creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

Master's Thesis of Engineering

Efficient simulation-based approaches for community-level probabilistic seismic risk assessment

지역단위 지진 리스크 평가를 위한 효율적 시뮬레이션 기반 접근법

February 2017

Graduate School of Engineering
Seoul National University
Civil and Environmental Engineering

Byeong-Seong Choi

Abstract

A research is required to evaluate seismic risk of community on the probabilistic framework. That is because (a) spatially distributed buildings/infrastructures are critical assets in urban community and (b) there are uncertainties in both natural hazard and structural behavior. While Monte Carlo Simulation (MCS) has been a solution to assess risk at community-level, MCS is not always efficient solution especially when severe hazardous scenarios should be investigated. Although Monte-Carlo simulation (MCS) provides straightforward environment to evaluate seismic risk on the urban assets, it requires a large computational cost to forecast rare event. Such catastrophic situations should be identified to keep our community being sustainable even after urban disasters. To overcome this issue, this thesis proposes alternative simulation-based approaches for probabilistic seismic risk assessment at community-level, (a) cross-entropy-based concurrent adaptive importance sampling (CE-CAIS) and (b) clustering-based approach. These new techniques are designed to establish computationally efficient frameworks for probabilistic seismic risk assessment on urban community. In Chapter 2, CE-CAIS is introduced to identify seismic risk on multi-state, large-scale systems with two dimensionality reduction techniques to expand the applicability of CE-CAIS. In Chapter 3, the clustering-based approach is demonstrated to forecast seismic risk on complex urban road networks with decreased computing resources. Several numerical examples are attached on each chapter to validate our proposals. Through this thesis, further researches are expected to produce other valuable achievements while increasing the communication with urban disaster and resilience.

Keyword : seismic risk, cross-entropy, principal component analysis, central limit theorem, feature selection
Student Number : 2015-21307

Contents

Abstract i
Chapter 1. Introduction 1
1.1. Study Background 1
1.2. Purpose of Research 2
Chapter 2. Probabilistic risk assessment of multi-state, large-scale systems using cross-entropy-based adaptive importance sampling 4
2.1. Introduction 4
2.2. Overview on fundamental methodologies 6
2.2.1. Cross-entropy-based adaptive importance sampling 6
2.2.2. Probabilistic risk assessment for urban community 8
2.3. Concurrent adaptive importance sampling 11
2.4. Post-hazard Traffic Flow Capacity of Hypothetical Road Network 12
2.4.1. Matrix-based System Reliability Method 13
2.4.2. Procedure of efficient sampling 14
2.4.3. Results 15
2.5. Dimensionality reduction techniques 20
2.5.1. Principle component analysis 20
2.5.2. Central limit theorem 22
2.6. Aggregated regional monetary loss in Shelby County 23
2.6.1. Damage factor and Loss Estimation for individual building 25
2.6.2. Procedure of efficient sampling 26
2.6.3. Results 28
2.7. Conclusion 34
Appendix 2A. Simulation of Seismic Hazard on Uncorrelated Standard Normal Space 34
2A.1. Magnitude-frequency relationship 35
2A.2. Realization of rupture surface 36
Appendix 2B. Updating Rules for CE-AIS-GM 37

2B.1. Optimal Importance Sampling Density 38
2B.2. Cross-entropy-based Updating Rule with Using Gaussian Mixture 38
2B.3. Initial parameters to implement CE-CAIS 41
Chapter 3. Feature selection and clustering-based approach for complex lifeline network 42
3.1. Introduction 42
3.2. Probabilistic Seismic Risk Assessment for Lifeline Network 45
3.2.1. Randomness in ground-motion prediction 45
3.2.2. Uncertainties in structural behavior 48
3.2.3. Network risk assessment in the context of utility 49
3.3. Feature Selection and Clustering-based Approach 52
3.3.1. Feature selection: proxy measures 53
3.3.2. Clustering-based approach for PRA 55
3.4. Post-hazard Traffic Flow of Hypothetical Road Network 56
3.5. Traffic Network on Bay Area, San Francisco 62
3.6. Conclusion 69
Appendix 3A. Scenario studies on Bay Area road network example 70
Chapter 4. Conclusion 75
Bibliography 77
Abstract 80

Tables

Table 2.1. Number of samples required to achieve target c.o.v 0.01……17
Table 2.2. Number of samples to obtain probabilities of events with original CE-AIS (c.o.v. 0.01) ……... 17
Table 2.3. Annual probabilites of reduced network traffic capacity ……... 17
Table 2.4. Mean values of damage factor ……... 26
Table 2.5. Standard deviation values of damage factor …….. 26
Table 3.1. Computational time of each step for risk assessement (Hypothetical network) ……... 59
Table 3.2. Error measurement for each k value (Hypothetical network) … 60
Table 3.3. Error measurement for each k value (Bay Area road network) ……. 64
Table 3.4. Computational time of each step for risk assessement (Bay Area road network) …….. 65
Table 3A.1. Evaluated proxy measures and target performance measure of the three scenarios ……... 70

Figures

Figure 2.1. Hypothetical traffic network facing with 9 active faults …….. 13
Figure 2.2. Annual probabilities by brute-force MCS/CE-AIS-GM/CE-CAIS-GM ……... 18
Figure 2.3. Effect of spatial correlation in intra-event residuals …… 18
Figure 2.4. Convergence of brute-force MCS …… 19
Figure 2.5. Convergence of CE-CAIS-GM …… 19
Figure 2.6. Scenario earthquake event (Mw 7.9; Blytheville) ……25
Figure 2.7. Eigen-values of principal components …… 30
Figure 2.8. Spatial profile of principal components ……... 31
Figure 2.9. Effect of calibration strategy ……31
Figure 2.10. A realization of ground-motion intensities with spatial correlation …….. .32
Figure 2.11. Conditional loss curve and its approximation by CLT …….. 32

Figure 2.12. Comparison of results by brute-force MCS/CE-CAIS-GM ... 33
Figure 2.13. Coefficient of variation of exceedance probability …….33
Figure 2A.1. Demonstration of sampling procedure for earthquake rupture ……. 37
Figure 3.1. Demonstration of sampling procedure for earthquake rupture ……..48
Figure 3.2. Realizations of the network to help understand proxy measures ……. 55
Figure 3.3. Illerstration of the hypothetical road network …….. 58
Figure 3.4. Annual exceedance probability curves (Hypothetical network)……... 61
Figure 3.5. Scatter plots of data on the feature space (Hypothetical network) ……. 62
Figure 3.6. Illustration of San Francisco Bay Area road network …….. 65
Figure 3.7. Super-distinct nodes and demands of the network ……... 66
Figure 3.8. Annual exceedance probability curves ……. 68
Figure 3.9. Scatter plots of data on the feature space ……. 68
Figure 3A.1. Scenario study on earthquake event with moment magnitude of 6.9 on Calaveras (Central) 2011 ……. 72
Figure 3A.2. Scenario study on earthquake event with moment magnitude of 6.9 on Green Valley 2011 …… 73
Figure 3A.3. Scenario study on earthquake event with moment magnitude of 7.5 on San Andreas (Peninsula) 2011 ……. 74

Chapter 1. Introduction

1.1. Study Background

Today's urban community is represented by its characteristics of being highly complex systems with large numbers of infra-structures and their interdependencies (Der Kiureghian and Song, 2008). Therefore, the risk of an urban community against natural and man-made hazards should be understood and assessed at the levels of not only individual damage of infrastructures but also its utility/transportation of urban networks and the entire community. Because of various uncertainties in potential hazards, structural damage, post-hazard performance of infrastructures, and socio-economic impacts, the risk assessment should be performed using probabilistic analysis methods. Such a probabilistic risk assessment provides a core basis for so-called resilience-based decision making, which requires accurate assessment of robustness, redundancy, resourcefulness, and rapid recovery of the community (Bruneau and Reinhorn 2006).

A main challenge in the probabilistic risk assessment (PRA) of an urban community comes from statistical dependence between failures of the infrastructures and its significant impact on the system-level risk (Song and Ok 2010, Lim and Song 2012). Thus, the PRA method needs to be able to incorporate the likelihood of joint failure patterns accurately into the procedure of assessing the post-hazard performance of a utility or transportation network, and the aggregated regional loss of the community. Besides that challenge, complex mechanism of the network performance makes the development and application of non-simulation-based approach limited (See examples of non-simulation approaches in Lim and Song 2012, Lee et al. 2011, Lim et al. 2015).

Because of these challenges, Monte-Carlo Simulation (MCS) has been widely accepted for the purpose of PRAs at the levels of networks and the entire community (Bommer et al. 2002, Shiraki et al. 2007, GEM 2016, Jayaram and Baker 2010, Miller and Baker 2015). However, a brute-force application of MCS has a limitation because it requires a huge computational time to identify low probability of post-hazard event. The low-probability event is still important to increase our communication with urban disasters in the context of risk mitigation and resilience-based decision making. As an effort to save computing resources, importance sampling (IS) approach has been employed by Jayaram and Baker (2010) to analyze post-hazard traffic flow for simplified (or aggregated) road

network. The research also exploited k-means clustering on the ground-motion intensities to recognize similarities between realized hazard scenarios. Later, an optimization-based approach was developed to select the appropriate number of damaged scenarios (damage-maps) among the generated catalog (Miller and Baker 2015). The optimization is based on the ground-motion intensity profiles of several sites and the profile of an indirect indicator, as known as proxy-measure, for the target performance measure. Although these researches have shown sufficiently improved performance with the point of efficient PRA, the applications of such approaches have been still constraint. For example, Jayaram and Baker (2010) has limitations because (a) the post-hazard performance of the simplified network differs from that of real-scaled network, (b) heuristic IS densities in the research are not guaranteed to keep consistency not relying on the local properties of the network, and (c) the clustering, only focusing on the data about ground-motion intensities, cannot reflect the uncertainties in structural behavior of infrastructures. Also, (d) the clustering algorithm, suggested in the paper, requires plentiful random-access memory (RAM) and computing time to insure sufficient work space in the computer since the algorithm needs to be implemented on the high dimensional space if the data is collected from amount of spatially distributed sites. On the other hand, Miller and Baker (2015) has another drawback because the optimization algorithm in the research employs tuned parameters to select damage-maps among the catalog. Such factors, which are potentially sensitive to the problem, are not desirable for general applications to complex and localized urban networks.

1.2. Purpose of Research

In this thesis, computationally efficient simulation-based approaches are suggested for probabilistic seismic risk assessment on urban community.

In the Chapter 2, a new importance sampling (IS) based framework is established by developing cross-entropy-based concurrent adaptive importance sampling (CE-CAIS). Recently, an efficient and versatile IS algorithm called cross-entropy-based adaptive importance sampling (CE-AIS) has been proposed (Rubinstein and Kroese 2004, Kurtz and Song 2013, Wang and Song 2016). CE-AIS identifies a near-optimal sampling density through few rounds of pre-sampling to achieve efficient and accurate MCS for component and system reliability problems. The chapter aims to facilitate applications of CE-AIS to (a) multi-state network reliability problems by identifying a near-optimal density for multiple

network states, termed as cross-entropy-based “concurrent” adaptive importance sampling (CE-CAIS), and (b) regional loss assessment by using dimension reduction techniques. This chapter first provides an overview of CE-AIS and community-level PRA for spatially distributed structures (Section 2.2). Then, CE-CAIS is presented for multi-state network reliability problems (Section 2.3), and demonstrated by an example of post-hazard traffic capacity analysis (Section 2.4). Next, dimension reduction techniques developed for application of CE-AIS to regional loss assessment are introduced (Section 2.5), and demonstrated by an example of aggregated monetary loss of Shelby County, TN (Section 2.6). Finally, Section 2.7 provides a summary and discussions on findings and future research needs.

