Statistical Analysis of Seismic Motion Based on Hierarchical Bayesian Models

T. Hamada(1), T. Itoi(2), N. Sekimura(3)

(1) *Graduate student, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan, hamada@safety.n.t.u-tokyo.ac.jp*
(2) *Associate Professor, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan, itoi@n.t.u-tokyo.ac.jp*
(3) *Professor, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan, sekimura@n.t.u-tokyo.ac.jp*

Abstract

An attenuation relationship, i.e., a ground motion prediction equation, is a statistical equation, conventionally used to predict ground motion intensity measure in probabilistic seismic hazard analysis. Ground motion prediction consists of analysis of the source, propagation path and site amplification characteristics. The site amplification characteristics are considered to be important factors affecting precision of ground motion prediction, therefore a site-specific ground motion prediction is desirable for probabilistic seismic hazard analysis. A conventional attenuation relationship, however, is not a site-specific but a generic equation to predict ground motion intensity that is because of the assumption of ergodicity, because the number of records that are observed at specific site is limited. A Bayesian approach is considered efficient for constructing a site-specific attenuation relationship. This study proposes a methodology to develop a site-specific attenuation relationship by employing a hierarchical Bayesian model. Then, a site-specific attenuation relation for JMA (Japan Meteorological Agency) seismic intensity scale is developed for crustal earthquakes observed in Japan from 1997 to 2011. The developed attenuation relationship by the hierarchical Bayesian method is considered less biased and explicitly accounts for all the prevailing uncertainties that is because of the less assumption of ergodicity. The attenuation relationship by the least square method was considered to bias the mean value of the predicted JMA seismic intensity up to about one, and to overestimate the standard deviation of aleatory uncertainty around four-thirds.

Keywords: PSHA, Attenuation relation, Ergodic assumption, Ground motion, Hierarchical Bayesian model

1. Introduction

In order to assess seismic risk to a structure, uncertainties in future earthquake ground motions are quantified using probabilistic seismic hazard analysis (PSHA). Uncertainties in future earthquake ground motions arise from uncertainties in earthquake occurrence as well as those in predicted ground motions. In predicting ground motions, an empirical attenuation relationship (e.g., [1],[2]) is conventionally used that is constructed by a statistical regression, e.g., by the least square method or the maximum likelihood estimation, performed on observed ground motion records. Recently, ground motion simulation based on fault rupture modelling is frequently used in stead of an attenuation relationship. An attenuation relationship is also used to validate the ground motion simulation model even in such cases. Ground motion prediction consists of analysis of the source, propagation path, and site amplification characteristics. The site amplification characteristics are considered to be important factors affecting precision of ground motion prediction, therefore a site-specific ground motion prediction is desirable for probabilistic seismic hazard analysis. A conventional attenuation relationship, however, is not a site-specific but a generic equation to predict ground motion intensity that is because of the ergodic assumption where the time-variant characteristics at a specific site is assumed to be identical to the ensemble characteristics, i.e., the characteristics of spatial variation [3]. A Bayesian approach is considered to be an efficient way to construct a site-specific attenuation relationship with less ergodic assumption [4][5]. Therefore, this study proposes a methodology for constructing a site-specific attenuation relationship by employing a hierarchical Bayesian model. Then, a site-specific attenuation relation for JMA (Japan Meteorological Agency) seismic intensity scale is developed for crustal earthquakes observed in Japan from 1997 to 2011[6]. An attenuation relation is also developed by the least square method to discuss the advantage of a proposed method.

2. Compilation of ground motion records

Ground motion records during crustal earthquakes in Japan that were observed from 1997 to 2011 are compiled from K-NET, National Research Institute for Earth Science and Disaster Prevention. In this study, near-field earthquake ground motions are of concern. 1703 records from 44 earthquakes are collected that were observed at 571 sites in total where the shortest distance from fault to site, i.e., the shortest fault distance is less than 100 km [7]. Moment magnitude of compiled earthquakes are from 5.1 to 6.9. Fig. 1 shows the characteristics of compiled ground motions with respect to the relationship between the fault distance and the moment magnitude.

