Seismic performance analysis based on Bayesian network

Hao Li [1, a] , Lang Wu[2,3], Jian Tang[4], XuCong Liu[5]

[1,2,4,5]Architecture and Civil Engineering Department, Jiang Xi Science & Technology Normal University, Nan Chang, China

[3]Nanchang Institute of Technology, Nan Chang, China

[a] cyberraider@163.com

Keywords: ductility demand；Bayesian network; Markov Chain Monte Carlo; posterior distribution

Abstract. Based on the regression analysis for 1918 seismic records, ductility demand probabilistic relationship with given earthquake intensity is formulated, and corresponding Bayesian network is established. With the given observation of earthquake intensity, the ductility demand posterior distribution is updated by using Markov Chain Monte Carlo method, thus the computation of the ductility demand is localized. Case study shows how the earthquake intensity parameters and the given observed values effect on the calculation result.

Introduction

Pacific Earthquake Engineering Research Center of USA has proposed the guiding principles for the probabilistic assessment of structural performance[1]: Three random variables: Intensity measure(IM), Damage Measure(DM) and Decision Variables(DV), are adopted, and the assessment is break down into three parts: probabilistic seismic hazard analysis $\lambda(IM)$, probabilistic seismic demand analysis $G(DM\,|\,IM)$ and probabilistic structural capacity analysis $G(DV\,|\,DM)$.

The guiding principles proposed by the PEER can be considered as a simplified three-level causal network described by probabilistic relations, that is, Bayesian network. The Bayesian network is a topological structure which composed of a directed acyclic graph and associated probability distribution functions [2]. As an artificial intelligence technique, it has been widely used for describing complex causal relationships. In the field of civil engineering, its applications are usually in the form of Bayesian inference to process single-level causal relationship [3-5].

In this paper, based on the regression analysis for 1918 seismic records, an intensity measure about seismic spectral characteristic, which has great effect on ductility demands, is constructed, and the ductility demand conditional probability distribution $G(DM\,|\,IM)$ is established based on it. One important feature of Bayesian network is that posterior distribution can be obtained given observed values. In this paper, collected seismic intensity parameters of specific region are used as observed values to update the posterior distribution of seismic ductility demand, thus realize the localization of seismic ductility demand calculation.

Analysis of Influencing Factors of Structure Ductility Demand

The structure ductility demands are affected by two aspects: the structure itself and the seismic actions. For the calculation of structure seismic ductility demands, the seismic actions have great uncertainty, while the uncertainty of structural characteristics is relatively small. So in the following analysis, the structural characteristic parameters are treated as constants rather than as variables.

Reasonable selection of seismic intensity measures is the basis for establishing the conditional probabilistic relationship $G(DM\,|\,IM)$ and reducing the dispersion [6, 7]. In this paper, Peak Ground Acceleration (PGA) is used as intensity measure, and through regression analysis for 1918 ground motion records, another new earthquake intensity measure F_{PGA} is constructed and the probabilistic relationship of $\eta - T - \mu$ is established.

Probabilistic Relationship of $\eta - T - \mu$

Consider the following single degree of freedom system:

$$\ddot{\mu} + \frac{4\pi}{T}\xi\dot{\mu} + \left(\frac{2\pi}{T}\right)^2 q(\mu) = -\frac{4\pi^2}{T^2\eta}\tau(t) \tag{1}$$

In which, ξ is the damping ratio, $q(u)$ is the normalized resilience model, $\tau(t)$ is normalized ground acceleration time history, T is the first structural period and η is the yield strength coefficient. The maximum value of μ obtained by time-history analysis based on the above equation is the structural ductility demand for specific ground motion recorder. For the simplicity, bilinear model with post yield stiffness coefficient 0.05 is adopted as the resilience model, and the damping ratio is 0.05.

Nonlinear time-history analysis is carried out for SDOF systems with T varied from 0.1 to 5s and yield strength coefficient η varied from 0.1 to 1.2. Based on these time-history analysis, simplified probabilistic relationship of $\eta - T - \mu$ can be obtained by multivariate nonlinear regression:

$$\mu = \mu_0 \left(\frac{0.1}{T}\right)^{r1} \left(\frac{0.1}{\eta}\right)^{r2} \tag{2}$$

In which, μ_0 can be seen as the ductility demand with 0.1s structural period and 0.1 of yield strength coefficient η. Where $r1$ and $r2$ can be seen as decreasing coefficients of structural period and yield strength coefficient. The fitting results of 1918 ground motion records show that the coefficient of decision of Equation 2 is more than 0.9, so Equation 2 can be considered as a good reflection of the $\eta - T - \mu$ relationship.

