Basic Study on Seismicity Smoothing Based on Quantification of Uncertainty Due to Lack of Knowledge

Minoru Matsubara[a*]
[a]Taisei Corporation, Tokyo, Japan

Abstract: In seismic hazard assessment, seismic activity prediction is performed using past earthquake records, and maximum likelihood estimation of G-R law parameters is typical. For example, Frankel's (1995) Smoothed seismicity does not assume regional divisions, but instead estimates future earthquake occurrence models based on the distribution of historical earthquakes and its smoothing. This smoothing process is unnecessary if the data is sufficient. In other word, it is considered to be a process to compensate for the lack of data. Frankel (1995) sets the smoothing correlation distance to 50 km, but the reason is not quantitatively stated in the paper. In this study, I evaluate smoothed seismicity based on quantitative evaluation of uncertainty due to lack of observation records.

Keywords: PRA, SHA, NPP, Smoothed Seismicity

1. INTRODUCTION

Predicting future seismic activity, seismic hazard analysis, involves two types of uncertainty: aleatory uncertainty and epistemic uncertainty. Epistemic uncertainty is an uncertainty caused by a lack of knowledge (observation records needed to determine model parameters). Seismic hazard analysis takes these uncertainties into account and evaluates the variability of future seismic hazard. In particular, epistemic uncertainty is modeled as a branch of a logic tree.

For example, the spatial distribution of the frequency of background earthquakes can be evaluated correctly when there are sufficient records of past earthquake occurrences. However, when the frequency of earthquakes is low, and the observation period is short, the epistemic uncertainty arises in evaluating the spatial distribution.

One of the methods for dealing with the spatial distribution of earthquake frequency is to define some regions in which the spatial distribution is uniform based on the seismotectonic map (for example, In Japan, Hagiwara (1991) and Kakimi et al. (2003) are often referenced). In this method, the spatial distribution is set with an emphasis on geological knowledge and not on the spatial distribution of past observation records.

On the other hand, there is a method called “smoothed seismicity”. Frankel (1995) proposed that the frequency of earthquakes observed for each mesh be smoothed by the following equation to give the frequency of earthquakes in the mesh. Here, $\tilde{n}_i$ is the number of earthquakes in the i-th mesh after smoothing, n_j is the number of earthquakes that occurred in the j-th mesh, Δ_{ij} is the distance between the i-th mesh and the j-th mesh, and c is the correlation distance.

$$\tilde{n}_i = \frac{\sum_j n_j \exp\left(-\frac{\Delta_{ij}^2}{c^2}\right)}{\sum_j \exp\left(-\frac{\Delta_{ij}^2}{c^2}\right)} \tag{1}$$

While the method of Frankel (1995) uses a constant correlation distance regardless of the location, A. Helmstetter et al. (2014) proposed the smoothing distance d_i associated with earthquake i as the horizontal distance between event i and the n_vth closest neighbor.

There are some methods for evaluating spatial distribution, but it is difficult to estimate true distribution when the observation records are few. In consideration of such uncertainty, multiple methods are often incorporated into hazard assessment by making branches of a logic tree. This branches are mainly based on the qualitative judgement, not on the quantitative evaluation.

This study proposes the method for quantifying the epistemic uncertainty to verify the logic tree or to establish the new way for constructing the branches of the logic tree based on the quantified uncertainty.

2. Quantification of the Epistemic uncertainty

2.1. Formulation of epistemic uncertainty

In this study, epistemic uncertainty is quantified using Bayes' theorem. Bayes' theorem is expressed by the following equation.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \propto P(B|A)P(A) \tag{2}$$

A is the parameter whose probability distribution is to be evaluated, and B is the observed records. $P(A)$ is the probability distribution of A before observation, and is called the prior distribution. $P(B|A)$ is called the likelihood function. The distribution $P(A|B)$ represents the epistemic uncertainty.

In this study, the epistemic uncertainty of the spatial distribution of occurrence frequency is quantified by the following procedure.

