

RESEARCH LETTER

Open Access



Statistical characterization of full-margin rupture recurrence for Cascadia subduction zone using event time resampling and Gaussian mixture model

Katsuichiro Goda^{1,2*}

Abstract

Earthquake occurrence modeling of large subduction events involves significant uncertainty, stemming from the scarcity of geological data and inaccuracy of dating techniques. The previous research on statistical modeling of full-margin ruptures of the Cascadia subduction zone attempted to address these issues. However, the adopted resampling method to account for the uncertain marine turbidite age data from the Cascadia subduction zone was not sufficient in the sample size. This study presents a statistical approach based on the Gaussian mixture model applied to significantly larger resampled Cascadia age data. The results suggest that the 3-component Gaussian mixture model outperforms the 2-component Gaussian mixture model and the 1-component renewal models by capturing the long gap and short-term clustering. The developed Gaussian mixture model is well suited to apply to probabilistic seismic and tsunami hazard analysis and the calculation of long-term probability of the future full-margin Cascadia events by considering the elapsed time since the last event.

Keywords Cascadia subduction earthquakes, Gaussian mixture model, Inter-arrival times, Data uncertainty

Introduction

Earthquake occurrence modeling of large magnitude events from major subduction zones and faults is a critical element in probabilistic seismic and tsunami hazard analysis, but involves significant uncertainty (Davies et al. 2018; Baker et al. 2021; Behrens et al. 2021; Mori et al. 2022). The main challenges of this modeling stem from the scarcity and the uncertainty/ambiguity of the data (Rhoades et al. 1994; Headquarters for Earthquake Research Promotion 2019). For the former, despite various efforts for identifying large paleo-seismic events from

geological data in major subduction zones (Philibosian and Meltzner 2020), the number of large earthquakes is rather small compared to the number of small-to-moderate earthquakes. For the latter, event times are often estimated based on radiocarbon dating of sediment samples and geological cores, and estimated event times are uncertain with a possible margin of error of several decades to a few hundred years.

By recognizing non-Poissonian and quasi-periodic recurrence characteristics of major subduction zones and matured faults, various earthquake occurrence models have been proposed in the literature (e.g., Ogata 1999; Sykes and Menke 2006). Both physics-inspired occurrence models (e.g., Shimazaki and Nakata 1980; Console et al. 2008) and statistical renewal models (e.g., Cornell and Winterstein 1988; Matthews et al. 2002; Abaimov et al. 2008) have been adopted in seismic and tsunami hazard assessments. A renewal process can characterize

*Correspondence:

Katsuichiro Goda
kgoda2@uwo.ca

¹ Department of Earth Sciences, Western University, London, Canada

² Department of Statistical and Actuarial Sciences, Western University, London, Canada

the evolution of occurrence probability with time by specifying an inter-arrival time (IAT) distribution of earthquakes. Popular IAT distributions include the log-normal distribution, Brownian Passage Time distribution (Matthews et al. 2002), and Weibull distribution (Abaimov et al. 2008), noting that the homogeneous Poisson model corresponds to the exponential distribution. Typically, the IAT distribution is characterized by three parameters: mean recurrence time, coefficient of variation (CoV), and elapsed time since the previous event. Although the concept of renewal-type earthquake occurrence models is appealing, when the number of earthquake occurrence time data is small, the superiority of the time-dependent renewal model to the time-independent Poisson model is not always obvious (Williams et al. 2019; Griffin et al. 2020).

The Cascadia subduction zone (CSZ) is located at the convergent boundary where the oceanic Juan de Fuca Plate, Gorda Plate, and Explorer Plate and the continental North American Plate interact (Fig. 1). The CSZ extends 1100 km from British Columbia to Northern California and has a convergence rate of 30 to 45 mm/year (DeMets et al. 2010). In the past, the CSZ ruptured in full-margin earthquakes of moment magnitude (M_w) greater than 8.5 with an average recurrence period of 530 years (Goldfinger et al. 2012) and the last event in 1700 (Satake et al. 2003). The sedimentary records of the CSZ events can be observed in the form of marine turbidite deposits (Goldfinger et al. 2012) and buried

soil/marsh data (Atwater et al. 2015). Using the Cascadia turbidite records produced by Goldfinger et al. (2012), Kulkarni et al. (2013) carried out statistical modeling and seismic gap analysis of full-margin CSZ ruptures, whose magnitude ranges were between 8.49 and 9.13, by using a hierarchical clustering method. They pointed out that there may be a relationship between the earthquake size and the inter-arrival time, noting that the earthquake size can be correlated with the fault rupture length and the thickness of the turbidite layers or turbidite volumes. Recognizing the significant uncertainty associated with the radiocarbon dating of deep-sea geological cores and synchronous turbidite event analysis, Monte Carlo resampling of the CSZ age data was conducted; however, the size of the resampled CSZ age data was small (20 simulated catalogs; each contains 18 events). They proposed that the Cascadia full-rupture records can be divided into several clusters, punctuated by longer gaps. Their investigation was motivated to present statistical evidence of clustering and seismic gap as plausible earthquake occurrence patterns in the CSZ.

