JGR Solid Earth

RESEARCH ARTICLE
10.1029/2021JB021996

Age Uncertainty in Recurrence Analysis of Paleoseismic Records

Philipp Kempf[1] **and Jasper Moernaut**[2]

[1]Institute of Geological Sciences, Freie Universitt Berlin, Berlin, Germany, [2]Institute of Geology, University of Innsbruck, Innsbruck, Austria

Key Points:

- Low number of recurrence intervals in paleoseismic records underestimates aperiodicity
- High age uncertainty relative to the mean recurrence interval in paleoseismic records overestimates aperiodicity
- For calculating coefficient of variation and burstiness it matters how recurrence intervals are sampled from records

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

P. Kempf,
philipp.kempf@fu-berlin.de

Citation:

Kempf, P., & Moernaut, J. (2021). Age uncertainty in recurrence analysis of paleoseismic records. *Journal of Geophysical Research: Solid Earth*, *126*, e2021JB021996. https://doi.org/10.1029/2021JB021996

Received 9 MAR 2021
Accepted 11 JUL 2021

Abstract Determining the aperiodicity of large earthquake recurrences is key to forecast future rupture behavior. Aperiodicity is classically expressed as the coefficient of variation of recurrence intervals, though the recent trend to express it as burstiness is more intuitive and avoids minor inaccuracies. Due to the underestimation of burstiness in records with a low number of recurrence intervals, the paradigm is to obtain long paleoseismic records with many events. Here, we present a suite of synthetic paleoseismic records designed around the Weibull and inverse Gaussian distributions that demonstrate that age uncertainty relative to the mean recurrence interval causes overestimation of burstiness. The effects of overestimation and underestimation interact and give complex results for accurate estimates of aperiodicity. Furthermore, we show that the way recurrence intervals are sampled from a paleoseismic record can have strong influences on the resulting statistic and its implication for probabilistic seismic hazard assessment. Comparing values of burstiness between paleoseismic records should therefore be done with caution.

Plain Language Summary To forecast future earthquake activity, paleoseismologists aim to have many events in a single sedimentary record to estimate the periodicity of an earthquake sequence with as little uncertainty as possible. This focus on the number of events is not wrong, but event age uncertainty is another—often neglected and not yet described—source of uncertainty that can interfere in estimating periodicity correctly. In this study, we show in what way and by how much event age uncertainty affects the uncertainty in periodicity. We create a model of many different artificial earthquake sequences. For our model setup, we choose: (1) two types of patterns; (2) six degrees of periodicity; (3) 10 different levels of event age uncertainty; and (4) a wide range of number of events (from 4 to 101 events). Then we create 50,000 earthquake sequences for each unique combination within this spectrum and analyze the variability in periodicity. We find that low number of events underestimates periodicity and high age uncertainty overestimates periodicity. Having many events in a record is more important, if the earthquake sequence is not periodic. Having accurately dated events is more important, if the earthquake sequence is periodic.

1. Introduction

The ultimate goal of paleoseismology is to estimate the magnitude, location, and timing of future earthquakes by reconstructing past earthquakes over multi-millennia timescales. Besides constraining the mean recurrence rate, paleoseismologists characterize the aperiodicity in earthquake sequences (e.g., Goes & Ward, 1994). Aperiodicity plays a key role in characterizing the main conceptual models of seismic behavior in major tectonic fault zones, being the elastic rebound theory (for periodic behavior, Reid, 1910), characteristic earthquakes (for weakly periodic behavior, Schwartz & Coppersmith, 1984), earthquakes as a Poissonian process with constant hazard rates (for random recurrences, e.g., Gomez et al., 2015) and earthquake supercycles (Sieh et al., 2008). In most cases, the aperiodicity is expressed by the coefficient of variation (*CoV*; mean-normalized standard deviation) of recurrence intervals (RIs), that is, the time between two consecutive events. Since the proposition of the term burstiness (*B*) (Goh & Barabási, 2008), authors sometimes use *B* instead of the *CoV* (Chen et al., 2020; Griffin et al., 2020; Salditch et al., 2020). *B* is the result of a one-to-one transformation of the *CoV* from the domain of 0 to $+\infty$ to *B*'s codomain of −1 to +1.

Because *CoV* and *B* do not take into account the order of events, Goh and Barabási (2008) also introduced a memory (*M*) term, which is the correlation coefficient of consecutive RIs, also called auto-correlation. *M*

ranges from −1 to +1, is negative when short (long) RIs follow long (short) RIs, and positive if short (long) RIs follow short (long) RIs. Because of the logical physical connection of large stress releases, reloading of the fault and potential stress transfer to other faults, M finds use in large earthquake literature. An extensive survey of mostly instrumental records of large earthquake sequences in major seismic regions and on isolated fault zones produced a range of M between −0.36 and +0.35 (Chen et al., 2020). However, it is unclear how reliable estimates of M are, when derived from paleoseismic records with typically low event numbers. We calculate M for synthesized records that are conceptually designed without memory ($M = 0$) and discuss the usefulness of M in large earthquake paleoseismology.

At the center of the statistics CoV, B and M are the RIs, which are based on event ages (EAs). Most onshore studies rely on sedimentary sequences in dynamic environments, where sedimentation rate is likely highly variable (e.g., Berryman et al., 2012) and cannot be used as a constraint (Bayesian prior) in age-modeling (Efron, 2013). In these onshore studies, EAs are often determined by dates constraining the EA to a minimum and maximum age and making use of the stratigraphic order of event deposits. Marine and lacustrine paleoseismology often benefit from a more stable sedimentation rate, which can be implemented in age-depth models (e.g., Howarth et al., 2014, 2016; Wils et al., 2020), effectively decreasing age uncertainty in age-depth models. The stratigraphic distance between event deposits in marine or lacustrine records given an average sedimentation rate is thus a reasonable estimator of the RI. The low uncertainty in relative age of events in paleoseismic records can be crucial for cross-site correlation of large earthquakes when determining spatiotemporal patterns in a seismically active region (Howarth et al., 2021). For these reasons, we focus our analyses on aquatic records with stable sedimentary environments.

As a result of different age model techniques, RIs are calculated in various ways. There are three common methods:

1. The "best" age approach, where a single "best" age (often the median of an age distribution) is assigned to each event and the difference is the RI (e.g., Kulkarni et al., 2013; Moernaut et al., 2018).
2. The age-depth model iteration approach, where RIs are calculated based on the age differences within Markov Chain Monte Carlo (MCMC) iterations (simulated accumulation histories) that make up the age-depth model. This method gives a mathematically more correct estimate for RIs and keeps the probabilistic nature of the EA. The age model is based on a constrained Bayesian prior of sedimentation rate, then this model produces valid RIs with less relative age uncertainty than other methods. The age-depth model iteration approach can be applied through, for example, the use of the Difference()-function on P_Sequences in OxCal (e.g., Bronk Ramsey & Lee, 2013; Sawai et al., 2009) or through the subtracted iteration ages of consecutive events from Bacon (Blaauw & Christen, 2011) age-depth models (e.g., Kempf et al., 2018).
3. The sampling EA approach, where the EA distributions are randomly sampled (Monte Carlo method) to calculate RIs. Age reversals, that is, negative RIs, can occur in this approach, if two consecutive events have overlapping EA distributions, and it matters how these are treated. Recent global compilations (e.g., Williams et al., 2019) use the sampling EA approach, regardless of the type of paleoseismic record, affecting relative age uncertainty. The differences in the way RIs are calculated and the preferred method used for a given type of paleoseismic record may lead to inconsistent interpretations of CoV and hamper the reliability of hazard assessments.

Here, we quantify the effect of age uncertainty on estimates of aperiodicity of earthquake sequences under the assumption of a stationary series, that is, aperiodicity is a stable statistic over all event occurrences. We begin with estimating the uncertainty in estimates of aperiodicity (CoV and B) over the number of events from various synthetic earthquake recurrence patterns without added age uncertainty (control). Then we simulate the effects that age uncertainty in paleoseismic records can have on earthquake sequence statistics. Beyond the obvious "more and better data is best," we evaluate when decreasing age uncertainty (better data) and when the traditional paradigm of looking for longer records with more events (more data) is the best way to improve the estimate of aperiodicity. We discuss various natural cases from the literature and show that even the method how to calculate measures of aperiodicity can affect the result and thus the interpretation of earthquake recurrence data.

2. Methods

2.1. Considerations for Synthetic Record Creation

The first half of this study is based on synthetic paleoseismic records that are designed to mimic existing paleoseismic records. The building blocks of these records are time-spans between two consecutive events, that is, RIs. The synthetic records are created using the Monte Carlo method, that is, they contain samples from a true probability density function (true PDFs) of RI distributions. The shape and scale of true PDFs follow a set of rules:

Rule 1: The mean RI of the true PDF is 100 years. We chose 100 years, because it is easy to scale to real records and has the advantage of simplifying the thinking in percentages. Note that the sample mean of any single synthetic record will likely differ from 100 years, especially for synthetic records with low numbers of RIs (n) (Kempf et al., 2018).

Rule 2: The number of RIs (n) in synthetic records ranges from 3 to 100. This encompasses almost all existing paleoseismic records, with prominent exceptions, for example, the $n = 118$ paleoseismic record of 7 kyr from Lake Tutira near the Alpine Fault in New Zealand (Gomez et al., 2015) or the $n = 413$ paleoseismic record of 220 kyr from the Dead Sea (Lu et al., 2020). The most relevant bracket of n is from 3 to 30, because this is where most paleoseismic records fall into and where most change in descriptive statistics is anticipated.

