A Procedure Estimating and Smoothing Earthquake Rate in a Region with the Bayesian Approach

J.P. Wang
The Hong Kong University of Science and Technology, Kowloon, Hong Kong

Abstract: Reliable instrumentation earthquake data are considered limited compared to the long return period of earthquakes, especially the major events. As a result, earthquake rate estimating could become unrealistic based on the limited earthquake observation with classical statistics algorithms. For example, given no $M \geq 6.0$ earthquakes were recorded in the past 50 years, a best-estimate for the earthquake rate around the region should be zero, and such a zero estimate is considered unconvincing owing to the short observation period and long earthquake return periods. In this paper, a Bayesian calculation is proposed to earthquake rate estimating and smoothing given a reliable, but relatively short, earthquake catalog compiled with instrumentation data recorded since the last century. The key to this Bayesian application to engineering seismology is to utilize the observed rates in neighboring zones as the prior information, then updated with the likelihood function governed by the earthquake observation in a target zone.

Keywords: The Bayesian approach, earthquake rate estimating and smoothing

1. INTRODUCTION

Earthquake frequency or annual rate is an important parameter for earthquake analyses such as seismic hazard assessment. Understandably, the most reliable approach to estimate the parameter is based on sufficient instrumentation data, such as an earthquake catalog compiled with instrumentation data in the past 100 years. However, the data is considered a limited observation, compared to the long return period of earthquakes, especially the major events. As a result, based on the classical statistics algorithms, the estimates on earthquake frequency would become unrealistic because the observation data are limited. For example, given no $M \geq 6.0$ events recorded in the past 100 years, a best-estimate annual rate for such an event would be equal to zero, which is somehow unrealistic and not convincing.

Different from classical statistics algorithms relying on samples only, the Bayesian approach is to develop an estimate considering both the information from samples and from general prior knowledge. In the Bayesian terminology, the method is to update the prior probability density function (PDF) with the likelihood function from samples to develop a posterior PDF. Next, a Bayesian estimate can be developed according to the posterior PDF, a result of general prior information and site-specific observations from samples.

The Bayesian method is commonly used in a variety of studies, such as estimating dissolved oxygen in a river [1], evaluating the reliability of pile foundations [2], reconstructing a Synthetic Aperture Radar image [3], and quantifying the risk of offshore drilling [4]. In addition to those applications, the Bayesian approach was also used in earthquake engineering and engineering seismology, such as evaluating earthquake-induced slope failures [5], characterizing the structure's vulnerability against earthquakes [6], and estimating the parameters of an active fault [7]. As a result, it is understood that the Bayesian approach is in a general framework, and the data used as prior information and likelihood function can vary case-by-case.

This paper presents a new Bayesian application to engineering seismology, estimating and smoothing earthquake frequency in a region given a relatively short earthquake observation in a comparison to the long return period of earthquakes. In addition to the methodology, a case study was also given as a demonstration example to this Bayesian application to earthquake rate estimating and smoothing.

2. THE OVERVIEW OF THE BAYESIAN APPROACH

As mentioned previously, the Bayesian approach is to develop a best estimate with multiple sources of data. Take a discrete case for example (i.e., the prior probability density function is discrete), the algorithm of the Bayesian approach can be expressed as follows [8]:

$$P''(\theta_i) = \frac{P'(\theta_i) \times P(\varepsilon \mid \theta_i)}{\sum_{i=1}^{n} P'(\theta_i) \times P(\varepsilon \mid \theta_i)} \tag{1}$$

where $P'(\theta_i)$ is the prior probability for an estimate θ_i in the prior probability mass function (PMF), $P(\varepsilon \mid \theta_i)$ is the likelihood function given observation ε, $P''(\theta_i)$ is the posterior probability after updating, and *n* is the number of estimates in the prior PMF.

An example was demonstrated in the following to help describe the Bayesian approach. First, the function shown in Fig. 1 is the prior PMF for $M \geq 6.0$ earthquakes in a region, and the probabilities for annual rate *v* equal to 0.01 and 0.02 per year are 30% and 70%, respectively. Second, the observation data indicate no such events in the past 50 years were observed around the region. As a result, in this case the likelihood function for *v* = 0.01 per year in this example can be calculated as follows:

$$\begin{aligned} & P(\varepsilon = zero\ event\ in\ 50\ years \mid mean\ rate = 0.01\ per\ year) \\ & = P(\varepsilon = zero\ event\ in\ 50\ years \mid mean\ rate = 0.5\ per\ 50\ years) \\ & = \frac{e^{-0.5} 0.5^0}{0!} = 0.61 \end{aligned} \tag{2}$$

It is worth noting that the calculation shown in Eq. 2 is on a customary presumption that earthquake occurrence follows the Poisson distribution [8], with its probability mass function defined as follows [8]:

$$P(X = x \mid v) = \frac{e^{-v} v^x}{x!} \tag{3}$$

where *v* is the mean rate or the mean value of the Poissonian random variable *X*.