- The random variable considered in this study is an N-dimensional random variable $\boldsymbol{X}$, which is a combination of random variables $\boldsymbol{X}^i$ (i is the mesh number) relating to the frequency of each of N meshes. Here, $\boldsymbol{x}_{obs}$ is the spatial distribution of the observed frequency, and $\boldsymbol{x}^*$ is the spatial distribution of the true frequency. In this study, it is assumed that $\boldsymbol{x}^*$ is not significantly different from $\boldsymbol{x}_{obs}$, and is $\boldsymbol{x}_{obs}$ or a smoothed distribution of $\boldsymbol{x}_{obs}$. In other words, the set $\mathfrak{X}$ of $\boldsymbol{x}$ considered in this study is represented as follows using the Gaussian smoothing function $g(x_{obs}, r_0)$:

$$\mathfrak{X} = \{g(\boldsymbol{x}_{obs}, c) | c > 0 \in R\} \tag{3}$$

  $\boldsymbol{x} \in \mathfrak{X}$ can be expressed as a function of the correlation distance c, so I will discuss c below. Frankel (1995) proposes 25km and 50km for the distribution of c, but here it is assumed that there is no useful information on the distribution shape, and the following non-informative prior distribution is assumed (uniform distribution, Shown in Figure 1. b and a are constants).

$$p_C(c) = \frac{1}{b-a} \qquad (a \le c \le b) \tag{4}$$

- Based on the smoothed frequency distribution $\boldsymbol{x} = g(\boldsymbol{x}_{obs}, c)$, a probability model is set for each mesh, assuming a Poisson distribution. Note that $\boldsymbol{x}$ can be expressed as a function of c, the likelihood function is expressed as follows:

$$p(\boldsymbol{x}_{obs}|c) = \left\{ \prod_i \frac{\exp\left(-x(c)^i\right) x(c)^{i^{x_{obs}^i}}}{x_{obs}^i\,!} \right\}^{\left(\frac{N'}{N}\right)} \tag{5}$$

- Here, $x^i(c)$ is the frequency of the i-th mesh after smoothing with the smoothing distance c, and x_{obs}^i is the number of earthquakes observed in the i-th mesh. The mesh to be multiplied here is within about 20 km that has an impact on the site (for a 10 km mesh, the mesh containing the site and the 8 meshes surrounding it).

- Based on the above, the epistemic uncertainty can be quantified using the following formula.

$$p(c|\boldsymbol{x}_{obs}) \propto p(\boldsymbol{x}_{obs}|c) p_C(c) \tag{6}$$

2.2. Trial analysis under a virtual earthquake environment

To confirm how epistemic uncertainty is quantified by the proposed method, an analysis was conducted for a hypothetical earthquake environment. Figure 2 shows the assumed earthquake environment. In this study, I consider two cases. Case A, where the maximum frequency is 0.45×10^{-4}/year/km^2, corresponds to an area with relatively high seismic activity in Japan. And Case B, where the maximum frequency is 0.45×10^{-5}/year/km^2 , corresponds to an area with relatively low seismic activity.

Based on the occurrence frequency distribution of Figure 2, I generated earthquake occurrence records for 20, 100, 500, and 10,000 years. The generated earthquake occurrence records (M5 or greater) are shown in Figure 3.

For the records in Figure 3, I calculate the posterior distribution of the smoothed distances by Equation (6). But the lower and upper limits (a, b) of the smoothed distance were not specified in Equation (3). Figure 4 shows the smoothed distribution for the 100-year observation period of Case B. The distribution of the correlation distance 25km or 50km is affected strongly by randomness of generating records and it appears to differ from the true spatial distribution. When the correlation distance is 100km, the smoothed distribution is almost uniform. So it is assumed that the lower limit of the smoothing distance is 10km, and the upper limit is 100km.

The posterior distributions are shown in Figure 5. In Figure 5, the maximum value of the posterior probability is normalized to 1. In both cases A and B, the posterior distribution is updated more significantly as the observation period is longer. Also, the distribution is updated more significantly in case A, which has a higher occurrence frequency, than in case B.

