Space and time interaction modeling of earthquake rupture occurrence

Luis Ceferino[a], Anne Kiremidjian[a] and Gregory Deierlein [a]
[a]Civil and Environmental Engineering, Stanford University

Abstract: This paper introduces an innovative probabilistic model for assessing earthquake rupture occurrence. This model is able to account for the complexity of time and space interactions of ruptures. The rupture occurrence is modeled as a Multivariate Bernoulli that is updated as a function of the time since the last rupture at different locations of the fault. A correlation model is introduced to account for the likelihood of rupture propagation. As a result, significant improvement from current methods is obtained: the ”inconsistency” problem of current models is overcome by using the Multivariate Bernoulli model. The applicability of the model is successfully tested on the subduction zone of Peru. This model was able to closely match the average release of energy and the histogram of the earthquake magnitudes on the fault.

1 Introduction

Earthquakes have caused major disruption of cities. The 2010 Haiti Earthquake caused more than 300,000 deaths, and the 2011 New Zealand Earthquake caused damage of $ 20 billion. The study of the earthquake rupture process is the first key step to analyze any earthquake consequence, from casualties to economic losses. Probabilistic models have been developed to assess the earthquake's occurrence, rupture size, magnitude, and location. Poisson models, a subset of the probabilistic models, have been widely used to assess earthquake rupture occurrence, however, they are not able to represent the interactions over time (time-dependent models) or space (i.e. rupture propagation) that characterize earthquakes [12].

Even though several models have been proposed to assess both time and space interaction of earthquake ruptures, the development of comprehensive models that assesses both interactions is still an active area of research. There are several models that consider only time interactions (e.g., Brownian Time Passage distributions (BTP) [10]). They are often used to model a characteristic earthquake of fixed or random magnitude; the magnitude is considered independent of the time between earthquake occurrences. These models built on previous methodologies that used the slip-predictable [8] or time-predictable hypothesis [1]. However, these models lack the treatment of uncertainty and interaction of ruptures occurring at different locations.

More comprehensive models integrated the time and space interactions of the earthquake ruptures (e.g. Semi-Markov process [9], (Uniform California Earthquake Rupture Forecast, version 2) UCERFv2 [3], UCERFv3 [4]). These models are based on the discretization of the fault into small segments. The Semi-Markov process initiates the rupture at one segment and propagated it to neighboring fault's segments [9]. While this model incorporates time and space interactions, after selecting the nucleation location, the rule adopted for rupture propagation is based only on the accumulated slip at the segment where the rupture starts. UCERFv2's models first estimated probabilities of any possible rupture on the faults (i.e. any possible segment rupture

configuration). Then, all rupture occurrences are sampled independently to estimate probability distributions resulting from the rupture model. Results from this procedure revealed a mistmatch between the initial and final distribution of earthquake interarrival times resulting in "inconsistency" of the model. Modifications introduced in UCERFv3 reduced the inconsistency, but did not completely remove it.
This paper proposes an innovative probabilistic model to represent the earthquake occurrence in a seismic fault that includes the time and space interactions of earthquake ruptures. An important contribution of this paper is the treatment of the earthquake rupture occurrence as a joint probability distribution of many small fault segments. As a result, this methodology completely overcomes the "inconsistency" problem found in UCERFv2 and UCERFv3. The model is tested by analyzing the rupture occurrences of large earthquakes on the subduction zone along the Coast of Lima, Peru.

2 Time and space dependent earthquake rupture model

A probabilistic model that evaluates earthquake occurrence is presented herein. This model considers the interaction over time (i.e. elastic-rebound theory) and space (i.e. multi-segment rupture and rupture propagation) of earthquake rupture occurrences in faults. Therefore, the model relies on the fundamental elastic rebound theory for earthquake occurrence [12], that states that earthquakes occur when the amount of energy accumulated between tectonic plates reaches the internal capacity. This theory implies that rupture probabilities decrease after the occurrence of a large earthquake and then increase as tectonic stresses reaccumulate over time. The probabilistic model consist of the following steps: 1) discretize the fault into small segments, 2) find the interarrival time distribution of rupture occurrence of each segment, and 3) find a spatial correlation function that models the rupture propagation probability (i.e. the influence of rupture occurrence to neighboring segments). The following explains the probabilistic models used in this framework.