The same calculation can be applied to the likelihood function of the other estimate (i.e., *v* = 0.02 per year). Along with the two prior probabilities given in Fig. 1, therefore the prior probabilities and likelihood functions available, the two posterior probabilities for *v* = 0.01 and *v* = 0.02 can be updated with the Bayesian algorithm (i.e., Eq. 1) as follows:

$$P''(v = 0.01) = \frac{0.3 \times 0.61}{0.3 \times 0.61 + 0.7 \times 0.37} = 0.41 \tag{4}$$

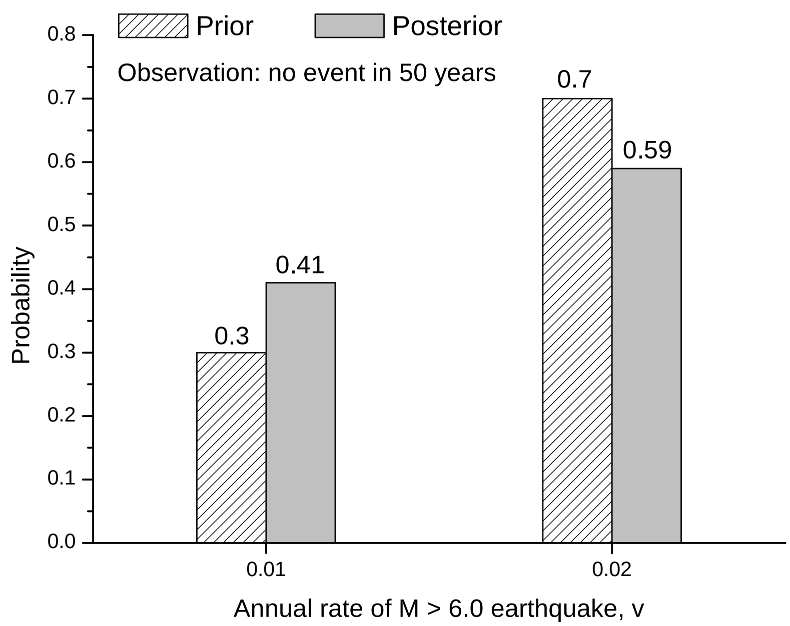
$$P''(v = 0.02) = \frac{0.7 \times 0.37}{0.3 \times 0.61 + 0.7 \times 0.37} = 0.59 \tag{5}$$

Fig. 1 shows the posterior function for this example after the Bayesian updating. Accordingly, the Bayesian estimate (i.e., the mean value of the posterior PMF) is 0.016 per year for $M \geq 6.0$ earthquakes after the Bayesian updating, given the prior information and the limited observation in the past 50 years.

3. EARTHQUAKES WITHIN SUB-REGIONS

For developing a more refined earthquake analysis, sometimes a region is further divided into a few sub-zones in an application. However, this practice would cause a challenge in the earthquake rate estimating because earthquake observation is also divided into a smaller sample size. We made an example shown in Fig. 2 with the earthquake data around Taiwan to further explain this situation: For a pre-defined seismic source in southeastern Taiwan [9], the annual rate of $M_L \geq 6.5$ events (i.e., M_L = local magnitude) could be around 0.036 per year given a total of 4 events recorded in the past 110 years. However, after dividing the zone into nine sub-zones for some application, the rate for $M_L \geq 6.5$ earthquakes in sub-zones S1, S4, S5, S6 and S9 should be zero, because no such events were observed within those zones during the time. (Note that the earthquake data around Taiwan have been used in a few earthquake studies, including seismic hazard assessments and earthquake statistics analyses [9, 10]. More details about the seismic source model and the earthquake catalog are available in those publications.)