Seismic intensity measure F_{PGA}

In the Equation 2 of this paper, the value range of the parameters μ_0 is very large, which can be seen as an objective reflection of the huge difference of differe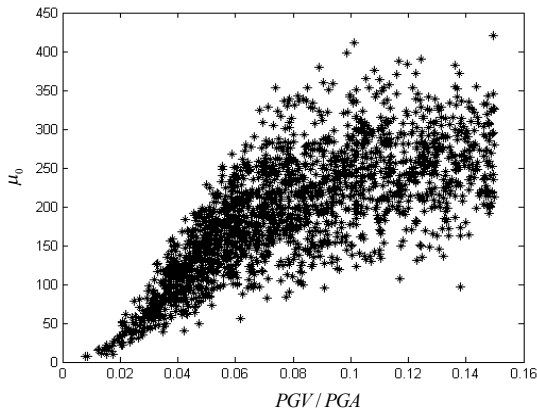nt spectral characteristics of earthquake. Unlike the intensity measures such as PGA, μ_0 cannot be measured, so it is necessary to construct a seismic spectrum characteristic parameter so as to relate with it. In this paper, such parameter directly related to μ_0 is constructed with considering the following two aspects:

First is to consider the velocity pulse characteristics. Related studies have shown that high energy velocity pulse will significantly increase the structural displacement [8, 9]. The velocity pulse can be represented by PGV/PGA. As shown in Fig.1, μ_0 is related to PGV/PGA, with increase of PGV/PGA, the dispersion of μ_0 is also increased. Therefore, 1918 records are divided

Fig.1 PGV/PGA vs μ_0

into three groups: less than 0.05 (411), 0.05 to 0.10 (942), 0.1 to 0.15 (565).

Second is to define a spectrum characteristic quantity which can directly reflect the parameter μ_0 and can be measured directly. This quantity is defined as peak ground acceleration (F_{PGA}) of normalized acceleration time history $\tau(t)$ obtained after low pass filtering (2.9Hz). Fig.2 shows that the two parameters are closely related. The regression relationship is shown below:

$$\mu_0 = (1 + \delta)(a F_{PGA} + b) \tag{3}$$

In which, δ is random-valued coefficient of variation, a and b is constant linear regression coefficients. The values of the coefficients for different groups are shown in table 1.

(a) group 1   (b) group 2   (c) group 3
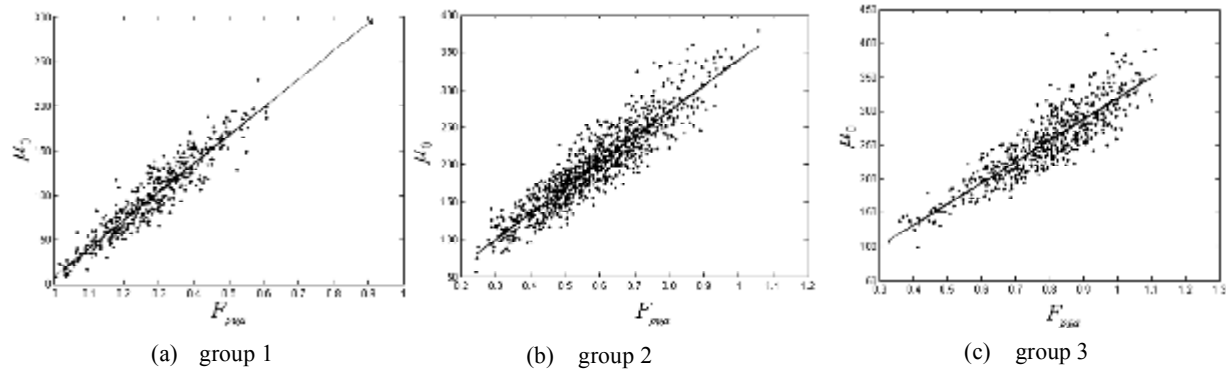
Fig.2 Relationship between F_{PGA} and μ_0