In the Chapter 3, a computationally efficient framework of PRA is introduced for lifeline network by succeeding two benchmark studies (Jayaram and Baker 2010, Miller and Baker 2015). For the purpose of convenient data analysis, indirect (but computationally reasonable) measures, as known as proxy measures, are employed again to cluster the date based on the similarity. The k-means clustering is executed to make groups and to assign a representative post-hazard performance value to each cluster. The framework is less sensitive to the localized network because there is no heuristic IS density or parameter unlike benchmark studies, which should be tuned according to the characteristics of problems; the suggestion can be comprehended on the extension of unsupervised learning with extracted features and clustering approach to develop computationally efficient framework of PRA for urban network. To achieve our research goal, we first review the framework of PRA for spatially distributed network under seismic hazard; the review provides investigations on uncertainties in ground-motion intensities, fragility of infrastructures, and network analysis for its functionality (Section 3.2). In Section 3.3, we extract features of realized damage-scenarios (damage-maps) to evaluate the similarity between samples, and k-means clustering algorithm is demonstrated to develop the computationally efficient framework. After instructing methods for feature extraction and clustering algorithm, the proposal is applied to two numerical examples in Section 3.4 (a hypothetical road network consisting of 100 bridges) and 3.5 (the San Francisco Bay Area road network composed by 1,656 bridges in morning trip time). Finally, Section 3.6 presents a summary and discussions on findings and future research needs.

Chapter 2. Probabilistic risk assessment of multi-state, large-scale systems using cross-entropy-based adaptive importance sampling

A research is required to assess seismic risk on community with probabilistic concepts. That is because (a) spatially distributed buildings/infrastructures are critical assets in urban areas and (b) there are uncertainties in both natural hazard and structural behavior. Also, catastrophic situations should be identified to keep our community being prepared for urban disasters in spite of its low-probability. Although Monte-Carlo simulation (MCS) provides straightforward environment to assess community-level seismic risk, it requires a large computational cost to forecast rare event. To overcome this issue, authors propose alternative simulation-based approach, cross-entropy-based "concurrent" adaptive importance sampling (CE-CAIS). This new sampling technique is designed to provide fair well convergences for multiple level of events, so that this concurrent converging strategy is much more efficient to estimate a variety of hazardous events and their probabilities. An example about post-hazard flow capacity shows outstanding efficiency of our proposal. Also, to apply CE-CAIS to realistic community-level risk assessment, two dimensionality reduction techniques, principal com-ponent analysis (PCA) and central limit theorem (CLT), are provided, especially for aggregated monetary loss in community. PCA represents high dimensional intra-event residuals in ground-motion with only a few essential variables, and CLT contributes on simple approximation to evaluate conditional loss distribution with a given intensity map. Another example about aggregated monetary loss in urban area is introduced to show effectiveness of dimensionality reduction techniques and their extendibility with CE-CAIS.

2.1. Introduction

Today's urban communities are highly complex systems characterized by large numbers of infra-structures and their interdependencies (Der Kiureghian and Song, 2008). Therefore, the risk of an urban community against natural and man-made hazards should be understood and assessed at the levels of its utility/transportation networks and the entire community rather than in terms of individual infrastructures. Because of various uncertainties in potential hazards,

structural damage, post-hazard performance of infrastructures, and socio-economic impacts, the risk assessment should be performed using probabilistic analysis methods. Such a probabilistic risk assessment provides a core basis for so-called resilience-based decision making, which requires accurate assessment of robustness, redundancy, resourcefulness, and rapid recovery of the community (Bruneau and Reinhorn 2006).

A main challenge in the probabilistic risk assessment (PRA) of an urban community comes from statistical dependence between failures of the infrastructures and its significant impact on the system-level risk (Song and Ok 2010, Lim and Song 2012). Thus, the PRA method needs to be able to incorporate the likelihood of joint failure patterns accurately into the procedure of assessing the post-hazard performance of a utility or transportation network, and the aggregated regional loss of the community. Besides that challenge, complex mechanism of the network performance makes the development and application of non-simulation-based approach limited (See examples of non-simulation approaches in Lim and Song 2012, Lee et al. 2011, Lim et al. 2015).

Because of these challenges, Monte-Carlo Simulation (MCS) has been widely used for the purpose of PRAs at the levels of networks and the entire community (Bommer et al. 2002, Shiraki et al. 2007, GEM 2016). However, a brute-force application of MCS requires a huge number of simulations to obtain a reliable result for low probability events. In efforts for hazard mitigation and resilience-based decision making, low-probability events are important because they may have a severe impact on urban communities and thus require establishing appropriate design targets and emergency action plans based on accurate assessment of their likelihoods and impacts. To overcome this limitation of brute-force MCS, importance sampling (IS) approaches have been used. For example, Jayaram and Baker (2010) employed a data reduction technique called k-means clustering for post-hazard traffic flow analysis. Later, a map-selection technique was also proposed to reduce the number of simulations and computational cost (Miller and Baker 2015). This approach uses an optimization algorithm to select representative scenarios of spatial distributions of ground motion intensities and damage maps for given target return periods. Although these research efforts significantly improved the efficacy of PRA of complex large-scale systems, simulation-based approaches still have weakness in terms of computational cost and a large number of random variables needed to represent uncertainties in spatial distribution of hazards and, states of individual infrastructures.

Recently, an efficient and versatile IS algorithm called cross-entropy-based adaptive importance sampling (CE-AIS) has been proposed (Rubinstein and

Kroese 2004, Kurtz and Song 2013, Wang and Song 2016). CE-AIS identifies a near-optimal sampling density through few rounds of pre-sampling to achieve efficient and accurate MCS for component and system reliability problems. This chapter aims to facilitate applications of CE-AIS to (a) multi-state network reliability problems by identifying a near-optimal density for multiple network states, termed as cross-entropy-based "concurrent" adaptive importance sampling (CE-CAIS), and (b) regional loss assessment by using dimension reduction techniques.

This chapter first provides an overview of CE-AIS and community-level PRA for spatially distributed structures (Section 2.2). Then, CE-CAIS is presented for multi-state network reliability problems (Section 2.3), and demonstrated by an example of post-hazard traffic capacity analysis (Section 2.4). Next, dimension reduction techniques developed for application of CE-AIS to regional loss assessment are introduced (Section 2.5), and demonstrated by an example of aggregated monetary loss of Shelby County, TN (Section 2.6). Finally, Section 2.7 provides a summary and discussions on findings and future research needs.

2.2. Overview on fundamental methodologies

This section reviews elementary components of simulation-based approach to achieve efficient framework for PRA of urban community. First, an adaptive IS is reviewed, which is known as CE-AIS. The merit of CE-AIS is that this methodology provides a near-optimal IS density asymptotically through a few rounds of pre-sampling without any subjective intuition. This merit of CE-AIS guides us to efficient framework for any general reliability problem, but this chapter only focuses on applying CE-AIS method to community-level PRA; it may be possible to modify the proposal, which is introduced in this chapter, to extend its applicability to solve other class of reliability problems. After that, existing simulation-based framework of PRA is demonstrated, especially for seismic hazard. This basic interpretation on community-level PRA is a key element to improve its computational efficiency through CE-AIS, which will be discuss on next sections.

2.2.1. Cross-entropy-based adaptive importance sampling

IS is a variance reduction technique which replaces the original sampling density in MCS with an IS sampling density. Consider a numerical integration and its MCS approach:

$$I_t = \int H(\mathbf{x}) f(\mathbf{x};\mathbf{v}) d\mathbf{x} \cong \frac{1}{N}\sum_{i=1}^{N} H(\mathbf{x}_i). \tag{2.1}$$

where $H(\mathbf{x})$ is a general function of variables $\mathbf{x}$, and $f(\mathbf{x};\mathbf{v})$ is the probability density function (PDF) of $\mathbf{x}$ with parameters $\mathbf{v}$. For most structural reliability analysis, $f(\mathbf{x};\mathbf{v})$ represents uncorrelated standard normal distribution.
This integration in (2.1) can be represented as (2.2) by introducing IS sampling density $h(\mathbf{x};\mathbf{w})$ with the distribution parameters $\mathbf{w}$, thus

$$I_t = \int \left[\frac{H(\mathbf{x}) f(\mathbf{x};\mathbf{v})}{h(\mathbf{x};\mathbf{w})}\right] h(\mathbf{x};\mathbf{w}) d\mathbf{x} = E_v\left[\frac{H(\mathbf{x}) f(\mathbf{x};\mathbf{v})}{h(\mathbf{x};\mathbf{w})}\right] \cong \frac{1}{N}\sum_{i=1}^{N} \frac{H(\mathbf{x}_i) f(\mathbf{x}_i;\mathbf{v})}{h(\mathbf{x}_i;\mathbf{w})}. \tag{2.2}$$

With appropriate IS sampling density in (2.2), the estimation by IS converges to I_t much more quickly than the estimation of brute-force MCS; in other word, our interests to exploit IS are the cases that variance of an estimator for I_t is much smaller than that of brute force MCS, $\mathrm{Var}_h[H(\mathbf{x}_i) f(\mathbf{x}_i;\mathbf{v})/h(\mathbf{x}_i;\mathbf{w})] \ll \mathrm{Var}_f[H(\mathbf{x}_i)]$.

Using Cauchy and Schwarz inequality, Rubinstein (1981) showed that $p^*(\mathbf{x})$ is the "best" function for IS density $h(\mathbf{x};\mathbf{w})$ which is expressed as:

$$p^*(\mathbf{x}) = \frac{|H(\mathbf{x})| f(\mathbf{x};\mathbf{v})}{\int |H(\mathbf{x})| f(\mathbf{x};\mathbf{v}) d\mathbf{x}} \tag{2.3}$$

However, this best IS density in (2.3) is not feasible without obtaining the final result of MCS because the denominator in (2.3) is corresponding to I_t if $H(\mathbf{x})$ is a positive function; it is an often case in practice. To overcome this problem, an adaptive IS approach is suggested so that one can find a near-optimal IS density, which is approximated version of the best IS density. This density is effectively obtained with a non-parametric distribution by using the Kullback-Leibler cross-entropy (CE) as a measure of the functional distance between two PDFs. In Cross-entropy-based adaptive importance sampling (CE-AIS) approach, the difference between the optimal sampling density and the current IS density model is evaluated by CE. Through a few rounds of pre-sampling, density model parameters of the previous round, denoted by $\mathbf{w}_{\mathrm{pre}}$, are updated to $\mathbf{w}$ such that the estimated CE between $p^*(\mathbf{x})$ and the sampling density $h(\mathbf{x};\mathbf{w})$ is minimized. This condition is mathematically described as

$$\frac{1}{N}\sum_{i=1}^{N} |H(\mathbf{x}_i)| W(\mathbf{x}_i;\mathbf{u},\mathbf{w}_{\mathrm{pre}}) \nabla_{\mathbf{w}} \ln h(\mathbf{x}_i;\mathbf{w}) = 0 \tag{2.4}$$

where $W(\mathbf{x};\mathbf{u},\mathbf{w}_{\text{pre}})$ is so-called likelihood ratio, $f(\mathbf{x};\mathbf{u})/h(\mathbf{x};\mathbf{w}_{\text{pre}})$. Closed-form updating rules representing (2.4) were derived for parametric unimodal distribution models in Rubinstein and Kroese (2004).

