α_j pairs and γ). The steps of the random walk were tested under different sizes. Finally, random step sizes with standard deviation of 12.5 for all μ_j, 0.1 for all α_j, and 17.5 for γ generated an acceptance rate of 30%. The acceptance rate measures the relative number of times that each step is accepted as a new state according to the relative likelihood gains. This level of acceptance rate is indicative of good sample "mixing" (i.e., effective exploration of the high-probability region of the parameter space) (see Chib and Greenberg (1995)). Figure 4a shows the log-likelihood gains (Y-axis) of each step in the MCMC of the first 2,000 samples (X-axis). The log-likelihood gains are measured as the logarithm of a function proportional to the posterior distribution (see Equation (7)). From the plot, the MCMC starts in points with low probability and needs nearly 300 steps of burning period to reach a stable distribution. Figure 4b shows a 2-D slice of the 17-D random walk. μ_4 and γ are shown in the X and Y axis, respectively. The red dot indicates the starting point of the random walk, the black indicates the position of the walk at step 300 (immediately after the burning period), and the green indicates the position at step 2,000. This plot shows that the random walk goes through a low-probability region at the beginning (in the path between the red and the black dots). After the burning period, the walk stays in the high probability region. Therefore, the initial 300 samples were discarded since the walk reaches the stability after the 300 first steps.



a) Log-likelihood gains b) 2-D slice of the random walk

Figure 4. Random walk in MCMC.

Figure 5a and 5b show the prior and posterior of the parameters μ_4 and γ, respectively. The blue curves are the PDFs of the prior of the parameters, the red bars correspond to the histogram of the posterior estimated using the samples from the MCMC, and the black curves are the corresponding PDF posterior estimations computed using a Gaussian kernel. It can be seen that the earthquake catalog effectively updates the distribution of the parameter and reduces the initial epistemic uncertainty from the prior.



a) μ_4 (in years) a) γ (in km)

Figure 5. Bayesian updating of the parameters of the model. Prior distribution in blue and posterior in black.

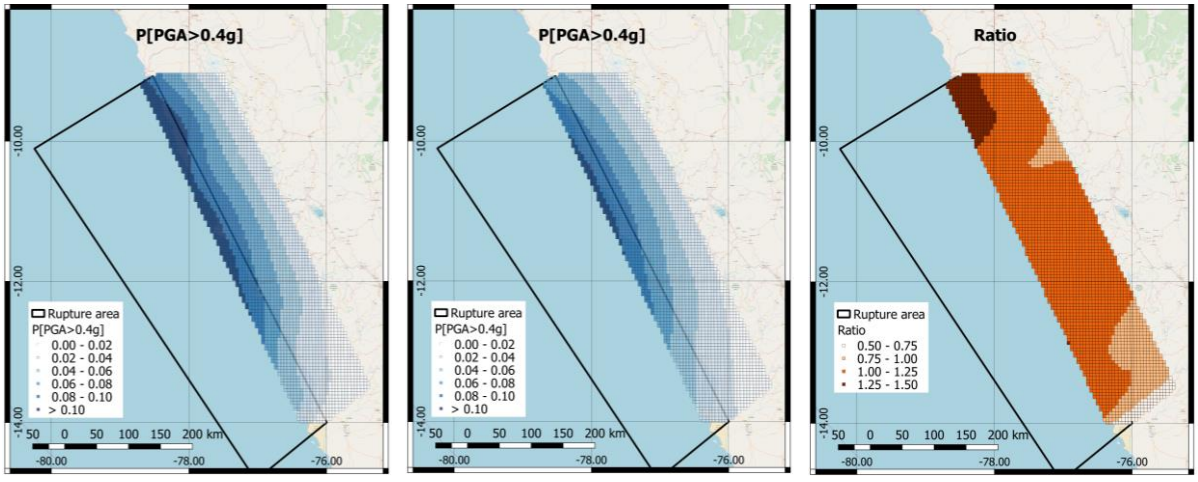
Table 1 shows the maximum a posteriori (MAP) of μ, α, and γ in columns 5, 6, and 7, respectively. The