Figure 6 shows the seismic hazard after smoothing for each correlation distance. The seismic hazard evaluation is performed according to Noda et al. (2002) (acceleration response spectrum value of 0.02s and 5% damping). The depth of the hypocenter was fixed at 5km. Figure 7 shows the average and variance of the seismic hazard calculated taking into account the posterior probability distribution in Figure 3 and epistemological uncertainty. In Case A, the variation was estimated to be small based on records spanning about 100 years, but in Case B, records spanning about 100 years were insufficient to evaluate the distribution of earthquake occurrences, and the variation in earthquake hazard was estimated to be large.

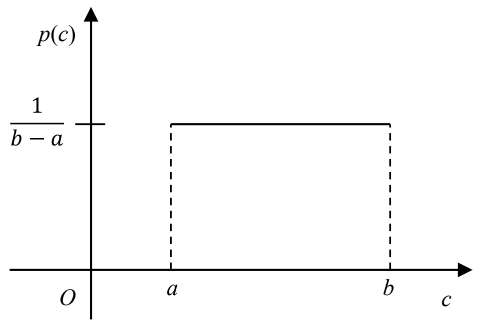


Figure 1. non-informative prior distribution

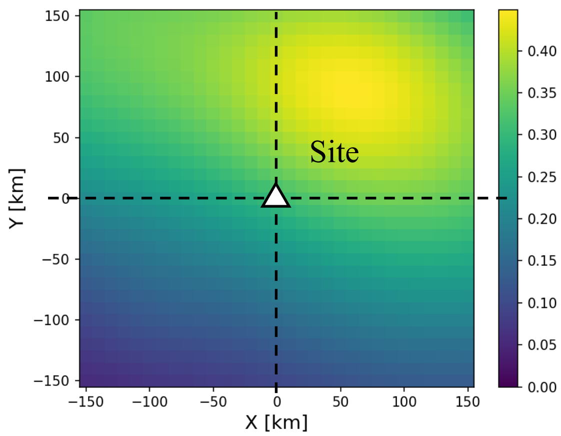
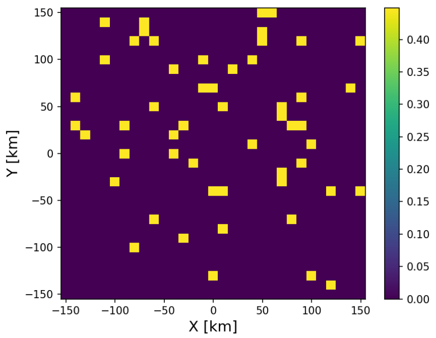
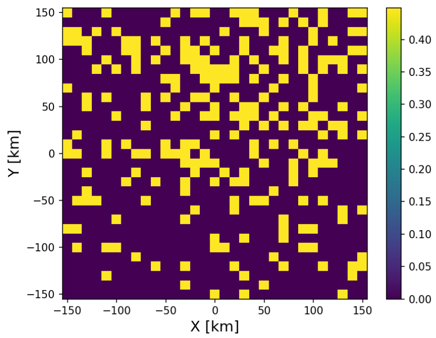


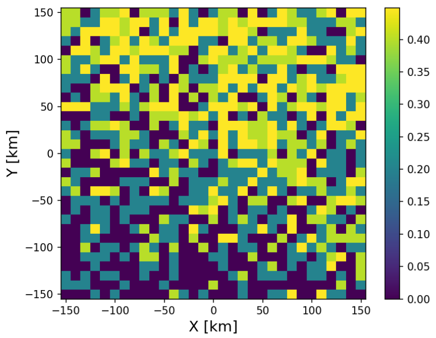
Figure 2. The spatial distribution of occurrence frequency
(In Case A: $\times 10^{-4}$ / year/km^2,
In CaseB: $\times 10^{-5}$ / year/km^2)