To facilitate the use of the CE-AIS approach for complex failure domains of component and system reliability problems, Kurtz and Song (2013) introduced a new CE-AIS approach using a nonparametric multimodal distribution model called the Gaussian mixture (GM). This CE-AIS-GM method showed superb efficiency for component problems with multiple design points or important domains as well as various system reliability problems including series, parallel and general systems. Extensive numerical tests confirmed that the accuracy and efficiency of the proposed method are not sensitive to the level of probability, or nonlinearity of limit-state surfaces. The method is robust against numerical noises in limit-state surfaces as well. Currently, Wang and Song (2016) improved the performance of CE-AIS to deal with high dimensional reliability problems by adopting the von Mises–Fisher mixture density model instead of a Gaussian mixture. See Kurtz and Song (2013) and Wang and Song (2016) for more details.

2.2.2. Probabilistic risk assessment for urban community

The uncertainties to implement PRA for spatially distributed structures can be listed as: (a) Magnitude of earthquake, (b) Location of rupture, (c) Intra-event residuals in ground-motion prediction equation (GMPE), (d) Inter-event residuals in GMPE, and (e) Uncertainty existing on damage/loss of individual structure. In probabilistic seismic hazard analysis or PRA, ground-motion prediction equation (GMPE) is an empirical model to estimate ground motion intensities at the site. A general form of GMPE is expressed as

$$\ln Y_{ij} = f(M_j, R_{ij}, \theta_i) + \sigma_{ij}\varepsilon_{ij} + \tau_j\eta_j \tag{2.5}$$

where Y_{ij} denotes the selected ground-motion intensity measure, -e.g. peak ground acceleration or spectral acceleration at site i for the earthquake event j-, $f(M_j, R_{ij}, \theta_i)$ is attenuation law to predict mean value of Y_{ij} given as a function of M_j (magnitude of event j), R_{ij} (seismological distance to site i in earthquake j), and θ_i (a set of other explanatory variables assigned at site i). ε_{ij} is the intra-event residual to represent site-to-site uncertainty within the same event j while η_j is the inter-event residual to represent event-to-event uncertainty that is shared by all sites, and σ_{ij} and τ_j are deterministic values describing the standard deviations of the intra-event and inter-event residuals, respectively. Here, ε_{ij} and η_j are standard normal random variables, i.e. zero mean and unit standard

deviation. While η_j is a constant at all sites for a given event, ε_{ij} varies from site to site, and has statistical dependence on those of other sites, i.e. ε_{kj}. This spatial correlation is often described by a selected auto-correlation model, which is given as a function of the distance between sites.

For PRA, the probability that loss of event exceeds a certain value l_i is

$$P(L \geq l_i) = \int P(L \geq l_i|\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} \tag{2.6}$$

where $\mathbf{y}$ represents the ground-motion intensities at sites of individual structures, and $f_{\mathbf{Y}}(\mathbf{y})$ is to describe the joint probability density function (PDF) of ground-motion intensities. This joint PDF of ground-motion intensities is obtained by combining source model and appropriate GMPE; while the former forecasts a random event of earthquake based on preliminary research on seismicity for the region, the latter predicts attenuated ground-motion intensities at sites under the occurred earthquake. Regarding this manner, (2.6) can be rewritten as

$$P(L \geq l_i) = \sum_{s=1}^{S} \left[\lambda_s \int P(L \geq l_i|m, \boldsymbol{r}, \eta, \boldsymbol{\varepsilon}) f_{\boldsymbol{X}|s}(m, r, \eta, \boldsymbol{\varepsilon}) dm d\boldsymbol{r} d\eta d\boldsymbol{\varepsilon} \right] \tag{2.7}$$

where λ_s is the annual occurrence rate of the event at source s, for $s = 1, \dots, S$, and $f_{\boldsymbol{X},s}(\cdot)$ is corresponding joint PDF of magnitude m, source-to-site distance r, inter-event residual η, and intra-event residual $\boldsymbol{\varepsilon}$ for s-th source.

To apply MCS-based approach on PRA, the conditional joint PDF $f_{\boldsymbol{X}|s}(\cdot)$ for source s can be substituted by its marginalized joint PDF $f_{\boldsymbol{X}}(\cdot)$ using total probability theorem with respect to seismic source

$$P(L \geq l_i) = \left(\sum_{s=1}^{S} \lambda_s \right) \cdot \int P(L \geq l_i|m, \boldsymbol{r}, \eta, \boldsymbol{\varepsilon}) f_{\boldsymbol{X}}(m, \boldsymbol{r}, \eta, \boldsymbol{\varepsilon}) dm d\boldsymbol{r} d\eta d\boldsymbol{\varepsilon} \tag{2.8}$$

This marginalized PDF $f_{\boldsymbol{X}}(\cdot)$ in (2.8) is exactly the mixture model of each conditional PDF with normalized weight of $\lambda_s / (\sum_{w=1}^{W} \lambda_w)$

$$f_{\boldsymbol{X}}(m, r, \eta, \boldsymbol{\varepsilon}) = \sum_{s=1}^{S} \frac{\lambda_s}{(\sum_{w=1}^{W} \lambda_w)} \cdot f_{\boldsymbol{X}|s}(m, \boldsymbol{r}, \eta, \boldsymbol{\varepsilon}) = \sum_{s=1}^{S} q_s \cdot f_{\boldsymbol{X}|s}(m, \boldsymbol{r}, \eta, \boldsymbol{\varepsilon}) \tag{2.9}$$

where q_s is an weight of source s, $\lambda_s / \left(\sum_{t=1}^{S} \lambda_t \right)$, that is normalized occurrence rate of s-th source. To simulate earthquake event from this mixture model, it is useful to adopt indicator random variables $\boldsymbol{z} = \{z_1, z_2, \dots, z_S\}$ of which only one element z_s is one and else are zero with probability of q_s. Considering the dependence between random variables in (2.7) ~ (2.9), the PDF $f_{\boldsymbol{X}|s}(\cdot)$ becomes

$$f_{\boldsymbol{X}|s}(m,r,\eta,\boldsymbol{\varepsilon}) = f_{R,M|s}(r|m)f_{M|s}(m)f_{\mathrm{H}}(\eta)f_{\mathbf{E}}(\boldsymbol{\varepsilon}), \tag{2.10}$$

so that (2.10) yields

$$f_{\boldsymbol{X}}(m,r,\eta,\boldsymbol{\varepsilon}) = \left[\sum_{s=1}^{S} q_s \cdot f_{R,M|s}(r|m)f_{M|s}(m)\right] \cdot f_{\mathrm{H}}(\eta)f_{\mathbf{E}}(\boldsymbol{\varepsilon}). \tag{2.11}$$

Please remind that residuals in above equations are not source-dependent variables; that is because usual empirical GMPEs have already achieved homoscedasticity in their models. Since large magnitude of the event tends to yield mega-size of rupture on the crust, it can be comprehended that source-to-site distance r is a causal variable of magnitude m although it is random. For example, Joyner-Boore distance r_{jb} is typical parameter in many GMPEs to define source-to-site distance, and this input value represents the shortest distance from site to surface projection of rupture plane on the ground. Therefore, it is necessary to reflect this causality from event-magnitude on the risk analysis while still considering spatial correlations between intra-event residuals at different sites.

Finally, the probability of exceedance for a certain value can be estimated by brute-force MCS-based approach

$$P(L \geq l_i) \cong \left(\sum_{s=1}^{S} \lambda_s\right) \cdot \frac{1}{N} \sum_{n=1}^{N} I(L \geq l_i), \tag{2.12}$$

where $I(L \geq l_i)$ is indicator function which is one if the loss with given realization of random variables exceeds a certain value l_i, otherwise zero. It is necessary to realize loss/damage state of individual structures based on simulation-based approach because, as long as the probabilistic concept is adopted, the consequence of hazard would be different even after the same event takes place.

To analysis seismic risk for a region where it is expected that earthquake events is rarely occurred, -e.g, stable continental tectonic area-, it is prone to adopt an earthquake scenario -scenario-based approach- that is a rarely probable but still has a severe effect on that area to establish a risk mitigation plan or emergency action plan instead of considering all of the possible events -event-based approach. For scenario-based seismic risk analysis, (2.7) turns to be more simplistic form as (2.13).

$$P(L \geq l_i) = \int P(L \geq l_i|\eta,\boldsymbol{\varepsilon}) f_{\mathrm{H}}(\eta) f_{\mathbf{E}}(\boldsymbol{\varepsilon}) d\eta d\boldsymbol{\varepsilon}. \tag{2.13}$$

2.3. Concurrent adaptive importance sampling

Risk assessment of hazard for a given urban area requires to attain a loss profile with its corresponding probability, but applying brute-force MCS to obtain multiple probabilities of loss has disadvantage when a rare (but severe) scenario is included. Therefore, to apply IS into risk analysis for the system which has multiple statuses of performance, it is good idea to find a best IS density that guarantees concurrently-well converging estimators for probabilities of corresponding events.

For multiple events of reliability analysis, a vector of functions, $\mathbf{H}(\mathbf{x})$ can be introduced, and its IS formulation can be written for the numerical integration in (2.14), i.e.

$$\mathbf{I_t} = \begin{Bmatrix} I_{t,1} \\ I_{t,2} \\ \vdots \\ I_{t,K} \end{Bmatrix} = \int_{\mathbf{x}} \mathbf{H}(\mathbf{x}) \frac{f(\mathbf{x})}{h(\mathbf{x};\mathbf{w})} h(\mathbf{x};\mathbf{v}) d\mathbf{x} = \int_{\mathbf{x}} \left[\begin{Bmatrix} H_1(\mathbf{x}) \\ H_2(\mathbf{x}) \\ \vdots \\ H_K(\mathbf{x}) \end{Bmatrix} \frac{f(\mathbf{x})}{h(\mathbf{x};\mathbf{w})} \right] h(\mathbf{x};\mathbf{w}) d\mathbf{x} \qquad (2.14)$$

Now, the optimal IS density function is defined as the one providing the least value for the sum of the squares of the coefficient of variations (c.o.v) of the elements in $\mathbf{I_t}$. The c.o.v is used instead of the variance because the elements in $\mathbf{I_t}$ can have different scales especially when the scenarios have different level of probabilities. With similar procedure to obtain (2.3), the optimal IS density $p^*(\mathbf{x})$ is now derived as (2.15) to handle the system with multiple consequences.

$$p^*(\mathbf{x}) = \frac{|\Psi(\mathbf{x})| f(\mathbf{x};\mathbf{u})}{\int |\Psi(\mathbf{x})| f(\mathbf{x};\mathbf{u}) d\mathbf{x}} \text{ where } \Psi(\mathbf{x}) = \sqrt{\sum_{k=1}^{K} [\varpi_k \cdot H_k(\mathbf{x})]^2} \qquad (2.15)$$

where ϖ_k is the k-th element of a unit vector, which is proportional to the inverse of the estimated $I_{t,k}$, $k = 1, \ldots, K$. Although the exact calculation of the optimal $\varpi_\mathbf{k}$ requires the final (unknown) outputs $I_{t,k}$, reasonable estimates would be possible with small number of pre-samples for efficient sampling. It may be more suitable when a simulation provides a conditional probability for each systemic event with certain random variables; in this study, $H_k(\mathbf{x})$, is usually defined as the conditional probabilities of the k-th events with given outcomes of ground-motion intensities.

In final, CE-CAIS is adaptable for the purpose of reliability analysis for multiple outcomes of system instead of the original CE-CAIS. Employing GM as non-parametric distribution to obtain near-optimal IS density, substitute $\Psi(\mathbf{x})$ in

(2.6) into $H(\mathbf{x})$ in cross-entropy based updating rules, which are listed in (2B.12 – 2B.15) of Appendix 2B.