2.1 Probabilistic models

Figure 1 represents an idealized fault discretized into N elements. This line can represent an strike-slip fault, or a rectangular subduction fault idealized as a line parallel to the fault strike. The elements represent the rupture units in the model, and their lengths can be associated with the smallest earthquake magnitude that can cause damage. Thus the model approximates the real rupture of earthquakes to the rupture of a certain number of units of the idealized fault. For example, the red zig-zaged line in the graph represents an earthquake rupture over three units of the fault.

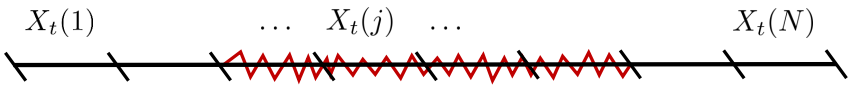

Figure 1: Discretization of the fault into small segments. The red zig-zag line represents a rupture on the fault.

Let X_t be the rupture vector at year t, where $X_t \in \{0,1\}^N$, N is the number of elements in the earthquake fault, and t is a one-year time index. Although in this paper the earthquake occurrences are tracked in one-year time intervals, the model can take different time intervals depending on the objective of the analysis. Each element in Figure 1 will have a corresponding element of the vector X_t, where $X_t(j)$ is the rupture state of the j-th element. $X_t(j)$ is equal 1 if there is a rupture (i.e. if the segment is in red in Figure 1) during year t, or 0 otherwise, where

$j = 1, 2, \cdots, N$.
Additionally, let T_t be the vector containing the time since the last earthquake up to year t on each segment, where $T_t \in \mathbb{N}^N$, N is the number of elements in the fault, and t is the year index. $T_t(j)$ is the j-th element of the vector T_t corresponding to the time since the last earthquake in the j-th element, where $j = 1, 2, \cdots N$. According to these definitions, Equation 1 holds. This equation resets the time since the last earthquake to 1 for year $t+1$ if there is a rupture on the j-th element during year t, otherwise, the the time since the last earthquake is increased by 1.

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

Using these definitions, then the statistical model is described as follows. The probability of X_t conditioned on T_t is estimated using a Multivariate Bernoulli distribution as shown equation 2, where p_t is a vector containing the probabilities of rupture during year t given T_t for each fault's element. $p_t(j)$ is the j-th element of p_t and is the marginal probability of rupture occurrence at the j-th fault's element during the year t given that the last rupture occurred $T_t(j)$ years ago. Therefore, $p_t(j)$ can be calculated as shown in equation 3, where t_j is the interarrival time between ruptures at the j-th element.

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

$$p_t(j) = P[X_t(j) = 1 | T_t(j)] = P[T_t(j) + 1 \geq t_j | t_j > T_t(j)] \tag{3a}$$

$$p_t(j) = \frac{P[T_t(j) + 1 \geq t_j > T_t(j)]}{1 - P[T_t(j) \geq t_j]} \tag{3b}$$

The rupture interarrival time t_j is modeled by a Brownian Time Passage (BTP) distribution. BTP processes represent an incremental stress accumulation on the fault that is result of a fixed stress rate and a random noise; whenever the stress reaches certain stress threshold, a rupture is generated. This model has been applied by USGS to estimate the time-dependent seismic hazard of California [4]; a more comprehensive review of the applicability of this distribution for modeling earthquake occurrences can be found in [10]. The BTP Probability Density Function (PDF) for the j-th element is given in Equation 4. The PDF is defined by the parameters μ_j and α_j.

$$P_{t_j}(t) = \Big(\frac{\mu_j}{2\pi\alpha_j t^3}\Big)^{1/2} \exp\Big(-\frac{(t-\mu_j)^2}{2\mu_j\alpha_j^2 t}\Big) \tag{4}$$