Figure 1: The prior and posterior PMFs in a demonstration Bayesian example



4. A BAYESIAN APPLICATION TO EARTHQUAKE RATE ESTIMATING AND SMOOTHING

As mentioned previously, the scope of this technical note is to provide a Bayesian application to estimate and smooth earthquake rates given limited data. In the following sections, we continued using the example (Fig. 2) around Taiwan to help explain and demonstrate this Bayesian calculation.

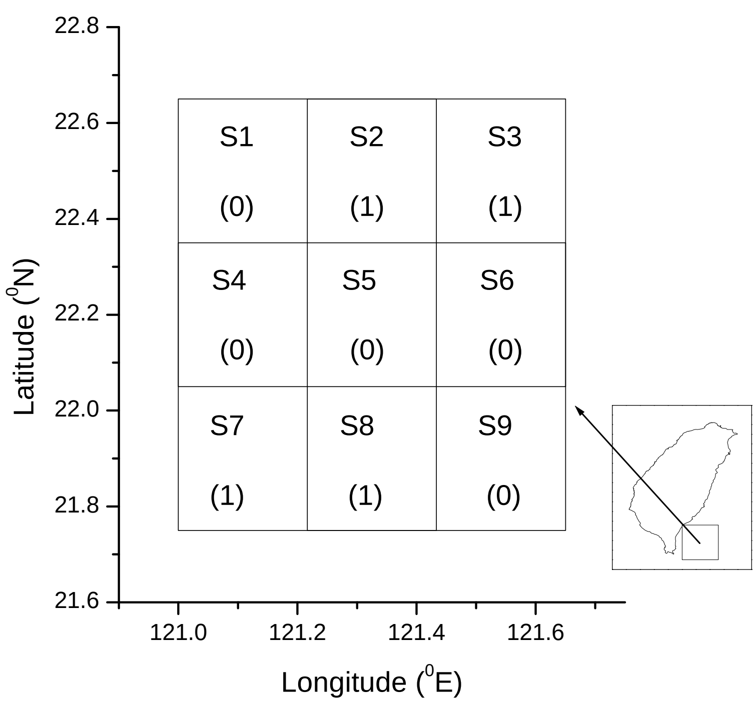
4.1. The methodology overview

Essentially, any of a Bayesian calculation is to integrate prior information with observation data to develop a Bayesian estimate. Because observation data are only available in this study, the key task in this Bayesian calculation is to develop some prior information and to integrate it with the observation. Under the circumstances, the analytical presumption in the development of the Bayesian calculation is to use the observations in neighboring areas as a source of prior information, one of the key presumptions in this study.

4.2. The definition of "neighbors"

Because the prior information is based on the observed earthquake rates in neighboring zones, the definition of "neighbors" needs clarified in the first place. In short, two areas are considered "neighbors" as long as they are in contact with each other in any form. Take Fig. 2 for example, S2, S4 and S5 are the neighbors of S1; S1, S3, S4, S5 and S6 are the neighbors of S2; and so on so forth….

Figure 2: The observed $M_L \geq 6.5$ earthquake rates within a seismic source in southeastern Taiwan based on the seismicity in the past 110 years; the value in the parenthesis is the observed earthquake rate during the time [9, 10]



4.3. The prior probability mass function

With the observed rates (denoted as v) in a target zone and in its neighboring zones, a prior PMF about earthquake rates in the target zone could be developed. Take Fig. 2 for example, the observed rates for $M_L \geq 6.5$ earthquakes within S1, S2, S4 and S5 are 0, 1, 0, and 0 in the past 110 years. Therefore, Fig. 3a shows the prior PMF for $M_L \geq 6.5$ earthquake rates in the target zone S1 based on the information from the four zones. Accordingly, the prior probabilities are 75% and 25% for the two estimates $v = 0$ and $v = 1$, respectively. In contrast, Fig. 4a shows the prior PMF for the center zone S5, with a prior probability of 56% and 44% for $v = 0$ and $v = 1$, respectively.