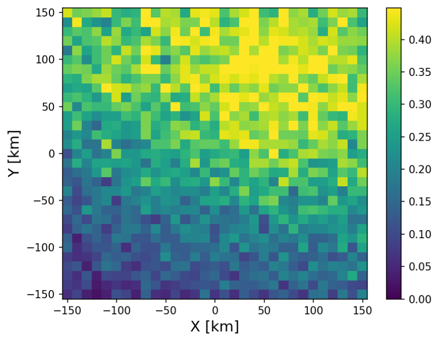
(a) 20 years



(b) 100 years

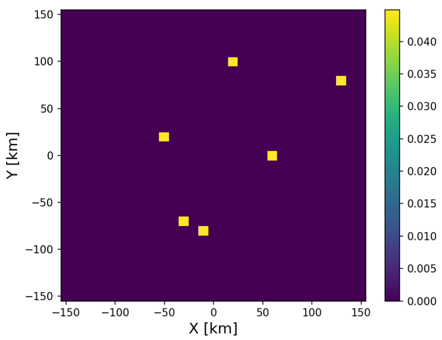


(c) 500 years

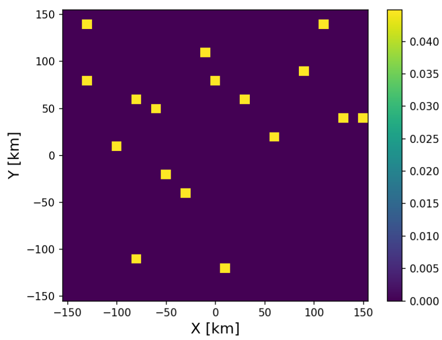


(d) 10000 years

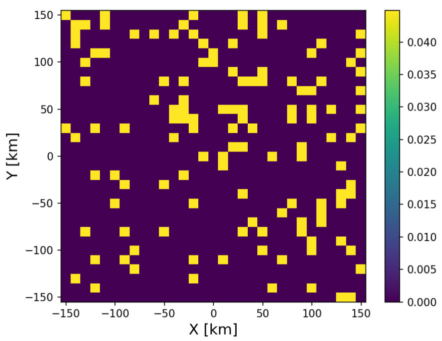
(1) Case A



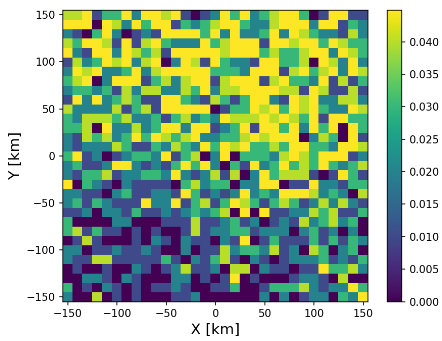
(a) 20 years



(b) 100 years



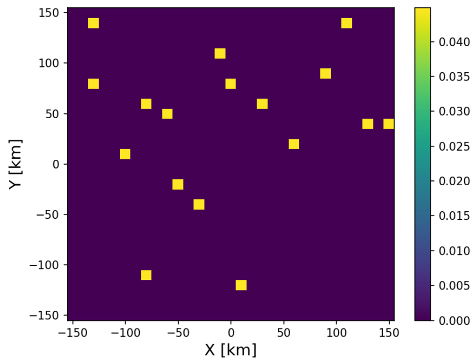
(c) 500 years



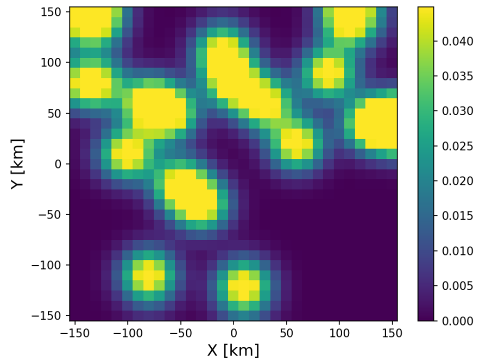
(d) 10000 years

(2) Case B

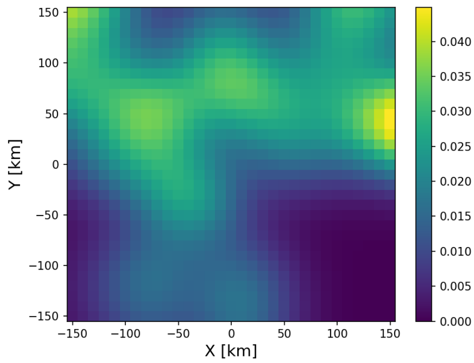
Figure 3. Generated earthquake records （$\times 10^{-4}$/year/km^2）