2.4. Post-hazard Traffic Flow Capacity of Hypothetical Road Network

Figure 2.1 illustrates a hypothetical traffic network, modified from Lee et al. (2011), Ukkusuri (2005), and Kim (2007), surrounded by 9 active faults; geometric and seismic properties of active faults are from Field et al. (2013) and modified to facilitate numerical tests in this example. Figure 2.1 (c) shows the details of the network visualizing the locations of the bridges. For simplicity, it is assumed that damage of bridges is the only factor that decreases the corresponding link's traffic capacity, which eventually cut down the traffic flow capacity of the network between the source and terminal nodes. The fragility model of each bridge was adopted from FEMA-443 (2003), and the bridge types were randomly selected. The reduced link capacity is assumed to be 100%, 75% and 50% of full capacity, which signifies "none or slight", "moderate or extensive", and "collapse" damage states, respectively while reduced capacity in bridge 6, 10 is 100%, 75% , 50%, 25%, and 0% for each of the five separate states. Spectral accelerations at the natural period 1.0 sec are required to compute the conditional probability of damage as in FEMA-443 (2003); Magnitude-scaling relationship in Shaw (2009) with parameters used in appendix E of Field et al. (2013), and GMPE in Boore and Atkinson (2008) are used in this example. It is assumed that the decision-maker is interested in the reduced flow capacity from nodes {13, 23, 24} to nodes {2, 6, 7} after an earthquake event; this multi-sources/sinks problem is handled by introducing two subjunctive node and link (dashed lines); node 100 is assigned as a master source to nodes {13, 23, 24}, and node 200 as a mater sink from nodes {2, 6, 7} in this example.

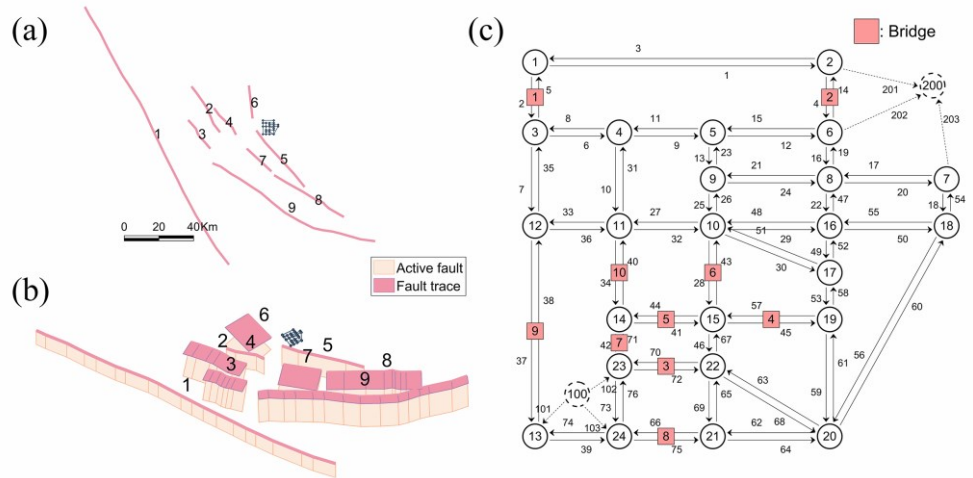

Figure 2.1. Hypothetical traffic network facing with 9 active faults

2.4.1. Matrix-based System Reliability Method

A sophisticated matrix-operation based approach, which is also known as Matrix-based system reliability method (MSR), is adopted to facilitate an efficient reliability analysis for this hypothetical traffic network. Consider a system reliability problem of which i-th component has d_i distinct states, $i = 1, \dots, I$ where I is the number of components. Then, the system event expressed by combinations of component events can be divided into $N = \prod_i^I d_i$ number of MECE events, denoted by e_j, $j = 1, \dots, N$. One can construct the probability vector $\mathbf{p}$ for each MECE event by iterative matrix procedure.

$$\mathbf{p}_{[1]} = \begin{bmatrix} P_{1(1)} \\ P_{1(2)} \\ \vdots \\ P_{1(d_1)} \end{bmatrix}, \text{ and } \mathbf{p}_{[i]} = \begin{bmatrix} \mathbf{p}_{[i-1]} \cdot P_{i(1)} \\ \mathbf{p}_{[i-1]} \cdot P_{i(2)} \\ \vdots \\ \mathbf{p}_{[i-1]} \cdot P_{i(d_i)} \end{bmatrix}, \quad i = 1, 2, \quad \dots, \quad I \tag{2.16}$$

where i-th component can have distinct d_i state and $P_{i(j)}$ denotes the probability that ith component is on the jth state, $j = 1, \dots, d_i$. From the probability vector $\mathbf{p}$, the probability of system failure can be calculated by collecting the MECE events which yields failure event and by summing up the probabilities of these MECE events. From the previous researches on MSR method in Lee et al. (2011), Kang et al. (2008), Song and Kang (2009), and Kang et al (2012), this collecting and summing up procedure can be replaced by inner product of event vector $\mathbf{c}$ with probability vector $\mathbf{p}$ for the sake of simplicity.

$$P(E_{sys}) = \sum_{j:e_j \in \{E_{sys}\}} \mathbf{p}_{[I]}(j) \text{ or } \mathbf{c}^{\mathrm{T}}\mathbf{p}_{[I]} \tag{2.17}$$

where e_j is the MECE component-level event that results the system event E_{sys}. In the case of dependent status (or failure) between components, one can achieve conditional independence between component events given outcomes of the random variables representing the sources of 'environmental dependence', or 'common source effects' in Kang et al. (2008), 'common source random variables (CSRVs)' in Song and Kang (2009). Thus, the total probability vector $P(E_{sys})$ is calculated by integrating conditional probability vector $P(E_{sys}|\boldsymbol{\zeta})$ with respect to CSRVs, $\boldsymbol{\zeta}$, and their joint PDF. (2.18) represents CSRVs concept and the corresponding sampling-based description.

$$P(E_{sys}) = \int_{\mathbf{s}} P(E_{sys}|\boldsymbol{\zeta}) f_{\mathbf{Z}}(\boldsymbol{\zeta}) d\boldsymbol{\zeta} \approx \frac{1}{N}\sum_{n=1}^{N} P(E_{sys}|\boldsymbol{\zeta}_n) \tag{2.18}$$

In addition, MSR method provides a way to compute the sensitivity of the system reliability with respect to its parameters. See Lee et al. (2011), Kang et al. (2008), Song and Kang (2009), and Kang et al. (2009) for more details.

2.4.2. Procedure of efficient sampling

Since we accepted MSR for network reliability, deterministic network analysis was preceded on main procedure of reliability analysis for all combinations of component-level event to identify the system events and their corresponding network performance. Thus, equation (2.6) is modified to form of probability mass function (PMF) as the following (2.19). To apply MSR to (2.6), $P(E_{sys})$ in (2.18) is corresponding to $P(L = l_i)$ in (2.19), which means post-hazard performance of network is decreased to l_i, where CSRVs $\boldsymbol{\zeta}$ in (2.18) is a ground-motion intensities vector $\mathbf{y}$ in (2.19).

$$P(L = l_i) = \int P(L = l_i|\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} \approx \frac{1}{N}\sum_{n=1}^{N} P(L = l_i|\mathbf{y}_n) \tag{2.19}$$

As a basic step to implement CE-CAIS in reliability analysis, it is recommended to perform the simulation on the uncorrelated standard normal space to avoid scaling and suitability issues. Appendix A demonstrates how to transform these random seismic source variables into uncorrelated standard normal variables in detail. According to appendix A, the conditional probability density function of

random seismic source variables on seismic source s has a relationship with PDF of uncorrelated standard normal variables $\mathbf{u}$ as (2.20).

$$f_{\boldsymbol{X}|s}(m, \boldsymbol{r}, \eta, \boldsymbol{\varepsilon}) d\boldsymbol{X} = f_{\boldsymbol{U}}(\mathbf{u}) d\mathbf{u} \tag{2.20}$$

where $f_{\boldsymbol{U}}(\cdot)$ is PDF of uncorrelated standard normal variables. By substituting (2.20) into (2.19), PMF of post-hazard network performance becomes (2.21).

$$P(L = l_i) = \left(\sum_{s=1}^{S} \lambda_s \right) \cdot \int \left[\sum_{t=1}^{S} q_t \cdot P_t(L = l_i | \mathbf{u}) \right] f_{\boldsymbol{U}}(\mathbf{u}) d\mathbf{u} \tag{2.21}$$

where q_t is an weight of source t, $\lambda_t / \left(\sum_{s=1}^{S} \lambda_s \right)$, that is normalized occurrence rate of t-th source again. $P_t(L = l_i | \mathbf{u})$ is conditional probability where status of the system is at l_i with given CSRVs $\mathbf{u}$ as a consequence of earthquake event that is occurred on source t. MSR method provides a comfortable way to obtain this system probability of event through matrix operation. For a consistency with CE-CAIS approach, $\sum_{t=1}^{S} q_t \cdot P_t(\cdot)$ in (2.21) is interpreted as $\Psi_k(\cdot)$ in (2.15). In final, (2.21) is modified to (2.22) formulating the IS description with distribution parameters $\mathbf{w}$. By employing GM as non-parametric distribution for CE-CAIS, parameters of IS density are relative weight of π_o, mean vector of $\boldsymbol{\mu}_o$ and covariance matrix of $\boldsymbol{\Sigma}_o$ for o-th Normal distribution, where $0 \leq \pi_o \leq 1$, $\sum_{o=1}^{O} \pi_o = 1$, and $o = 1, \ldots, O$ in (2B.11); this method is named as CE-CAIS-GM from this sub-section.

$$P(L = l_i) = \left(\sum_{s=1}^{S} \lambda_s \right) \cdot \int \left\{ \left[\sum_{t=1}^{S} q_t \cdot P_t(L = l_i | \mathbf{u}) \right] \frac{f_{\boldsymbol{U}}(\mathbf{u})}{h_{\boldsymbol{U}}(\mathbf{u}; \mathbf{w})} \right\} h_{\boldsymbol{U}}(\mathbf{u}; \mathbf{w}) d\mathbf{u} \tag{2.22}$$

The whole procedure in this section is summarized as follows:

Step 1: Deterministic maximum flow analyses to identify possible network performance scenarios (in terms of reduced flow capacity) and the corresponding component damage combinations.

Step 2: First pre-sampling to estimate the weights ϖ_k in (2.6).

Step 3: Update IS density model using CE-AIS-GM with respect to (2.6); that is CE-CAIS-GM.

Step 4: Repeat Step 3 until converged.

Step 5: Final sampling using the near-optimal IS density found from Step 4.

2.4.3. Results

Before to listing results of our proposed reliability analysis, it is proper to

define what the term “brute-force MCS” in this subsection means since there are two ways to state brute-force MCS; first one is based on indicating function as (2.12), and the other is based on conditional sampling accepting MSR method as (2.21). At this *Results* section, the term “brute-force MCS” represents the later (2.21) except for cases with any additional explanation.

In Step1, a total of 38×52=164,025 combinations of damage states by 10 bridges result in only 11 possible network flow capacity values, i.e. $K = 11$ in (2.14) including non-damaged case of which flow capacity is 12,600 veh/hr. In Steps 2-4, a total of 7,000 pre-samples were enough to find a near-optimal IS density for overall convergence of the probabilities of 11 states – 1,000 samples in Step2, and every 1,000 samples for 5 updates in Step 3 and 4. Setting our target c.o.v as 0.01, the identified near optimal IS density achieved the convergence using only 47,896 samples. This is considered superb improvement in efficiency because brute-force MCS demanded 5,781,842 samples. See Table 2.1 and Figure 2.2 for the details of the results.