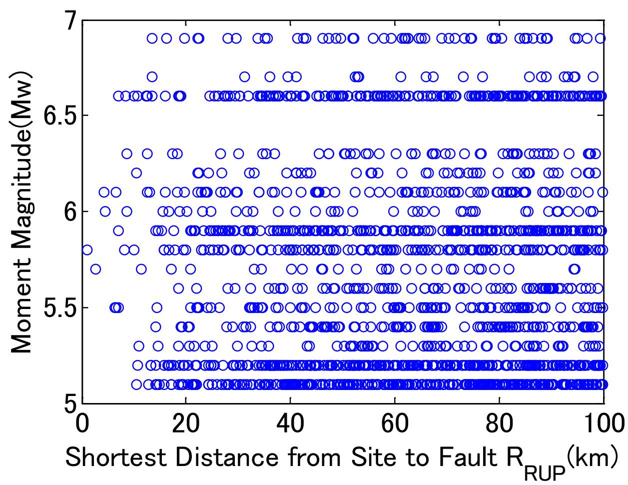


Fig.1 Shortest distance and moment magnitude of compiles ground motion records [7]

3. Ground motion prediction model

In this study, JMA seismic intensity is assumed to be predicted as a function of moment magnitude (source term), the shortest fault distance (propagation path term) and a site amplification term as follows:

$$I = aM_W - 2\log_{10}(X + b \times 10^{0.5M_W}) - cX + d + f_s + \sigma\varepsilon \tag{1}$$

where, I, M_W and X are JMA seismic intensity, moment magnitude, and site-to-fault distance (km) respectively. ε and σ are a standard normal random variable and standard deviation of aleatory uncertainty in predicted ground motion respectively. a, b, c and d are coefficients, and f_s is the site amplification term that is typically a function of 30m average shear-wave velocity in a conventional attenuation equation that is described in the next chapter.

4. Attenuation relationship based on least square method

The least square method is one of standard approaches to obtain a statistical relationship between variables. In an attenuation model based on the least square method, the site amplification term f_s in Equation (1) is assumed to be a function of $\log_{10}V_{S30}$ as follows:

$$I = aM_W - 2\log_{10}(X + b \cdot 10^{0.5M_W}) - cX + d + e \cdot \log_{10}V_{S30} + \sigma\varepsilon \tag{2}$$

where, e is a coefficient. V_{S30} (m/s) is a 30m average shear-wave velocity, i.e., time averaged Vs in top 30m, that is conventionally considered to be a reasonable explanatory variable to describe the site amplification characteristics due to the shallow soil characteristics. In the process of regression, in order to avoid the mutual influence, i.e., trade-off, between the coefficients in the model, the two-step stratified regression method [2] is employed. The coefficient e for V_{S30} is obtained by analyzing the residuals by another regression. An attenuation relationship obtained is as follows:

$$I = 1.36M_W - 2\log_{10}(X + 0.00550 \cdot 10^{0.5M_W}) - 0.00670X - 1.63\log_{10}V_{S30} + 3.30 + \sigma\varepsilon \quad (3)$$

where, standard deviation σ is estimated to be 0.608 as a standard deviation of the residuals, i.e., difference between observed and predicted. Fig.2 shows the comparison between an attenuation relation ($M_W = 6.0$, $V_{S30} = 400\mathrm{m/s}$) and observed JMA seismic intensities. Observed JMA seismic intensities in Fig.2 are converted to those for the same condition ($M_W = 6.0,\ V_{S30} = 400\mathrm{m/s}$) using Equation (3).

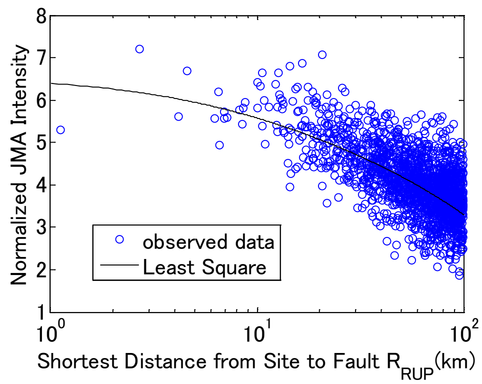


Fig.2 Comparison between the observed data and attenuation relationship obtained by the least square method ($M_W = 6.0,\ V_{S30} = 400\mathrm{m/s}$)

5. Attenuation relationship based on hierarchical Bayesian regression model

5.1. Overview of hierarchical Bayesian method

In Bayesian updating, the posterior distribution is obtained by multiplying the prior distribution and the likelihood function as follows:

$$f(y|x) \propto L(x|y)\pi(\mathrm{y}) \quad (4)$$

where, $f(y|x)$, $L(x|y)$ and $\pi(\mathrm{y})$ are the posterior distribution, the likelihood function of observed data x and the prior distribution respectively. The prior and posterior distrubution, respectively, quantify the degree-of-belief, i.e., the uncertainties in each parameter, before and after new data are observed.