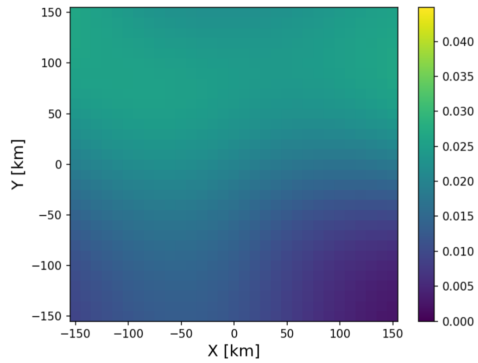
(1) Before smoothing



(2) c=25km

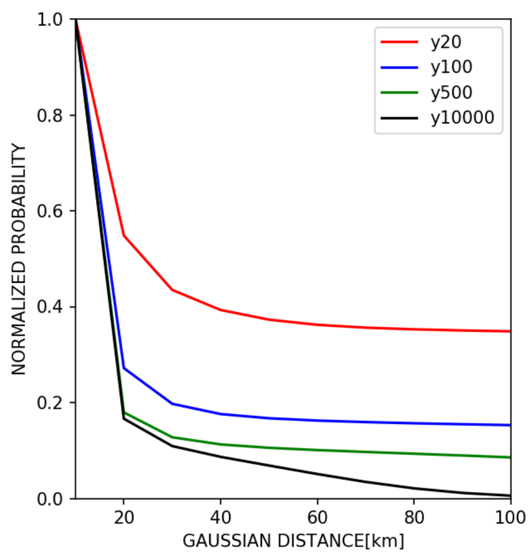


(3) c=50km

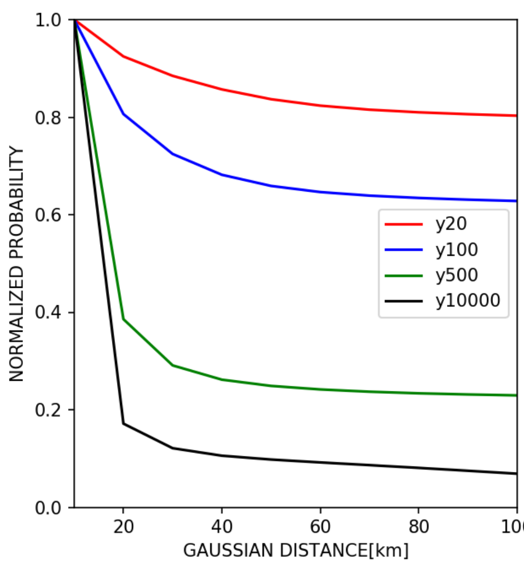


(4) c=100km

Figure 4. Smoothed distributions (Case B, 100years)