A correlation model is introduced for the Multivariate Bernoulli in order to evaluate the interaction over space of ruptures. This is the influence that a rupture at one specific segment can have on neighboring segments, or in other words, the propagation effect of ruptures to neighboring segments. The spherical correlogram, shown in Equation 5, is used for this model. This function gives the correlation of rupture occurrence between the elements i and j of the fault, and it is function of the distance $dist(i, j)$ between them. The correlogram model is defined by the parameter γ.

$$\rho_{i,j} = \exp\Big(-\Big(\frac{dist(i,j)}{\gamma}\Big)^2\Big) \tag{5}$$

Using these formulations, the model for multi-segment rupture occurrence at each year is complete. However, there is no close-form solution for the correlated Multivariate Bernoulli in terms

of the marginal probabilities of rupture occurrence p_t and the correlation model $\rho_{i,j}$. Therefore, sampling earthquake occurrences from this distribution is not possible. In order to overcome this issue, an approximated multivariate Bernoulli model is developed. First, a vector Z_t of normally distributed random variables is introduced as given in Equation 6. The mean is a zero-valued vector, and the covariance matrix is extracted from the correlogram function in Equation 5. Then, a transformation is applied to Z_t as shown in Equation 7. The inverse normal CDF function is applied to each element of Z_t. $X_t(j)$ will equal 1 if the inverse CDF of $Z_t(j)$ is smaller than $p_t(j)$, or 0 otherwise. $p_t(j)$ is obtained from Equation 3, which uses the CDF of the BPT distribution. This approximated distribution, defined in terms of Z_t, will preserve the marginal probability distributions of rupture occurrence p_t, but will slightly modify the final correlations of multi-segment ruptures. Even though the resulting distribution is an approximated version of the original, this procedure has been successfully applied numerously [6].

$$Z_t \sim \mathcal{N}(0,\Sigma), \text{where } \Sigma_{i,j} = \rho_{i,j} \tag{6}$$

$$X_t(j) = 1\{\Phi^{-1}(Z_t(j)) < p_t(j)\} \tag{7}$$

Using this procedure, the samples of multi-segment rupture occurrence can be easily obtained for each year. The next section will show a case study, where all the parameters of this model are estimated using historic earthquake data, and then applied to evaluate probabilities of future earthquake ruptures in a subduction zone.

3 Model Applied to Subduction Zones

The model is demonstrated through an application to assess the occurrence of large earthquakes on the subduction zone along the Coast of Lima, Peru.

3.1 Rupture Earthquake Data

Figure 2 shows the subduction zone off the coast of Peru. The cyan plane shown in the graph is the region of study. This plane has approximately 770 km along the fault trench, and it has been identified as one of the main sources of earthquakes in Peru [13]. This plane was idealized into a line parallel to the trench of the fault (solid blue line), and it was subdivided into ten equal-length elements. Each element has a length of approximately 77 km, which corresponds roughly to an earthquake magnitude of 7.5 according to the scaling relationship proposed in [14]. Therefore, this is the minimum magnitude that the model will analyze. This magnitude was chosen since it corresponds to the minimum magnitude of the historic earthquake catalog that was used to calibrate the model.

The seismicity of approximately 450 years was included in the analysis. This catalog is considered complete for earthquakes larger than 7.5 Mw. Table 1 shows this catalog and the sources from which this information was extracted. Figure 2 shows the rupture areas of earthquakes that were recorded by seismic instrumentation. To the right of this graph, a time line of the rupture events is shown with the corresponding rupture lengths of each event. For the pre-instrumentation events, estimates of their location were used [2]. For the post-instrumentation events, the reported areas were projected into the fault trench direction. Previous to the seismic-instrumentation time, the estimates of rupture length were based on earthquake damage [2].

Table 2 shows the years of rupture occurrence at each fault's element. Since earthquake rupture lengths do not match the discretization perfectly, the following rule was used: whenever the events ruptured less than 50% of a segment length, no rupture was considered to occur on the segment; otherwise, a rupture was considered to occur on the segment. This is not considered a limitation of the model as higher accuracy can be obtained by defining smaller element lengths

Table 1: Catalog of large earthquakes' M_w and rupture lenghts

Year	1586	1619	1664	1678	1687	1725	1746
M_w	8.1	7.85	7.5	7.85	8.4	7.5	8.6
$L_{rup}(km)$	175	125	75	125	300	75	350
Source	[2]	[2]	[2]	[2]	[2]	[2]	[2]

Year	1940	1966	1970	1974	1996	2007
M_w	8.2	8.1	7.9	8.1	7.5	8.0
$L_{rup}(km)$	230	185	115	250	80	160
Source	[7]	[7]	[7]	[7]	[11]	[11]

along the source than those considered herein. The objective here is to demonstrate how the models works.