This study conducts statistical modeling of full-rupture CSZ recurrence using event time resampling and Gaussian mixture method. The same information of the past CSZ age data suggested by Goldfinger et al. (2012) is used, and the same resampling approach as in Kulkarni et al. (2013) is adopted to dealing with the uncertainty of the CSZ age data. Each event can be represented by multiple possible CSZ age data, and the uncertainty of each age data, associated with radiocarbon dating, can be approximated by a triangular distribution based on the best, lower, and upper estimates. The statistical modeling carried out in this study differs from that of Kulkarni et al. (2013) in two major ways. Firstly, the size of the resampled CSZ age data is significantly larger (5000 vs. 20 simulated catalogs). By treating the simulated catalogs as possible realizations of the underlying original age data, more stable characteristics of the CSZ age data can be examined. Secondly, the Gaussian mixture model is fitted to the IAT distribution based on the simulated CSZ age data, which exhibit multi-component features. A major advantage of the proposed method is that when the elapsed time since the last event is considered in probabilistic seismic and tsunami hazard analysis, the mixing proportions of the multiple normal distributions can be updated with this information. The important contributions of this research include the development of a viable statistical method for the full-margin CSZ ruptures by capturing the multi-component characteristics of the IAT data and the implementation of the developed method in the context of probabilistic seismic and tsunami hazard analysis.

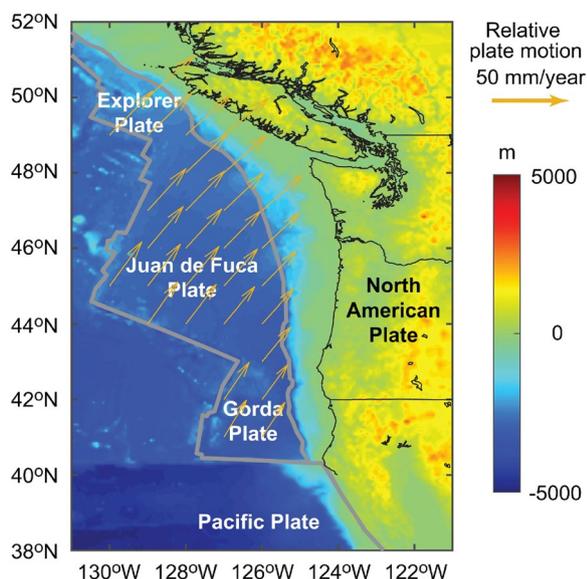


Fig. 1 Tectonic plates surrounding the Cascadia subduction zone. The relative plate motions between the Juan de Fuca, Gorda, and Explorer Plates and the North American Plate are based on DeMets et al. (2010)

Data

Event times of full-margin ruptures in the Cascadia subduction zone

The CSZ extends from Vancouver Island to Cape Mendocino (Fig. 1), and multiple submarine channel-canyon systems are distributed along the continental margin of the CSZ. During major storms, earthquakes, and tsunamis, turbidity currents can be triggered by river-carried sandy and silty sediments, sliding down the continental shelf. From branching tributaries, turbidity currents merge into the main channel and form a large turbidite. By contrast, small-to-moderate storms and far-field tsunamis are unlikely to induce synchronized turbidity currents along the entire continental margin of the CSZ.

To develop a catalog of megathrust subduction earthquakes in the CSZ, Goldfinger et al. (2012) carried out extensive coring surveys spanning the entire margin of the CSZ and analyzed collected samples using marine radiocarbon dating and stratigraphic correlation techniques. Due to the synchronous occurrence of turbidite currents along the 500-km coast, the northern CSZ events are best explained by the paleo-seismic events that can trigger a CSZ-wide event. For the southern CSZ, most of the turbidite samples are well-corresponded and correlated with the spatial extent of onshore paleo-seismic records, while there are uncorrelated turbidites that were likely to be depositional products after smaller earthquakes, local storms, or far-field tsunamis. The northern portion of the CSZ ruptured less frequently with mean recurrence periods of 500 to 530 years and has a strong spatial correlation with the southern half, thus leading to a synchronized full or near-full rupture which could result in M_w 9-class megathrust events. The southern portion of the CSZ, in addition to the whole-region ruptures, experienced additional smaller earthquakes (M_w 8-class events) according to the turbidite records with mean recurrence periods of 240 to 320 years. According to Goldfinger et al. (2012), 19 well-dated turbidites that ruptured the northern CSZ were identified and denoted by T1 to T18, noting that there is T17a which was considered as a separate full-margin rupture from T17. T1 corresponds to the most recent 1700 event, whereas T18 corresponds to the oldest event

in 9795 calibrated years before the present. Among the 19 turbidite events, T2 was eventually excluded from the list of full-margin events because this event was not consistently recorded on buried soil/marsh data on land, implying that T2 might have been caused by non-seismic sources.