Since the original CE-AIS approach needs to find a different near-optimal IS density for each of 11 network states, the pre- and final-samplings are performed separately for each state. Therefore, the number of samples for CE-AIS in Table 2.1 is the sum of 11 runs in Table 2.2. As shown in the Table 2.1, CE-CAIS achieved the same level of c.o.v. with only 0.8% of the simulations by brute-force MCS and 24% of CE-AIS. Figure 2.2 and Table 2.3 confirms that the estimates converge to the same probabilities. Figure 2.3 confirms that the spatial correlation in intra-event residuals obviously influences on the results of risk assessment; for the systemic events with low probability, neglecting these correlations underestimates probability of occurrence. Figure 2.4 shows the convergence histories of the brute-force MCS, and Figure 2.5 is for the proposed CE-CAIS. The plots confirm that CE-CAIS converge much faster and evenly in spite of great variability in probabilities for multiple states.

For a more detailed discussion, the required number of samples N_s is evaluated by (2.23) to achieve target c.o.v. $\bar{\delta}$ to consider the case that the probabilities are evaluated by indicator function-based brute-force MCS as (2.12).

$$N_s = \frac{1 - P}{\bar{\delta}^2 \cdot P} \tag{2.23}$$

The probability P is the lowest probability among those of possible consequences, and target c.o.v is set as 0.01 which is the same with the other applications. By using the relationship in (2.23), the number of samples to achieve the same target level (c.o.v 0.01) is calculated as 531,904,894 with $P \approx 1.88 \times 10^{-5}$; P is not

multiplied with annual occurrence rate of earthquake, $\sum_{s=1}^{S} \lambda_s$ (= 0.151), yet.

Table 2.1. Number of samples required to achieve target c.o.v 0.01

	Brute-force MCS (indicator-based)	Brute-force MCS (conditional sampling)	CE-AIS-GM	CE-CAIS-GM
Pre-sampling	-	-	61,000	7,000
Final-sampling	531,904,894	5,781,842	136,780	40,896
Total	531,904,894	5,781,842	197,780	47,896

Table 2.2. Number of samples to obtain probabilities of events with original CE-AIS (c.o.v. 0.01)

Reduced flow capacity (veh/hr)	None	300	450	600	750	900
Pre-sampling	1,000	6,000	5,000	8,000	5,000	5,000
Final-sampling	-	12,355	5,463	10,299	19,134	7,036
Total	1,000	18,355	10,463	18,299	24,134	12,036
Reduced flow capacity (veh/hr)	1,200	1,500	1,800	2,100	2,400	Total
Pre-sampling	6,000	7,000	6,000	7,000	6,000	61,000
Final-sampling	10,680	13,374	8,335	16,063	34,041	136,780
Total	16,680	20,374	14,335	23,063	40,041	197,780

Table 2.3. Annual probabilites of reduced network traffic capacity

Reduced flow capacity (veh/hr)	None	300	450	600	750	900
Brute-force MCS	0.148	2.668×10^{-5}	2.465×10^{-4}	1.132×10^{-5}	0.615×10^{-5}	1.819×10^{-4}
CE-AIS-GM	0.151	2.717×10^{-5}	2.508×10^{-4}	1.141×10^{-5}	0.612×10^{-5}	1.845×10^{-4}
CE-CAIS-GM	0.148	2.647×10^{-5}	2.403×10^{-4}	1.130×10^{-5}	0.612×10^{-5}	1.802×10^{-4}
Reduced flow capacity (veh/hr)	1,200	1,500	1,800	2,100	2,400	
Brute-force MCS	0.659×10^{-5}	1.630×10^{-5}	5.717×10^{-5}	0.284×10^{-5}	1.734×10^{-5}	
CE-AIS-GM	0.665×10^{-5}	1.647×10^{-5}	5.872×10^{-5}	0.289×10^{-5}	1.732×10^{-5}	
CE-CAIS-GM	0.658×10^{-5}	1.628×10^{-5}	5.717×10^{-5}	0.284×10^{-5}	1.727×10^{-5}	

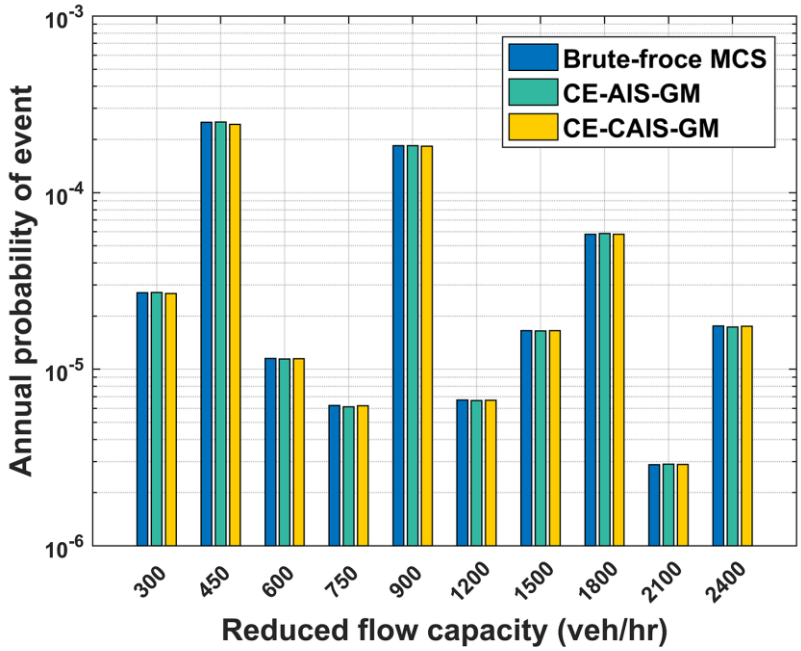

Figure 2.2. Annual probabilities by brute-force MCS/CE-AIS-GM/CE-CAIS-GM

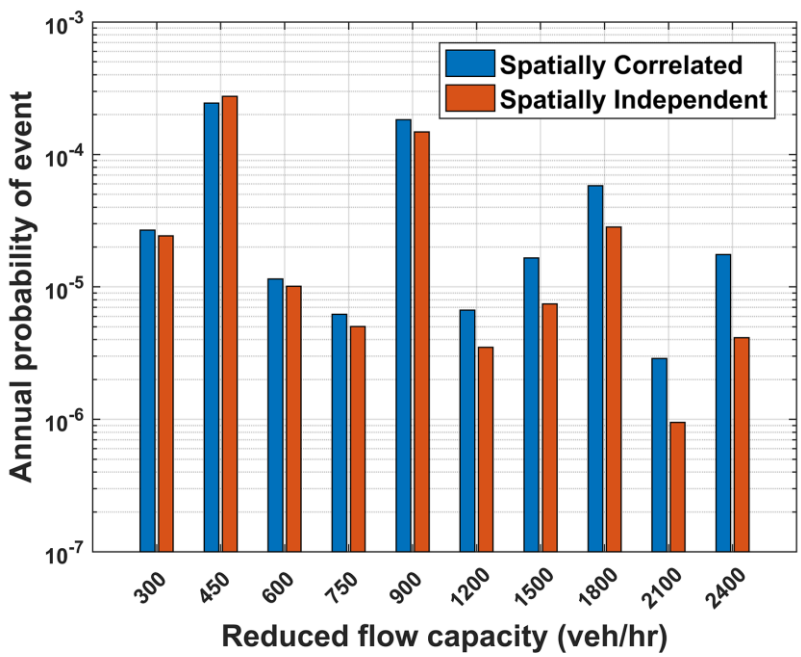

Figure 2.3. Effect of spatial correlation in intra-event residuals

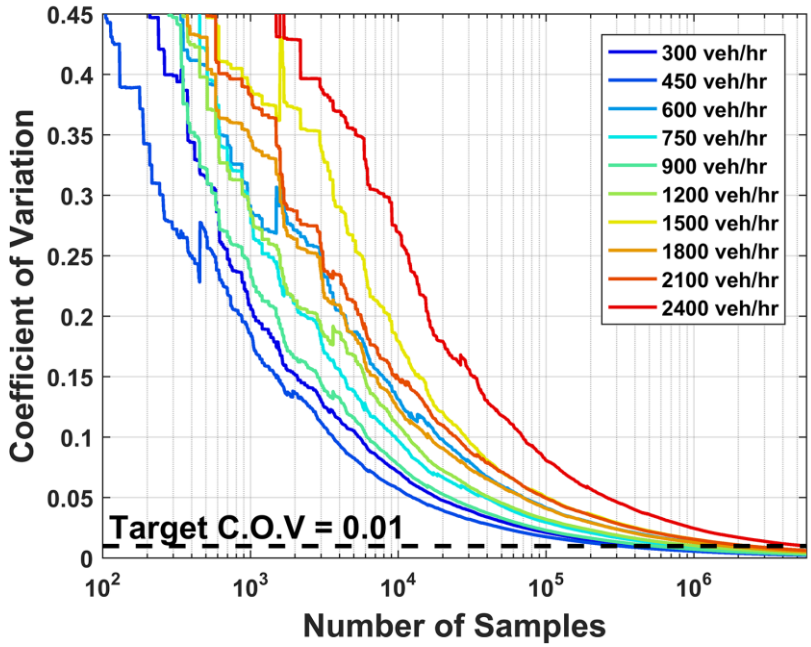

Figure 2.4. Convergence of brute-force MCS

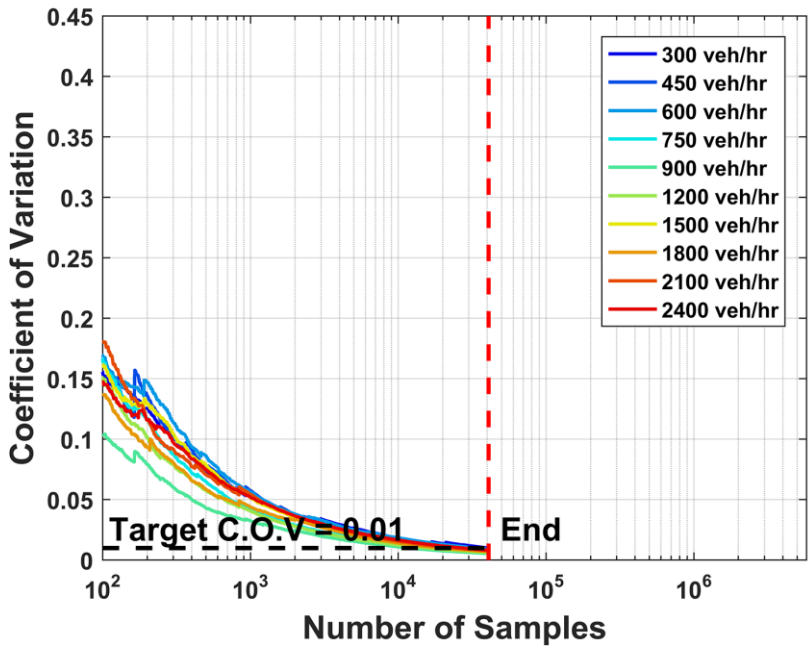

Figure 2.5. Convergence of CE-CAIS-GM

2.5. Dimensionality reduction techniques

To implement a more precise seismic risk analyse on a community, it is essential to appropriately consider intra-event residuals in GMPE. These spatially distributed random variables should be assigned to each site properly, so that it is feasible to describe ground-motion intensities over a region with consideration of the correlation among themselves. In many applications, a representative ground-motion intensity is allocated on each cell after target region is discretized into finite number of cells (but sizes of these cells are still small enough). However, the problem is that the dimensionality of these random variables is still too high, which makes it hard to apply CE-CAIS to regional seismic risk analysis. It is almost impossible to search for such a near-optimal importance sampling (IS) density on the high dimensional space, which is dissimilar with the case of low dimension. Therefore, it is good idea to reduce the dimensionality of random variables to facilitate IS for regional PRA.

