Challenges in Model Identification for Seismic Performance Evaluation of Building Accounting for Uncertainties

Sangwon Lee
PhD student, Dept. of Architecture, The University of Tokyo, Tokyo, Japan

Yuma Matsumoto
PhD student, Dept. of Architecture, The University of Tokyo, JSPS Research Fellow, Tokyo, Japan

Taro Yaoyama
Project Assist. Prof., Dept. of Architecture, The University of Tokyo, Tokyo, Japan

Takenori Hida
Assoc. Prof., Major in Urban and Civil Engineering, Ibaraki University, Ibaraki, Japan

Tatsuya Itoi
Assoc. Prof., Dept. of Architecture, The University of Tokyo, Tokyo, Japan

ABSTRACT: While there have been various attempts to account for uncertainties in evaluating the performance of a building, there have been insufficient studies to assess the uncertainties associated with model identification of buildings. This paper discussed the challenges of taking into consideration the uncertainties in model identification of existing buildings and proposed a novel method for assessing uncertainties in observation data using Bayesian statistics.

1. INTRODUCTION

In evaluating the seismic performance of a building, it is important to properly treat various types of uncertainties involved in the evaluation process. In the field of performance-based earthquake engineering, several attempts have been made to assess the uncertainties involved at each stage. For example, recently published FEMA P-58 (Applied Technology Council (2018)) proposed a framework for probabilistically treating *Intensity Measures (IM)*, *Engineering Demand Parameter (EDP)*, *Damage Measure (DM)*, and *Decision Variable (DV)*, based on the performance-based design method (Moehle and Deierlein (2004)) proposed by The Pacific Earthquake Engineering Research Center. In FEMA P-58, the annual frequency of exceedance of the *DV* is obtained by the triple integral of the conditional probability shown in equation (1).

$$\nu(DV) = \int\int\int p(DV|DM)\, p(DM|EDP)\, p(EDP|IM)\, \nu(IM)\; dDM\, dEDP\, dIM \tag{1}$$

In general, $p(EDP|IM)$ is calculated by Monte Carlo simulation using a response analysis model. While *IM*, *EDP*, *DM*, and *DV* are all treated as random variables considering uncertainties, it is common to use a deterministic model as a response analysis model to simulate the dynamic behavior of a target building.

The process of constructing a numerical model of a building always involves errors, referred to as modeling errors. There are many sources of modeling errors, such as variations introduced during the construction process or unmodeled features like ignored non-structural components (Katafygiotis and

Beck (1998)).

To compensate for these modeling errors, model updating methods are conducted for seeking the appropriate model to reproduce the observed response of a building. However, such a deterministic model fit to a limited observation is not always optimal because the dynamic characteristics of structures can fluctuate according to the season and aging. When predicting future building behavior by using the updated building's response analysis model, "uncertainty" should be considered along with "best estimation".

This paper discussed the challenges of taking into consideration the uncertainties in model identification of existing buildings and proposed a novel method for assessing uncertainties in observation data using Bayesian statistics.

2. ISSUES OF ASSESSING UNCERTAINTIES IN MODEL IDENTIFICATION

Some studies have been made to consider uncertainties in model identification using Bayesian updating. Such studies have been validated with respect to the maximum a posteriori estimate, but they have not sufficiently examined the shape of the posterior distribution that represents uncertainties. In some cases, response analysis models with maximum a posteriori estimates are used deterministically to predict response and damage for future earthquake.

Uncertainties should be properly and quantitatively assessed. Their effects should be reflected as part of the response analysis model. By using a so-called "fully Bayesian" model identification method, uncertainties are quantitatively assessed by using a probabilistic model expressing uncertainties of a response analysis model as random variables of the model parameters to assess damage and repair costs including associated uncertainties.

In order to identify models with quantified uncertainties, the following issues need to be addressed:

1. How to quantitatively assess the uncertainties based on observed data including observation noise.
2. How to properly consider uncertainties in future that are not included in observed data.
3. How to combine uncertainties estimated from observed data and uncertainties identified from existing information of engineering knowledge.