Rule 3: The type of distribution of the true PDF needs to be relevant for paleoseismology. We chose Weibull and inverse Gaussian distributions for reasons discussed in Matthews et al. (2002). However, because results are similar for both Weibull and inverse Gaussian Distributions, we refer to the Supporting Information for all analyses beyond the introduction of the scenarios.

Rule 4: All RIs within one synthetic record are independent from another. This means all synthetic records are randomly sampled from their true PDF (Monte Carlo method).

Rule 5: The coefficient of variation (CoV) is the mean-normalized standard deviation and defined as

$$\begin{aligned} CoV &= \frac{\sigma}{\mu} \\ \overline{CoV} &= \frac{s}{\overline{x}} \end{aligned} \tag{1}$$

Where $\overline{CoV}$ is the sample CoV, σ (s) is the (sample) standard deviation and μ ($\overline{x}$) is the (sample) mean. We aim for a range of realistic CoVs between 0.25 and 1.46 as obtained in published lacustrine paleoseismic records in different tectonic settings (cf. Chen et al., 2020; Griffin et al., 2020; Moernaut, 2020). The one-to-one transformation from CoV to B is (Goh & Barabási, 2008)

$$\begin{aligned} B &\equiv \frac{\sigma \,/\, \mu - 1}{\sigma \,/\, \mu + 1} = \frac{CoV - 1}{CoV + 1} \\ \overline{B} &\equiv \frac{s \,/\, \overline{x} - 1}{s \,/\, \overline{x} + 1} = \frac{\overline{CoV} - 1}{\overline{CoV} + 1} \end{aligned} \tag{2}$$

where $\overline{B}$ is the sample burstiness. The parameterization of the true PDFs is chosen accordingly.

2.2. Side Note on Terminology

CoVs of approximately 1 ($B \approx 0$) are often called "time-independent." because Poissonian processes, which have a time-independent (constant) hazard rate, have a CoV of 1. We prefer "aperiodic," because there are many other distributions that can produce CoVs of 1, but have time-dependent hazard rates.

Paleoseismic records with $CoV > 1.1$ ($B > 0.05$) are sometimes called "clustered." However, this terminology can be confused with being the result of a cluster analysis, in which series of short RIs are mathematically delimited from long quiescent phases. We follow the reasoning of Salditch et al. (2020), who noted that paleoseismic records with $CoV < 1$ ($B < 0$) can appear clustered, too, and use the term "bursty" instead for records with $CoV > 1.1$ ($B > 0.05$).

Table 1
Parameters for Creating True PDFs Used to Produce Weibull Synthetic Records

Input parameters	True PDF characteristics					
Distribution type	Shape (k_{wei})	Scale (λ_{wei})	Mean (μ_{wei})	CoV	B	Level of aperiodicity
Weibull	4	≈ 110	100	≈0.281	≈– 0.563	Strongly periodic
Weibull	2	≈113	100	≈0.523	≈– 0.313	Weakly periodic
Weibull	1.3	≈108	100	≈0.776	≈– 0.126	Weakly periodic
Weibull (exp. dist.)	1	100	100	1	0	Aperiodic
Weibull	0.85	≈92	100	≈1.181	≈0.083	Bursty
Weibull	0.7	≈79	100	≈1.462	≈0.188	Bursty

We acknowledge that the boundaries are somewhat arbitrary, however, we refer to a record with a $CoV < 0.5$ ($B < -\frac{1}{3}$) as "strongly periodic," from 0.5 ($B = -\frac{1}{3}$) to 0.9 ($B \approx -0.05$) as "weakly periodic," from 0.9 ($B \approx -0.05$) to 1.1 ($B \approx +0.05$) as "aperiodic" and >1.1 ($B \approx +0.05$) as "bursty."

2.3. Choosing Parameters for Probability Distribution Functions

The parameterization of the Weibull and inverse Gaussian distributions are chosen in accord with the five rules specified above. The parameters are listed in Tables 1 and 2. The resulting PDFs (Figure 1) are the priors for synthetic records. For the equations, we refer the reader to the Supporting Information in this article.

2.4. Age Uncertainty Generator

To analyze the effect of age uncertainty in paleoseismic records on descriptive statistics like the CoV or B, we subject the synthetic records to what we termed an age uncertainty generator. The age uncertainty generator adds or subtracts time from the RI. Time is added or subtracted according to a random sample from a truncated normal distribution that is centered on a mean of 0 years and its standard deviation parameter is a chosen input parameter. The randomly assigned age uncertainty is then scaled depending on the RI that it is applied to. The application of the age uncertainty generator changes the original synthetic record, which is replaced.

Table 2
Parameters for Creating True PDFs Used to Produce Inverse Gaussian Synthetic Records. Note That the Inverse Gaussian Distribution of CoV = 1 (B = 0) Does Not Have the Shape of an Exponential Distribution, Nor Does it Have a Uniform Hazard Rate

Input parameters	True PDF characteristics				
Distribution type	Shape (λ_{inv})	Mean (μ_{inv})	CoV	B	Level of aperiodicity
Inverse Gaussian	1,600	100	0.25	–0.6	Strongly periodic
Inverse Gaussian	400	100	0.5	≈– 0.333	Strongly periodic
Inverse Gaussian	≈178	100	0.75	≈– 0.143	Weakly periodic
Inverse Gaussian	100	100	1	0	Aperiodic
Inverse Gaussian	≈69	100	1.2	≈0.091	Bursty
Inverse Gaussian	≈51	100	1.4	≈0.167	Bursty

The necessity to scale the added age uncertainty to the RI arises from the following: When the RI is long ($RI > 100$ years), the age uncertainty is constrained by the age uncertainty of the age model of the paleoseismic record. To estimate the age uncertainty over age, we compiled reported age uncertainties of event ages from numerous marine and lacustrine Holocene paleoseismic records mainly based on radiocarbon dates (Figure 2). From –10 years BP (youngest event collection of records) to ≈300 years BP the mean uncertainty rises from near 3 to ≈70 years. Beyond ≈300 years BP, the counter-intuitive result is that the age uncertainty remains nearly constant over age or increases only slowly with age. This is rather fortunate, because this means that the age uncertainty over age (Figure 2) approximates age uncertainty over RI reasonably well. We adopted this knowledge by scaling the age uncertainty of the RI with a relatively flat logarithmic function for the domain where $RI > 100$ years.

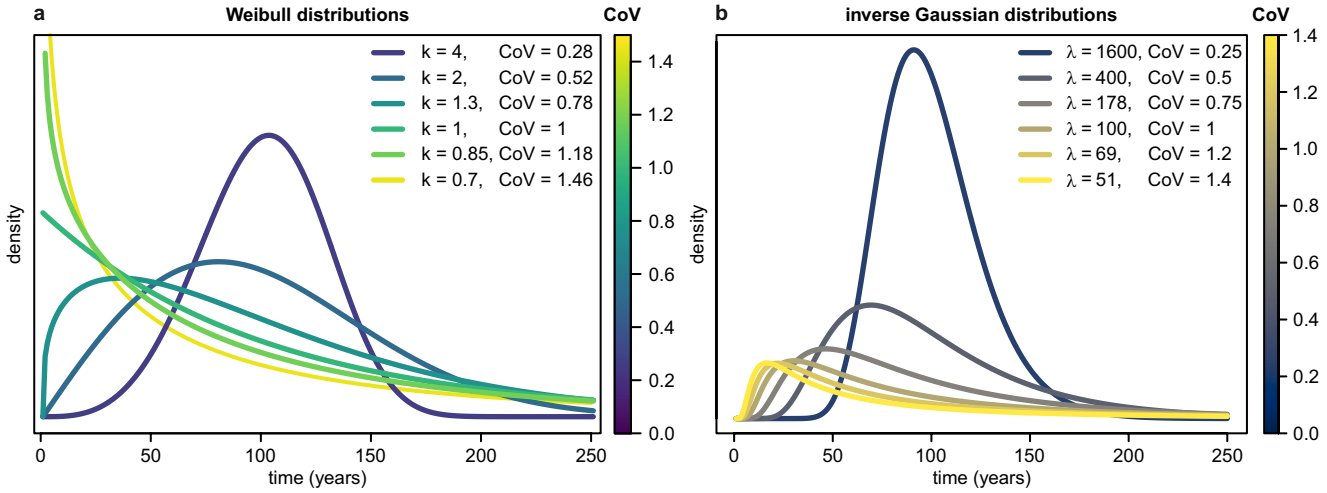

Figure 1. True probability density functions of (a) the Weibull distributed variables detailed in Table 1 and (b) the inverse Gaussian distributed variables detailed in Table 2, each color-coded for their respective aperiodicity (CoV).