Table 2: Years of rupture occurrence of each fault element of Figure 2

Section Label									
1	2	3	4	5	6	7	8	9	10
1687	1664	1586	1586	1746	1746	1678	1678	1996	1619
2007	1687	1687	1687	1940	1940	1746	1725		
	2007	1974	1746	1974	1966	1966	1970		
			1940						
			1974						

3.2 Model calibration

First, the marginal probabilities of rupture interarrival times for all the ten segments of the fault are found. For elements with at least three ruptures in the catalog (i.e. at least two interarrival time samples), the parameters μ_j and α_j of Equation 4 are calculated using the Maximum Likelihood Estimator (MLE) method. For elements with two ruptures (i.e. only one interarrival time sample), the MLE can not be applied to estimate both parameters. Consequently, the Equal Moment method is used. The mean is set up to the only existing sample time, and the coefficient of variation is set to the average coefficient of variation of the elements with at least three ruptures; this coefficient is 0.92. For the elements that ruptured only once in catalog, i.e. no intearrival time sample, a mean of 450 years and a coefficient of variation of 0.92 is assumed.
Different values of γ of the correlogram model, shown in Equation 5, are studied to determine the ones that best approximate the earthquake data. The following values of γ were analyzed: 96, 193, 289, 385, and 481 km. Figure 3 shows the correlogram functions for each of these values.

3.3 Results

This study uses a 100,000-year simulation of the rupture process using Monte Carlo analysis. The sampling procedure of rupture occurrence for a general n_{years} number of years is shown in Algorithm 1. This algorithm uses the calibrated parameters found in the previous subsections.
Figure 4a and b show one realization of the rupture process for the next 10,000 years using a γ values equal to 96 and 289 km, respectively. The graphs indicate that as the value of γ increases, the number of events decreases and rupture lengths increase. This occurs since large γ values