2.5.1. Principle component analysis

Consider a linear relationship $\mathbf{A}(N \times R)$ that is connecting N-dimensional normally distributed random variables $\boldsymbol{\varepsilon} = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N\}^T (N \times 1)$, which has a significant correlation between themselves, e.g. intra-event residuals in this thesis, with R-dimensional Gaussian variables $\mathbf{e} = \{e_1, e_2, \dots, e_R\}^T (R \times 1)$ where R is much less than N and its error $\tilde{\boldsymbol{\varepsilon}} = \{\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \dots, \tilde{\varepsilon}_N\}^T$ $(N \times 1)$. If this error is negligible,

$$\boldsymbol{\varepsilon} = \mathbf{A}\,\mathbf{e} + \tilde{\boldsymbol{\varepsilon}} \approx \mathbf{A}\,\mathbf{e} \tag{2.24}$$

With assuming $\boldsymbol{\varepsilon}$ with zero-mean and unit standard deviation to neglect the scale of the variables, correlation matrix $\mathbf{R}_{\varepsilon\varepsilon}$ of $\boldsymbol{\varepsilon}$ and its eigen-decomposition is

$$\mathbf{R}_{\varepsilon\varepsilon} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}} \tag{2.25}$$

where $\boldsymbol{\Lambda}$ is diagonal matrix of which diagonal elements are descending ordered eigen-values while $\mathbf{U}$ is orthogonal matrix that consist of corresponding eigen-vectors. This eigen-decomposition, which so-called principal component analysis (PCA), guarantees to find orthogonal components according to scales of their variance.

To be consistent with the purpose of dimensionality reduction, linear relationship $\mathbf{A}$ and reduced variables $\mathbf{e}$ in (2.24) is represented through PCA

$$\mathbf{A} = \mathbf{U}_{(:,1\sim R)}\sqrt{\mathbf{\Lambda}_{(1\sim R,1\sim R)}},\ \text{and}\ e_i \sim N(0,1) \tag{2.26}$$

where $\mathbf{U}_{(:,1\sim R)}$ is $(N \times R)$ matrix of which elements are extracted from 1st to R-th column of orthogonal matrix $\mathbf{U}$, and $\sqrt{\mathbf{\Lambda}_{(1\sim R,1\sim R)}}$ is the square root of diagonal matrix $\mathbf{\Lambda}$ up to its R-th eigen-components. The e_i is the i-th element of $\mathbf{e}$, and it is uncorrelated standard normal random variables. What is noteworthy about this eigen-mapping is that it is the best linear transformation in the senses of least squared sum of error between original variables and reconstructed variables.

Although PCA provides an effective dimensionality reduction technique in the least squared sense, the risk analysis on the reduced space loses the information about uncertainties since the variance of approximated ε_i by (2.24) becomes smaller than the original one. For the large scale problems, like regional PRA, this effect is obvious since negligible errors at individual structure-level are significantly cumulated; with a little different point of view, the loss of the information about uncertainties in the hazard means that less diverse ground-motion scenarios is considered. Therefore, it is necessary to compensate this loss when exploiting PCA on full scaled regional risk analysis, and the following calibration technique is introduced.

The fragility model is just to describe how likely the structure is damaged from a given hazard intensity. For this purpose, conditional probability of exceeding k-th damage state (ds_k) is represented with resisting capacity A_m of the structure against given hazard-intensity Y. Since it is common for point (node) type structures -for example, building, bridge, and station- to adopt fragility model under assumption of log-normal distribution, the fragility curve is:

$$P(DS > ds_k|Y) = \Phi\left(\frac{\ln Y - \ln A_m}{\beta_{DS}}\right) \tag{2.27}$$

where $\Phi(\cdot)$ is cumulative density function of standard normal distribution, and β_{DS} is randomness of structural capacity. For PRA, the fragility can be expressed with using GMPE neglecting the subscribed indices for the simplicity. This standard fragility is modified to refill the lost information by dealing with neglected parts in PCA as independent random variables as (2.28).

$$\begin{aligned} P(DS > ds_k|Y) &= P\left(z \le \frac{\ln Y - \ln A_m}{\beta_{DS}}\right) \\ &= P\left(z \le \frac{f(M,R,\lambda) + \sigma\varepsilon + \tau\eta - \ln A_m}{\beta_{DS}}\right) \end{aligned} \tag{2.28}$$

$$= P\left(z \le \frac{f(M,R,\lambda) + \sigma\bar{\tilde{\varepsilon}} + \tau\eta + \sigma\tilde{\varepsilon} - \ln A_m}{\beta_{DS}}\right)$$
$$= P\left(z \le \frac{\overline{\overline{\ln Y}} + \sigma\tilde{\varepsilon} - \ln A_m}{\beta_{DS}}\right)$$
$$= \Phi\left(\frac{\overline{\overline{\ln Y}} - \ln A_m}{\sqrt{\beta_{DS}^2 + \sigma^2 \cdot \mathrm{Var}[\tilde{\varepsilon}]}}\right)$$

where z is standard normal random variable according to the definition of cumulative density function $\Phi(\cdot)$. $\bar{\tilde{\varepsilon}}$ is the reconstructed intra-event residual through PCA, which is equal to $\mathbf{A} \cdot \mathbf{e}$ in (2.24), and $\tilde{\varepsilon}$ is the error between original variable ε and its reconstructed one $\bar{\tilde{\varepsilon}}$. $\overline{\overline{\ln Y}}$ is ground-motion intensity at the site with considering its aleatory uncertainties under the reduced dimension. One can analytically obtain the variance of $\tilde{\varepsilon}$, $\mathrm{Var}[\tilde{\varepsilon}]$, with using the relationship of $\mathbf{R}_{\varepsilon\varepsilon} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}$, i.e, $\mathrm{Var}[\tilde{\varepsilon}]$ is a corresponding diagonal component of $\left[\mathbf{R}_{\varepsilon\varepsilon} - \mathbf{U}_{(:,1\sim R)}\boldsymbol{\Lambda}_{(1\sim R,1\sim R)}\mathbf{U}^{\mathrm{T}}_{(:,1\sim R)}\right]$.

2.5.2. Central limit theorem

For regional PRA, one of the most practical measures for regional damage is evaluated by regionally aggregated monetary loss over an urban area to repair structural or non-structural damage from the earthquake event. It is common to assume the damage of spatially distributed individual structures is statistically independent neglecting the dependence in capacities between structures when ground-motion intensities are given outcomes of realization, which is as known as the ground motion field. On the extension of this assumption, central limit theorem (CLT) is applicable conditionally to regionally summated repair costs on individual buildings or infrastructures with a realization of ground-motion field.

Consider independent random variables $X_1, X_2, \ldots, X_N$ where *i*-th random variable X_i has mean of μ_i and standard deviation of σ_i; it is not necessary for random variables to be identically distributed with significant size of N, but moment of $|X_i|$ should exist with order of larger than 2. CLT imposes that the sum of these variables follows normal distribution asymptotically with sufficiently ncreasing number of N

$$\frac{1}{(\sum_{i=1}^{N}\sigma_i^2)}\sum_{i=1}^{N}(X_i-\mu_i)\sim N(0,1)\ \text{when}\ n\to\infty \tag{2.29}$$

Although this version of CLT is weaker version than classical CLT that assumes independent and identical distribution of X_i, this expression is still effective to implement community-level PRA except for the only case that replacement costs is extraordinarily expensive for a only few buildings/infrastructures. Thus, the conditional probability that the aggregated loss exceeds certain loss level is effectively approximated with drastically small order of differences

$$P(L \geq l_i|\mathbf{y}) \cong 1-\Phi\left(\frac{l_i-\mu_{L|\mathbf{y}}}{\sigma_{L|\mathbf{y}}}\right) \tag{2.30}$$

where $\mu_{L|\mathbf{y}}$ is conditional mean of regional loss while $\sigma_{L|\mathbf{y}}$ is conditional standard deviation with a given ground-motion field. This conditional central moments $\mu_{L|\mathbf{y}}$ and $\sigma_{L|\mathbf{y}}$ are conveniently evaluated through simple operation with mean and standard deviation of each loss. What is better to adopt CLT is that, when following simulation-based approach, it is not required to generate random numbers to indicate individual damage states and losses of spatially distributed structures. Hence, this dimensionality reduction technique with CLT itself improves the efficiency of simulation; for example, when there are amounts of buildings and infrastructures in the target community, large scale of random number generation is replaced by simple calculation to obtain sum of individual mean values and variances.

2.6. Aggregated regional monetary loss in Shelby County

A regional risk analysis on Shelby County is introduced with scenario-based approach that assumes point source of moment magnitude M_w 7.9 event at Blytheville, AR (35.927ºN, 89.919ºW). The source of earthquake event is selected by reflecting patterns of seismicity in the New Madrid Seismic Zone. Data for 305,694 buildings on Shelby County was drawn from default data set embedded in open-source software ERGO from National Center for Supercomputing Application (NCSA) after eliminating data of buildings that are matched with abnormal fragility. This data includes locations, structural type, replacement cost with respect to {structural}/{drift-sensitive non-structural}/{acceleration-sensitive non-structural} damage, and other information about buildings. The fragility model

for these buildings is also exported from the same source, and the uncertainties on structural type are not considered in this application, Figure 2.6 illustrates the seismic source and building assets in Shelby County. GMPE suggested by Atkinson and Boore (1995) is selected for this numerical example, but the residual model is partially adopted from Boore and Atkinson (2008) since Atkinson and Boore (1995) did not distinguish the contributions by intra- and inter- residual. Either attenuated peak ground acceleration (PGA) or spectral acceleration at 0.2sec ($Sa_{0.2sec}$) is required according to structural type and its fragility as defined in Steelman et al. (2007). The target area (Shelby County) is divided into 5,267 cells of which each size is 500m×500m with grids to assign representative ground-motion intensities for each cell. Since there are 4,487 cells of which assets not only PGA but also $Sa_{0.2sec}$ as inputs for their fragility models (structural damage), the total 9,754 numbers of random variables is demanded to describe intra-event residuals. The correlation between inter-event residuals for PGA and $Sa_{0.2sec}$ is set to 0.6 based on subjective determination by authors while the spatial correlation for intra-event residuals is adopted from Loth and Baker (2011) to consider the spatial correlations between different intensities, PGA at site i and $Sa_{0.2sec}$ at site j. All the soil properties at sites are assumed to be $Vs_{30} = 760\mathrm{m/s}$ for the simplicity. However, our framework is still applicable to different model parameters either for correlations of residuals or for characteristics of sites.