However, for the domain $0 < RI \leq 100$ years, RIs are short enough for the absolute age and the age uncertainty of the age model to become less relevant. Instead, the RI and its uncertainty is better estimated by the sedimentary interval between the event deposits and the sedimentation rate and its variability. This argument is especially powerful for paleoseismic records from environments with stable sedimentation rates, for example, marine and lacustrine environments, and becomes increasingly useless the higher the relative sedimentation rate variability, that is, in floodplains or marshes. With both domains having reasonably

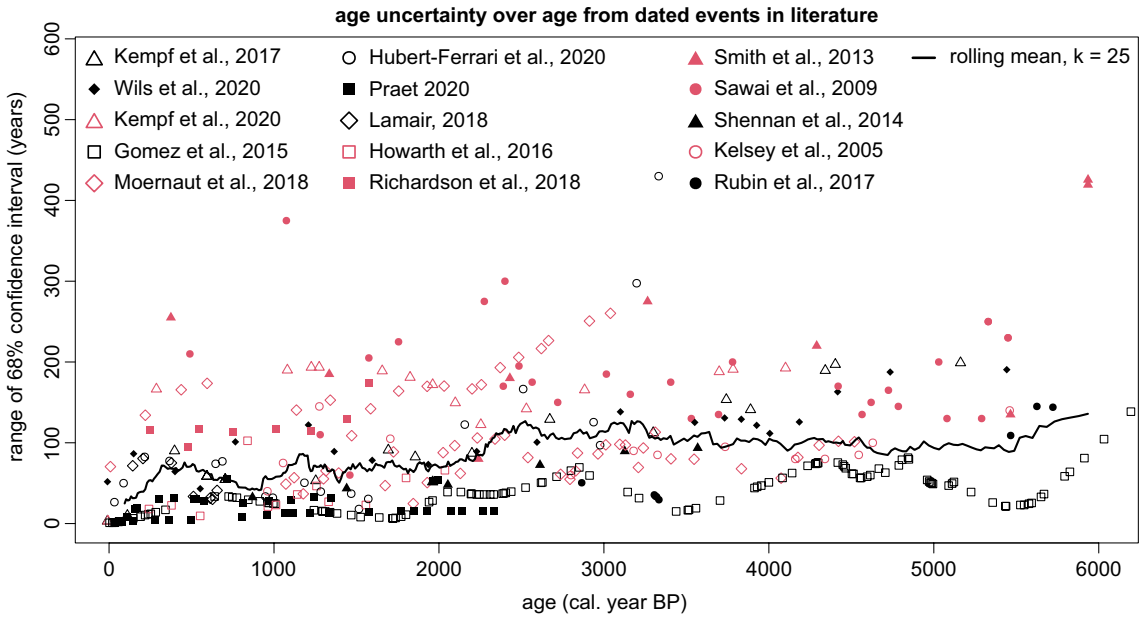

Figure 2. Reported age uncertainty over age of dated events from a selection of real paleoseismic records with $n > 8$ from literature. The age uncertainty is expressed as the width of the 68% confidence interval, because this is a similar parameter that we use to generate age uncertainty in our model. The black solid line is a rolling average with a rolling window of 25. Note the nearly constant or only slowly increasing rolling mean age uncertainty in age models with age. The records of Rubin et al. (2017), Sawai et al. (2009), and Shennan et al. (2014) are the only non-subaqueous records in this selection.

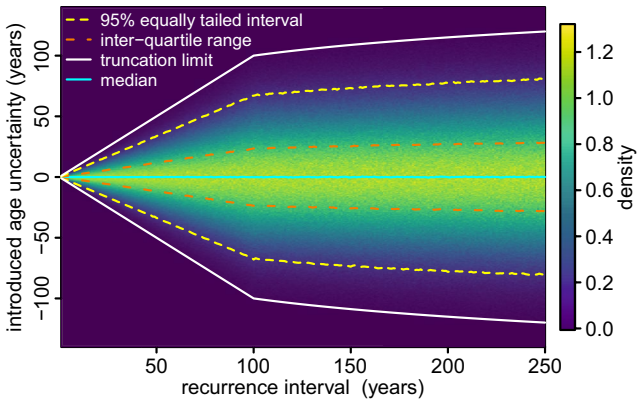

Figure 3. Density map of the scaled age uncertainty that is added to the synthetic records in order to simulate age uncertainty in subaqueous paleoseismic records, created after the piecewise function in Equation 3.

logical foundations in the considered paleoseismic records, we argue that the application of the age uncertainty should be a two-domain piecewise function.

For example, Kempf et al. (2017) studied tsunami deposits in Lake Huelde, Chile. Two tsunami deposits, hO and hP, were dated by the age-depth model with a 95% equally tailed interval (95% ETI) between 4,156 to 4,535 cal. years BP and between 4,209 to 4,603 cal. years BP, with an age uncertainty at 379 and 394 years, respectively. However, a relatively constant sedimentation rate in Lake Huelde considerably lowered the uncertainty of the relative age between hO and hP to ≈ 85 ± 50 years (Kempf et al., 2018).

We chose the scaling of the age uncertainty in the domain of $RI \leq 100$ to be linear, being 0 when $RI = 0$ years, increasing to ≈ 1 at $RI = 100$ years (Equation 3).

$$x_u \sim \begin{cases} \mathcal{N}\left(x, \frac{x}{\bar{x}}\sigma_u^2\right); & for\ x \leq 100 \\ \mathcal{N}(x, \log_{\bar{x}}(x)\,\sigma_u^2); & for\ x > 100 \end{cases} \quad (3)$$

Where the RI with the generated age uncertainty (x_u) is distributed as (~) a normal distribution $\mathcal{N}(\ldots)$ with the RI (x) as the mean and the variance (σ_u^2). The variance is scaled linearly with $\frac{x}{\bar{x}}$ for $x \leq 100$ and logarithmically with $\log_{\bar{x}}(x)$ for $x > 100$.

We truncate the normal distribution of the age uncertainty to $-0.99\ \mu < \mathcal{N}(\ldots) < 0.99\ \mu$. The lower truncation limit is necessary in order to prevent the rare case of illogical negative RIs. The upper truncation limit is to keep the age uncertainty symmetrical. For $\sigma_u = 35$ years this means that < 1% of the randomly assigned age uncertainties lies outside of the truncation limits and needs re-sampling from the normal distribution.

We ran the model with standard deviations $\sigma_u = 0$ (no added age uncertainty), 5, 10, 15, 20, 25, 30, 35, 40, and 45 years. The results presented in figures are from a model run with $\sigma_u = 35$ years, as is the visualization of the age uncertainty generator shown in Figure 3. Results for other model runs with different σ_u are in the Supporting Information.

To give an example for the age uncertainty generator, if the RI before adding age uncertainty, x, is 11 years, the sample mean of the specific synthetic record, $\bar{x}$, is 121 years and the random sample from the truncated normally distributed age uncertainty, with $\sigma_u = 35$ years, is −22 years, then the first domain of the piecewise function for the age uncertainty generator is applied, because $0 < x \leq 100$ and then $x_u = 11 + (-22)\frac{11}{121} = 9$ years. In this case, the scaling decreases the random age uncertainty from −22 to −2, because of the relatively short RI.

2.5. Memory as a Correlation Coefficient of Consecutive RIs

The memory term (M) is a descriptive statistic for records of RIs introduced by Goh and Barabási (2008). It is the correlation coefficient of consecutive RIs, or the "auto-correlation," and is defined as:

$$M \equiv \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{\left(x_i - m_1\right)\left(x_{i+1} - m_2\right)}{\sigma_1\sigma_2} \quad (4)$$

where m_1 (m_2) and σ_1 (σ_2) are the sample mean and sample standard deviation of the RIs x_i (x_{i+1}) for $n = 1,\ldots,n-1$, respectively (Goh & Barabási, 2008). Note that order of RIs matters for M. Different permutations of the same set of RIs have different M. M ranges from −1 to 1 and is negative for sequences where short RIs follow long RIs and vice versa and M is positive where short RIs follow short RIs and long RIs follow long RIs. $M = 0$ is a sequence without memory, where the preceding RI does not appear to influence the next RI in the sequence. Because of rule 4 (independence of samples), the synthetic records should be at or close to $M = 0$.

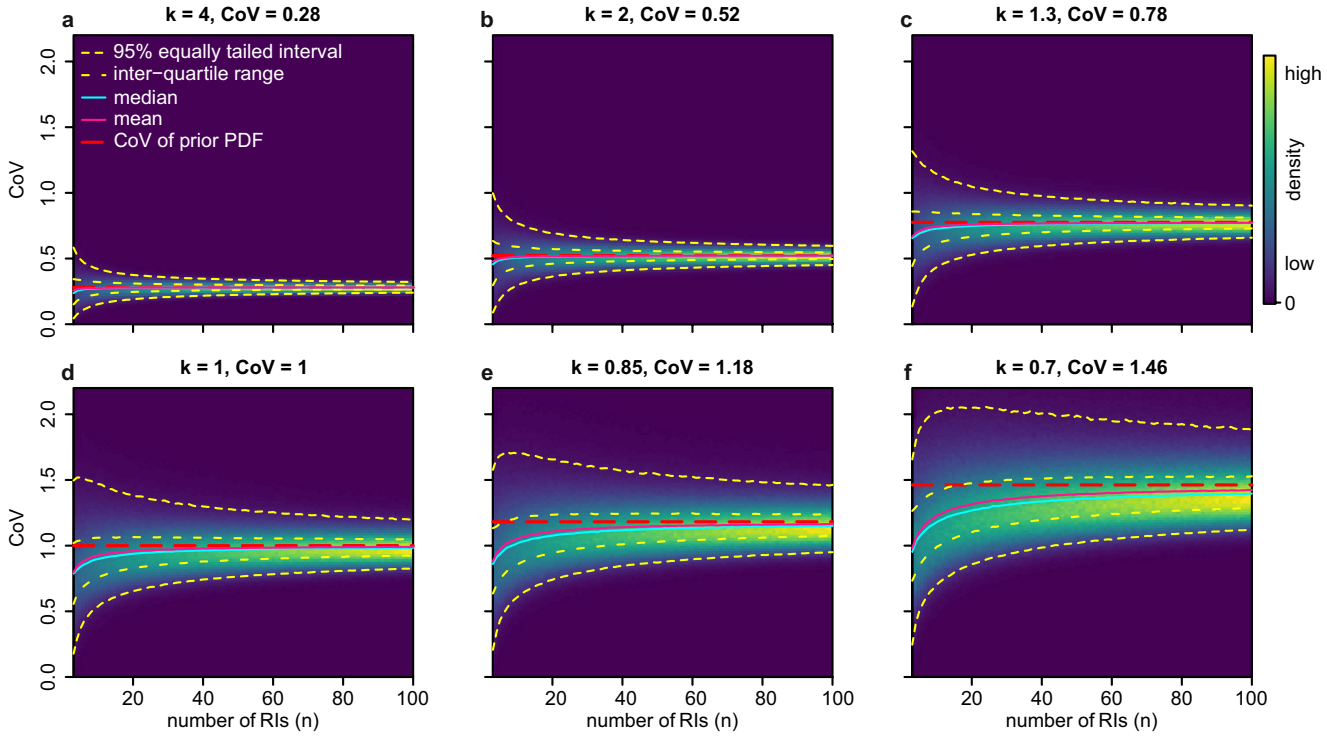

Figure 4. Density maps of coefficients of variation (*CoV*s) from synthetic records based on Weibull distributed variables over the number of recurrence intervals (*n*).