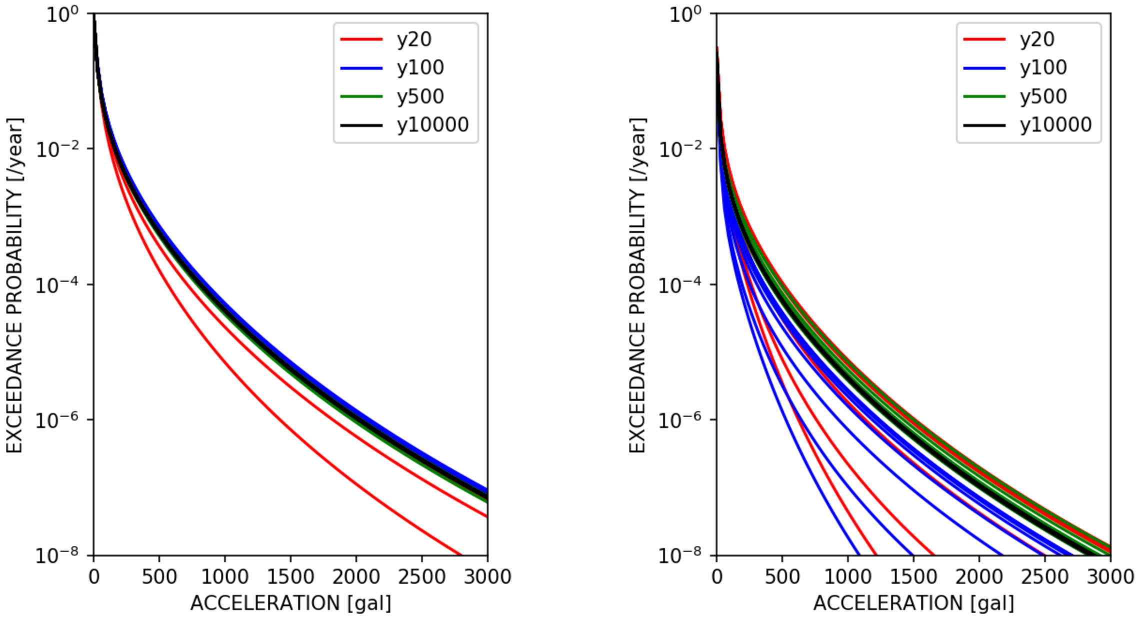


(1) Case A



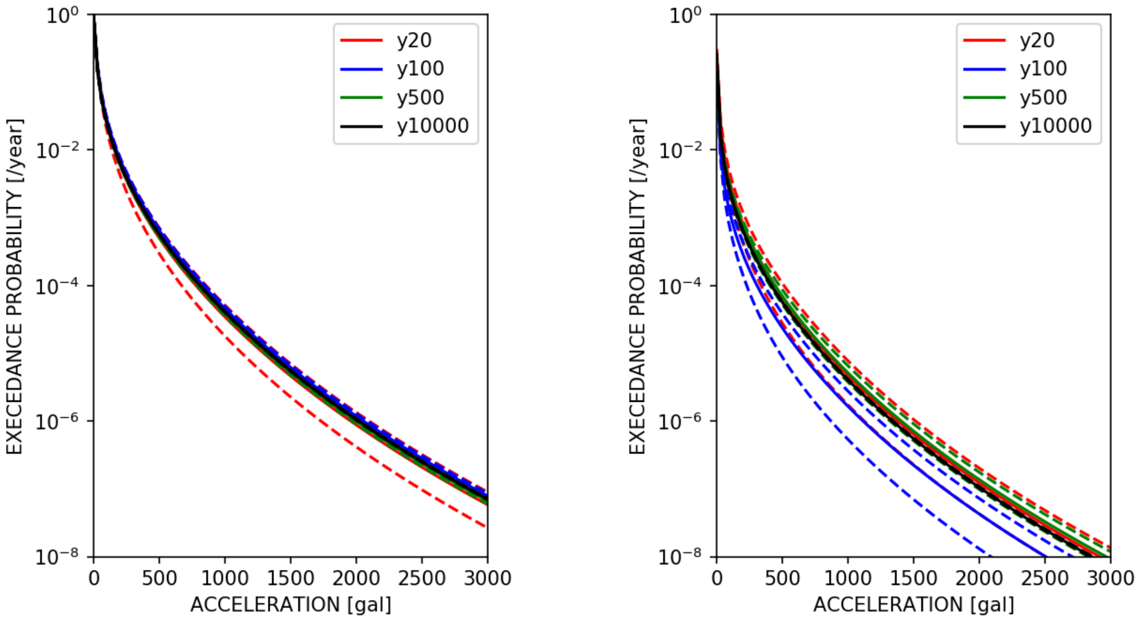
(2) Case B

Figure 5. Posterior distribution



(1) Case A (2) Case B

Figure 6. Hazard curves for smooth distances of 10km-100km (in 10km increments)



(1) Case A (2) Case B

Figure 7. Mean and variance ($\mu \pm \sigma$) of earthquake hazard considering posterior probability

3. Logic tree modeling based on the quantification of epistemic uncertainty

In Chapter 2, I performed an evaluation of the posterior probability distribution with smooth distance as a parameter and a hazard analysis, and quantified the epistemic uncertainty. However, this procedure requires a large number of hazard analyses, which is cumbersome, so it's necessary to consider how to incorporate it into the hazard analysis. Here, I consider a method to introduce the equivalent epistemic uncertainty of earthquake hazards into the logic tree using a point estimation method.