Bayesian network model

In this paper, Bayesian network, as shown in fig.3, is constructed to calculate the seismic ductility demand distribution. The prior distribution and conditional distribution or values of the Bayesian network nodes and the parameters involved are shown in Table 1.
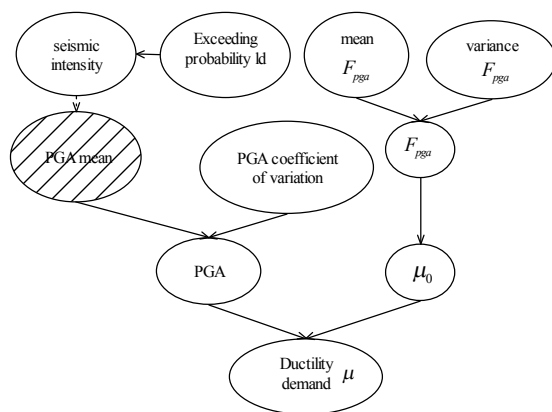

Fig.3 Baysian network

Table 1: Parameter values and distributions of Bayesian network

parameter name	Distribution type or value
Exceeding probability	$\log norm(-3.45, 0.58)$
Earthquake intensity ld	Extreme value type III truncated distribution
PGA coefficient of variation	$\log norm(-0.36, 0.10)$
PGA Mean value	$10^{ld \cdot \log_{10} 2 - 0.01}$
PGA	Extreme value type I distribution
F_{PGA} Mean value	Logarithmic normal distribution
F_{PGA} variance	Logarithmic normal distribution
F_{PGA}	Logarithmic normal distribution
μ_0	Formula 3
Ductility demand μ	Formula 2
δ	Group 1：$norm(2.20e-3, 0.18)$ Group 2：$norm(4.26e-4, 0.11)$ Group 3：$norm(1.83-4, 0.10)$
a	Group 1：318.05 Group 2：342.93 Group 3：312.60
b	Group 1：8.11 Group 2：-2.90 Group 3：5.97
$r1$	Group 1：$norm(1.60, 0.21)$ Group 2：$gamma(35.51, 0.04)$ Group 3：$extm\_typeI(1.04, 0.13)$
$r2$	Group 1：$norm(1.30, 0.17)$ Group 2：$norm(1.37, 0.08)$ Group 3：$norm(1.38, 0.05)$

The node "exceeding probability" in the network represents exceeding probability of rarely occurred strong earthquake, it can be regarded as a random variable varied between 0.02~0.03. "Seismic intensity ld" obeys the right truncated distribution of seismic intensity under given exceeding probability. Shadow node "mean value of peak acceleration" is a non-probabilistic node, which can be obtained by calculating with the given seismic intensity. The node "PGA coefficient of variation" is a priori node, and the average coefficient of variation is 0.7. "Peak acceleration PGA" follows the extreme value type I distribution under a given mean coefficient of variation [10]. The parameter F_{PGA} uses the mean and the variance as its priori nodes.

When there is no observation value, the Monte Carlo simulation of Bayesian network can be used to calculate the prior distribution of ductility demand. When there are observed values of ground motion

record of specific region, all the random variable nodes in Bayesian network can be updated to posterior distribution. Given F_{PGA}, the posterior distribution of μ_F and σ_F can be expressed as:

$$P(\mu_F,\sigma_F \mid \ddot{F}_{pga}) = \frac{P(\ddot{F}_{pga} \mid \mu_F,\sigma_F)P(\mu_F)P(\sigma_F)}{\iint P(\ddot{F}_{pga} \mid \mu_F,\sigma_F)P(\mu_F)P(\sigma_F)d\mu_F d\sigma_F} \tag{4}$$

MCMC algorithm

The denominator of the Bayesian posterior probability is usually a complicated multi-dimensional integral as shown in Equation 4, it's difficult to solve by using the analytical method or the general Monte Carlo numerical simulation, therefore, Markov Chain Monte Carlo simulation is used to solve the problem.

Markov chain and MH algorithm
Markov chain is a sample generation sequence which the current value of the variable depends only on the value at the previous time and does not depend on that at other time. The probability of shifting from one state value to another is called the transition probability P. If the Markov chain is ergodic, there must be a stable static distribution π^* which meets the relationship of $\pi^* = \pi^* P$.