4.4. The updating

With the prior and observation data available, then the Bayesian estimate can be developed. In this case study demonstration, the target zone S1 has a prior PMF like Fig. 3a, and a zero-event observation for $M_L \geq 6.5$ earthquakes in the past 110 years. As a result, considering earthquake occurrence follows a Poisson distribution, the likelihood function for the estimate v = 0 can be calculated as follows:

$$P(\varepsilon = zero\ event\ observation \mid v = 0) = \frac{e^{-0}0^0}{0!} = 1 \tag{6}$$

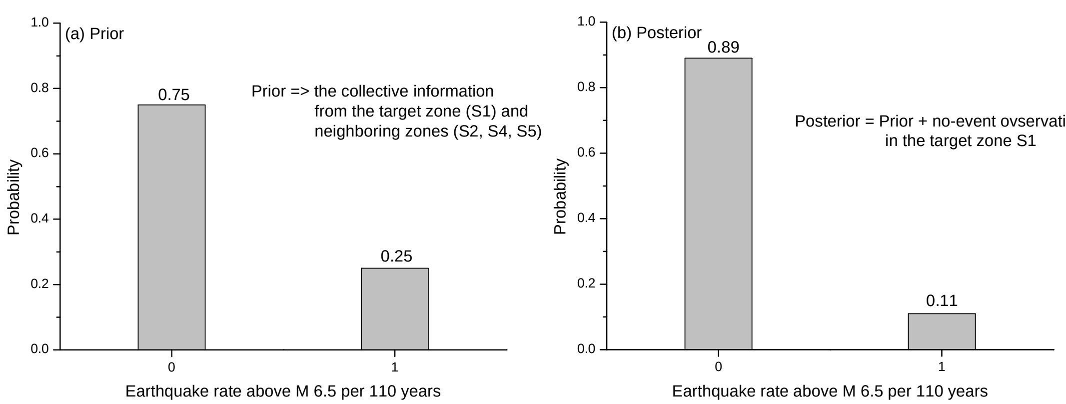
The same calculation was then applied to the other estimate v = 1, and with the two prior probabilities and likelihood functions available, the posterior probability for v = 0 can be updated with the underlying Bayesian algorithm (i.e., Eq. 1) as follows:

$$P''(v=0) = \frac{P'(v=0) \times P(\varepsilon \mid v=0)}{P'(v=0) \times P(\varepsilon \mid v=0) + P'(v=1) \times P(\varepsilon \mid v=1)} = \frac{0.75 \times 1}{0.75 \times 1 + 0.25 \times 0.37} = 0.89 \tag{7}$$

where ε here denotes the observation that is no $M_L \geq 6.5$ earthquakes in S1 during the time.

Repeating the calculations for the other estimate v = 1, its posterior probabilities can be calculated as well. As a result, Fig. 3b shows the posterior PMF for $M_L \geq 6.5$ earthquakes within S1, and the Bayesian estimate on the earthquake rate is equal to 0.09 per 110 years, or 0.0008 per year, according to the posterior PMF. On the other hand, Fig. 4b shows the posterior PMF for the center zone S5. According to the posterior PMF, the Bayesian estimate on the rate of $M_L \geq 6.5$ earthquakes within S5 is 0.18 per 110 years, equivalent to 0.002 per year.

Figure 3: a) the prior probability mass function for two $M_L \geq 6.5$ earthquake rates for zone S1 (see Fig. 2), and b) the posterior function after the Bayesian updating with no $M_L \geq 6.5$ observed in the target zone S1 in the past 110 years



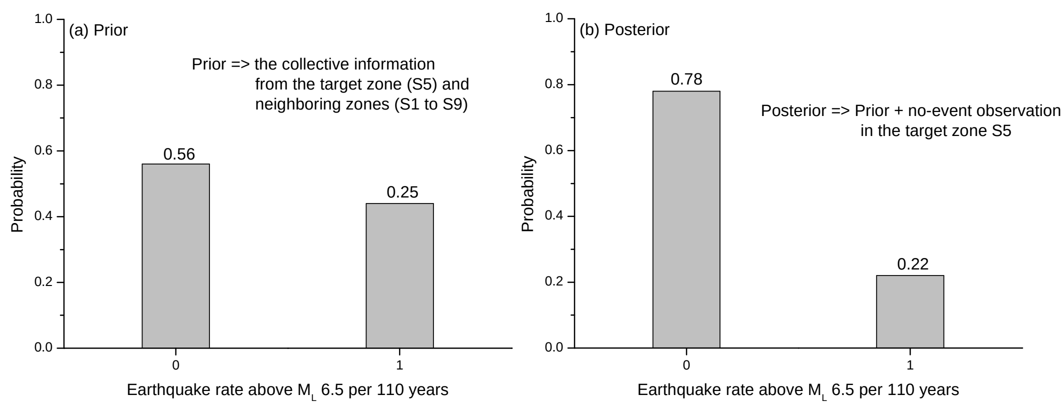
In the two Bayesian calculations, we can see although the observed earthquake rates in S1 and S5 are both equal to zero, the updated Bayesian estimates are varied because the different prior information, or different "neighbors" of the two target areas.