In a Bayesian updating, the prior distribution for each unknown parameter is determined based on generic datasets, expert opinions, etc. A Bayesian hierarchical modeling, on the other hand, is a statistical model consisting of multiple levels, i.e., a hierarchical form. In a hierarchical Bayesian model a parameter, e.g. variance, of the prior distribution that is called a hyperparameter is assumed to be also uncertain and to follows a probability distribution. The prior distribution for a hyperparameter is assumed to be a non-informative prior distribution. By using a hierarchical Bayesian model, the subjective setting for the prior distribution can be avoided.

5.2. Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) is one of the methods for sampling from a certain probability distribution that is conveniently used for the hierarchical Bayesian models [8],[9]. The Gibbs sampler is one of the widely used algorithms for simulating Markov chains that generates a multi-dimensional Markov chain conditional on the most recent values of all other components. The algorithms proceed as follows:

Let the vector of unknowns $\boldsymbol{\theta}(=[\theta_1\ \theta_2\ \cdots\ \theta_n]')$ consists of n components, where subscripts denote components of θ.

1. Assume arbitrary starting values $\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_n^{(0)}$ for each component and superscripts denote the number of iterations.
2. Sample new realizations for each element of θ by cycling through the following steps:
   - Sample a new realization for θ_1, i.e., the first component of θ, from the conditional distribution of θ_1 given the most recent values of all other elements of θ and the data $\boldsymbol{x}$:

$$\theta_1^{(i+1)} \sim f(\theta_1 | \theta_2^{(i)}, \theta_3^{(i)}, \ldots, \theta_n^{(i)}, \boldsymbol{x})$$

   - Sample a new realization $\theta_2^{(i+1)}$, i.e., the second component of θ, from its conditional distribution $\theta_2^{(i+1)} \sim f(\theta_2 | \theta_1^{(i+1)}, \theta_2^{(i)}, \theta_3^{(i)}, \ldots, \theta_n^{(i)}, \boldsymbol{x})$
   - Sample a new realization $\theta_k^{(i+1)}$ for the *k*-th component of θ up to the last component $\theta_n^{(k+1)}$
3. Repeat "Step 2" for m times, conditioning on the most recent realization of other parameters, to obtain a sequence of dependent realizations of the vector of unknowns $\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \ldots, \theta^{(m)}$

Fig.3 shows the schematic illustration of a strucuture of the hierarchical Bayesian model for parameters of an attenuation relationship. In a proposed model, it is assumed that a site amplification term f_S in Equation (1) is site-specific and the zero-mean normal distribution is assumed for the prior distribution where the noninformative inverse gamma distribution (mean 1, variance 10^5) is assumed for its variance. The coefficients a, b, c and d in Equation (1) are also assumed to be the zero-mean normal distribution (variance 10^6) as a non-informative distribution that are assumed not to be different between sites. The variance σ^2 in Equation (1) is assumed to be uniform among records and to be the inverse gamma distribution (mean 1, variance 10^6) for the prior distribution. MCMC is used to obtain the posterior distribution. The procedure uses the Gibbs sampler to generate random numbers as explained above.