Kulkarni et al. (2013) adopted the full-margin rupture data by Goldfinger et al. (2012) to develop an earthquake clustering model for the CSZ (see Table 1 of Kulkarni et al. (2013)). The data consist of 83 turbidite ages for the 19 events (i.e., T1 to T18, including T2 and T17a); different events have different numbers of age data (for instance, T3 has 8 age data). Each of the turbidite data comes with three age estimates, i.e., best, +2 sigma bound, and -2 sigma bound. Kulkarni et al. (2013) considered that the probability distribution of the individual age data can be represented by the triangular distribution with the best estimate as the mode and the two-sigma bounds as the upper and lower limits. The same interpretations of the turbidite data are adopted in this study. The data used by Kulkarni et al. (2013) are shown in Fig. 2a. Kulkarni et al. (2013) observed several long gaps after large-magnitude events that produced large turbidite masses (see Fig. 2 of Kulkarni et al. (2013)) and hypothesized that a large earthquake magnitude is associated with a longer inter-arrival time. They were motivated to develop an earthquake occurrence model that distinguishes inter-cluster data and gap data using a hierarchical clustering method. After excluding T2 from the dataset, they obtained the 17 IAT data from the 18 age data. They indicated that a renewal model with the normal IAT distribution performs well with the central portion of the IAT data but is not suitable for characterizing the lower and upper tails. This can be seen in Fig. 2b. More details of the seismic gap analysis for the CSZ can be found in Kulkarni et al. (2013).

Monte Carlo resampling of the Cascadia age data

Due to nonnegligible uncertainty of the CSZ age data from Goldfinger et al. (2012), Kulkarni et al. (2013) conducted 20 Monte Carlo resampling of the CSZ age data and used for their hierarchical clustering analysis. The procedure of the resampling is summarized as:

Table 1 Parameters of the fitted Gaussian mixture models

Number of components	AIC	Component 1 [Proportion, mean, standard deviation]	Component 2 [Proportion, mean, standard deviation]	Component 3 [Proportion, mean, standard deviation]
1	1,194,100	[1.0, 561.6, 271.8]	–	–
2	1,186,767	[0.880, 493.6, 204.6]	[0.120, 1059.3, 162.5]	–
3	1,185,057	[0.646, 503.5, 139.2]	[0.240, 905.6, 223.8]	[0.114, 166.8, 95.0]

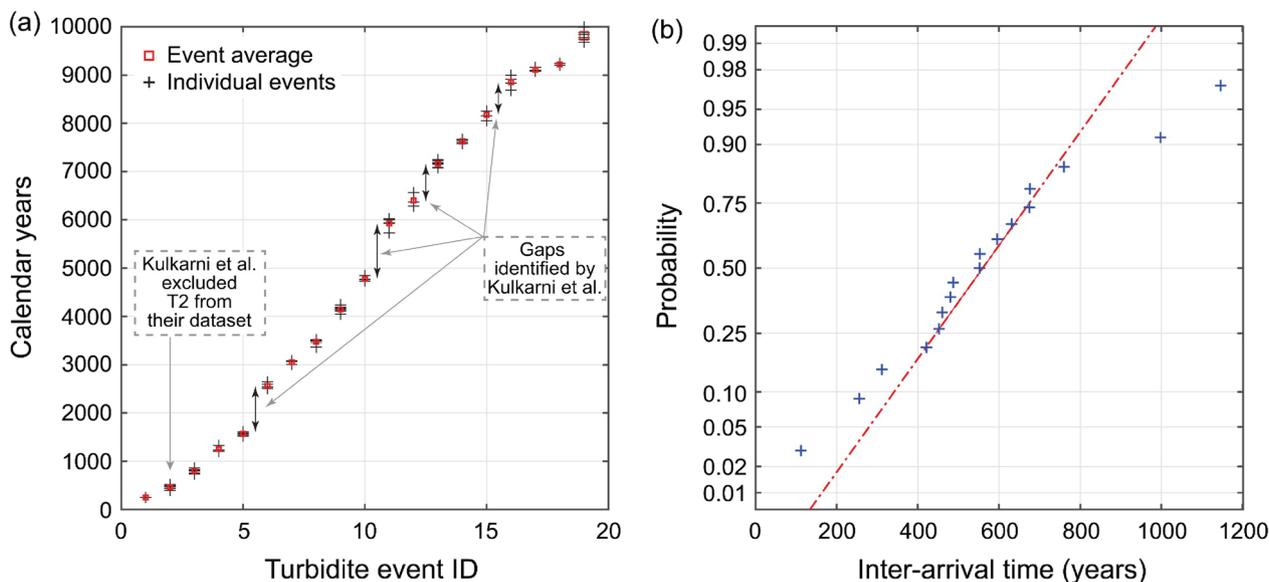


Fig. 2 **a** Cascadia subduction zone age data by Goldfinger et al. (2012). **b** Normal probability plot of the inter-arrival time data; the dataset excludes T2 shown in **a**, following the same approach taken by Kulkarni et al. (2013)

1. Set the number of Monte Carlo resampling.
2. For each turbidite event, choose one of the data randomly with equal chance (i.e., all listed data in Table 1 of Kulkarni et al. (2013) are regarded as equally reliable).
3. Sample the age of the turbidite data chosen in Step 2 from the triangle distribution, which is determined by the best and ± 2 sigma bounds.
4. Repeat Steps 2 and 3 for all turbidite data. The IAT data can be obtained for each catalog.
5. Repeat Steps 2 to 4 for the resampling number specified in Step 1.