A confusion may arise here, because all Poisson processes are exponentially distributed and are inherently memoryless, that is, both in the sense that time since the last event does not affect the expected time until the next event (constant hazard rate) and in the sense that the length of the previous RI does influence the length of the next RI. However, not all exponentially distributed processes are Poissonian and do not necessarily share the independence of consecutive RIs. Records may be without memory ($M = 0$) and not be exponentially distributed. Conversely, records may have non-zero M and be exponentially distributed. As "memorylessness" of Poissonian processes is a well-established concept in statistics, we suggest to specify when memory, like in M, is recurrence interval specific.

3. Results

3.1. Statistics of Recurrence Intervals – Without Added Age Uncertainty

To analyze CoV, B and M over the number of RIs we extracted the statistics from the synthesized records without the added age uncertainty. These statistics generally show large uncertainty for records with low n and develop into narrower distributions toward higher n (Figures 4, 5, and 7). Because the results for Weibull and inverse Gaussian true PDFs show generally similar trends, we only show the Weibull synthetic records. The results for inverse Gaussian records are in the Supporting Information.

All synthetic records' means and medians underestimate the CoV and B at low n and asymptotically approach the true CoV and B, that is, the respective population parameter. The underestimation of true CoV and B is more pronounced in records that are less periodic. Records with strongly periodic true PDFs produce considerably less uncertainty in CoV and approach the true CoV and B at lower n than records with less periodic and bursty true PDFs. It is worth noting that the underestimation of the CoV and B for low n is described here for the mean and median. Overestimation also exists within the distribution of CoV and B. Depending on n, the overestimation can be extreme even within the 95% ETI.

In addition, random and bursty records have positively skewed distributions of CoV, with the exception of $n = 3$ for strongly bursty records (Figure 6). This can be important, because when the median CoV

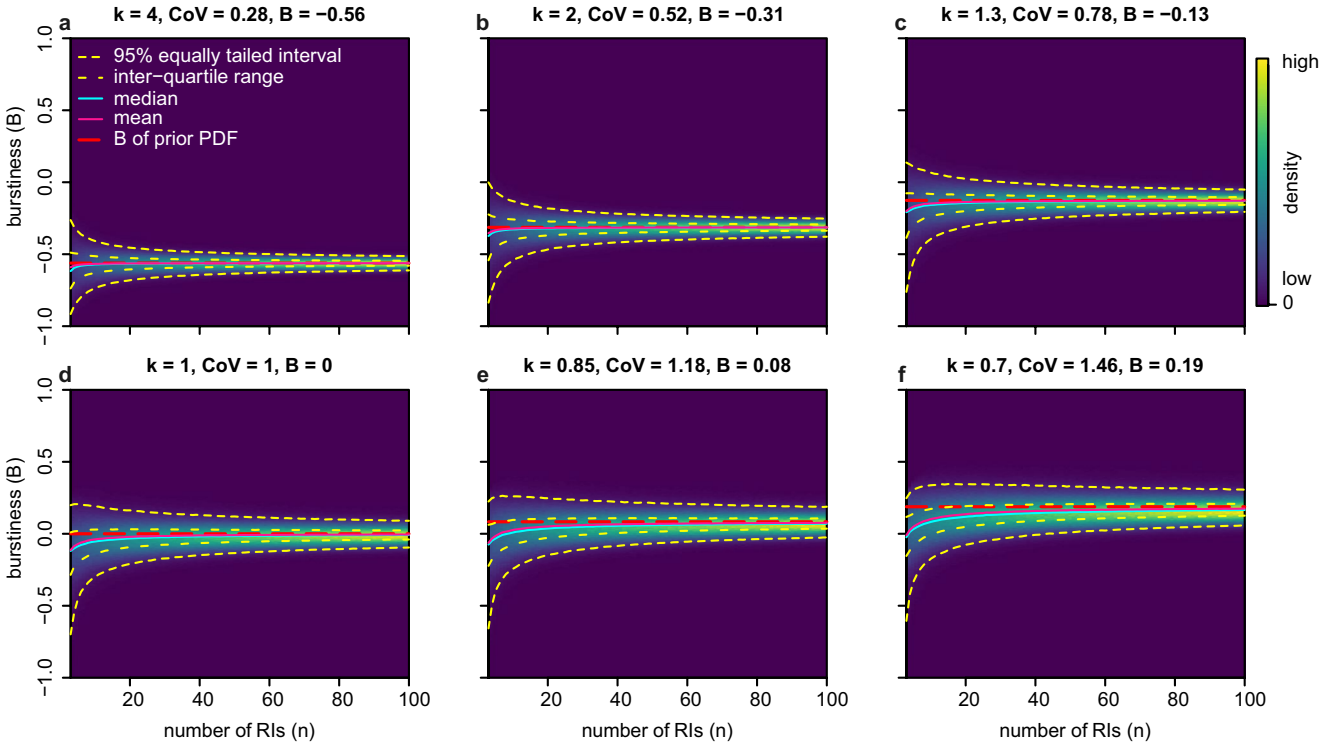

Figure 5. Density maps of burstiness (B) from synthtic records based on Weibull distributed variables over the number of recurrence intervals (n). In essence, this is Figure 4 after the CoV to B transformation.

underestimates the true CoV at low n, then the underestimation can be stronger for the mode, that is, the most likely outcome, when the distribution of CoV is positively skewed. The positive skewness is most pronounced for $n \approx 20$, if the true PDF has a $CoV \geq 0.75$. For records with true PDFs with $CoV < 0.75$ the positive skewness decreases with increasing n. Once transformed into the space of B the skewness for most records is negative and enters a range of low skewness between –0.2 and +0.2 by $n > 7$ and stays within this range up to $n \approx 22$ (Figure 6). Records with strong burstiness develop positive skewness at higher n. Periodic records approach non-skewed, that is, symmetric, distributions at higher n.

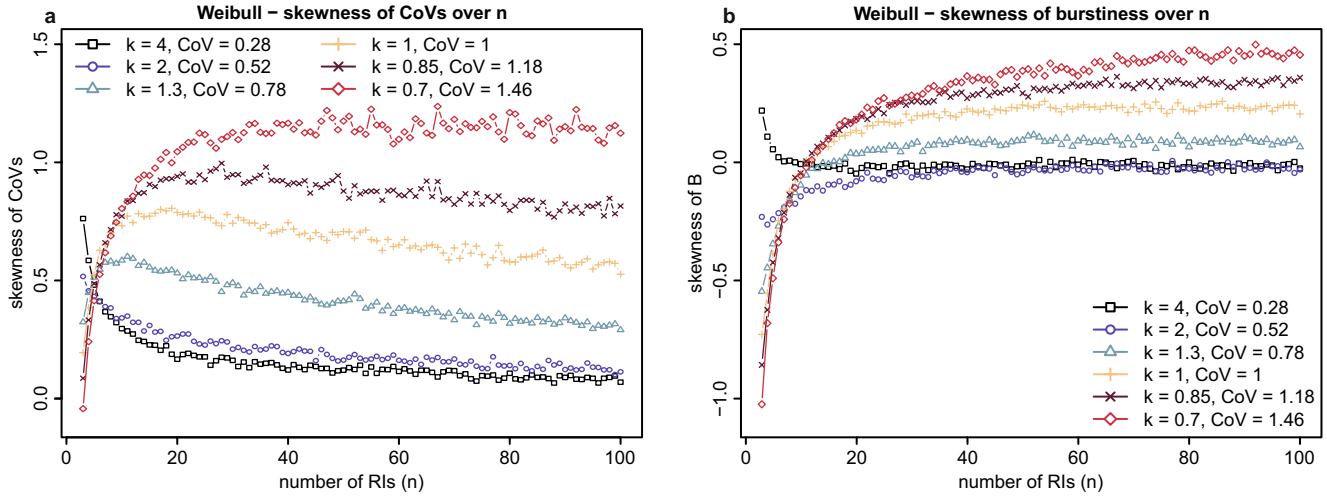

Figure 6. Skewness of the coefficient of variation (CoV) and burstiness (B) of Weibull distributed variables over recurrence intervals (n). (a) for $4 < n < 100$ all records show positive skewness and the skewness at high n is stronger the burstier the record is.(b) when transformed into the space of B skewness is weaker, much weaker for $n > 8$. Skewness of CoV and B over n of inverse Gaussian scenarios is in the supplementary information.