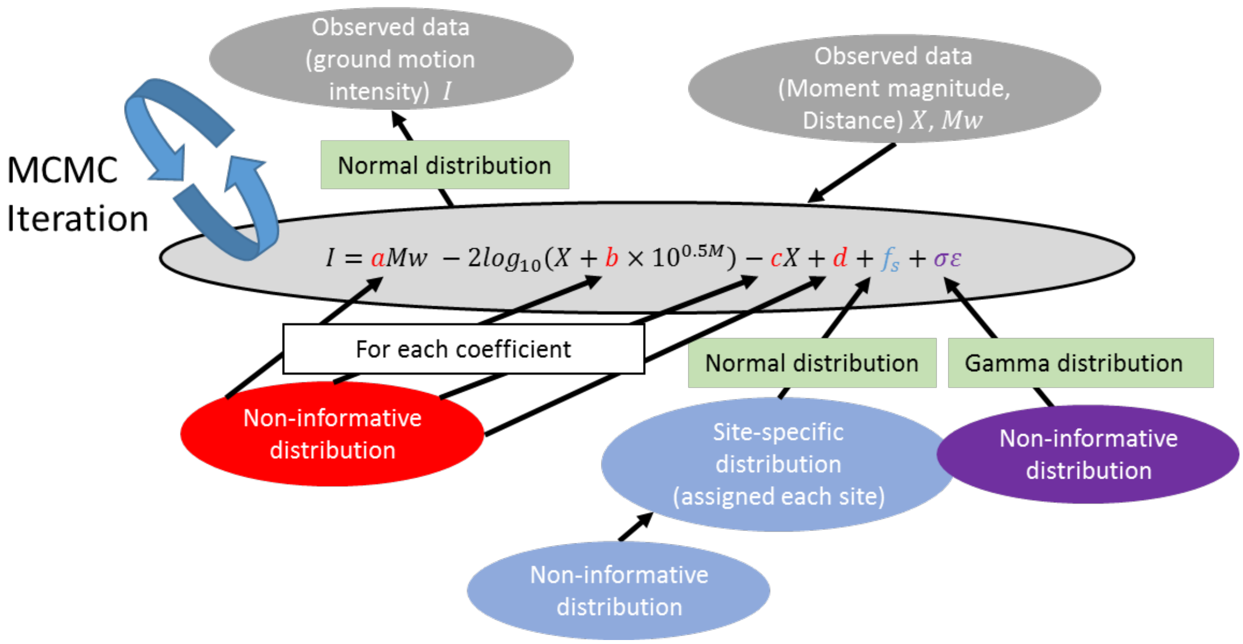


Fig.3 Hierarchical Bayesian model for site-specific attenuation relation

MCMC sampling is performed on using a open source program called JAGS [9] operated on a statistical analysis tool “R”. The number of iterations is 55000, while the first 5000 samples are discarded (burn-in) and the thinning interval is 100 respectively.

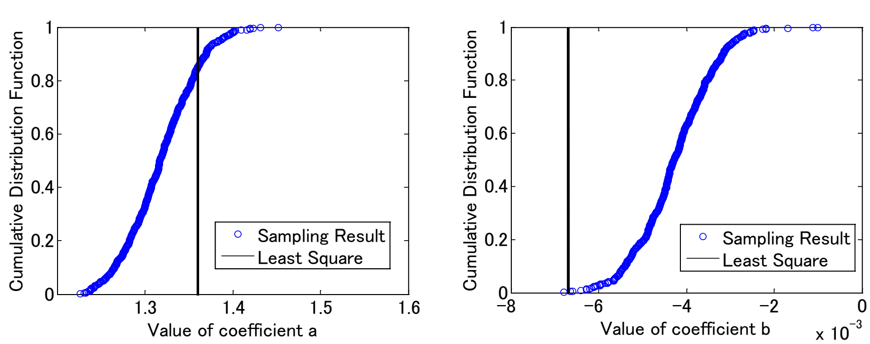
5.3. Results

The attenuation relationship obtained by the hierarchical Bayesian method is as follows:

$$\begin{aligned} I = 1.32(\pm 0.0390)M_W - 2\log_{10}(X + 0.00499(\pm 0.00201) \times 10^{0.5M_W}) \\ - 0.00430(\pm 0.00201)X + f_s - 0.806(\pm 0.187) \end{aligned} \tag{5}$$

where, $\pm$ in the equation indicates the standard deviation of the posterior distribution of each coefficient.

Fig.4. shows the cumulative distribution for the posterior distribution of each coefficient in Equation (1). The vertical line shows the value obtained by the least square method for comparison. The difference between the two methods is that a hierarchical Bayesian method can evaluate the posterior distribution, i.e., statistical uncertainty. The results from the two different methods are harmonic with each other.