Now, it's possible to calculate the mean, standard deviation, and skewness from the posterior probability distribution of the smooth distance *c*. Using the two-point estimation method, the two smooth distances c^+, c^- and their weights p^+, p^- whose mean, standard deviation, and skewness match the original distribution can be calculated, I set these as a branch of the logic tree (see Figure 8).

Figure 9 shows the hazard curve at the sample points of the two-point estimation method. Figure 10 shows the mean and variance of the hazard curve calculated from the result of Figure 9. Although the variance is somewhat large, the results are generally consistent with Figure 7. Since the effect of the smoothing distance on earthquake hazard is gradual and the nonlinearity is not noticeable, it is expected that the two will generally match well.

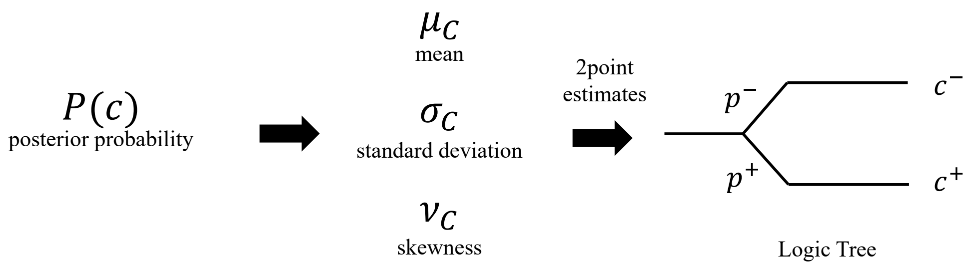
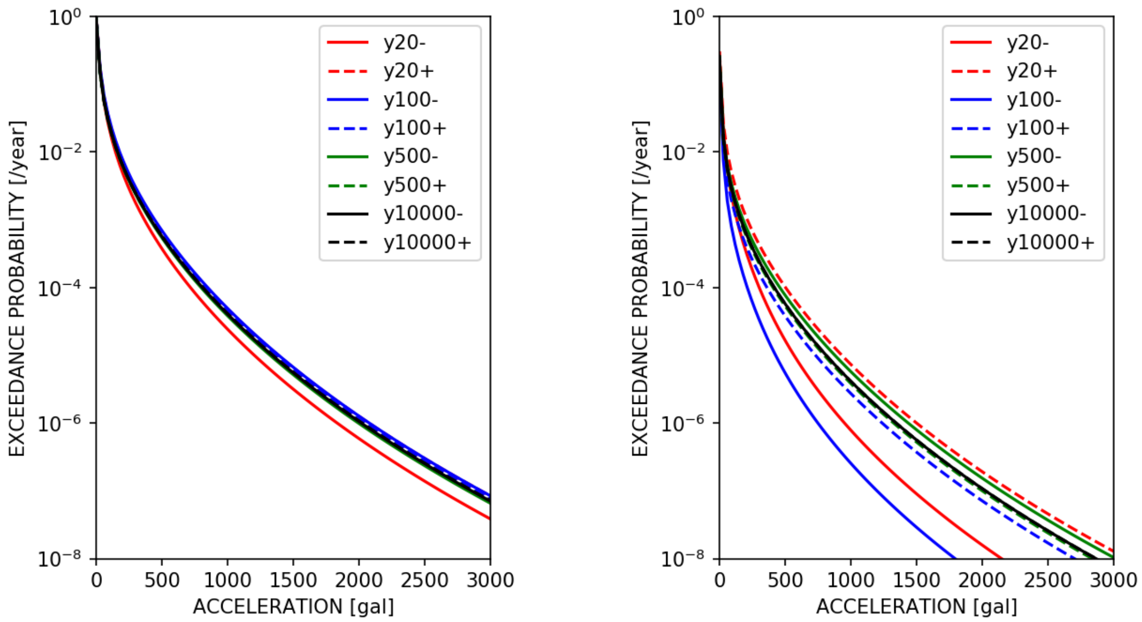
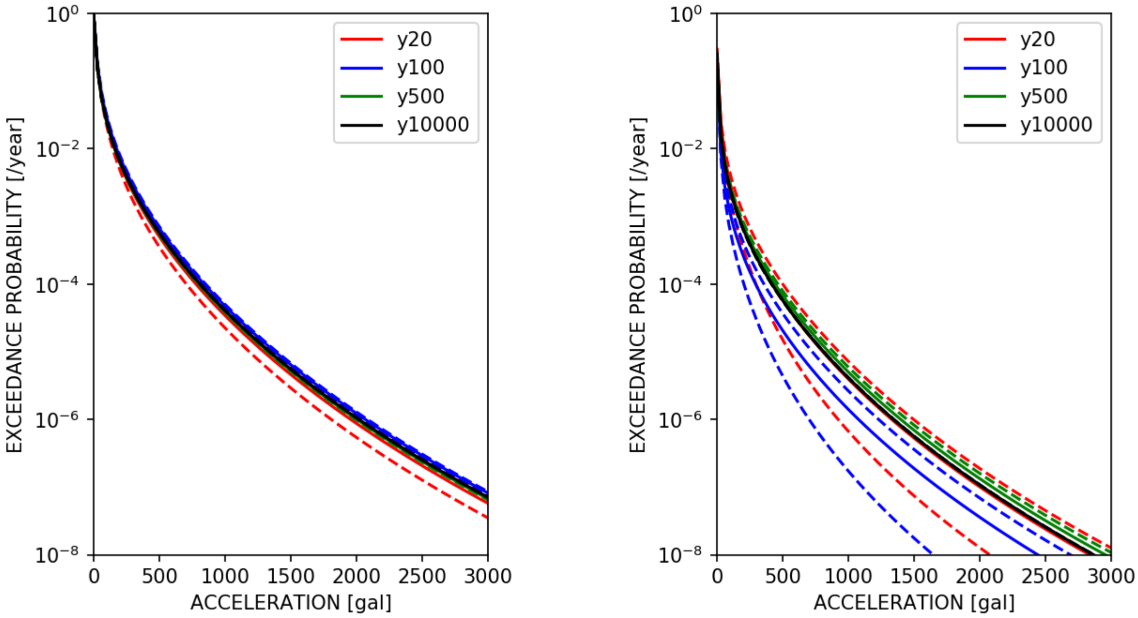