Among these issues, the first issue (item 1) is considered to be the first step to be tackled. By properly assessing the uncertainties based on observed data, then, it will also be possible to include the appropriate level of consideration for uncertainties in future not contained in observed data (item 2).

3. METHODOLOGY

In this chapter, we now formulate a method (Lee et al. (2022), Lee and Itoi (2022)) that can quantitatively assess the uncertainties regarding the first issue mentioned above.

It is assumed in this study that observation data $\mathscr{D} = \{u,\ {}_o\mathrm{Y}\}$ consisting of the input seismic ground motion u and response of building ${}_o\mathrm{Y}$ have been obtained, we assess the model parameters $\boldsymbol{\theta}$ such as the initial stiffness and yield displacement that constitute the response analysis model. We propose a method to obtain the distribution of the conditional probability density of the model parameters given observation data $p\left(\boldsymbol{\theta}|\mathscr{D}\right)$.

By applying Bayes' theorem and considering that $p\left(\mathscr{D}\right)$ is a constant, $p\left(\boldsymbol{\theta}|\mathscr{D}\right)$ can be expressed as in equation (2).

$$p\left(\boldsymbol{\theta}|\mathscr{D}\right) \propto p\left(\mathscr{D}|\boldsymbol{\theta}\right)\ p(\boldsymbol{\theta}) \tag{2}$$

Here, $p\left(\boldsymbol{\theta}|\mathscr{D}\right)$, $p\left(\mathscr{D}|\boldsymbol{\theta}\right)$, and $p(\boldsymbol{\theta})$ denote the posterior distribution of the model parameters $\boldsymbol{\theta}$, the likelihood of the observation data $\mathscr{D}$, and the prior distribution of the model parameters $\boldsymbol{\theta}$ respectively.

If the right-hand side of equation (2) is given, samples from the posterior $p\left(\boldsymbol{\theta}|\mathscr{D}\right)$ can be obtained using Markov chain Monte Carlo (MCMC) methods. Assuming that prior distribution $p(\boldsymbol{\theta})$ is given from prior information, only the likelihood $p\left(\mathscr{D}|\boldsymbol{\theta}\right)$ needs to be known.

Then, using some multidimensional random variable $\boldsymbol{z}$ that satisfies $p\left(\mathscr{D}|\boldsymbol{z},\ \boldsymbol{\theta}\right) = p\left(\mathscr{D}|\boldsymbol{z}\right)$, the likelihood is given as in equation (3).

$$p\left(\mathscr{D}|\boldsymbol{\theta}\right) = \int p\left(\mathscr{D}|\boldsymbol{z}\right) p\left(\boldsymbol{z}|\boldsymbol{\theta}\right) d\boldsymbol{z} \tag{3}$$

Transforming $p(\mathscr{D}|\boldsymbol{z})$ by Bayes' theorem and considering that $p(\mathscr{D})$ is a constant, it is expressed as in equation (4).

$$p(\mathscr{D}|\boldsymbol{\theta}) \propto \int \frac{p(\boldsymbol{z}|\mathscr{D})\, p(\boldsymbol{z}|\boldsymbol{\theta})}{p(\boldsymbol{z})} d\boldsymbol{z} \qquad (4)$$

As long as there is a way to obtain the probability distribution of $\boldsymbol{z}$ and the conditional probabilities of $\boldsymbol{z}$, it is possible to express the uncertainties of the response analysis model as the variability of the model parameters, which are random variables.

4. APPLICATION OF VAE

As an example of application of the method formulated in the previous chapter, we show that the posterior distributions of the model parameters is estimated using the approximated probability distribution obtained by Variational Auto-Encoder (Lee et al. (2022)).

4.1. Overview of VAE

Variational Auto-Encoder (VAE) is a sort of deep generative model that approximates the probabilistic structure from which observed data are generated by compressing data features into low-dimensional random variables (Kingma et al. (2014)). The structure of VAE is shown in Figure 1. The encoder of the neural network compresses the input data X into latent variables $\boldsymbol{z}$. The neural network decoder, which reconstructs data from latent variables, generates output data $\hat{\mathrm{X}}$ of the same size as the input data.