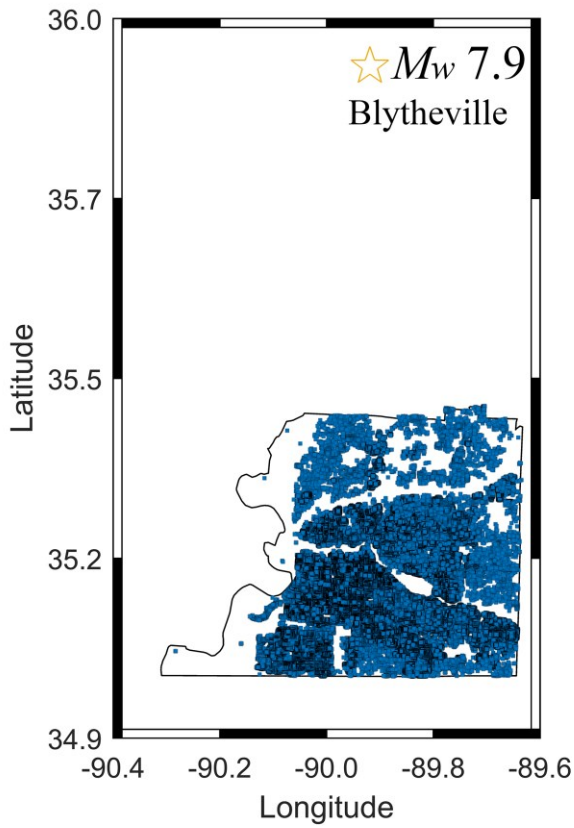

Figure 2.6. Scenario earthquake event (M_w 7.9; Blytheville)

2.6.1. Damage factor and Loss Estimation for individual building

We accepted the concept of damage factor which was introduced in Bai et al. (2009) to match fragility curves with monetary loss. Then, the mean value of direct economic loss for individual building μ_{L_i} is calculated as (2.31).

$$\mu_{L_i} = M_i \sum_{j=1}^{3} \alpha_j \mu_j \quad \text{where } \mu_j = \sum_{k=1}^{4} P(DS = ds_k) \mu_{DF_{k,j}} \tag{2.31}$$

where M_i is replacement cost for i-th building, and α_j is ratio of replacement cost of damage type j to total cost M_i while μ_j is mean damage factor for damage type j. Since we do not consider loss of contents in the building due to lack of data, there are three damage types: {structural}, {acceleration-sensitive non-structural}, and {drift-sensitive non-structural}; thus, $j = 1, 2, 3$. To calculate

μ_j in (2.31), it is required to obtain $P(DS = ds_k)$ from fragility curve, and that is probability of k-th damage state for j-th damage type. $\mu_{DF_{k,j}}$ is the mean damage factor of damage type j when the damage state is indicated as k.

Variance of direct economic loss for individual building, $\sigma_{L_i}^2$, is also obtained without difficulty.

$$\sigma_{L_i}^2 = M_i^2 \sum_{j=1}^{3} \alpha_j^2 \sigma_j^2$$
$$\text{where } \sigma_j^2 = \sum_{k=1}^{4} \left[P(DS = ds_k) \cdot \left(\sigma_{DF_{k,j}}^2 + \mu_{DF_{k,j}}^2 \right) \right] - \mu_j^2 \quad (2.32)$$

where σ_j^2 is variance of damage factor for damage type j, and $\sigma_{DF_{k,j}}^2$ is variance of damage factor when damage state for type j is indicated as k. Steelman and Hajjar (2008) provides tables of $\mu_{DF_{k,j}}$ and $\sigma_{DF_{k,j}}$ with corresponding damage types and damage states as Table 2.4 and 2.5. The dependence between losses of different damage type is neglected.

Table 2.4. Mean values of damage factor

$\mu_{DF_{k,j}}$	Damage Type (j)			
Damage state (k)	Structural	Non-structural (Drift-sensitive)	Non-structural (Accel.-sensitive)	Contents
Insignificant	0.005	0.03	0.03	0.015
Moderate	0.155	0.18	0.13	0.09
Heavy	0.55	0.525	0.425	0.2625
Complete	0.9	0.875	0.825	0.4375

Table 2.5. Standard deviation values of damage factor

$\sigma_{DF_{k,j}}$	Damage Type (j)			
Damage state (k)	Structural	Non-structural (Drift-sensitive)	Non-structural (Accel.-sensitive)	Contents
Insignificant	0.002	0.02	0.02	0.01
Moderate	0.058	0.08	0.0467	0.04
Heavy	0.1	0.15	0.15	0.075
Complete	0.04	0.0833	0.117	0.0417

2.6.2. Procedure of efficient sampling

The magnitude and source-to-site distance are not uncertain anymore for a scenario-based risk analysis, so that the probability of loss exceeding a certain

value is

$$P(L \geq l_i) = \int P(L \geq l_i|\boldsymbol{\eta}, \boldsymbol{\varepsilon}) f_{\mathrm{H}}(\boldsymbol{\eta}) f_{\mathbf{E}}(\boldsymbol{\varepsilon}) d\boldsymbol{\eta} d\boldsymbol{\varepsilon} \tag{2.33}$$

where $\boldsymbol{\eta}$ is a vector of which elements are inter-event residuals of ground-motion intensities that are required as input of fragility models, i.e, η_{PGA} and $\eta_{\mathrm{Sa0.2sec}}$ in this paper, while $\boldsymbol{\varepsilon}$ is a vector of which element is intra-event residuals at each site.

After accepting PCA, the intra-event residual $\boldsymbol{\varepsilon}$ is described with essential principal components (PCs), and they are realized with uncorrelated standard normal random variables $\mathbf{e}$ $(\mathrm{R} \times 1)$; recall $\boldsymbol{\varepsilon} \approx \mathbf{A}\,\mathbf{e} = \mathbf{U}_{(:,1\sim R)}\sqrt{\boldsymbol{\Lambda}_{(1\sim R,1\sim R)}}\mathbf{e}$ and $e_i \sim N(0,1)$ from (2.24) and (2.26). After calibrating fragility models with our suggestion in (2.28), our risk analysis is formulated as:

$$P(L \geq l_i) = \int P(L \geq l_i|\boldsymbol{\eta}, \mathbf{e}) f_{\mathrm{H}}(\boldsymbol{\eta}) f_{\mathbf{E}}(\mathbf{e}) d\boldsymbol{\eta} d\mathbf{e} \tag{2.34}$$

Substituting (2.30) into (2.34), (2.34) becomes as (2.35)

$$P(L \geq l_i) = \int \left[1 - \Phi\left(\frac{l_i - \mu_{L|\boldsymbol{\eta},\mathbf{e}}}{\sigma_{L|\boldsymbol{\eta},\mathbf{e}}}\right)\right] f_{\mathrm{H}}(\boldsymbol{\eta}) f_{\mathbf{E}}(\mathbf{e}) d\boldsymbol{\eta} d\mathbf{e} \tag{2.35}$$

where $\mu_{L|\boldsymbol{\eta},\mathbf{e}}$ is summation of conditional mean of losses at individual buildings, and $\sigma_{L|\boldsymbol{\eta},\mathbf{e}}$ is squared root of sum of conditional variance of losses at individual buildings; the conditional mean and variance of loss at individual building can be evaluated by (2.31) and (2.32), respectively, with a given realization of $\boldsymbol{\eta}, \mathbf{e}$.

By using Cholesky decomposition, inter-event residual vector $\boldsymbol{\eta}$ is easily transformed into uncorrelated standard normal variables while $\mathbf{e}$ has already achieved uncorrelated standard normal distribution due to orthogonality in PCA.

$$f_{\mathrm{H}}(\boldsymbol{\eta}) f_{\mathbf{E}}(\mathbf{e}) d\boldsymbol{\eta} d\mathbf{e} = f_{\boldsymbol{U}}(\mathbf{u}) d\mathbf{u} \tag{2.36}$$

Thus, our final problem statement becomes

$$P(L = level_i) = \int \left[1 - \Phi\left(\frac{level_i - \mu_{L|\mathbf{u}}}{\sigma_{L|\mathbf{u}}}\right)\right] f_{\boldsymbol{U}}(\mathbf{u}) d\mathbf{u} \tag{2.37}$$

For a consistency with CE-CAIS approach, $1 - \Phi(\cdot)$ in (2.37) is interpreted as $H_k(\cdot)$ in (2.5). Again, this problem statement is based-on modified fragility model in (2.28) that is to refill the lost information on aleatory uncertainties. Finally, the IS formulation is (2.38).

$$P(L = level_i) = \int \left\{ \left[1 - \Phi\left(\frac{level_i - \mu_{L|\mathbf{u}}}{\sigma_{L|\mathbf{u}}} \right) \right] \frac{f_U(\mathbf{u})}{h_U(\mathbf{u};\mathbf{w})} \right\} h_U(\mathbf{u};\mathbf{w}) d\mathbf{u} \tag{2.38}$$

For the case of CE-CAIS-GM, there parameters are relative weight of π_o, mean vector of $\boldsymbol{\mu}_o$ and covariance matrix of $\boldsymbol{\Sigma}_o$ for o-th Normal distribution, where $0 \le \pi_o \le 1$, $\sum_{o=1}^{O} \pi_o = 1$, and $o = 1, \dots, O$ in (2B.11) as discuss in Section 2.4.

To implement seismic risk analysis on the reduced dimension, PCA is applied to intra-event residuals to eliminate (eigen-) components of which eigen-values are neglible. There are several ways to determine these non-informative components, and two representative approaches is (a) to include up to i-th PC, where the sum of eigen-values from first PC to i-th component occupies 85~95 % of whole summation and (b) just to truncate from $(i+1)$-th PC by setting the component that is corresponding to negligible eigen-value. In this paper, authors recommend second approach for the purpose of dimension reduction to implement CE-CAIS, and this recommendation is valid since we have an effective calibration strategy in (2.28).

The procedure of PRA focusing on aggregated monetary loss is summarized as follows:

Step 1: Construction of correlation matrix for intra-event residuals, and eigen-decomposition ment PCA.
Step 2: Truncation of eigen-components which do not have significant eigen-value, and reconstruction of correlation matrix to evaluate reduced variance of intra-event residuals at sites.
Step 3: Fragility Modification for individual structures according to (2.28).
Step 4: First pre-sampling to estimate the weights ϖ_k in (2.14) with CLT on the reduced random variables space that is obtained from Step 2.
Step 5: Update IS density model using CE-AIS-GM with respect to (2.14) also with CLT on the reduced random variables space as Step 4.
Step 6: Repeat Step 3 until converged.
Step 7: Final sampling using the near-optimal IS density found from Step 4 also with CLT on the reduced random variables space as Step 4~6.

2.6.3. Results

To start our discussion on results, we define the term “brute-force MCS” as the simulation-based approach with indicator function as (2.12) instead of conditional sampling in (2.37); that is different notation with previous sub-section *2.5.3*. Figure 2.7 illustrates eigen-values of corresponding PCs (eigen-components) that are sorted in a descending order. For this numerical example, authors truncated

components over than 40th PCs; in other word, only eigen-components up to 40 is considered to realize intra-event residuals for overall sites when applying dimension reduction technique with PCA. Several spatial profile of PCs are illustrated in Figure 2.8 with its relative amplitude that is proportional to the maximum value in eigen-components multiplied with squared root of its corresponding eigen-value. As observed in the figure 2.7 and 2.8, higher-order of components show negligible amplitudes and locally focused correlations. In Figure 2.8, the maximum amplitudes of each profile have relative value of 1 for 1st PC, 0.502 for 10th PC, and 0.146 for 100th PC. Since intra-event residuals both for PGA and $Sa_{0.2sec}$ have spatial correlation, each mode of PC has both elements of PGA and $Sa_{0.2sec}$; each PC is a $(9{,}754 \times 1)$ vector, 5,267 elements is for PGA while 4,487 is for $Sa_{0.2sec}$.

It tends to underestimate overall risk if the calibration strategy in (2.28) is not included in the procedure, and that is because realized hazard without significant residuals does not cover sufficient uncertainties in GMPE; again, reconstructed intra-event residuals only with (2.26) show reduced variance. The calibration strategy is to solve this problem works effectively by adjusting the fragility curve. To discuss the effect of calibration we proposed in (2.28), Figure 2.9 is provided. This figure shows the results by different sampling approaches with the same number of samples – 100,000; brute-force MCS with full residuals, MCS with residuals by only 40 PCs without calibration, MCS with residuals by 40 PCs with calibrations. The result demonstrates that calibration tactics recovers the information on uncertainties through modified fragility curve.