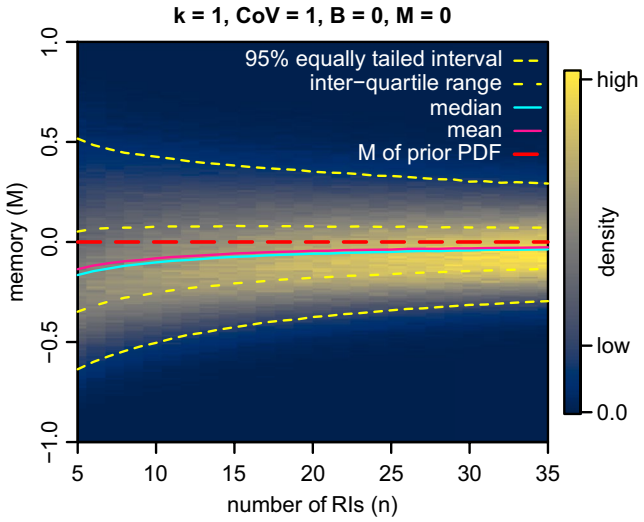

Figure 7. Density map of memory (M) from a synthetic record based on an exponentially distributed variable (Weibull distributed with shape parameter $k_{wei} = 1$) over the number of recurrence intervals (n). Note that neither the type of underlying distribution (Weibull or inverse Gaussian) nor the coefficient of variation of the true PDF nor age uncertainty appeared to have any influence on this pattern. All other density maps of M had seemingly identical density maps.

The true M of all synthetic records is 0, because of the independent sampling method that we used to construct the synthetic records. Means and medians of M of the sample show underestimation of the true M of the RI sequence, which is more pronounced at lower n (Figure 7). The underestimated, weakly negative M wrongly suggests that short (long) RIs have a tendency to follow long (short) RIs. The inter-quartile range and the 95% ETI narrows with increasing n, but remains relatively wide with $M \approx \pm 0.30$ for $n = 35$. The distribution of M is positively skewed even at relatively high n ($n > 35$). In contrast to the CoV and B, the asymptotic approach of the mean and median M to the true M appears to be independent from the aperiodicity of the true PDF and age uncertainty.

3.2. Statistics of Recurrence Intervals – With Added Age Uncertainty

To analyze the effect of age uncertainty in paleoseismic records we compare the results with and without added age uncertainty. With added age uncertainty the distributions of CoV and B are shifted toward higher values. This shift is more pronounced for strongly periodic records and the shift decreases with increasing true CoV and true B. This causes the mean and median of strongly and weakly periodic synthetic records to overestimate the true CoV and B regardless of the number of recurrence intervals (Figure 8). However, the added age uncertainty can partially mitigate the underestimation at low n. The overestimation decreases with

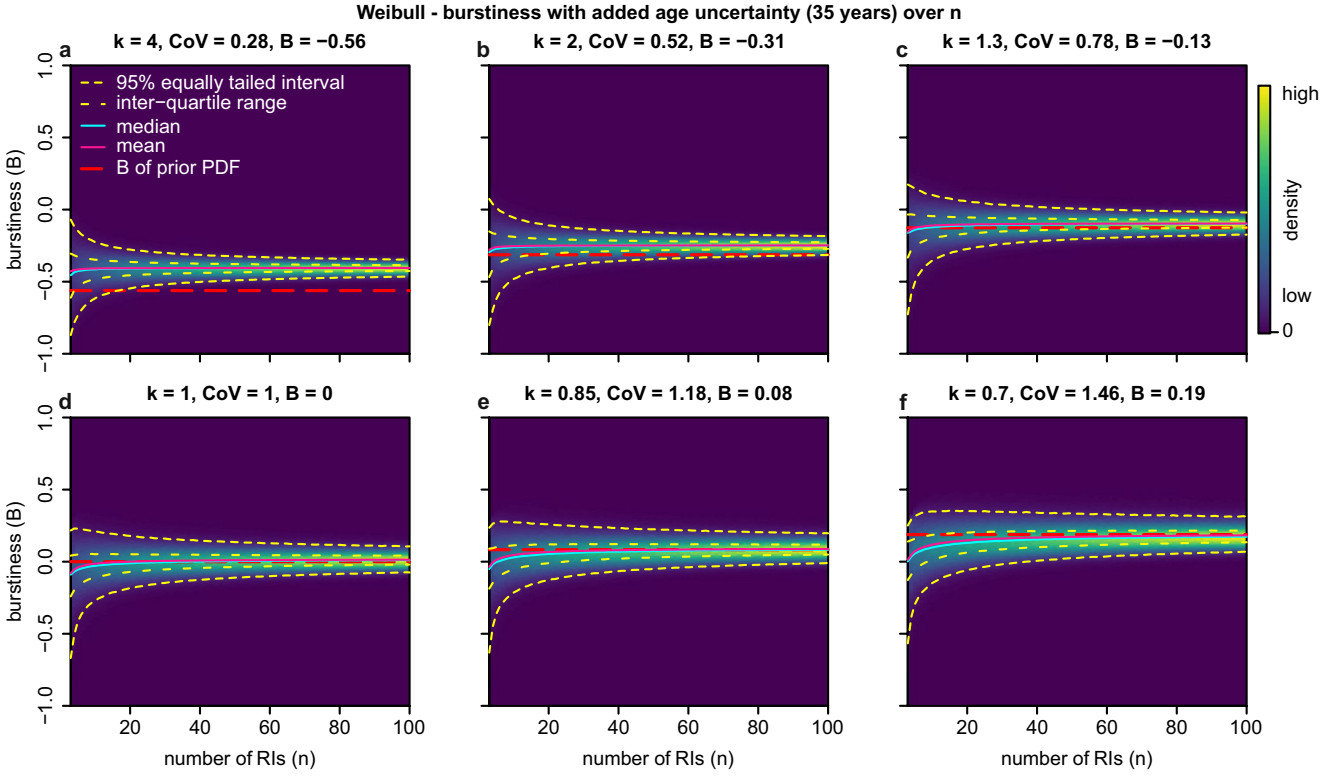

Figure 8. Density maps of burstiness (B) from synthetic records based on Weibull distributed variables with added age uncertainty over the number of recurrence intervals (n). The age uncertainty generation is based on a truncated normal distribution that is scaled to the recurrence interval that it is applied to with $\mu = 0$ and $\sigma = 35$ years (see Equation 3). Note that for low B (a), significant overestimation occurs for all n. For B between −0.13 and 0.08 (c–f), the B value is better estimated without a significant systematic offset. For the same type of figure for all computed Weibull distributed scenarios and all inverse Gaussian distributed scenarios, for all computed age uncertainties and for both CoV and B we refer to the supplementary information.

increasing CoV and B, to the point where records with added age uncertainty and $CoV \geq 1$ ($B \geq 0$) are almost indistinguishable from records without added age uncertainty.

The width of the 95% ETI and the skewness of all three statistics (CoV, B and M) behaves comparable to those in synthetic records without age uncertainty, that is, the 95% ETI decreases as n increases and the statistics are negatively skewed at low n, which in the case of CoV and B is overcome at high n.

4. Discussion

4.1. Memory (M) of Synthetic Records

In Goh and Barabási (2008), memory (M) was proposed for, among other types of data, an instrumental earthquake record from Japan which includes all earthquakes of magnitude >2. Paleoseismic records are fundamentally different from instrumental records, because they record only the strongest and much fewer events than the instrumental record.

In this study, the synthetic paleoseismic records are conceptually built with $M = 0$, yet the bounds of the 95% ETI are ±0.30 for $n = 35$ (Figure 7). For records with lower n the 95% ETI is even wider. The wide 95% ETI for records with $n < 35$ affects most paleoseismic records of large earthquakes. Given this insensitivity to randomly sampled synthetic paleoseismic records, we suggest to interpret weakly positive or weakly negative M with great caution in paleoseismology of rare large earthquakes, where usually $n < 35$. Based on a statistical analysis of a global compilation of 80 long-term records of earthquakes Griffin et al. (2020) come to a similar conclusion, where ca. 90% of the 80 records of large earthquakes are records with M indistinguishable from 0.

4.2. Underestimation of B in Paleoseismic Records Due to Low n

In real paleoseismic records, the estimates of true CoV and B generally suffer from low n, a common problem of inferential statistics in paleoseismology. The high uncertainty at low n in statistical analyses of paleoseismic records is intuitively understood and generally handled well in literature (e.g., McCalpin, 2009; Nishenko, 1985). Aperiodicity (CoV or B) becomes better constrained at lower n the more periodic the record is (e.g., Kempf et al., 2018; Moernaut, 2020).

In contrast to the high uncertainty, the bulk underestimation of CoV and B at low n is less intuitive and disproportionately affects bursty paleoseismic records. Although nicely demonstrated for the exponential distribution (e.g., Williams et al., 2019), this underestimation is commonly neglected or forgotten, when CoV or B are reported. This underestimation can be neglected in records with $n \geq 5$ and true $CoV < 0.3$ (true $B < -0.5$, Figures 4a and 5a). The burstier the record the stronger the underestimation becomes (Figures 4d–4f and 5d-5f).

The skewness of both CoV and B is almost always neglected in literature with few exceptions (e.g., Griffin et al., 2020). The positive skewness of the CoV at low n adds to the underestimation of CoV, because the mode, that is, the most likely result, of the CoV distribution underestimates the true CoV even more than the mean or median (Figure 6a). This issue is partially resolved after transformation of the CoV into the space of B, where the negative skew partially makes up for the underestimation (Figure 6b).

4.3. Overestimation of B in Paleoseismic Records Due to Age Uncertainty

Adding age uncertainty to a synthetic paleoseismic record bulk shifts the mean and median B and can thus lead to overestimation of B (Figure 8). The overall pattern of B over n is unchanged. The magnitude of this overestimation of B depends on how large the age uncertainty is relative to the mean RI and on the aperiodicity of the record. This effect disproportionately affects strongly periodic records (Figure 8a) to a degree that for most strongly periodic records ($CoV \leq 0.5$ and relative age uncertainty > 20%) the CoV and B are overestimated regardless of the number of RIs in the record. The same effect is negligible for aperiodic and bursty records (Figures 8d–8f). This is unsurprising when viewing it as adding the disorder of age uncertainty to either a highly ordered, that is, strongly periodic, sequence or to already disordered sequences such as aperiodic and bursty records.