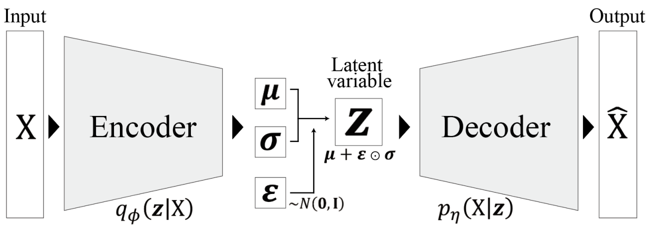


Figure 1: Structure of VAE

Due to space limitations, the reader is referred to the original paper for more details on VAE (Kingma et al. (2014)). Note that VAE is characterized by the fact that the outputs of the encoder and decoder are random variables. In particular, the encoder outputs z following a normal distribution, which is convenient when applied to the proposed method.

4.2. Formulation using VAE

Using $q(\bullet)$, which approximates $p(\bullet)$, the likelihood can be approximated as in equation (5) instead of equation (4).

$$p(\mathscr{D}|\boldsymbol{\theta}) \approx q(\mathscr{D}|\boldsymbol{\theta}) \propto \int \frac{q(\boldsymbol{z}|\mathscr{D})\, q(\boldsymbol{z}|\boldsymbol{\theta})}{q(\boldsymbol{z})} d\boldsymbol{z} \qquad (5)$$

Here, we use the output of the encoder of the learned VAE $q_\phi(\bullet)$ as $q(\bullet)$, where $\boldsymbol{z}$ and ϕ correspond to the latent variables and the parameters that constitute the network of encoders of the VAE, respectively. Since the encoder output $q_\phi(\boldsymbol{z}|\bullet)$ is a normal distribution and $q_\phi(\boldsymbol{z})$ can also be assumed to be a standard normal distribution, the right-hand side of equation (5) can be obtained analytically.

5. NUMERICAL EXPERIMENT

Numerical experiment is conducted to discuss the effectiveness of the proposed method.

5.1. Experiment Conditions

In this numerical experiment, the north-south component of El-Centro earthquake wave during 1940 Imperial Valley earthquake is used. Assuming that the building response during a medium earthquake (El-Centro wave in this case) is observed, the posterior distributions of the model parameters are estimated. The estimation results are then used to make probabilistic predictions of the peak response to a large earthquake (El-Centro wave amplified to 200% in this case), and the effectiveness of the proposed method is discussed.

A single degree-of-freedom model with bilinear restoring force characteristics is used. The set values of the model parameters are: natural frequency of 2 Hz, elastic damping ratio of 0.05, yield displacement of 2cm, and degrading ratio, which is the ratio of stiffness after yielding to initial stiffness, 0.2. The input/output acceleration time history waveform is assumed to include Gaussian noise with mean of 0 gal and standard deviation of 1 gal.

5.2. Training of VAE

The dataset used to train the VAE was created by generating uniform random numbers in the range shown in Table 1. The yield displacement is expressed as a ratio to the peak displacement from a