Figure 8. The flowchart of making logic tree based on the quantification of epistemic uncertainty



(1) Case A (2) Case B

Figure 9. Hazard curve calculated using sample points of the two-point estimation method



(1) Case A (2) Case B

Figure 10. The mean and variability ($\mu \pm \sigma$) of the hazard curves evaluated from the results of the two-point estimation method

4. CONCLUSION

I formulated the quantification of epistemic uncertainty in the estimation of spatial distribution of frequency and confirmed the impact of this on earthquake hazard assessment using a hypothetical earthquake environment. It was found that this uncertainty has a large impact on earthquake hazard when seismic activity is low.

I also investigated a method for constructing a logic tree using a two-point estimation method as the way to incorporate this epistemic uncertainty into actual hazard assessment. The results of the two-point estimation were roughly equivalent to those of the total integral. The proposed method can reflect epistemic uncertainty in the estimation of spatial distribution in seismic hazard analysis.

Acknowledgements

I would like to thank Professor Tatsuya ITOI for useful discussions on this study.

References

[1] T. Hagiwara: Earthquakes in the Japan Archipelago, 1991 (in Japanese)
[2] T. Kakimi, T. Matsuda, I. Aida, Y. Kinugasa: A Seismotectonic Province Map in and around the Japanese Islands, Earthquake, The seismological society of Japan, 2003
[3] A. Frankel: Mapping Seismic Hazard in the Central and Eastern United States, Seismological Research Letters, Volume 66, 1995
[4] A. Helmstetter, Yan Y. Kagen, David D. Jackson: High-resolution Time-independent Grid-based Forecast for M ≧5 Earthquakes in California, Seismological Research Letters, Volume 78, 2007
[5] S. Noda, K. Takahashi, M. Tohdo, T. Watanabe: Response Spectra for Design Purpose of Stiff Structures on Rock Sites, OECD Workshop on the Relations Between Seismological DATA and Seismic engineering, Istanbul, 2002