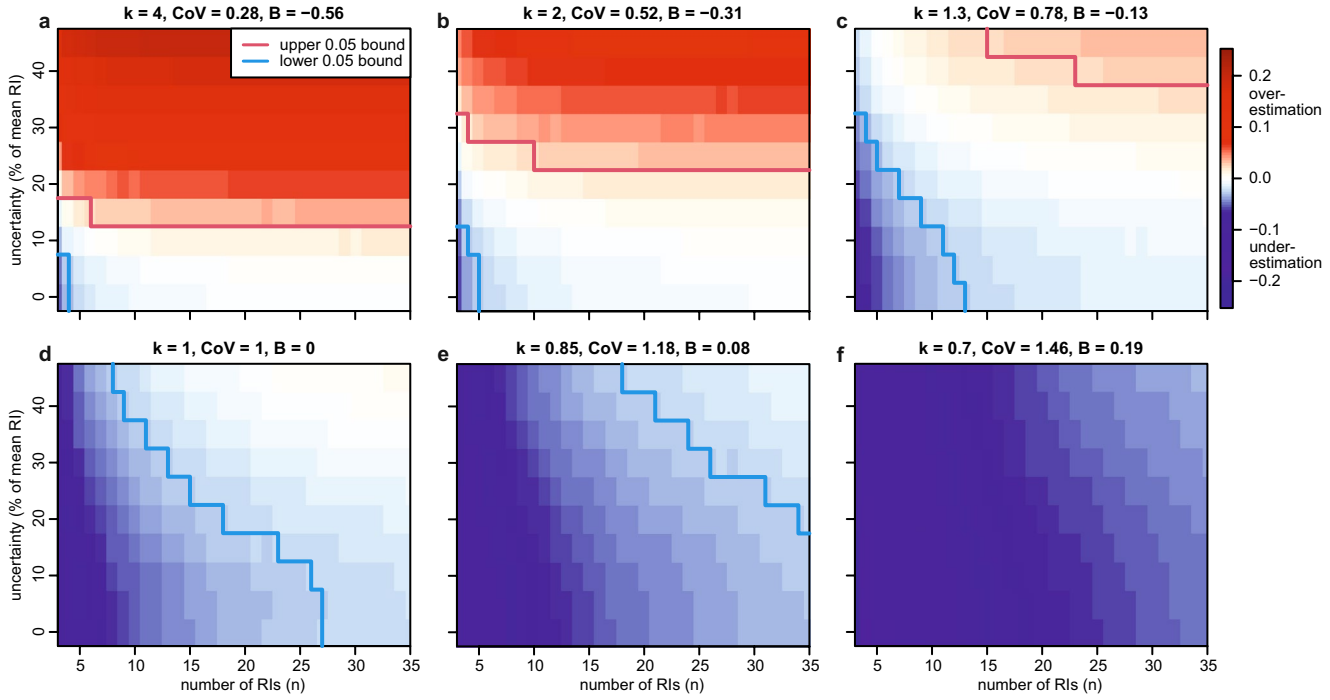

Figure 9. Overestimation and underestimation of the median burstiness depending on age uncertainty in the paleoseismic record over n. The arbitrary bounds of ±0.025 are equivalent to 95% at $CoV = 1$. The graphs show the difference between the median sample B of the synthetic records with the respective age uncertainty and the true B of the true distribution. In order to get accurate results for sample B of paleoseismic records, the age uncertainty and n should be below and left of the red lines and a above and right of the blue line, that is, in the white-colored areas. For low B, age uncertainty plays a major role (horizontal red lines in a and b), whereas for high B, the number of recurrence intervals becomes more important (subvertical blue lines in (d–f)). Here n ranges up to 35. The full studied range up to $n = 100$ can be downloaded from the supplementary information (Figures S5 and S6).

With both observations (overestimation and underestimation) in mind, we map the difference between the sample B and the true B depending on age uncertainty relative to the mean RI and n for the parametrizations described for the Weibull scenarios (Table 1, Figure 9) and inverse Gaussian scenarios (Table 2, Figure S2). In this figure we include a bracketed space of ±0.025 in the difference between sample B and true B that we call "accurate," which delimits situations where one would probably adequately estimate the true B from a paleoseismic record. This is an arbitrary boundary, which corresponds to 95% confidence interval, that is, ±0.05 to the CoV of a Poissonian process.

The importance of the discussed underestimation of B due to low n is evident. However, it is the interplay of this underestimation with the so far neglected overestimation of B due to age uncertainty in paleoseismic records that will decide whether the inferred aperiodicity of a paleoseismic record is likely accurate, overestimated or underestimated. For strongly periodic records, what matters most for an accurate assessment of B is the age uncertainty relative to the mean RI, as long as $n > 5$ (Figure 9a). For weakly periodic records, an accurate result is achieved with $n > 15$ as long as the age uncertainty remains at or under 25% of the mean RI (Figures 9b and 9c). For aperiodic and bursty records, age uncertainty plays a minor role. Instead, there is a need for high n to not underestimate B (Figures 9d–9f). Ironically, a uniform offset in the estimate of B of an aperiodic or bursty record is avoided, if age uncertainty in the record is relatively high (Figures 9e and 9f). Albeit mathematically true, this balancing effect cannot be used to improve the estimate of true CoV or B by increasing age uncertainty.

This model does not solve the challenge of trying to estimate a population parameter from imperfect data with limited sample size. Instead, the model can provide a better understanding of systemic biases in aperiodicity estimates and the associated uncertainties. To paleoseismologists who want to use this model's result, we suggest the following three-step plan:

1. Determine the number of recurrence intervals (n) and the mean relative age uncertainty in the record;

2. In Figures S1–S3 and/or S4 find the model, whose median value for the specific n best describes the paleoseismic record; and
3. Look at and understand the biases (overestimation or underestimation) and uncertainties (inter-quartile range and 95% ETI) associated with such a record and implement these in the interpretation of the paleoseismic recurrence statistics.

4.4. How Recurrence Intervals are Sampled Affects Burstiness

As there are differences in age models for paleoseismic records, there are also differences in how age and age uncertainty are assigned to event deposits and to the intervals between them. Furthermore, there are various approaches how to sample RIs from the event age distributions that the age model provides. Here we summarize six approaches of calculating burstiness, and demonstrate that they produce different results using five paleoseismic records from literature (Figure 10).

"Best age" approach: In the "best age" approach, a single age is derived from the event age distribution (often the median). This age is then used as the "best" and only age of the event. The result is that there is a single age for each event in the record and there is only a single correct solution to burstiness and no age uncertainty is propagated to the statistic (e.g., Kulkarni et al., 2013; Moernaut et al., 2018). This approach is enticing for its simplicity. However, it can produce extreme results compared to other methods (e.g., Figure 10a, records of Lake Calafquén and Lake Riñihue, purple vertical lines).

Age-depth model MCMC iterations: The MCMC iterations that make up many age-depth models can be used to calculate the RIs per iteration. When the MCMC iterations are based on sedimentation rate priors, this can substantially decrease relative age uncertainty. Usually, MCMC iterations that contain age reversals are impossible or are discarded in the process of age-depth modeling. Negative RIs are therefore avoided. In this approach, MCMC iterations that assigned a comparatively old age to one event are likely to assign a comparatively old age to the next youngest event in the record and vice versa. This stabilizes the resulting RI. The resulting burstiness of the RIs is often a narrower distribution compared to the random sampling methods discussed below (Figure 10a, black lines) (e.g., Kempf et al., 2017, 2020). The subjective informative priors used to run the age-depth model weakly influence how narrow this distribution is.

Random Sampling EAs and discarding negative RIs: Many onshore records cannot rely on sedimentation rate-based MCMC iterations. Instead, the MCMC iterations on the absolute event ages are only prevented from producing age reversals without a depth-dependent constraint, that is, the stratigraphic order of dates and event deposits is obeyed. This is the same as subtracting ages randomly sampled from the event age distributions of two consecutive event ages and discarding records that produced negative results (e.g., Scharer et al., 2011). The result is a distribution of burstiness. This method has a weak bias toward longer RIs and can be unlikely to produce valid solutions, if n is high and age uncertainty is relatively large in comparison to the RI (Figure 10a, record of Lake Tutira, dark red lines).

Random Sampling EAs and setting implausibly short or negative RIs to a geologically reasonable minimal value: Instead of discarding records that produce implausibly short RIs or illogical age reversals, RIs under a threshold value could be set to a geologically reasonable minimal value (e.g., Biasi et al., 2015; Scharer et al., 2011). Here, we use a threshold value of 20% of the mean RI, that is, 20 years. However, this value depends on the seismic region and the environment of the sedimentary record. This produces a polar bias toward RIs with the minimal duration. To our knowledge this is hypothetical, and has not been used this way when calculating the CoV or B for an earthquake sequence (Figure 10a, green-gray lines).

Random Sampling EAs and sorting EAs: If the event sequence is not perfectly stratigraphically ordered, for example, when the paleoseismic record is a regional composite record, where stratigraphic order is not evident, then the event ages can be randomly sampled and then sorted, before the RIs are calculated. Avoiding negative RIs by design. Solutions to such a composite record do not have a single correct solution for the order of events and can therefore produce a wide distribution for burstiness. This method causes underestimation of burstiness, if the record has a high age uncertainty relative to the mean RI, because then multiple stratigraphic orders become more likely (e.g., Figure 10a, record of Lake Tutira, red lines).