With the posterior probability being seen as the stable static distribution probability of the Markov chain, the Metropolis-Hasting algorithm generates sampling sequence as follows:
(1) Starting from the initial value x_0;
(2) According to the current value x_t, candidate sample x^* is generated according to the jump distribution $q(x^* \mid x_t)$
(3) The acceptance rate α of candidate samples is calculated as following:

$$\alpha = \min\left(\frac{f(x^*)q(x^* \mid x_t)}{f(x_t)q(x_t \mid x^*)},1\right) \tag{5}$$

In which, $f(x)$ is the kernel function of posterior probability.
(4) Accept the candidate sample with probability α, Return to step 2.
The algorithm eventually converges to the static distribution, that is, posterior distribution.
Convergence diagnosis
The entire MCMC simulation consists of sampling and testing. Since the Markov chain does not reach the static distribution at the beginning, the samples generated at the beginning must be discarded, which is called the burn-in period. Since the sample is generated based on the previous sample, the correlation of the sample sequence is large, so the periodic interval sampling method is used to reduce the correlation.

There are a lot of methods to check the convergence of MCMC algorithm[11-13]. The commonly used methods include Geweke test and Raftery-Lewis test. The Geweke test is mainly used to verify the convergence of the sample moments. Raftery-Lewis test provides more detailed information needed for sampling convergence: by converting the Markov process at the time of sampling into a 0-1 binarized Markov process, the required length of the burn-in period, the total sampling length, and the thinning period is calculated. In this paper, both test methods are used in MCMC simulation.
The computational architecture of this paper
The MCMC algorithm adopted in this paper includes three stages:
(1) Adjustment stage. In this stage, the initial sample and the proposed distribution parameters are adjusted. Simulated annealing method is used to adjust the parameters of the proposed distribution.
(2) burn-in and sampling stage. In this stage, Metropolis-Hasting algorithm is used for sampling.
(3) Convergence diagnostic stage. Raftery-Lewis test was used to estimate the length of the burn-in period, the total sampling length, and the thinning period. Then returns the second stage to iterate until it passes the Geweke test.

Case study

A structure is simplified by PUSHOVER analysis as a SDOF system: the restoring force model is bilinear, the yield stiffness ratio is 0.05, the mass is 1×10^6 kg, the damping ratio is 0.05, the period is 1.10s, the yield strength is 1738kN, and the yield displacement is 53.426mm. Consider the structure ductility demand for seismic fortification target of "No Collapsing in the Strong Earthquake" for 8 degree of earthquake intensity.

Assuming that the seismic velocity pulse meet the condition of third group, that is, $PGV / PGA \in [0.10, 0.15]$. In the absence of observations, the prior distributions of the nodes of interest are shown in Table 2. In table 2, the mean value of F_{PGA} obeys the logarithmic normal distribution with the mean value of 0.5 and the coefficient of variation of 0.1. The 50-years exceeding probability of rarely occurred strong earthquake obeys log normal distribution with a mean value of 0.025 and a coefficient of variation of 0.1. When F_{PGA} and PGA variation coefficient distributions changes while mean value of that are constant distributions, following conclusions can be obtained from table 2:

1. When the variability of PGA and F_{PGA} decreases, the variability of ductility demand decreases accordingly, especially when the coefficient of variation of the two is about 0.2, the coefficient of variation of ductility requirement is less than 1.

2. The variability of ductility demand μ is more sensitive to the variability of PGA than that of F_{PGA}.

3. The mean value of the ductile demand μ appears to be affected by the variability of PGA.

Assuming that: distribution of exceeding probability is shown in Table 1; the mean value of the distribution of the standard deviation of PGA is 0.7 and the standard deviation is 0.1; the mean value of F_{PGA} obeys the logarithmic normal distribution with the mean value of 0.5 and the coefficient of variation of 0.1, its variance obeys the logarithmic normal distribution with the mean value of 0.1 and the coefficient of variation of 0.1. Now consider the situation with observed values: the ground peak acceleration PGA is 8.8396 and 9.4864; F_{PGA} are 0.8425 and 0.8706.

The prior distribution and posterior distribution of the nodes of interest are shown in Table 3.