4.4. The earthquake rate after the Bayesian smoothing

With the updating repeated for the rest of the zones, the smoothing of the earthquake rate within a region could be achieved. For this example using the earthquake data around Taiwan, Fig. 5 shows the smoothed rates for each of the nine sub-zones with the Bayesian approach. It is worth noting that the total of the smoothed rates is equal to the observed rate, four events in total, based on the seismicity detected in the past 110 years.

From this case study, we can see that now the estimates on earthquake rates become more realistic with the Bayesian calculation, rather than a zero-event estimate with the classical statistical algorithms based on limited observations available. Therefore, the new Bayesian application to earthquake rate estimating and smoothing could be a useful option for earthquake studies, given the reliable earthquake instrumentation data are limited compared to the long return period of earthquakes, especially the major events.

Figure 4: a) the prior probability mass function for two $M_L \geq 6.5$ earthquake rates for S5, and b) the posterior function after the Bayesian updating with no $M_L \geq 6.5$ events observed during the time



5. DISCUSSIONS

5.1. Analytical presumption behind the Bayesian calculation

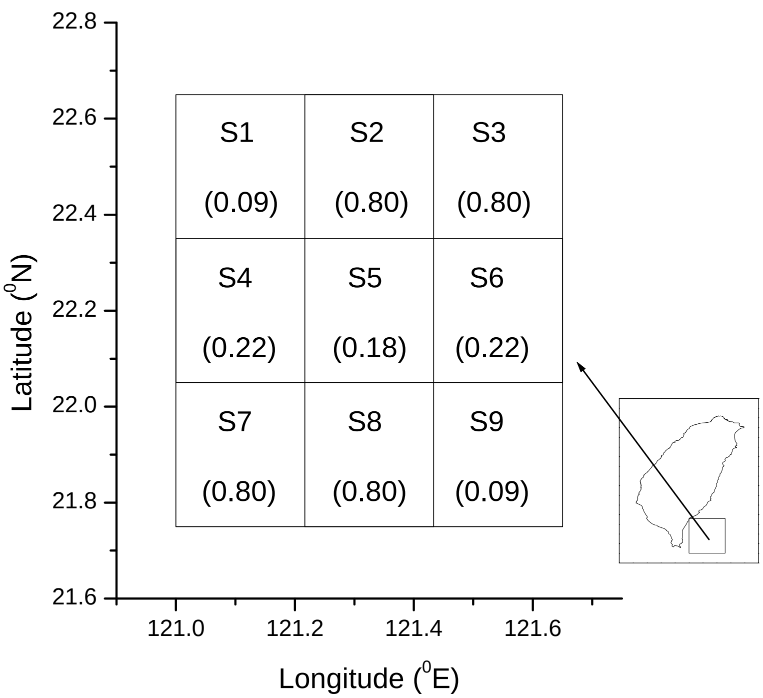
An implicit analytical presumption behind the development of this Bayesian smoothing is that we weighted the observations differently for a given target zone. That is, we considered the observed rate in the target zone is more "reliable," and it should play a heavier role in the calculation than those observed rates from neighboring zones. As a result, we utilized the observed rate of the target zone in both the prior PMF and the likelihood function also governed by that observation. In other words, the observed rate of the target zone was "double counted" in this Bayesian calculation for smoothing the earthquake rate in a region.

5.2. On the proper use of the Bayesian approach

Explicitly, observation data are more reliable than prior information from judgment, experience, etc. In other words, when observation data are in a large sample size, it is not necessary to take general prior information into account for developing a best estimate, also pointed out in other Bayesian studies in geotechnical site characterizations [11, 12]. To sum up, the purpose of using the Bayesian approach was to utilize some prior information to compensate the sample-size issue when encountered. In other words, when the site was well investigated with sufficient site-specific data, the classical algorithms relying on reliable observation data should be able to develop a representative estimate, without the involvement of general prior information.

The same basics on the proper use of the Bayesian approach should be applied to this Bayesian calculation for earthquake rate estimating and smoothing. For estimating large-earthquake rates like the demonstration example, the 110-year-long instrumentation earthquake data could be too limited to estimate the rates with classical statistics algorithms, considering the long return period of the large earthquakes. By contrast, if the problem is to estimate small-earthquake rates such as $M_L \geq 3.0$ around Taiwan, the 110-year-long observation with 50,000 $M_L \geq 3.0$ events should be sufficient to develop a reliable estimate using the classical algorithms.