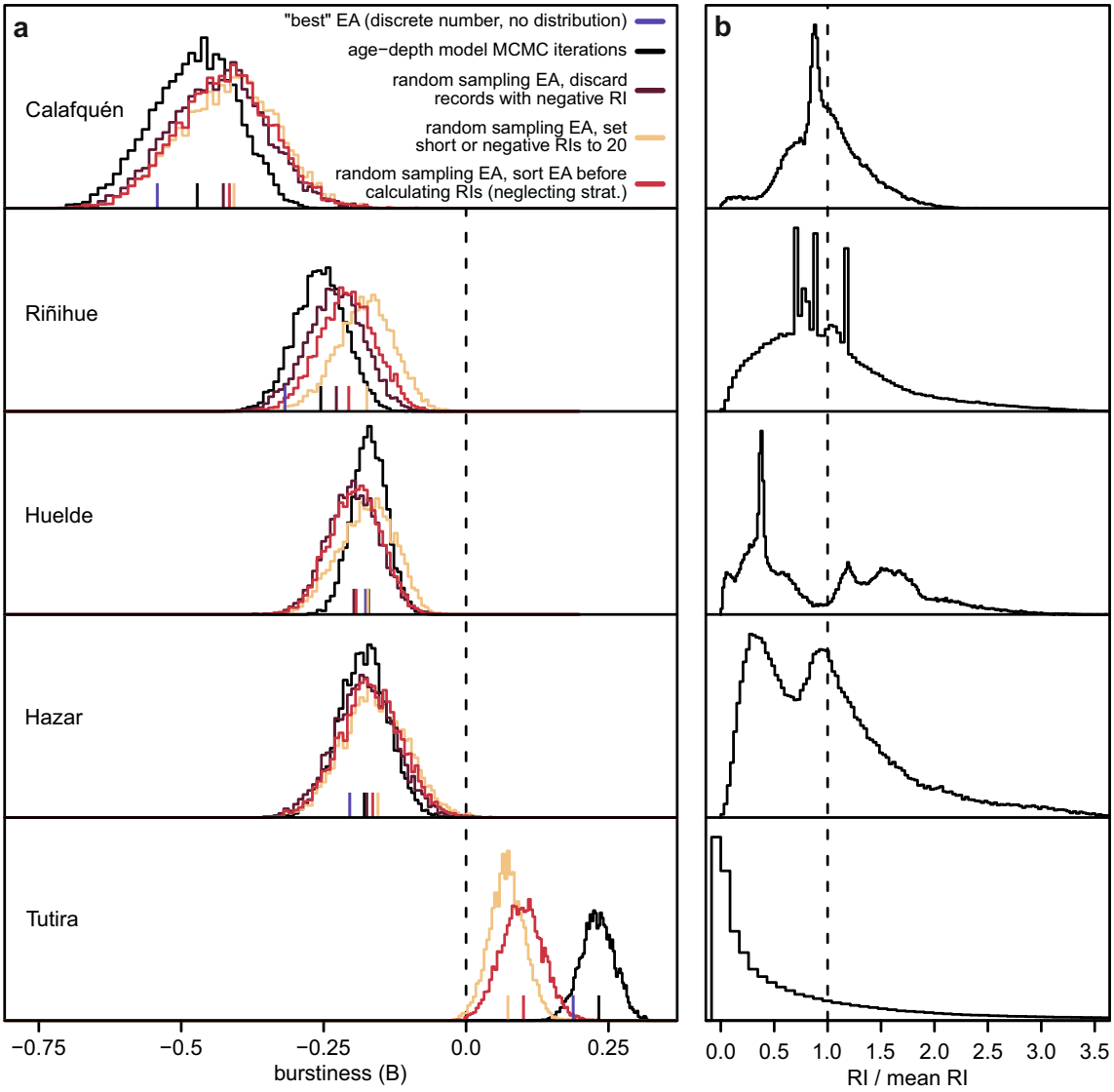

Figure 10. Example paleoseismic records of (top to bottom) Lake Calafquén (n = 12) and Lake Riñihue (n = 34) (Moernaut et al., 2018), Lake Huelde (n = 16) (Kempf et al., 2017), Lake Hazar (n = 20) (Hubert-Ferrari et al., 2020), and Lake Tutira (n = 118) (Gomez et al., 2015). (a) shows the results of various ways to sample recurrence intervals from a paleoseismic record and demonstrates that it matters for calculating burstiness (B). The vertical lines underneath the distributions are the arithmetic means of the distributions of the same color. (b) shows the overall recurrence pattern of the respective record. Spikes in this distribution are from historically known events that narrow the distribution of single RIs, for example, the Lake Calafquén, Lake Riñihue, and Lake Huelde records. Note that the method of the brown line did not produce a result for the Lake Tutira record.

CoV and B differ depending on the way the RIs are sampled from the paleoseismic record. This highlights a problem in comparing aperiodicities of paleoseismic records, for example, is the Lake Riñihue record (Moernaut et al., 2018) more periodic than the Lake Huelde record (Kempf et al., 2017) (Figure 10)?

Depending on which sampling method is used to determine the aperiodicity of either sedimentary record, one could come to the conclusion that either the Lake Riñihue record is more periodic than the Lake Huelde record (for "best age" approach or age-depth model MCMC iterations) or the Lake Riñihue record is more or less equally periodic to the Lake Huelde record (for the other three random sampling EA methods).

The difference in methods to calculate measures of aperiodicity prompts a discussion of when to use which method. If a Bayesian age-depth model consisting of MCMC iterations (simulated accumulation histories) is available, we suggest making use of the data with the MCMC iterations approach. It is mathematically sound and propagates the uncertainty correctly to the resulting statistic. The MCMC iterations approach

is especially powerful for marine and lacustrine records, where sedimentation rate is relatively constant or where changes in lithology can hint at changes in sedimentation rate. However, there are cases, where the age model does not rely on MCMC iterations and this method cannot be applied. The next best approach is to randomly sample the EA and discarding negative RIs. This method calculates aperiodicity and propagates uncertainties correctly. The only disadvantage is that it does not make use of the relative age information, that is, available in most age-depth models. This method should be applicable for onshore records from floodplains, marshes, and similar settings. However, discarding negative RIs and resampling the EAs can lead to excessively long sampling loops, if n is high and/or if many EAs overlap in time, for example, Lake Tutira record. In this case, we would advocate to follow Biasi et al. (2015) and set short or negative RIs to a minimal and geologically plausible period, for example, 20 years (depending on the seismic region and the sedimentary record), if n is low and the overlaps in EAs are significant. Or if n is high, use the "best age" approach.

5. Conclusions

With the help of synthetic records, with and without added age uncertainty, this study demonstrated five key findings:

1. The use of the auto-correlation coefficient of recurrence intervals, called memory (M), has probably little use in paleoseismology of large earthquakes, where often $n < 35$. At $n = 35$ the bounds of the 95% equally tailed interval of the synthetic records in this study are at ±0.30, when by design the synthetic records' true $M = 0$.
2. Low numbers of recurrence intervals (n) in paleoseismic records lead to underestimation of the coefficient of variation (CoV), which is increasingly relevant the more aperiodic the record. This underestimation is less pronounced once CoV is transformed to burstiness (B). Furthermore, the distributions of B are less skewed than distributions of CoV. We recommend the use of B, because it is more intuitive, less prone to underestimation at numbers of events that are common in paleoseismology and at the same time serves all the same purposes as the CoV.
3. Age uncertainty leads to overestimation of the CoV and B, especially for strongly periodic records. This effect has only a minor impact on random and bursty records.
4. It is the interplay between the underestimation of B due to low n and the overestimation of B due to age uncertainty relative to the mean recurrence interval that decides if the estimate of B is accurate. Ironically, there are realistic cases in which weak relative age control may counteract biases in the estimation of B that are introduced by low n, especially in aperiodic or bursty records. However, increasing age uncertainty cannot be used to improve the accuracy of the estimate of B. Rather, this result indicates how a paleoseismic record can be improved the most.
5. There are different types of age models and there are different ways to calculate burstiness from event age data. The results can differ strongly and one should carefully choose the most appropriate method for the information available and the characteristics of the record's sedimentary environment.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All data and code are stored and available with the Zenodo data and code repository with the DOI/URL: https://doi.org/10.5281/zenodo.4589987. Some of the data were compiled from publications by Moernaut et al. (2018), Kempf et al. (2017), Hubert-Ferrari et al. (2020), and Gomez et al. (2015), and are made available in the data and code repository entry is used here.

Acknowledgments
Basil Gomez, Alan Orpin, Aurelia Hubert-Ferrari are all thanked for sharing their age model information and further guidance on their paleoseismic records. Thanks are due to constructive reviews by Jonathan Griffin and Randolf Williams that substantially improved the quality of the manuscript.
Open access funding enabled and organized by Projekt DEAL.

References

Berryman, K. R., Cochran, U. A., Clark, K. J., Biasi, G. P., Langridge, R. M., & Villamor, P. (2012). Major earthquakes occur regularly on an isolated plate boundary fault. *Science*, *336*(6089), 1690–1693. https://doi.org/10.1126/science.1218959

Biasi, G. P., Langridge, R. M., Berryman, K. R., Clark, K. J., & Cochran, U. A. (2015). Maximum-likelihood recurrence parameters and conditional probability of a ground-rupturing earthquake on the Southern Alpine Fault, South Island, New Zealand. *Bulletin of the Seismological Society of America*, *105*(1), 94–106. https://doi.org/10.1785/0120130259

Blaauw, M., & Christen, J. A. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis*, *6*(3), 457–474. https://doi.org/10.1214/11-BA618

Bronk Ramsey, C., & Lee, S. (2013). Recent and planned developments of the program OxCal. *Radiocarbon*, *55*(2–3), 720–730. https://doi.org/10.2458/azu_js_rc.55.16215

Chen, Y., Liu, M., & Luo, G. (2020). Complex temporal patterns of large earthquakes: Devil's staircases. *Bulletin of the Seismological Society of America*, *110*(3), 1064–1076. https://doi.org/10.1785/0120190148

Efron, B. (2013). Bayes' theorem in the 21st century. *Science*, *340*(6137), 1177–1178. https://doi.org/10.1126/science.1236536

Goes, S. D. B., & Ward, S. N. (1994). Synthetic seismicity for the San Andreas fault. *Annali di Geofisica*, *37*(6), 1495–1513.