Because the prior distribution of PGA has a large variance, it shows a good "compatibility" with the observed values. The changes of its posterior distribution, especially the changes of quintile value are not significant; on the other

Table 2 Node information of Bayesian network

mean value of CV	PGA mean variance	μ_0 mean variance	μ mean variance
0.7,0.7	5.8517 4.6845	162.4579 110.6908	3.0456 4.9100
0.7,0.2	5.8286 4.7373	161.3827 38.3434	3.0000 4.0323
0.2,0.7	5.8161 2.1785	165.2831 121.0101	2.7468 2.9565
0.2,0.2	5.7963 2.1189	162.1613 38.2385	2.6776 1.9402

Note: Mean value of CV are PGA and F_{PGA} from left to right.

Table 3 Node information of Bayesian network

Node	prior distribution			
	mean	variance	Quantile 0.5	Quantile 0.95
PGA	5.810	4.616	4.882	14.261
μ_0	162.33	27.385	160.29	210.90
μ	3.017	3.863	1.853	9.788

Node	Posterior distribution			
	mean	variance	Quantile 0.5	Quantile 0.95
PGA	5.547	4.094	4.743	13.168
μ_0	227.60	35.76	225.40	290.23
μ	3.90	4.510	2.517	12.273

hand, the prior distribution of F_{PGA} has a large deviation from the observed values; however, the posterior distribution corrects this deviation. As can be seen from the table 3, posterior distribution of has a significant "right shift" compared with the prior distribution, which indicates that due to the addition of the observed value, μ_0 should have a larger value. The distribution of the ductility demand μ is also shifted to the right of the prior distribution, and the coefficient of variation is reduced from 1.28 to 1.16.

Conclusion

Based on the 1918 collected ground motion records, the nonlinear time-history analysis of the SDOF systems with different yield strength coefficient and periods are performed. By means of regression analysis, a new parameter F_{PGA} is constructed to describe the influence of seismic spectral characteristics on ductility demand. On this basis, the corresponding Bayesian network is formed. The Markov Chain Monte Carlo (MCMC) simulation method is used to update the posterior distribution of the ductility demand under the given earthquake intensity observed values.

Case study shows that intensity parameters, F_{PGA} and PGA all have great influence on seismic ductility demand, and the Bayesian network constructed based on them can better reflect the quantitative logic relationship between ductility and intensity. The Bayesian network can adjust the posterior distribution according to observed values, thus "localize" the ductile demand calculation.

Acknowledgements

References

[1] Cornell C A, Krawinkler H. Progress and challenges in seismic performance assessment [OL]. PEER Center News, 2000, http://peer.berkeley.edu/news/2000springs

[2] Jordan M I. Learning in Graphical Models [M]. MIT Press, 1998.

[3] Geyskens P, Kiureghian A D, Monterio P. Bayesian prediction of elastic modulus of concrete[J]. Journal of structural engineering, 1998, 124 (1): 89-95.

[4] Beck J L, Katafygiotis L S. Updating models and their uncertainties. I: Bayesian statistical framework[J]. Journal of Eng. Mech., 1998, 124 (4): 455-461.

[5] Yuen K V, Siu K A, Beck J L. Two-stage structural health monitoring approach for phase I benchmark studies[J]. Journal of Eng. Mech., 2004, 130 (1):16-33.

[6] Shome N, Cornell C A, Bazzurro P, Carballo J E. Earthquake, Records, and Nonlinear Responses[J]. Earthquake Spectra, 1998, 14 (3): 469-497.

[7] Baker J W, Cornell C A. A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon[J]. Earthquake Engineering & Structural Dynamics, 2005, 34 (10): 1193-1217.

[8] Malhotra P K. Response of buildings to near-field pulse-like ground motions[J]. Earthquake engineering and structural dynamics, 1999, 28: 1309-1326.

[9] Huang Jianwen, Zhu Xi. Inelastic response analysis for the SDOF structures under near-field earthquakes[J]. China Safety Science Journal, 2003, 13 (11): 59-65 (in Chinese)

[10] Ou Jinping, Duan Yubo, Liu Huiyi. Structural random earthquake action and its statistical parameters[J]. J. Harbin Archit.& Civ. Eng. Inst., 1994, 27 (5): 1-10 (in Chinese).

[11] Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments[G]. In Bayesian Statistics 4. Oxford: Oxford Univ. Press, 1992: 169-193.

[12] Raftery A E, Lewis S. How many iterations in the Gibbs sampler[G]. In Bayesian Statistics 4. Oxford: Oxford Univ. Press, 1992: 763-773.

[13] Cowles M K, Carlin B P. Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review[J]. Journal of the American Statistical Association, 1996, 91(434): 883-904.