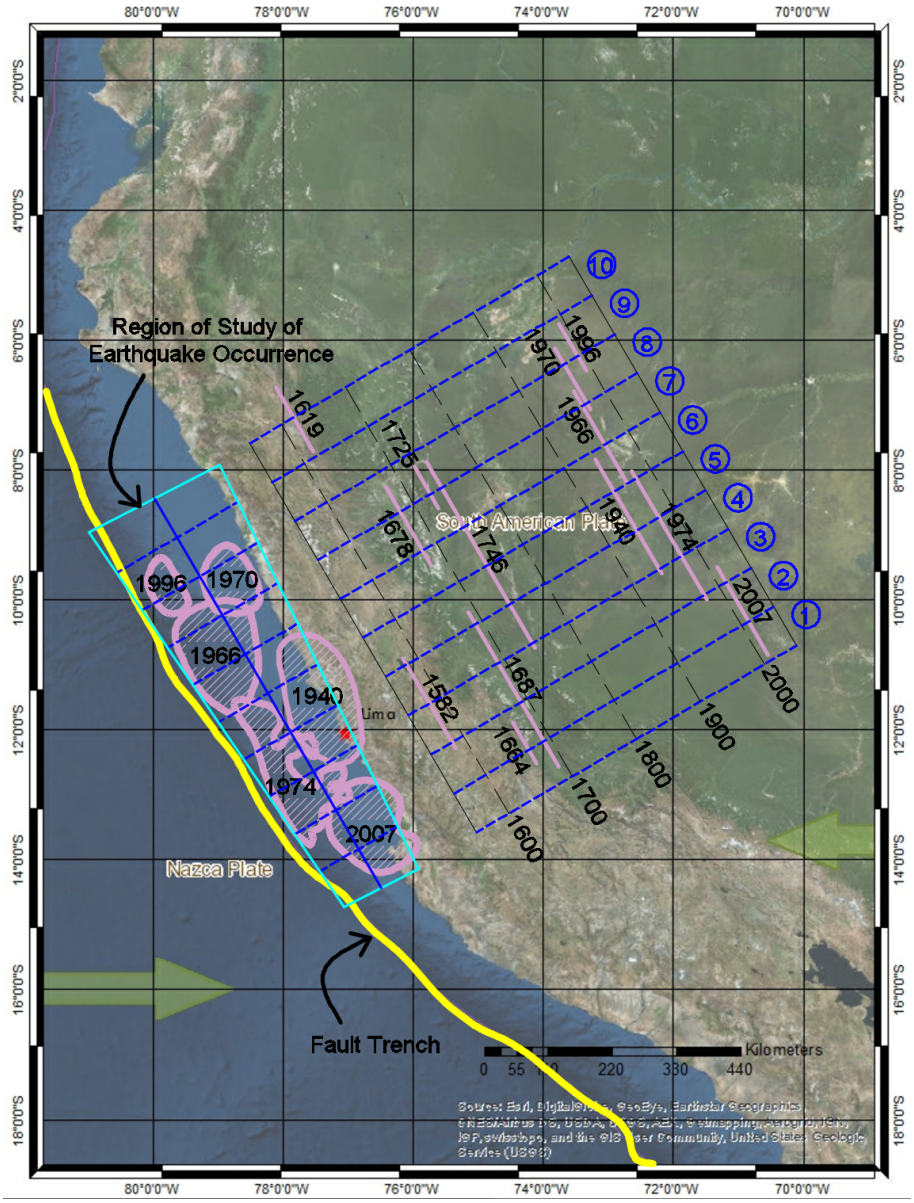

Figure 2: Region of study of large M_w earthquake occurrence (cyan quadrilateral), and discretization of this region in 10 equal-lenght sections (dashed blue lines). Rupture lengths of previous large earthquakes are shown (pink lines) on their respective sections. The label of each section is shown inside of the pink shapes.

generate stronger correlations as seen in Figure 3, and therefore, the probability of earthquake propagation to neighboring elements increases so that the elements tend to rupture together rather than individually.

In order to verify that the model can replicate the complexity of the rupture process in this sudbuction zone, the average annual release of energy in the historical catalog is compared to the ones obtained from the model for the different γ values. The energy release of the catalog is calculated by uniformly distributing the seismic moment, M_0, over each of the elements that rupture. M_0 is calculated from the relationship proposed by Hanks and Kanamori (Equation 8, where M_0 is in units of $10^{-7}Nm$) [5] using the M_w values from Table 1. At each element, all releases of M_0 were added and divided over the 450 years. This is shown in black in Figure 5a. An identical procedure is followed to estimate the average annual release of M_0 for the

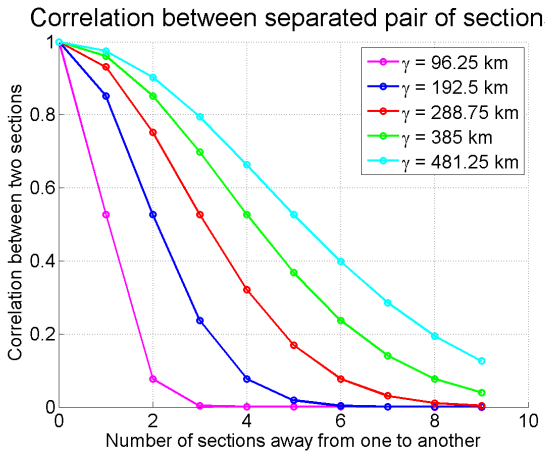

Figure 3: Exponential variogram for $\gamma = 96$, 193, 289, 385, and 481 km .