Figure 5: The estimates on $M_L \geq 6.5$ earthquake rates for each of the nine sub-zones within a seismic source in southeastern Taiwan with the Bayesian calculation, based on the prior information from the neighboring zones and the likelihood function governed by the earthquake observation in the target zone



6. SUMMARY

Owing to the long return period of earthquakes, one challenge in earthquake parameter estimating is the lack of a representative sample size. As the demonstration examples using earthquake data around Taiwan, the limited observation could lead to an unrealistic "zero" estimate for earthquake occurrences, because of the relatively short period of observation compared to the long return period of earthquakes, especially the major events.

Different from classical statistics algorithms, the Bayesian approach is to utilize both prior information and observation data to develop a best estimate, and it has been practiced in many studies especially when observation data are limited. Like those Bayesian applications, this paper introduces a new Bayesian calculation for earthquake rate estimating and smoothing, given reliable, but limited, instrumentation earthquake data available.

The key to this Bayesian calculation includes the use of the observed earthquake rates in neighboring zones as the prior information, and "double counting" the earthquake observation in a target zone during the Bayesian updating. In terms of the result, the Bayesian calculation can develop a more realistic estimate on earthquake rates rather than zero, with limited instrumentation earthquake data.

Acknowledgements

The author is very thankful for the provision of the earthquake data from Prof. Yih-Min Wu of National Taiwan University.

References

[1] A. Patil and Z. Q. Deng. "*Bayesian approach to estimating margin of safety for total maximum daily load*," Journal of Environmental Management, 92, pp. 910-918, (2011).
[2] L. M. Zhang, D. Q. Li and W. H. Tang. "*Impact of routine quality assurance on reliability of bored piles*", Journal of Geotechnical and Geoenvironmental Engineering, 132, pp. 622-630, (2006).
[3] D. Vu, M. Xue, X. Tan and J. Li. "*A Bayesian approach to SAR imaging*," Digital Signal Processing, 23, pp. 852-858, (2013).
[4] N. Khakzad, F. Khan and P. Amyotte. "*Quantitative risk analysis of offshore drilling operations: A Bayesian approach*," Safety Science, 57, pp. 108-117, (2013).
[5] Y. Song, J. Gong, S. Gao, D. Wang, T. Cui and B. Wei. "*Susceptibility assessment of earthquake-induced landslides using Bayesian network: A case study in Beichuan, China*," Computers & Geosciences, 42, pp. 189-199, (2012).
[6] P. S. Koutsourelakis. "*Assessing structural vulnerability against earthquakes using multi-dimensional fragility surfaces: A Bayesian framework*," Probabilistic Engineering Mechanics, 25, pp. 49-60, (2010).
[7] M. Amighpey, B. Voosoghi and M. Motagh. "Deformation and fault parameters of the 2005 Qeshe earthquake in Iran revisited: A Bayesian simulated annealing approach applied to the inversion of space geodetic data", International Journal of Applied Earth Observation and Geoinformation, 26, pp. 184-192, (2014).
[8] A. H. S. Ang and W. H. Tang. "*Probability concepts in engineering; emphasis on applications to civil and environmental engineering*," John Wiley & Sons, 2007, New Jersey.
[9] C. T. Cheng, S. J. Chiou, C. T. Lee and Y. B. Tsai. "*Study on probabilistic seismic hazard maps of Taiwan after Chi-Chi earthquake*," Journal of GeoEngineering, 2, pp. 19-28, (2007).
[10] J. P. Wang, D. Huang, S. C. Chang and Y. M. Wu. "*New evidence and perspective to the Poisson process and earthquake temporal distribution from 55,000 events around Taiwan since 1900*," Natural Hazards Review ASCE, 15, pp. 38-47, (2014).
[11] Y. Wang, S. K. Au and Z. Cao. "*Bayesian approach for probabilistic characterization of sand friction angles*," Engineering Geology, 114, pp. 354-363, (2010).
[12] J. P. Wang and Y. Xu. "*Estimating the standard deviation of soil properties with limited samples through the Bayesian approach*," Bulletin of Engineering Geology and the Environment, doi: 10.1007/s10064-014-0619-5 in press (2014).