Long-term slip history discriminates among occurrence models for seismic hazard assessment - supplementary material

D.D. Fitzenz,[1] M.A. Ferry,[1] and André Jalobeanu,[1,2]

1. The Bayesian framework

In this section, we present the mathematical formalism used for data integration and model combination. By model, we mean the type of probability density function used to describe the inter-event time distribution for full-segment ruptures on a given fault-segment. In this study, all models have two parameters that we will refer to as a and b. They are:
The BPT distribution:

$$P(x\,|\,a,b) = \left(\frac{b}{2\pi\,x^3}\right)^{1/2} \times \exp\left(-\frac{b(x-a)^2}{2a^2x}\right) \tag{1}$$

The lognormal distribution:

$$P(x\,|\,a,b) = \frac{1}{xb\sqrt{(2\pi)}} \times \exp\left(-\frac{(\log(x)-a)^2}{2b^2}\right) \tag{2}$$

and the Weibull distribution:

$$P(x\,|\,a,b) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} \times \exp\left(-\left(\frac{x}{a}\right)^b\right) \tag{3}$$

1.1. Investigation of the parameter space for each model, as constrained by the observations

1. Integration of seismicity data: We compute the likelihood for the observed inter-event times.
It consists of the product over each inter-event time of the product of the functional form of the model (eq. 1 to 3) by the distribution of the observed time interval, integrated with respect to the unknown inter-event times. Because we use historical and archeological data (uncertainties on dates are below 50 years), we chose Dirac distributions for the observed time intervals, i.e., perfectly known dates with respect to our problem. The expression of the likelihood simplifies as shown by equation 4:

$$L_{CLE}(a,b) = \prod_{i=1,5}\int P(\Delta t_i\,|\,a,b) \times P(\Delta t_i^{obs}\,|\,\Delta t_i)\,d\Delta t_i = \prod_{i=1,5} P(\Delta t_i^{obs}\,|\,a,b) \tag{4}$$

However, we could choose other distributions reflecting the observational error to integrate other types of data, such as paleoseismological data.
To get $P_{CLE}(a,b)$, the CLE posterior pdf, proportional to $L_{CLE}(a,b) \times P(a,b)$ from Bayes' rule, one has to choose a prior distribution for a and b. We want priors including as little information as possible. We choose flat distributions for the mean recurrence time and for the other parameter b. This translates into a flat distribution for a for the BPT and the Weibull distributions, and into $\exp(a)$ for the lognormal distribution. Indeed, if a' follows a flat distribution (e.g., $P(a') = C$), and $a = \log(a')$, then $da' = da \times \exp(a)$. Since $P(a')da' = P(a)da$, $P(a) = C \times \exp(a)$. For mathematical convenience, we choose a lower bound of 1 year for the mean recurrence interval. The upper bound is set to 500,000 years in this study because it is high enough to represent infinity. For BPT and Weibul, $P(a) = C$ between 1 and 500,000 years, and for BPT, $P(a) = C\exp(a)$ between log(1) and log(500,000). The prior on b is chosen flat between $-\infty$ and $+\infty$. This choice has no consequence on model comparison as the CLE posterior will be used after normalization. What is shown on Fig2A are the normalized CLE posteriors for the 3 distributions.

2. Use of cumulative slip data.
We compute the likelihood from the square of differences between synthetic cumulative slips and observations at the time of the geomorphic markers, averaged over the number of realizations N (number of catalogs produced for a couple of parameters). We assume the measurement noise of the l-th cumulative slip follows a Gaussian distribution of standard deviation $\sigma_D[l]$, and n is the number of data points. We omit the multiplicative constant

since it does not depend on the inter-event time model.

$$L^{synth}_{CLE}(a,b) = P(D^{obs} \,|\, a,b) \propto \frac{1}{N}\sum_{k=1}^{N} \exp\left(-\sum_{l=1}^{n} \frac{(D^{synth}_k[l](a,b) - D^{obs}[l])^2}{2\,(\sigma_D[l])^2} \right) \tag{5}$$

3. Combined use of CLEs and cumulative offsets: We compute the posterior probability by using P_{CLE} as a prior and by multiplying it to the likelihood of the geomorphic markers.

$$P(a,b \,|\, D^{obs}, \Delta t^{obs}) \propto P_{CLE}(a,b) \times L^{synth}_{CLE}(a,b) = P(a,b \,|\, \Delta t^{obs}) \times P(D^{obs} \,|\, a,b) \tag{6}$$

This iterative process (the fact of using a posterior as a prior to integrate an additional dataset) is what is called Bayesian updating. Hence, we are ready to integrate more datasets as they become available.

1.2. Using the (seismicity) prior or the combined (CLE - synthetic CLE) posterior to compute the corresponding distribution of inter-event times and the cumulative density function.