Algorithm 1 Sample rupture realizations

procedure SAMPLING(n_{years})
  Initialize Σ as a NxN matrix
  for $j = 1, \cdots, N$ **do**
    Initialize $T_1(j)$ as the number of years since the last rupture at segment j until today.
    for $k = 1, \cdots, N$ **do**
      Calculate $\Sigma(j,k)$ as $\rho_{j,k}$ in Equation 5.
    end for
  end for
  for $t = 1, \cdots, n_{\text{years}}$ **do**
    Sample Z_t using Equation 6.
    for $j = 1, \cdots, N$ **do**
      Calculate $p_t(j)$ using Equation 3.
      Find X_t using Equation 7.
      Find T_{t+1} using equation 1
    end for
  end for
end procedure

simulation; M_0 is averaged over the 100,000 years of simulation. These results are also shown in Figure 5a. This plot indicates that the model can reproduce the shape of the annual average release of M_0. The γ values of 193 and 289 km give the results that match the catalog the best. While there is a "bump" at element 3 that the model cannot represent well, the energy releases at all other segments are well represented.

$$M_w = \frac{2}{3} log(M_0) - 10.7 \tag{8}$$

The histogram of M_w of the historic catalog is next compared to the histograms resulting from the model. Figure 5b shows this comparison. The histogram from the historic catalog is shown in black. These results indicate that γ values of 193 and 289 km match very well the distribution of M_w from the historic catalog. These γ values were the same ones that showed the best agreement in the average annual release of M_0 in Figure 5a, therefore, from both perspectives: preserving the average annual release of M_0 and the histogram of M_w, these values can be used to represent

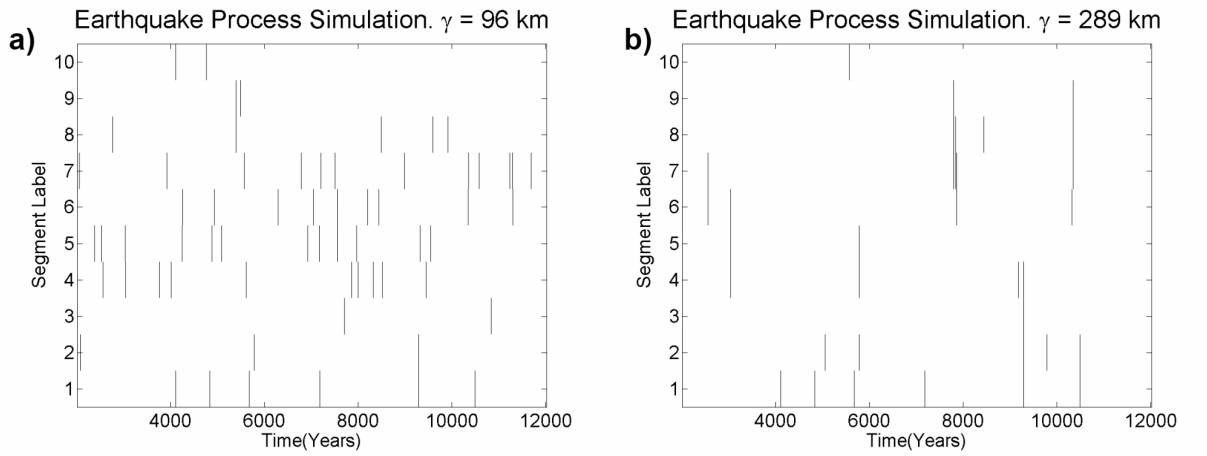

Figure 4: 10,000 years of simulated rupture occurrence. a) γ = 96, b) γ = 289 km.