Because the radiocarbon dated ages involve large uncertainty and some of the adjacent turbidite events are only separated by a few hundred years, the above-mentioned resampling procedure can result in the reversed order of the simulated events. In this study, when the simulated CSZ age catalog has negative IAT data (i.e., reversed order), the simulated trial is discarded; the rejection rate is approximately 20%. Note that Kulkarni et al. (2013) considered the simulated CSZ age catalogs that have IAT data longer than 100 years only because the shortest IAT in the original CSZ age catalog was about 100 years. This resampling method results in a high rejection rate of the simulated catalogs (approximately, 62% of the simulated catalogs are abandoned). Since the threshold of 100 years is arbitrary and this high rejection rate could cause bias in statistical modeling, all IAT data are required to be positive (i.e., data do not have to be

greater than 100 years). It is noteworthy that the analyses are also carried out by allowing negative IAT data (i.e., no rejection); the results are not significantly different from the baseline results. Although the simulation procedure adopted in this study is the same as Kulkarni et al. (2013), considering a larger resampling size and adopting a less subjective rejection criterion of the simulated turbidite age data will lead to robust characteristics of the simulated CSZ event time data.

Figure 3 shows a histogram of the Cascadia IAT data distribution from 5000 resampling simulations. Each simulated catalog consists of 18 events and 17 IAT data can be calculated (note: all IAT data are positive). Therefore, the histogram is based on 85,000 IAT data. The simulated IAT data are right skewed and exhibit heavy tails on both upper and lower sides, compared with the normal distribution (see Fig. 2b). The longer IAT data are associated with long gaps, whereas the shorter IAT data are related to short-time clustering. The mean and standard deviation of the simulated IAT data are 561 years and 272 years, respectively, and thus the CoV is calculated as 0.485.

Methods

A Gaussian mixture model comprised multiple Gaussian components, each captures clustering in the data. For one-dimensional data, such as the IAT data for the full-margin Cascadia rupture case, the model for K components can be expressed by:

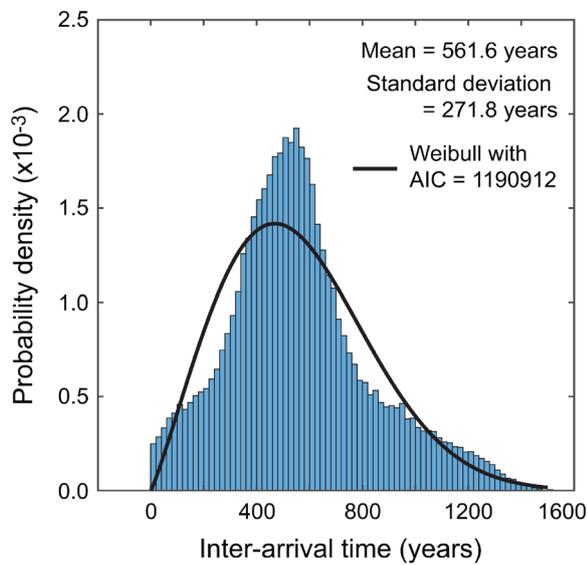


Fig. 3 Histogram of the inter-arrival time data from 5000 simulations and comparison with the 1-component Weibull-based renewal model. The simulated earthquake catalogs that contain negative inter-arrival times are excluded

is achieved or the iteration number reaches the specified trial value. Since the success of the EM algorithm depends on the data complexity and initial values and the solution may converge to a local minimum, multiple runs of the EM algorithm are made. The results with the highest log likelihood value is adopted as the final estimate. In this study, the parameter estimation is performed using the *fitgmdist* function in MATLAB, and the number of EM initializations is set to 50.

The Gaussian mixture model is fitted to the simulated IAT data by considering $K = 1, 2,$ and 3 . The inspection of the histogram shown in Fig. 3 indicates that the central, lower, and upper portions of the IAT data can be represented by different components. When $K = 1$, the Gaussian mixture model is equivalent to a renewal model with the normal IAT distribution. In this sense, the goodness-of-fit of renewal models with other IAT distributions (e.g., lognormal and Weibull) can be compared with the 1-component Gaussian mixture model. To facilitate the comparison of the Gaussian mixture models with different numbers of model parameters, the Akaike Information Criterion (AIC) can be used:

$$AIC = 2N_p - 2\ln L, \tag{2}$$

$$f(x) = \sum_{i=1}^K \pi_k N(x|\mu_k, \sigma_k) = \sum_{i=1}^K \frac{\pi_k}{\sqrt{2\pi\sigma_k}} \exp\left(-\frac{(x - \mu_k)^2}{2\sigma_k^2}\right), \tag{1}$$

where π_k is the mixing proportion of the k th component and its summation over all K components equals 1, and μ_k and σ_k are the mean and standard deviation of the k th component, respectively. The k th mixing proportions represents the probability of observing data that come from the k th Gaussian component. With the increase of K , the number of model parameters increases; for instance, for $K = 1, 2,$ and 3 , the number of parameters is 2, 5, and 8, respectively.