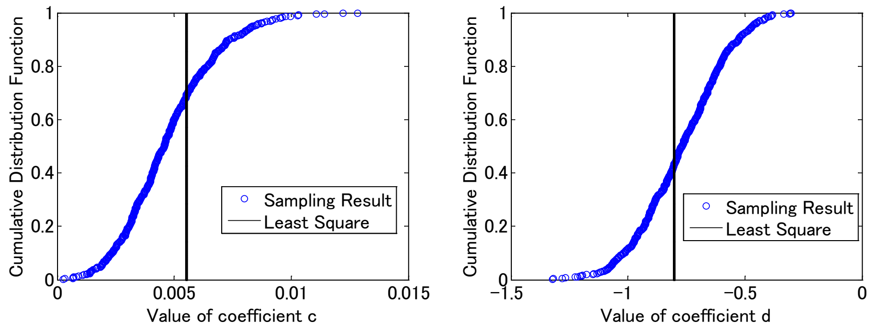


Fig.4. Comparison between each coefficient in Equation (1) obtained by the hierarchical Bayesian model and the least square method

Fig.6. shows the comparison of σ estimated both by the least-squares method and by the hierarchical Bayesian method whose cumulative density distribution is obtained by using the MCMC sampling. The standard deviation of an attenuation relation by the least-square method is 0.608 as obtained in Chapter 4, the mean value of σ from the hierarchical Bayesian method, on the other hand, is estimated 0.465 that is three-quarters of that from the least square method. The reason that an attenuation relationship obtained by the least square method overestimate the σ value four-thirds times is considered because it is developed based on the ergodicity assumption.

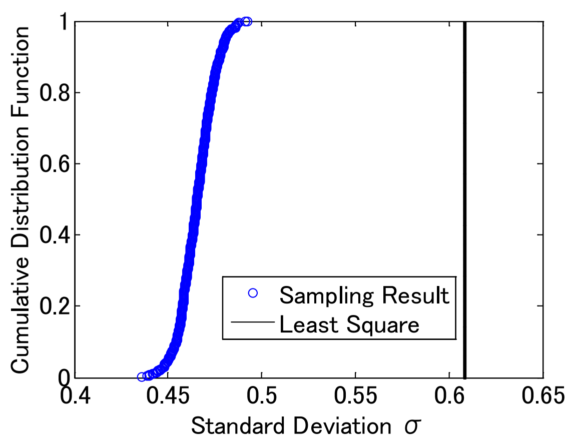


Fig.6. Comparison of standard deviation σ in Equation (1) obtained by the hierarchical Bayesian model and the least square method

Fig.7 shows the histogram of the mean value of the site amplification term for 571 sites obtained by the hierarchical Bayesian model. Fig.8 shows the site amplification term f_S for each observation station. The line in the figure shows the average relationship as a function of the 30m average shear-wave velocity V_{S30} obtained by the least square method. The line is harmonic with the result of the least square method, i.e., Equation (3), in term that f_S decreases as V_{S30} increases. The maximum difference between site amplification term and the average relationship is almost one. As described before, the standard deviation of aleatory uncertainty of the hierarchical Bayesian method reduces about three-quarters. This is because site-specific site amplification term f_S is employed replacing the 30m average shear-wave velocity V_{S30}.

Fig.9. shows the comparison between a attenuation relation ($M_W = 6.0,$ average site condition) and observed JMA seismic intensities. Observed JMA seismic intensities in Fig.2 are converted to those for the same condition ($M_W = 6.0,\ V_{S30} = 400\mathrm{m/s}$) using Equation (3). By employing a site-specific term for the site amplification characteristics, the scatter decreases compared with Fig. 2.

Fig.10 shows examples of comparison between observed JMA seismic intensity, predicted JMA seismic intensity by the two attenuation relationships developed in this study. As shown in the figures, the predicted intensity by the hierarchical Bayesian method agrees with the observed better than that by the least square method.