Goh, K.-I., & Barabási, A.-L. (2008). Burstiness and memory in complex systems. *Europhysics Letters*, *81*(4), 48002. https://doi.org/10.1209/0295-5075/81/48002

Gomez, B., Corral, Á., Orpin, A. R., Page, M. J., Pouderoux, H., & Upton, P. (2015). Lake Tutira paleoseismic record confirms random, moderate to major and/or great Hawke's Bay (New Zealand) earthquakes. *Geology*, *43*(2), 103–106. https://doi.org/10.1130/G36006.1

Griffin, J. D., Stirling, M. W., & Wang, T. (2020). Periodicity and clustering in the long-term earthquake record. *Geophysical Research Letters*, *47*(22), e2020GL089272. https://doi.org/10.1029/2020GL089272

Howarth, J. D., Fitzsimons, S. J., Norris, R. J., & Jacobsen, G. E. (2014). Lake sediments record high intensity shaking that provides insight into the location and rupture length of large earthquakes on the Alpine Fault, New Zealand. *Earth and Planetary Science Letters*, *403*, 340–351. https://doi.org/10.1016/j.epsl.2014.07.008

Howarth, J. D., Fitzsimons, S. J., Norris, R. J., Langridge, R., & Vandergoes, M. J. (2016). A 2000 yr rupture history for the Alpine fault derived from Lake Ellery, South Island, New Zealand. *GSA Bulletin*, *128*(3–4), 627–643. https://doi.org/10.1130/B31300.1

Howarth, J. D., Orpin, A. R., Kaneko, Y., Strachan, L. J., Nodder, S. D., Mountjoy, J. J., et al. (2021). Calibrating the marine turbidite palaeoseismometer using the 2016 Kaikura earthquake. *Nature Geoscience*, *14*(3), 1–7. https://doi.org/10.1038/s41561-021-00692-6

Hubert-Ferrari, A., Lamair, L., Hage, S., Schmidt, S., Çaatay, M. N., & Avşar, U. (2020). A 3800 yr paleoseismic record (Lake Hazar sediments, eastern Turkey): Implications for the East Anatolian Fault seismic cycle. *Earth and Planetary Science Letters*, *538*, 116152. https://doi.org/10.1016/j.epsl.2020.116152

Kempf, P., Moernaut, J., & De Batist, M. (2018). Bimodal recurrence pattern of tsunamis in south-central Chile: A statistical exploration of paleotsunami data. *Seismological Research Letters*, *90*(1), 194–202. https://doi.org/10.1785/0220180204

Kempf, P., Moernaut, J., Van Daele, M., Pino, M., Urrutia, R., & De Batist, M. (2020). Paleotsunami record of the past 4300 years in the complex coastal lake system of Lake Cucao, Chiloé Island, south central Chile. *Sedimentary Geology*, *401*, 105644. https://doi.org/10.1016/j.sedgeo.2020.105644

Kempf, P., Moernaut, J., Van Daele, M., Vandoorne, W., Pino, M., Urrutia, R., & De Batist, M. (2017). Coastal lake sediments reveal 5500 years of tsunami history in south central Chile. *Quaternary Science Reviews*, *161*, 99–116. https://doi.org/10.1016/j.quascirev.2017.02.018

Kulkarni, R., Wong, I., Zachariasen, J., Goldfinger, C., & Lawrence, M. (2013). Statistical analyses of great earthquake recurrence along the cascadia subduction zone. *Bulletin of the Seismological Society of America*, *103*(6), 3205–3221. https://doi.org/10.1785/0120120105

Lamair, L. (2018). *Holocene history of natural hazards in central Japan (Fuji five lakes): Imprints of earthquakes, typhoons and volcanic events in lacustrine sediments* (Doctoral dissertation). University of Liège. Retrieved from http://hdl.handle.net/2268/219597

Lu, Y., Wetzler, N., Waldmann, N., Agnon, A., Biasi, G. P., & Marco, S. (2020). A 220,000-year-long continuous large earthquake record on a slow-slipping plate boundary. *Science Advances*, *6*(48), eaba4170. https://doi.org/10.1126/sciadv.aba4170

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. https://doi.org/10.1785/0120010267

McCalpin, J. (2009). In R. Dmowska, D. Hartmann, & H. T. Rossby (Eds.), *Paleoseismology* (2nd ed.). Elsevier.

Moernaut, J. (2020). Time-dependent recurrence of strong earthquake shaking near plate boundaries: A lake sediment perspective. *Earth-Science Reviews*, *210*, 103344. https://doi.org/10.1016/j.earscirev.2020.103344

Moernaut, J., Van Daele, M., Fontijn, K., Heirman, K., Kempf, P., Pino, M., & De Batist, M. (2018). Larger earthquakes recur more periodically: New insights in the megathrust earthquake cycle from lacustrine turbidite records in south-central Chile. *Earth and Planetary Science Letters*, *481*, 9–19. https://doi.org/10.1016/j.epsl.2017.10.016

Nishenko, S. P. (1985). Seismic potential for large and great interplate earthquakes along the Chilean and Southern Peruvian Margins of South America: A quantitative reappraisal. *Journal of Geophysical Research*, *90*(B5), 3589–3615. https://doi.org/10.1029/JB090iB05p03589

Praet, N. (2020). *Towards the construction of a lacustrine paleoseismic record in south-central Alaska: A trembling tale of landslides and turbidites* (Unpublished doctoral dissertation). Ghent University.

Reid, H. F. (1910). The mechanism of the earthquake. In *The California earthquake of April 18, 1906, Report of the State Earthquake Investigation Commission* (Tech. Rep. No. 2). Carnegie Institution.

Richardson, K. N. D., Hatten, J. A., & Wheatcroft, R. A. (2018). 1500 years of lake sedimentation due to fire, earthquakes, floods and land clearance in the Oregon Coast Range: Geomorphic sensitivity to floods during timber harvest period. *Earth Surface Processes and Landforms*, *43*(7), 1496–1517. https://doi.org/10.1002/esp.4335

Rubin, C. M., Horton, B. P., Sieh, K., Pilarczyk, J. E., Daly, P., Ismail, N., & Parnell, A. C. (2017). Highly variable recurrence of tsunamis in the 7,400 years before the 2004 Indian Ocean tsunami. *Nature Communications*, *8*, 16019. https://doi.org/10.1038/ncomms16019

Salditch, L., Stein, S., Neely, J., Spencer, B. D., Brooks, E. M., Agnon, A., & Liu, M. (2020). Earthquake supercycles and long-term fault memory. *Tectonophysics*, *774*, 228289. https://doi.org/10.1016/j.tecto.2019.228289

Sawai, Y., Kamataki, T., Shishikura, M., Nasu, H., Okamura, Y., Satake, K., & Aung, T. T. (2009). Aperiodic recurrence of geologically recorded tsunamis during the past 5500 years in eastern Hokkaido, Japan. *Journal of Geophysical Research*, *114*(B01319). https://doi.org/10.1029/2007JB005503

Scharer, K. M., Biasi, G. P., & Weldon, R. J. (2011). A reevaluation of the Pallett Creek earthquake chronology based on new AMS radiocarbon dates, San Andreas fault, California. *Journal of Geophysical Research*, *116*(B12), B12111. https://doi.org/10.1029/2010JB008099

Schwartz, D. P., & Coppersmith, K. J. (1984). Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas Fault zones. *Journal of Geophysical Research: Solid Earth*, *89*(B7), 5681–5698. https://doi.org/10.1029/JB089iB07p05681

Shennan, I., Bruhn, R., Barlow, N., Good, K., & Hocking, E. (2014). Late Holocene great earthquakes in the eastern part of the Aleutian megathrust. *Quaternary Science Reviews*, *84*, 86–97. https://doi.org/10.1016/j.quascirev.2013.11.010

Sieh, K., Natawidjaja, D. H., Meltzner, A. J., Shen, C.-C., Cheng, H., Li, K.-S., & Edwards, R. L. (2008). Earthquake supercycles inferred from sea-level changes recorded in the corals of West Sumatra. *Science*, *322*(5908), 1674–1678. https://doi.org/10.1126/science.1163589

Smith, S. B., Karlin, R. E., Kent, G. M., Seitz, G. G., & Driscoll, N. W. (2013). Holocene subaqueous paleoseismology of Lake Tahoe. *The Geological Society of America Bulletin*, *125*(5–6), 691–708. https://doi.org/10.1130/B30629.1

Williams, R. T., Davis, J. R., & Goodwin, L. B. (2019). Do large earthquakes occur at regular intervals through time? A perspective from the geologic record. *Geophysical Research Letters*, *46*(14), 8074–8081. https://doi.org/10.1029/2019GL083291

Wils, K., Daele, M. V., Kissel, C., Moernaut, J., Schmidt, S., Siani, G., & Lastras, G. (2020). Seismo-turbidites in Aysén Fjord (Southern Chile) reveal a complex pattern of rupture modes along the 1960 Megathrust earthquake segment. *Journal of Geophysical Research: Solid Earth*, *125*, e2020JB019405. https://doi.org/10.1029/2020JB019405