The model parameters of the Gaussian mixture model can be estimated using the Expectation–Maximization (EM) algorithm (McLachlan and Peel 2000). The algorithm attempts to maximize the log likelihood function of the Gaussian mixture model for given data in two steps iteratively. For the specified value of K , initial values for component means, covariance matrices, and mixing proportions are generated through a k -means++ technique. In the Expectation step, the algorithm computes posterior probabilities of component memberships for each data point. Subsequently, in the Maximization step, with the component membership posterior probabilities, component means, covariance matrices, and mixing proportions are estimated based on the maximum likelihood method. The EM steps are iterated until the convergence

where N_p is the number of model parameters and $\ln L$ is the log likelihood value of the model. For the same data, a model with a smaller AIC value is superior. The AIC values shall be used as model performance indicator in a relative sense. Therefore, the difference of the AIC values for the two competing models (e.g., 2-component versus 3-component Gaussian mixture models) should be evaluated. Burnham and Anderson (2004) suggest that a model with the AIC value difference greater than 2 with respect to the competing model is considered to have weak support.

Results and discussion

Model performances of the Gaussian mixture models

To begin with the performance evaluations of different statistical models for the simulated IAT data of the full-margin CSZ ruptures, the 1-component Gaussian mixture model (i.e., renewal model with the normal IAT distribution) and other renewal models are considered. The AIC values with the Gaussian, exponential, gamma, lognormal, extreme value, Weibull, Rayleigh, and Brownian Passage Time distributions are obtained as 1,194,100, 1,246,239, 1,197,931, 1,222,593, 1,220,980, 1,190,912, 1,191,436, and 1,271,206, respectively. The

Weibull distribution (i.e., $AIC=1,190,912$) is superior to other models, including the Gaussian model (i.e., $AIC=1,194,100$). The fitted 1-component model based on the Weibull distribution is also shown in Fig. 3. Although the Weibull distribution captures the unsymmetrical feature of the simulated IAT data (i.e., longer upper tail), the fitted model reveals significant mismatches on both sides of the dominant mode of the simulated data. To improve the fit, more complex statistical models can be considered.

Next, the 2-component and 3-component Gaussian mixture models are used to characterize the simulated data. The results of the model fit are shown in Table 1 and Fig. 4. The visual inspection and AIC values of the fitted Gaussian mixture models indicate that the multi-component Gaussian models outperform all single-component models that are examined previously and that the 3-component Gaussian mixture model is superior to the 2-component Gaussian model ($AIC=1,185,057$ versus $1,186,767$). When the 3-component model is implemented, three components correspond to clusters with mean recurrence periods of 503, 905, and 167 years; the second component corresponds to the long gaps with the mixing proportion of 0.240, whereas the third component corresponds to the short-term clustering with the mixing proportion of 0.114. When the number of components is restricted to $K=2$, the long-gap component is retained in the model, while the less obvious short-term clustering component is merged with the most dominant central component. In short, the data-driven statistical

investigations carried out in this section indicate strong support for the 3-component cluster model for the full-margin CSZ ruptures, outperforming the 1- and 2-component clustering models.

Simulations of earthquake event times by considering the elapsed time

The fitted 3-component Gaussian mixture model can be used in probabilistic seismic and tsunami hazard analysis for the CSZ by considering the current elapsed time since the last 1700 event, i.e., $T_E=323$ years. The adjustment to the 3-component Gaussian mixture model (Fig. 4b) is straightforward by modifying the mixing proportions of the 3 components. First, for each component, calculate the probability that the IAT is longer than T_E , and then multiply this probability by the original mixing proportion of the component. Once the same calculations are performed for all three components, these quantities are

Table 2 Calculations of modified proportions for the elapsed time with $T_E=323$ years

Component (μ, σ)	Original mixing proportion	Probability of $IAT > T_E$	Column (2) \times Column (3)	Modified mixing proportion
1 (503.5, 139.2)	0.646	0.903	0.5833	0.705
2 (905.6, 223.8)	0.240	0.995	0.2387	0.288
3 (166.8, 95.0)	0.114	0.050	0.0057	0.007