MAP estimate is the one that equals the mode of the distribution, and it was taken as the step in the MCMC with the highest associated likelihood in the posterior distribution. The comparison of the Bayesian updating and the simple approach shows that both approaches give similar results on most of the parameters. Most of them differ by less than 25%. The Bayesian updating uses the whole catalog dataset to update all the parameters of the model, whereas the simple approach updates the parameters using only the ruptures occurring at the individual sections. Therefore, the Bayesian updating gave more reliable results to the sections with low number of data points (i.e. ruptures). For example, the Bayesian updating gave an estimate to the parameters of Section 1, which could not be obtained using the simple approach because it only has one data point. Similarly, the Bayesian approach estimated the mean interarrival time in Section 8 to be 156 years. As described previously, the simple approach estimated the mean in Section 8 to be 34 years due to three successive events happening in a short interval of time. However, on average there was 1 rupture each 150 years (three ruptures during the 450 years in Section 8), which is closer to the Bayesian estimate and largely mismatches the estimate from the simple approach. As mentioned previously, this happens due to the scarcity of data and that it is known that next interarrival time will be longer than 271 years (i.e. the last rupture was in 1746), but this information is not included in the simple approach since the next interarrival time is still unknown. The Bayesian approach overcomes this issue and gives results based on all the catalog for all parameters of the model.

Additionally, the Bayesian updating resulted in α values within a range between 0.4 and 1.0. The large α values found with the simple approach were reduced. Therefore, the resulting α values were more consistent with existing estimates from other tectonic regions (Sykes and Menke, 2006).

4.3. Hazard Comparison

Hazard calculations based on the parameters estimated from both approaches were compared. Though most of the parameter estimates were similar, the impact on the seismic hazard is explored here. The probability of exceeding a Peak Ground Acceleration (PGA) larger than 0.4g during the next 30 years due to earthquakes larger than Mw 7.5 was computed in Lima. The probability of rupture occurrence was computed using Monte Carlo on the model with parameters estimated from both approaches. Figure 6a and Figure 6b show the hazard results of the simple approach and the Bayesian updating, respectively.



a) Simple approach b) Bayesian updating c) Ratio of both approaches

Figure 6. Hazard results (P[PGA>0.4g] during the next 30 years) using the simple approach and the Bayesian updating, and the ratio between both hazard estimations.

Both hazard maps display smaller hazard in the South than in the North since the recent earthquake in 2007 ruptured the two southernmost sections in the tectonic region (in black polygon), resetting the time-dependent probability of rupture of the model. Figure 6c shows the ratio between the hazard calculated with the simple approach over the one with the Bayesian updating. The hazard differs in ± 25% in most of the region. The main difference is in the northern region, where the simple approach

overestimates the hazard by more than 25% when compared to the Bayesian updating. This occurs due to the very short mean interarrival time (mean of 34 years) from the simple approach. This mean generates more frequent ruptures in the North region than the Bayesian updating (mean of 156 years). Therefore, it can be seen that the simple approach can propagate unreliable hazard results in regions where there are only a few ruptures.

5. CONCLUSIONS

A summary of a probabilistic formulation for modeling of time and space interactions of earthquake mainshock occurrence was presented in this paper. The model discretizes a tectonic fault into small sections and uses multiple Brownian Passage Time (BPT) distributions to model rupture interarrival time. It also uses a spatial correlation function to model space interdependencies of earthquake rupture occurrence representing stress coupling and rupture propagation between neighboring sections of the fault. Additionally, two approaches for estimating the parameters of the rupture model were described in the paper. The first one is a simple approach that estimates the parameters of the BPT in each section separately. Then, it calibrates the correlation function in order to match both the annual exceedance rate of magnitudes in the whole tectonic region and the annual seismic moment release in each section. The second one is a Bayesian updating that estimates the posterior of the joint distribution of the model parameters through a Markov Chain Monte Carlo (MCMC).

This paper applies both parameter estimation approaches to a 450-year earthquake catalog in the subduction zone along the Coast of Lima. The earthquake catalog contained the location and the sizes of large interface ruptures with Mw larger than 7.5. The Bayesian updating showed to be effective at reducing the epistemic uncertainty of the parameters of the model. The comparison of the two approaches showed that both gave similar estimates (with a difference of less than ± 25% in most of the parameter estimates). However, the simple approach can give unreliable results in sections where only a few ruptures occurred. For example, in one of the fault sections, the simple approach estimated the mean interarrival time to be 34 years since three events occurred there in a very short period of time. Yet, there was one event every 150 years on average in the catalog. The Bayesian updating provided more reliable results (mean of 156 years in the section) since, unlike the simple approach, it included the complete information of the catalog to estimate all parameters simultaneously. Additionally, the Bayesian updating allowed to perform estimation for all the model parameters, whereas the simple approach was not able to estimate the parameters for fault sections with two ruptures or fewer. The effect of both parameter estimation approaches on the seismic hazard was also assessed. The hazard was estimated as the probability of exceeding a Peak Ground Acceleration (PGA) of 0.4g during the next 30 years. The hazard results were similar in most of the region of analysis (with a difference of less than ± 25%). However, where the rupture interarrival time was largely underestimated as described earlier, the hazard was largely overestimated by more than 25%.