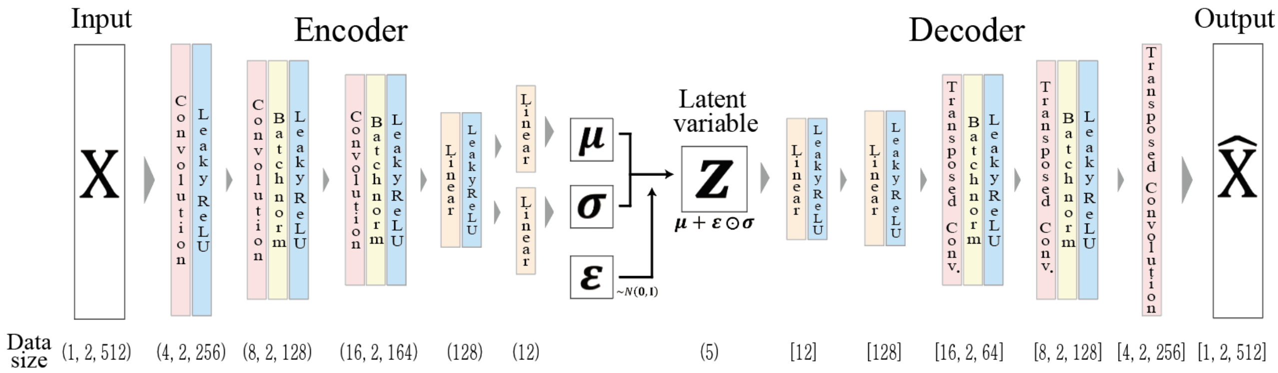


Figure 2: Encoder-decoder Architecture of VAE

linear response analysis (hereafter referred to as the yield displacement ratio). The range of yield displacement ratio, as in Table 1, is determined so that half of the dataset consists of nonlinear responses. The input ground motion was subjected to Gaussian noise with a mean of 0 gal and a standard deviation of 0.1 gal as the observation noise, and resampled to a sampling period of 0.001 s. Response analysis was performed using a damping ratio of 0.05, and 100,000 and 10,000 response analysis results were obtained for training and validation, respectively. The transfer functions of the response acceleration to the input acceleration were obtained in the frequency range of 0.2-9.7 Hz (512 points), and the real and imaginary parts were given to different channels.

Table 1: Range of model parameters for dataset

Natural Frequency (Hz)	Yield Disp. Ratio	Degrading Ratio
0.5~5	0.2 ~1.8	0 ~1

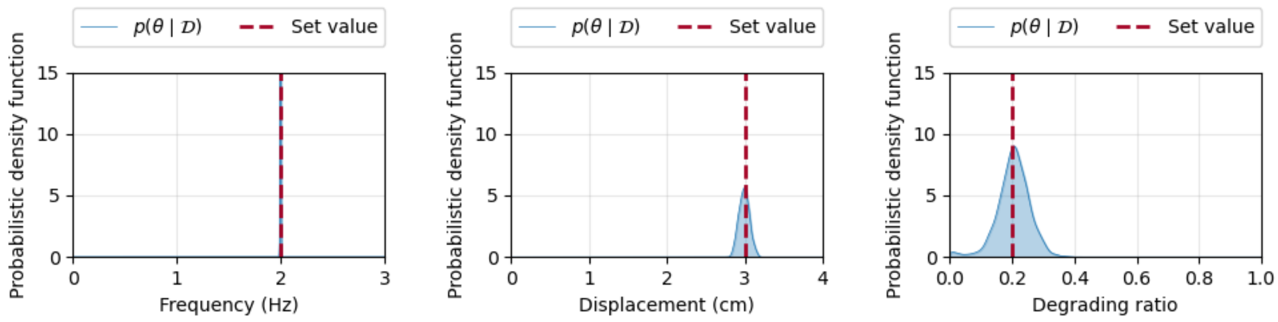
Figure 2 shows an encoder-decoder architecture of the VAE used for training. The data size in the figure represents (number of kernels, number of channels, data length). The (1, 2, 512)-dimensional transfer function in the input layer is compressed through the CNN and FC layers of the encoder into a 5-dimensional latent variable z, which is returned to the (1, 2, 512)-dimensional transfer function by the decoder in a symmetric configuration. The hyperparameters for training were set to a batch size of 1024, a learning rate of 0.0001, and 1000 epochs of training.

5.3. *Identification of Model Parameters*

Assuming an uninformed prior distribution, the Metropolis algorithm (Metropolis et al. (1953)), a MCMC method, was used to obtain samples of model parameters following the posterior distribution. Figure 3 shows the posterior distribution $p(\boldsymbol{\theta}|\mathscr{D})$ obtained by gaussian kernel density estimation with the burn-in interval set to 2,000 and the remaining 6,000 samples accepted. The set values of model parameters are indicated by the red dotted lines. The posterior distribution of each model parameter has a peak very close to the set value of model parameter. The natural frequency, which is easy to estimate from medium earthquake, have a narrow distribution, while the yield displacement and degrading ratio, which are difficult to estimate, have a wide distribution. These results are reasonable considering the feature of the bilinear restoring force characteristic that the nonlinear response with post-yield stiffness appears in a short period of time.