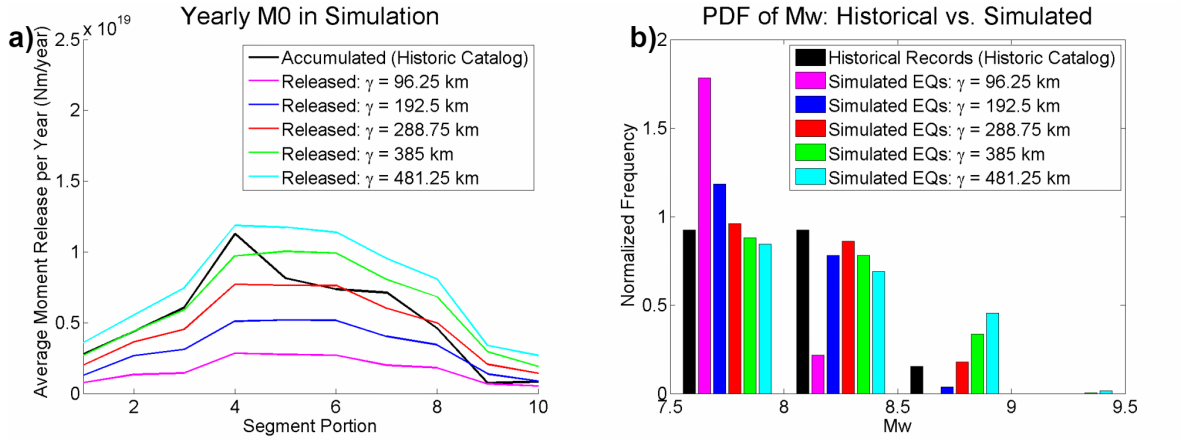

Figure 5: Historic catalog vs. model results. a) Average annual M_0 release, b) distribution of M_w

the rupture propagation behavior for large earthquakes on the subduction zone along the Coast of Lima.

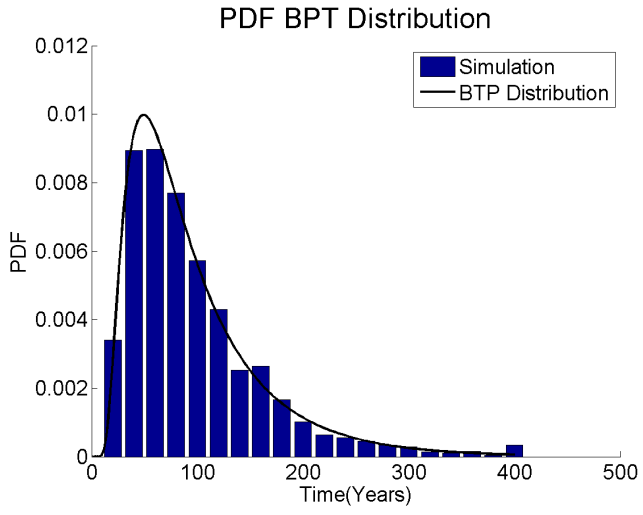

Figure 6: Comparison between initial and simulated distribution of rupture interarrival time of element 4 of the fault.

Finally, the interarrival time distribution of segment 4 is compared to the results from the simulations. As stated previously, one main challenge of the time-dependent statistical models is to overcome the ”inconsistency” of previous models: the initial assumed distributions of rupture interarrival time were different from the ones resulting from the simulations. Specifically, there was an overestimation of the probabilities at the lower tail of the distribution [3, 4]. In order to overcome this inconsistency, an intrinsic, ”consistent” model is applied. The multi-segment rupture occurrences are modeled using a joint probability distribution with a spatial correlation model scheme. Therefore, marginal distributions of interarrival time distributions at each segment remain the same in the simulations. Figure 6 shows the initial assumed distribution of rupture interarrival time for segment 4 in black. This distribution is compared to the histogram resulting from the simulation in Algorithm 1. Since the model is intrinsically consistent, it can be seen that this histogram converges to the black curve.

4 Conclusions

This paper presents a probabilistic model for assessing earthquake rupture occurrence. The aim of this model is to extend current probabilistic methodologies to comprehensively evaluate the interactions over time and space (i.e. rupture propagation) of earthquake ruptures. The ”inconsistency” of current methodologies is addressed by modeling multi-segment earthquake rupture using a multivariate Bernoulli distribution. The multivariate distribution is function of 1) marginal rupture probabilities at different fault segments estimated according to the time since the last rupture at each segment, and 2) a spatial correlation function that models rupture propagation from individual segments to their neighbors.

The case study evaluates earthquake rupture occurrence of large events in the subduction zone of Lima using the model presented herein. The correlation function is calibrated to find the parameters that best represented the rupture process of this subduction zone. Results from this calibration indicated that this model is capable of matching 1) the average seismic moment release on each segment of the fault with reasonable accuracy, and 2) the histogram of earthquake magnitude occurrence on the fault of the earthquake catalog of the last 450 years. Additionally, the ”consistency” of the methodology is displayed by showing that the proposed probability distribution of interarrival times at segments matched the simulations.