The prior and the posterior give the probability of the parameters knowing the data, for a given model. We can therefore go beyond picking the optimum (a_0, b_0) and tracing $P(\Delta t, a_0, b_0)$, with Δt ranging from 0 to 10,000 years. We can either 1) compute the unique pdf resulting from the weighted average of all $P(\Delta t, a, b)$ with the weights given by the probability of (a,b) given by the prior or the posterior (FigS1); or 2) we can use all possible $P(\Delta t, a, b)$ to calculate 95% confidence intervals at given times. In that last case, at given values for Δt, we can compute histograms for the value of $P(\Delta t, a, b)$, using prior(a,b) (resp., posterior(a,b)) to update the number of occurrences of each value. In both cases, with the added constraint that no earthquake happened since 1033, we truncate the pdf and keep only the part for $\Delta t > 977$. We renormalize. The cumulative density function

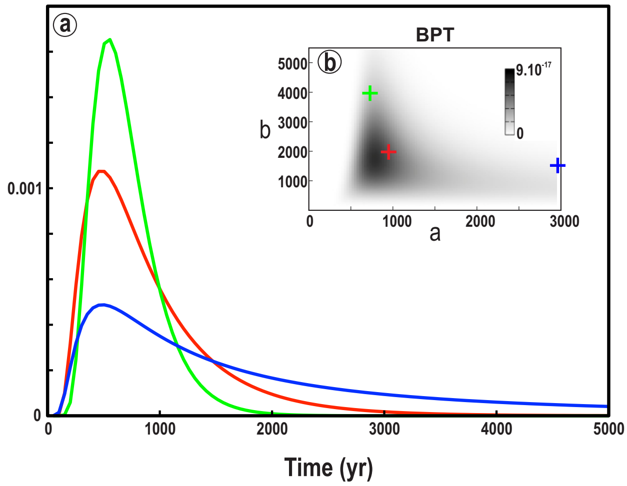

Figure 1. PDFs of the same BPT law with parameters chosen in the seismicity likelihood. The large a parameters in S1b yield a much heavier tail in S1a, and this is reflected in the shallow slope of the CDFs in Fig. 2D.

cdf(t,a,b) is the probability for an event to occur before t. It is the integral of $P(\Delta t, a, b)$ between $-\infty$ and t. We can compute a cdf for the weighted average, or for each individual $P(\Delta t, a, b)$. Fig 2D shows both types of results. The lines show the cdf of the weighted averages (one from the prior, one from the posterior) and the error bars show the 95% confidence intervals on the cdfs. They reflect the histograms of cdf(30 yr, a, b) and cdf(300 yr, a, b).

1.3. Model combination

From Bayes' theorem, the posterior probability of each model $P(model|data)$ is proportional to the evidence $\times$ the prior on the model. Assuming that we choose to assign equal priors to each model (we can not tell before getting the data which model is more plausible), the different models are ranked by evaluating the evidence [*MacKay*, 2003]. It is, by definition, the integral of the product of the prior and the cumulative slip likelihood $\int P(a,b \,|\, \Delta t^{obs}) \times P(D^{obs} \,|\, a,b)\, da db$ involving the un-normalized expression of step 3.

Note that the plausibility for competing models is deemed significantly different only if the evidences are different by at least one order of magnitude.

In the combination, the relative weights of the different models are given directly by the ratio between their evidences.

In the case of the JVF, our complete dataset yields the following relative weights for each model: 0.45 for Weibull, 0.33 for logN, and 0.22 for BPT, i.e., no model is significantly better than the others, and we need to use them all in a combination (we can not select one). The final model is the weighted average of the three composite laws (i.e., the three models, but for all their parameter values according to the posterior probability of the parameters).

1.4. Comment on the meaning of data integration

When we use the likelihood computed with the seismicity data to compute the probability density function for the next event (and go on to the cdf), we manipulate composite laws. We should therefore not expect to get the same properties as with individual laws taken at the optimum parameter values. BPT, Weibull and lognormal laws with their optimum parameters were found to be very similar to each other until we reach times that are 2 to 3 times the peak time [*Matthews et al.*, 2002]. However, we show here that for the composite laws (i.e., the ones which incorporate all the data) the cdfs are significantly different, as are their uncertainties, even within 300 years after a peak time 3 times as large. Finding the best model is therefore crucial even at relatively short time scales.

References

MacKay, D. J. C. (2003), *Information theory, inference, and learning algorithms*, Cambridge University Press.

Matthews, M., W. L. Ellsworth, and P. A. Reasenberg (2002), A brownian model for recurrent earthquakes, *Bull. Seis. Soc. Am.*, *92*(6), 2233–2250.