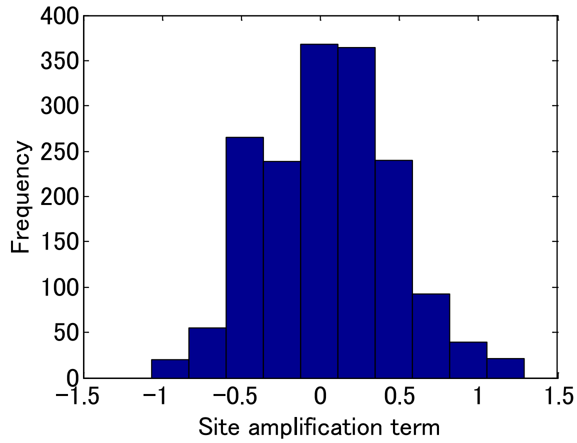


Fig.7. Frequency of site amplification term f_S

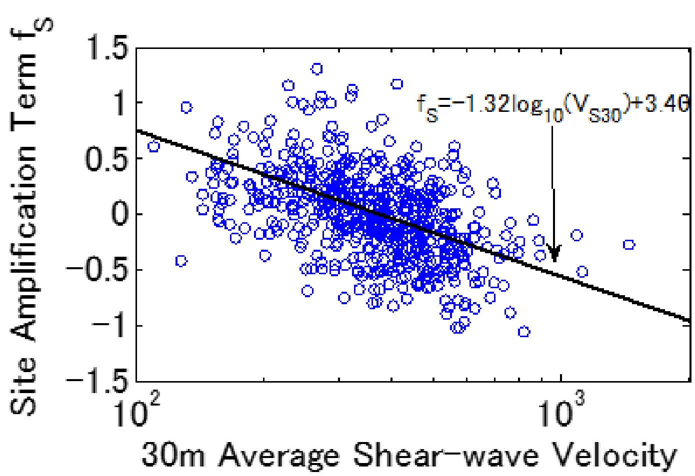


Fig.8. Relation between site amplification term f_S and V_{S30}

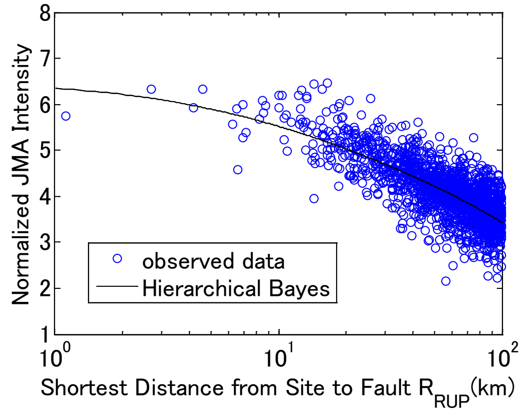
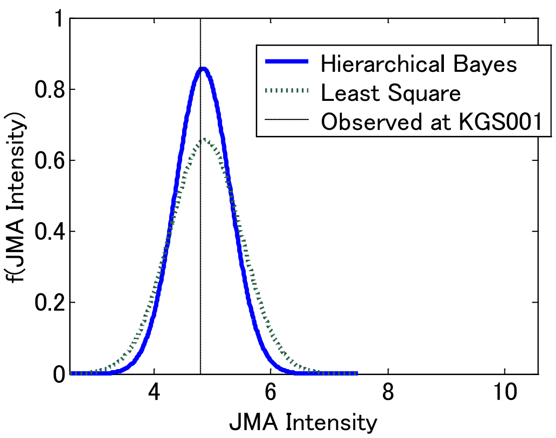
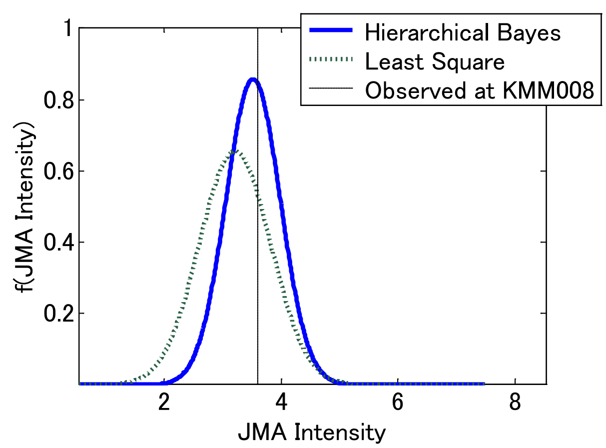


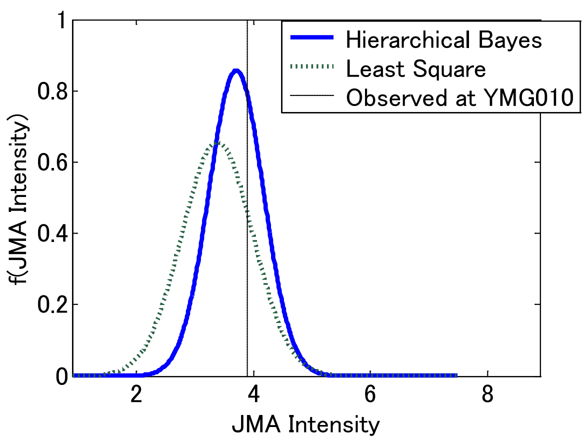
Fig.9. Comparison between the observed data and regression curve obtained by the hierarchical Bayesian method ($M_W = 6.0$, average site condition)