These findings show that the model can capture the complexities of the earthquake rupture process. Still, further work is needed to make this model more robust: 1) Physics-based models can be coupled with this probabilistic methodology to enlarge the timespan of the historic earthquake catalogs. Large earthquakes have interarrival times of hundreds, and sometimes even thousands of years, and therefore, this information cannot be estimated with statistical confidence by using historic catalogs that do not have a long timespan. Catalogs resulting from physics-based models can address this issue. 2) The functional form of the spatial correlation model of earthquake occurrence can be directly explored from data. In this paper, an spherical form was used, however, historical or physics-based catalogs can be used to evaluate the validity of this form.

Acknowledgement

This research was partially supported by NSF Grant 1645335 and the Shah Family Fellowship through the Department of Civil and Environmental Engineering at Stanford University. The authors are grateful for their generous support.

References

[1] T. Anagnos and A. S. Kiremidjian. “STOCHASTIC TIME-PREDICTABLE MODEL FOR EARTHQUAKE OCCURENCES”. In: *Bulletin of the Seismological Society of America* 74.6 (1984), pp. 2593–2611.

[2] L. Dorbath, A. Cisternas, and C. Dorbath. “Assessment of the size of large and great historical earthquakes in Peru”. In: *Bulletin of the Seismological Society of America* 80.3 (1990), pp. 551–576.

[3] E. H. Field and V. Gupta. *Conditional, Time-dependent probabilities for segmented Type-A faults in the WGCEP UCERF 2, Appendix N in The Uniform California Earthquake Rupture Forecast, version 2 (UCERF 2): U.S. Geological Survey Open-File Report 2007-1437N and California Geological*. Tech. rep. Pasadena, CA: United States Geological Survey, 2008, pp. 1–29.

[4] E. H. Field et al. “Long-Term Time-Dependent Probabilities for the Third Uniform California Earthquake Rupture Forecast (UCERF3)”. In: *Bulletin of the Seismological Society of America* 105.2A (2015), pp. 511–543.

[5] T. C. Hanks and H. Kanamori. “A moment magnitude scale”. In: *Journal of Geophysical Research* 84.B5 (1979), pp. 2348–2350. arXiv: ??.

[6] R. Jin et al. “Generating Spatial Correlated Binary Data Through a Copulas Method”. In: 3.4 (2015), pp. 206–212.

[7] H. Kanamori. “The energy release in great earthquakes”. In: *Journal of Geophysical Research* 82.20 (1977), p. 2981.

[8] A. S. Kiremidjian and T. Anagnos. “STOCHASTIC SLIP-PREDICTABLE MODEL FOR EARTHQUAKE OCCURRENCES”. In: *Bulletin of the Seismological Society of America* 74.2 (1984), pp. 739–755.

[9] K. A. Lutz and A. S. Kiremidjian. “A Stochastic Model for Spatially and Temporally Dependent Earthquakes”. In: *Bulletin of the Seismological Society of America* 85.4 (1995), pp. 1177–1189.

[10] M. V. Matthews, W. L. Ellsworth, and P. a. Reasenberg. “A Brownian model for recurrent earthquakes”. In: *Bulletin of the Seismological Society of America* 92.6 (2002), pp. 2233–2250.

[11] M. Motagh et al. “Coseismic slip model of the 2007 August Pisco earthquake (Peru) as constrained by Wide Swath radar observations”. In: *Geophysical Journal International* 174 (2008), pp. 842–848.

[12] H. F. Reid. “The Elastic-Rebound Theory of Earthquakes”. In: *Bulletin of the Department of Geology, University of California Publications* vol. 6, no (1911), pp. 413–444.

[13] H. Tavera et al. *Mapas de Peligro Sísmico para el Perú*. Tech. rep. Instituto Geofísico del Perú, 2014.

[14] D. L. Wells and K. J. Coppersmith. “New Empirical Relationships among Magnitude, Rupture Length, Rupture Width, Rupture Area, and Surface Displacement”. In: *Bulletin of the Seismological Society of America* 84.4 (1994), pp. 974–1002.