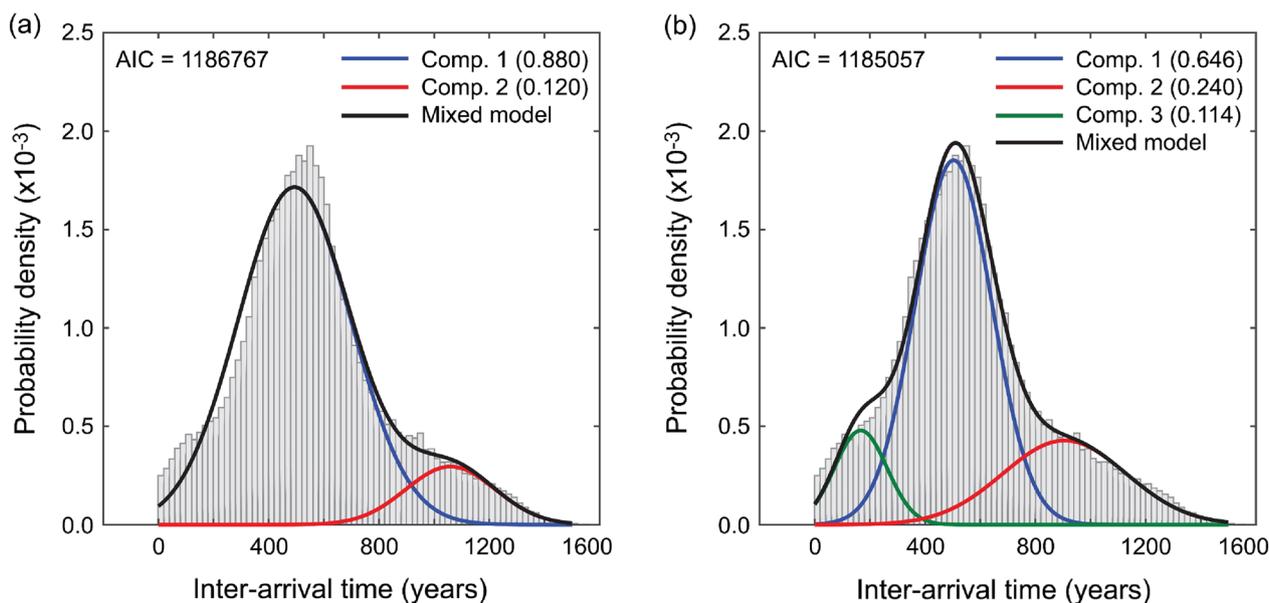


Fig. 4 Comparison of **a** the 2-component Gaussian mixture model and **b** the 3-component Gaussian mixture model. The proportions of the mixed components are indicated in the brackets

normalized to 1.0, and can be used as the updated mixing proportions. This calculation step is demonstrated in Table 2.

The simulation steps of stochastic event catalogs for non-zero elapsed times are as follows:

1. Considering the fitted 3-component Gaussian mixture model and the elapsed time T_E , calculate the modified mixing proportions for all components (Table 2).
2. Set the number of stochastic event catalogs to be simulated. Set the duration for the hazard analysis T_D .
3. For simulating the occurrence time of the first event in the catalog, T_E must be considered. Simulate a uniform random number between 0 and 1 to select a suitable mixture component by using the modified mixing proportions. Subsequently, simulate t_{IAT} from the conditional probability distribution of the selected component.
4. If t_{IAT} is less than T_D , the simulated event should be registered as $t_1 = t_{IAT}$ in a stochastic event catalog and proceed to the second event; otherwise, the simulation process for this catalog realization is stopped (i.e., no event occurs over a period of T_D).
5. For the second event, T_E is reset to 0 and generate t_{IAT} using the original proportions for the Gaussian mixture model (similar to Step 3). When $t_2 = t_1 + t_{IAT}$ is less than T_D , the second event is registered in the stochastic event catalog; otherwise, the simulation for this catalog realization is stopped.
6. Continue Step 5 until the updated time of the most recent event exceeds T_D .
7. Repeat Steps 3 to 6 as many times as required to generate stochastic event sets that reflect the multi-component earthquake occurrence characteristics and the data uncertainty of the full-margin CSZ ruptures.

The above-mentioned simulation procedure is implemented to develop a histogram of the 3-component Gaussian mixture model conditioned on the elapsed time of 323 years (Fig. 5). When the elapsed time is 323 years, the modified proportion for the short-term clustering is significantly decreased from 0.114 to 0.007. Consequently, the proportions for the dominant central component and the long-gap component are increased. The probability density function shown in Fig. 5 is renormalized (or conditioned) by considering no occurrence of the full-margin CSZ event in the last 323 years. Therefore, its height is taller than the probability distribution shown in Fig. 4b. It is noted that the long-term probability of the future full-margin CSZ events can be computed using

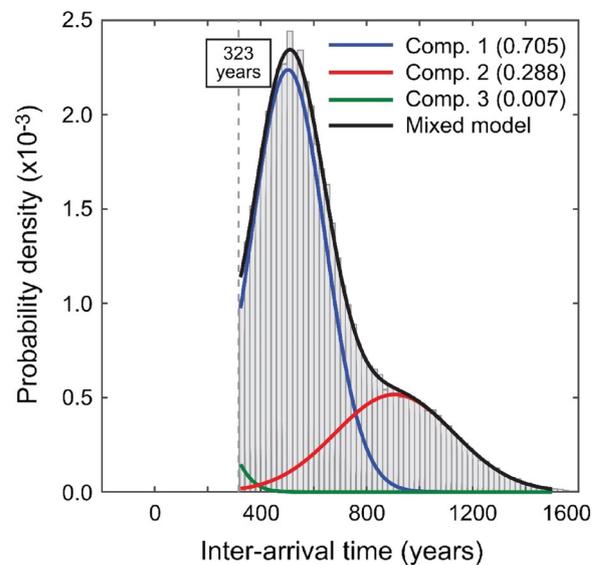


Fig. 5 3-component Gaussian mixture model conditioned on the elapsed time of 323 years. The proportions of the mixed components are indicated in the brackets, which are adjusted based on the elapsed time considered (Table 2)

the results shown in Fig. 5. For instance, by considering $T_D = 50$ years, the probability can be calculated as 6.6%.