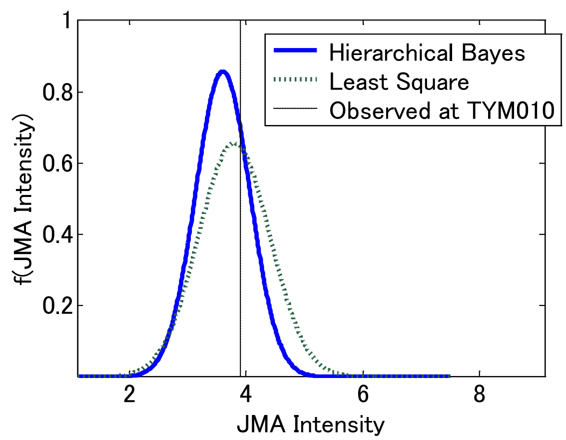
(a) KGS001, March 26, 1999 (M_W6.1)



(b) KMM008, May 13, 1997 (M_W6.0)



(c) YMG010, June 25, 1997 (M_W5.8)



(d) TYM010, March 25, 2007 (M_W6.7)

Fig.10. Examples of comparison between the predicted and observed JMA seismic intensity

6. Conclusions

In this study, by employing the hierarchical Bayesian method that is one of Bayesian updating methods, a site-specific attenuation relationship is developed to predict JMA seismic intensity for crustal earthquake in Japan. By comparing with conventional methods, i.e., the least square method, the influence of the ergodic assumption was discussed. The developed attenuation relationship by the hierarchical Bayesian method is considered less biased and explicitly accounts for all the prevailing uncertainties that is because of the less assumption of ergodicity. The attenuation relationship by the least square method was considered to bias the mean value of the predicted JMA seismic intensity up to about one, and to overestimate the standard deviation of aleatory uncertainty around four-thirds. As a future work, comparison with other approaches such as a reference site approach needs to be conducted to discuss the advantage of the proposed approach.

7. Acknowledgements

Authors thank K-NET and F-net (National Research Institute for Earth Science and Disaster Prevention) for providing data used in this study.

8. References

[1]. Si, H. & Mirodikawa, S. : New Attenuation Relationship for Peak Ground Acceleration and Velocity Considering Effects of Fault Type and Site Condition, J. Struct. Constr. Eng. 523, 63-70, 1999.(in Japanese with English abstract)

[2]. Fukushima, Y., & Tanaka, T.: A new attenuation relation for peak horizontal acceleration of strong earthquake ground motion in Japan, Bulletin of the Seismological Society of America, 80(4), 1990.

[3]. Anderson, J.G. & Brune, J.N.: Probabilistic Seismic Hazard Analysis without the Ergodic Assumption, Seismological Research Letters, 70(1), 19-28, 1991.

[4]. Wang, M & Takada, T.: A Bayesian Framework for Prediction of Seismic Ground Motion, Bulletin of the Seismological Society of America, 99(4), 2348-2364, 2009.

[5]. Kuehn, N. M & Scherbaum, F. : A Partially Non-Ergodic Ground Motion Prediction Equation for Europe, Bulletin of Earthquake Engineering, DOI 10.1007/s10518-016-9911-x, 2016.

[6]. Japan Meteorological Agency (JMA). On Seismic Intensity. Gyosei: Tokyo, 1996; 46–224 (in Japanese).

[7]. Itoi, T. et al. : Statistical Equations of Response Spectra of Crustal Earthquake for Assessment of Multiple Facilities Seismic Risk, Journal of Japan Association for Earthquake Engineering, Vol 15 (6), 6_126-6_141, 2015. (in Japanese with English abstract)

[8]. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin.: Bayesian Data Analysis, 2nd edition. Chapman and Hall/CRC, 2003

[9]. Lunn, D. et al.: The BUGS Book: A Practical Introduction to Bayesian Analysis, 2012.

[10]. Plummer, M: JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, Proceedings of the third international workshop on 'Distributed Statistical Computing' (DSC 2003), 2003.