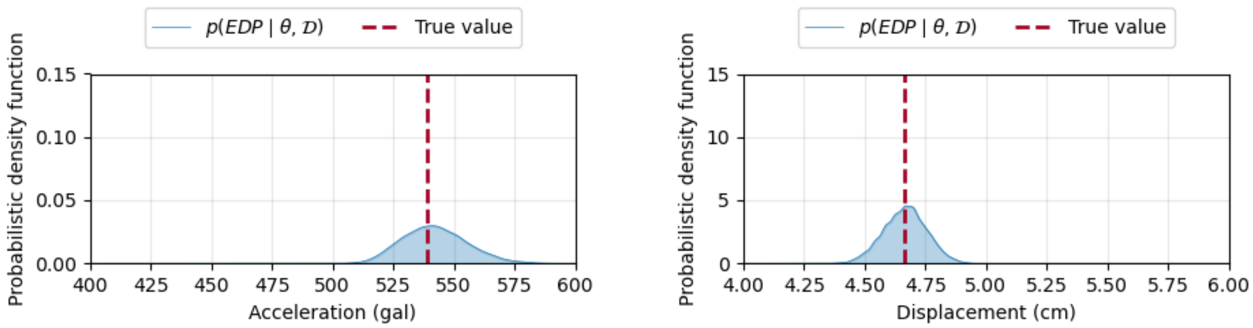
5.4. *Prediction for Large Earthquake*

The *EDP* (peak response acceleration, peak relative displacement) for a 200% amplified El-Centro wave was obtained by response analysis, using 1000 model parameters randomly selected from the samples obtained using MCMC. The probability distribution of the *EDP* obtained by gaussian kernel density estimation is shown in Figure 4. The true value of the peak response obtained by the response analysis using the set values of model parameters



(a) Natural Frequency (b) Yield displacement (c) Degrading ratio

Figure 3: Posterior distributions of model parameters θ



(a) Peak absolute acceleration (b) Peak relative displacement

Figure 4: Predictive distributions of EDP

are indicated by the red dotted line. The figure shows that the probability distribution of *EDP* has a peak in the posterior distribution near the correct value of *EDP*. Not only does the peak coincide with the true value, but it is also possible to assess the effect of the uncertainties in the model parameters on the *EDP* as a distribution.

6. CONCLUSIONS

This paper discussed the need to consider uncertainties in model identification and summarized the issues considering uncertainties in model identification. As a first step to tackle these issues, we demonstrated a fully Baysian method to assess the uncertainties in observed data. Numerical experiments showed that the proposed method could successfully estimate posterior distributions of the model parameters with appropriate level of uncertainties. It was also demonstrated that the posterior distribution of the model parameters can be used to predict the EDP for future earthquake.

7. REFERENCES

Applied Technology Council (2018). *Seismic Performance Assessment of Buildings. Volume 1 - Methodology*, second edition.

Katafygiotis, L. and Beck, J. (1998). "Updating models and their uncertainties. 1: Bayesian statistical framework." *Journal of Engineering Mechanics*, 124, 456–461.

Kingma, D. P., Rezende, D. J., Mohamed, S., and Welling, M. (2014). "Semi-supervised learning with deep generative models, <https://arxiv.org/abs/1406.5298>.

Lee, S. and Itoi, T. (2022). "From "structural health monitoring" to "building health care" - opportunities and challenges -." *The 8th International Symposium on Reliability Engineering and Risk Management*, 244–248.

Lee, S., Matsumoto, Y., Yaoyama, T., Hida, T., and Itoi, T. (2022). "Proposal for updating response analysis model using variational autoencoder for fully bayesian performance evaluation of building." *Summaries of Technical Papers of Annual Meeting*, 353–354.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). "Equation of state calculations by fast computing machines." *The journal of chemical physics*, 21(6), 1087–1092.

Moehle, J. and Deierlein, G. (2004). "A framework methodology for performance-based earthquake engineering." *13th World Conference on Earthquake Engineering*, number 679.