Magnitude distribution of the Cascadia full-margin ruptures

The magnitude frequency distribution is an important element for the earthquake occurrence modeling. In many cases, the temporal earthquake occurrence process and the earthquake size distribution are related. For this reason, it is useful to mention how the magnitude distribution of the full-margin CSZ ruptures can be characterized. Goldfinger et al. (2012) estimated the earthquake magnitudes of the full-margin CSZ events (see Table 8 of Goldfinger et al. (2012)). A direct approach is to adopt this empirical magnitude distribution. Alternatively, other magnitude distributions can be adopted. For instance, the (characteristic) uniform distribution may be used by specifying the lower and upper magnitude limits of the full-margin rupture scenarios. A truncated normal distribution with the lower and upper magnitude limits can also be adopted by specifying the mean and standard deviation based on the empirical magnitude data. The above-mentioned approaches essentially decouple the temporal occurrence model and the magnitude frequency model. If a coupled characterization of the temporal occurrence and earthquake size is desirable, a stronger hypothesis, such as a hierarchical earthquake occurrence model by

Kulkarni et al. (2013), which can capture the dependency between the waiting time and the earthquake magnitude, could be adopted in probabilistic seismic and tsunami hazard analyses. A discussion as to which of the hypothesized models is more suitable is beyond the scope of this study.

Conclusions

The earthquake occurrence modeling for major subduction events involves significant data uncertainty due to the scarcity of available geological data and the adopted dating techniques. Due to the complex nature of major subduction zones, the data may exhibit the clustering characteristics, separated by long gaps. The marine turbidite age data for the full-margin CSZ ruptures, which were compiled by Goldfinger et al. (2012), revealed such data features. Building upon the previous work by Kulkarni et al. (2013) which attempted to characterize such clustering behavior of the Cascadia full-margin events using a hierarchical clustering analysis, this study adopted the Gaussian mixture model as an alternative method. To fully capture the uncertainty of the underlying marine turbidite-based Cascadia age data, the number of Monte Carlo resampling was increased substantially (from 20 to 5000).

The statistical modeling results clearly showed that the 3-component Gaussian mixture model outperforms the 2-component Gaussian mixture model and the 1-component renewal models. In addition to the most dominant central component with the mean recurrence period of 500 years, the long-gap component, and the short-term clustering component were identified. The developed Gaussian mixture model can be used to generate stochastic event sets for probabilistic seismic and tsunami hazard analysis and to evaluate the long-term probability of the future full-margin CSZ events. The computer algorithm and code are provided to promote the use of the proposed Gaussian mixture method.

As a final remark, it is important to mention the limitations of the results presented in this study. Although the marine turbidite age data for the full-margin CSZ ruptures by Goldfinger et al. (2012) and Kulkarni et al. (2013) were taken as a valid dataset for statistical modeling of the full-rupture CSZ recurrence, a subset of the identified full-margin CSZ ruptures can be adopted by considering the age data to certain calibrated years before the present. For instance, onshore-based paleo-tsunami records are generally shorter than offshore-based paleo-tsunami records. The reduced data size will result in different recurrence models for the full-margin CSZ ruptures. To reflect epistemic uncertainty associated with the data selection of the earthquake occurrence modeling, such

different models can be implemented in a logic tree for probabilistic seismic and tsunami hazard analysis.

Abbreviations

AIC	Akaike Information Criterion
CoV	Coefficient of variation
CSZ	Cascadia subduction zone
EM	Expectation maximization
IAT	Inter-arrival time

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40562-023-00306-6>.

Additional file 1. A MATLAB code to fit the Gaussian mixture model to the inter-arrival time data of the Cascadia subduction zone. The Cascadia-turbidite age data suggested by Kulkarni et al. (2013) are included in the MATLAB code.

Funding

The work is funded by the Canada Research Chair program (950-232015) and the NSERC Discovery Grant (RGPIN-2019-05898).

Declarations

Competing interests

The author acknowledges that there are no conflicts of interest.