In summary, the simple approach is easier and quicker since there are closed-form expressions for estimating the BPT parameters, and the correlation calibration procedure is intuitive to implement. The Bayesian updating requires a slightly more complex implementation since all the parameters are updated simultaneously. Additionally, the Bayesian updating can require large numbers of samples (i.e., steps) from the MCMC and also requires calibration of the random walk for good ‘sampling mixing’. Based on the above observations, modeling tests, and the application of the model, it is found that simple approach can give an initial, quicker estimation of the parameters of the model, however, the Bayesian updating proves to be more reliable in estimating the parameters and assessing the time-dependent seismic hazard.

Future work will include (1) incorporating data from physics-based earthquake rupture simulation to further decrease the epistemic uncertainty of the model parameters in the Bayesian updating, and (2) propagating the epistemic uncertainty of the model parameters to the hazard results.

6. ACKNOWLEDGMENTS

This research was supported by NSF EAGER Grant No. 1196842 and the Shah Family Fellowship through the Department of Civil and Environmental Engineering at Stanford University.

8. REFERENCES

Ceferino, L., Galvez, P., Ampuero, J.-P., Kiremidjian, A., & Deierlein, G. (2018b). Bayesian Updating of Earthquake Rupture Model using Historic and Synthetic Physics-based Earthquake Catalogs. *Bulletin of the Seismological Society of America, In Preparation*.

Ceferino, L., Kiremidjian, A., & Deierlein, G. (2017). Space and time interaction modeling of earthquake rupture occurrence. In *12th International Conference on Structural Safety and Reliability* (pp. 694–703). Vienna, Austria.

Ceferino, L., Kiremidjian, A., & Deierlein, G. (2018a). Probabilistic Formulation for Modeling of Space and Time Interactions of Earthquake Rupture Occurrence. *Bulletin of the Seismological Society of America, In Preparation*.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, *49*(4), 327–335.

Ellsworth, W. L., & Beroza, G. C. (1995). Seismic Evidence for an Earthquake Nucleation. *SCIENCE-NEW YORK THEN WASHINGTON*, 851–851.

Field, E. H., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., … Zeng, Y. (2015). Long-Term Time-Dependent Probabilities for the Third Uniform California Earthquake Rupture Forecast (UCERF3). *Bulletin of the Seismological Society of America*, *105*(2A), 511–543.

Helmstetter, A., & Werner, M. J. (2014). Adaptive smoothing of seismicity in time, space, and magnitude for time-dependent earthquake forecasts for California. *Bulletin of the Seismological Society of America*, *104*(2), 809–822.

Jin, R., Wang, S., Yan, F., & Zhu, J. (2015). Generating Spatial Correlated Binary Data Through a Copulas Method. *Science Research*, *3*(4), 206–212.

Matthews, M. V., Ellsworth, W. L., & Reasenberg, P. a. (2002). A Brownian model for recurrent earthquakes. *Bulletin of the Seismological Society of America*, *92*(6), 2233–2250.

Reid, H. F. (1911). The Elastic-Rebound Theory of Earthquakes. *Bulletin of the Department of Geology, University of California Publications*, *6*, *no*, 413–444.

Sykes, L. R., & Menke, W. (2006). Repeat times of large earthquakes: Implications for earthquake mechanics and long-term prediction. *Bulletin of the Seismological Society of America*, *96*(5), 1569–1596.

Tweedie, M. C. K. (1957). Statistical Properties of Inverse Gaussian Distribution I. *The Annals of Mathematical Statistics*, *28*(2), 362–377.

Wells, D. L., & Coppersmith, K. J. (1994). New Empirical Relationships among Magnitude, Rupture Length, Rupture Width, Rupture Area, and Surface Displacement. *Bulletin of the Seismological Society of America*, *84*(4), 974–1002.