Received: 27 July 2023 Accepted: 27 October 2023

Published online: 13 November 2023

References

- Abaimov SG, Turcotte DL, Shcherbakov R, Rundle JB, Yakovlev G, Goltz C, Newman WI (2008) Earthquakes: recurrence and interoccurrence times. *Pure Appl Geophys* 165:777–795
- Atwater B, Musumi-Rokkaku S, Satake K, Tsuji Y, Ueda K, Yamaguchi DK (2015) The Orphan Tsunami of 1700—Japanese clues to a parent earthquake in North America, 2nd edn. University of Washington Press, Seattle, p 144
- Baker JW, Bradley B, Stafford P (2021) *Seismic Hazard and Risk Analysis*. Cambridge University Press, Cambridge, p 600
- Behrens J, Løvholt F, Jalayer F, Lorito S, Salgado-Gálvez MA, Sørensen M, Abadie S, Aguirre-Ayerbe I, Aniel-Quiroga I, Babeyko A, Baiguera M, Basili R, Belliazzi S, Grezio A, Johnson K, Murphy S, Paris R, Rafliana I, De Risi R, Rossetto T, Selva J, Taroni M, Del Zoppo M, Armigliato A, Bureš V, Cech P, Cecioni C, Christodoulides P, Davies G, Dias F, Bayraktar HB, González M, Gritsevich M, Guillas S, Harbitz CB, Kânoğlu U, Macías J, Papadopoulos GA, Polet J, Romano F, Salamon A, Scala A, Stepinac M, Tappin DR, Thio HK, Tonini R, Triantafyllou I, Ulrich T, Varini E, Volpe M, Vyhmeister E (2021) Probabilistic tsunami hazard and risk analysis: a review of research gaps. *Front Earth Sci* 9:628772
- Burnham KP, Anderson DR (2004) Multimodel inference understanding AIC and BIC in model selection. *Sociol Methods Res* 33:261–304
- Console R, Murru M, Falcone G, Catali F (2008) Stress interaction effect on the occurrence probability of characteristic earthquakes in Central Apennines. *J Geophys Res* 113:B08313
- Cornell CA, Winterstein SR (1988) Temporal and magnitude dependence in earthquake recurrence models. *Bull Seismol Soc Am* 78:1522–1537
- Davies G, Griffin J, Løvholt F, Glimsdal S, Harbitz C, Thio HK, Lorito S, Basili R, Selva J, Geist E, Baptista MA (2018) A global probabilistic tsunami hazard assessment from earthquake sources. *Geol Soc Spec Publ* 456(1):219–244
- DeMets C, Gordon RG, Argus DF (2010) Geologically current plate motions. *Geophys J Int* 181:1–80. <https://doi.org/10.1111/j.1365-246X.2009.04491.x>

- Goldfinger C, Nelson CH, Morey AE, Johnson JE, Patton J, Karabanov E, Gutierrez-Pastor J, Eriksson AT, Gracia E, Dunhill G, Enkin RJ, Dallimore A, Vallier T (2012) Turbidite event history: methods and implications for Holocene paleoseismicity of the Cascadia subduction zone. United States Geological Survey Professional Paper 1661–F. <http://pubs.usgs.gov/pp/pp1661f/>.
- Griffin JD, Stirling MW, Wang T (2020) Periodicity and clustering in the long-term earthquake record. *Geophys Res Lett*. <https://doi.org/10.1029/2020GL089272>
- Headquarters for Earthquake Research Promotion (2019) Evaluation of Long-term Probability of Earthquake Occurrence Along the Japan Trench, 144 p.
- Kulkarni R, Wong I, Zachariassen J, Goldfinger C, Lawrence M (2013) Statistical analyses of great earthquake recurrence along the Cascadia subduction zone. *Bull Seismol Soc Am* 103:3205–3221
- Matthews MV, Ellsworth WL, Reasenber PA (2002) A Brownian model for recurrent earthquakes. *Bull Seismol Soc Am* 92:2233–2250
- McLachlan G, Peel D (2000) *Finite Mixture Models*. John Wiley & Sons Inc, Hoboken, p 427
- Mori N, Satake K, Cox D, Goda K, Catalan P, Imamura F, Tomiczek T, Lynett P, Miyashita T, Muhari A, Titov V, Ho T, Wilson R (2022) Giant tsunamis: generation, propagation, early warning, and long-term assessment. *Nat Rev Earth Environ* 3:557–572
- Ogata Y (1999) Estimating the hazard of rupture using uncertain occurrence times of paleoearthquakes. *J Geophys Res Solid Earth* 104:17995–18014
- Philibosian B, Meltzner AJ (2020) Segmentation and supercycles: a catalog of earthquake rupture patterns from the Sumatran Sunda Megathrust and other well-studied faults worldwide. *Quatern Sci Rev* 241:106390
- Rhoades DA, Van Dissen RJ, Dowrick DJ (1994) On the handling of uncertainties in estimating the hazard of rupture on a fault segment. *J Geophys Res Solid Earth* 99:13701–13712
- Satake K, Wang K, Atwater BF (2003) Fault slip and seismic moment of the 1700 Cascadia earthquake inferred from Japanese tsunami descriptions. *J Geophys Res Solid Earth* 108(B11):2535. <https://doi.org/10.1029/2003JB002521>
- Shimazaki K, Nakata T (1980) Time predictable recurrence for large earthquakes. *Geophys Res Lett* 7:279–282
- Sykes LR, Menke W (2006) Repeat times of large earthquakes: implications for earthquake mechanics and long-term prediction. *Bull Seismol Soc Am* 96:1569–1596
- Williams RT, Davis JR, Goodwin LB (2019) Do large earthquakes occur at regular intervals through time? a perspective from the geologic record. *Geophys Res Lett* 46:8074–8081

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