Next, Figure 2.10 and 2.11 are to check the applicability of CLT to approximate conditional loss curve, which is the exceedance probability curve for total regional monetary loss. Figure 2.10 represents a realized ground-motion fields both for PGA and $Sa_{0.2sec}$ with their spatial correlations; Figure 2.10 (a) is a realization for PGA, and Figure 2.10 (b) is the corresponding realization for $Sa_{0.2sec}$. For a given realization, the conditional loss curve is obtained by brute-force MCS with 100,000 samples, and the curve is approximated by CLT with two distribution parameters, mean and standard deviation. From Figure 11, both results show a great similarity, but CLT has a merits in a computational point; the only thing required to approximate conditional loss curve is only mean and standard deviation to employ CLT.

In our full application of CE-CAIS, PCA, and CLT, 100,000 pre-samples were generated to estimate the initial weights of ϖ_k in Step 4, and 25,000 pre-samples were used to update IS density model to obtain a near-optimal IS density - every 5,000 samples for 5 updates in Step 5 and 6; total 125,000 pre-samples are

generated. After simulating 100,000 final samples in Step 7, we are successful in obtaining loss curve that is almost identical to the result by a million number of brute-force MCS (Figure 2.12). Furthermore, this result by our suggestion is much more efficient than that by brute-force MCS since its c.o.v's are almost less than 0.03 (Figure 2.13). The result by brute-force MCS is not on a comparable level with our proposal.

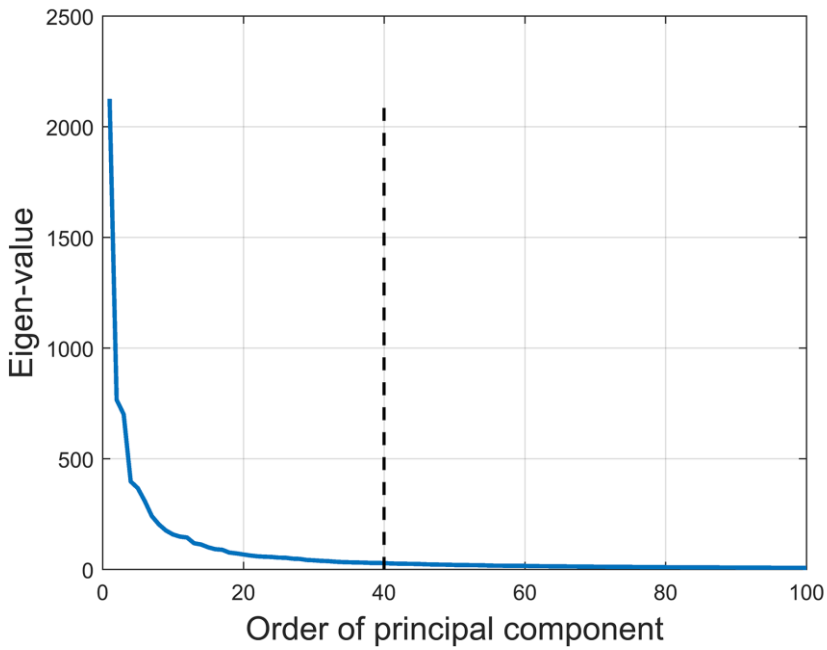

Figure 2.7. Eigen-values of principal components

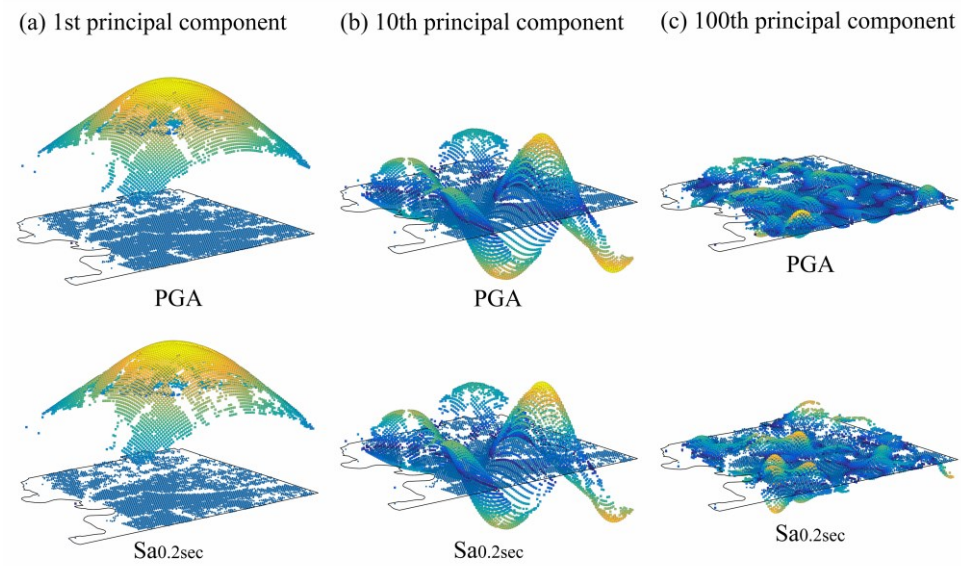

Figure 2.8. Spatial profile of principal components

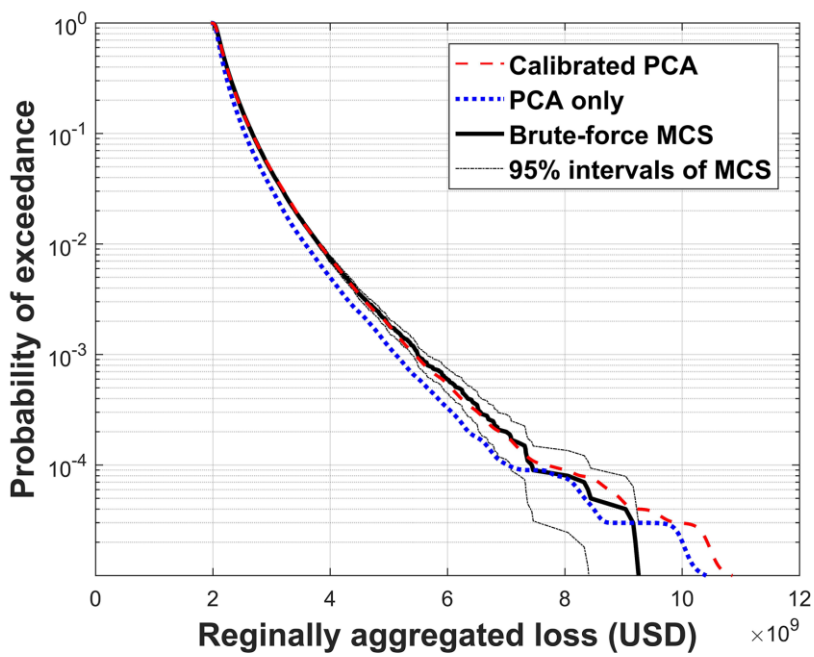

Figure 2.9. Effect of calibration strategy

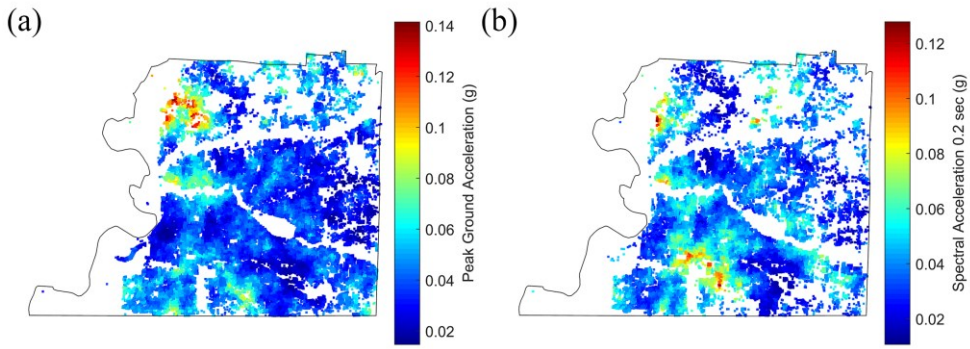

Figure 2.10. A realization of ground-motion intensities with spatial correlation

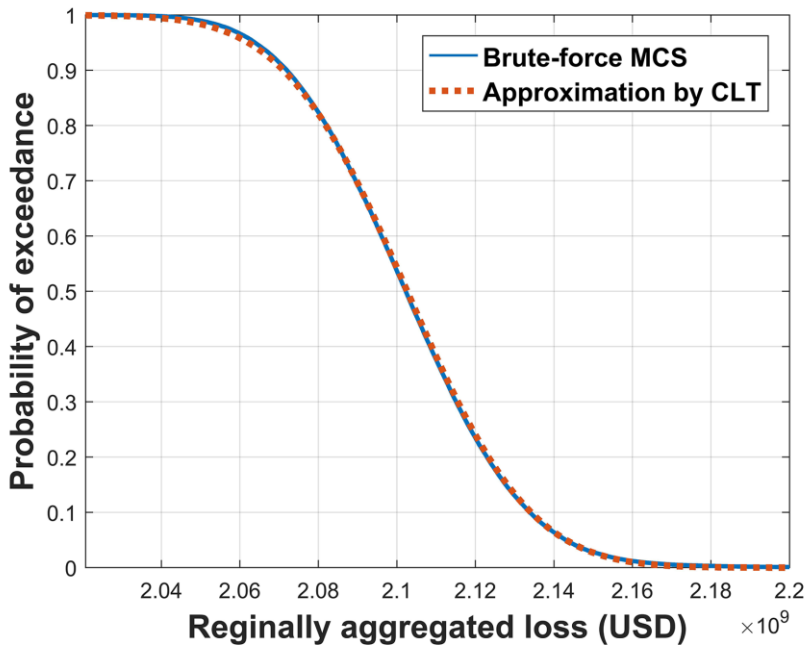

Figure 2.11. Conditional loss curve and its approximation by CLT

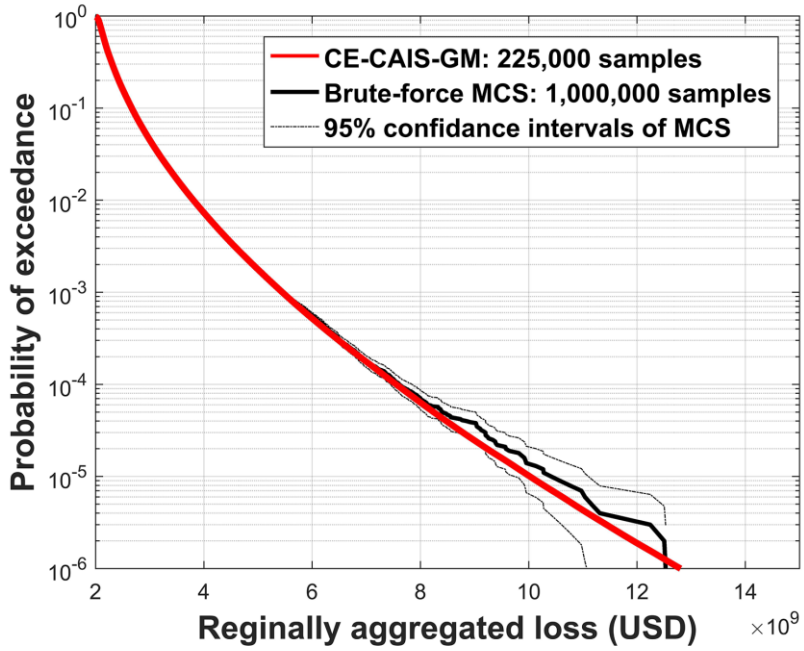

Figure 2.12. Comparison of results by brute-force MCS/CE-CAIS-GM

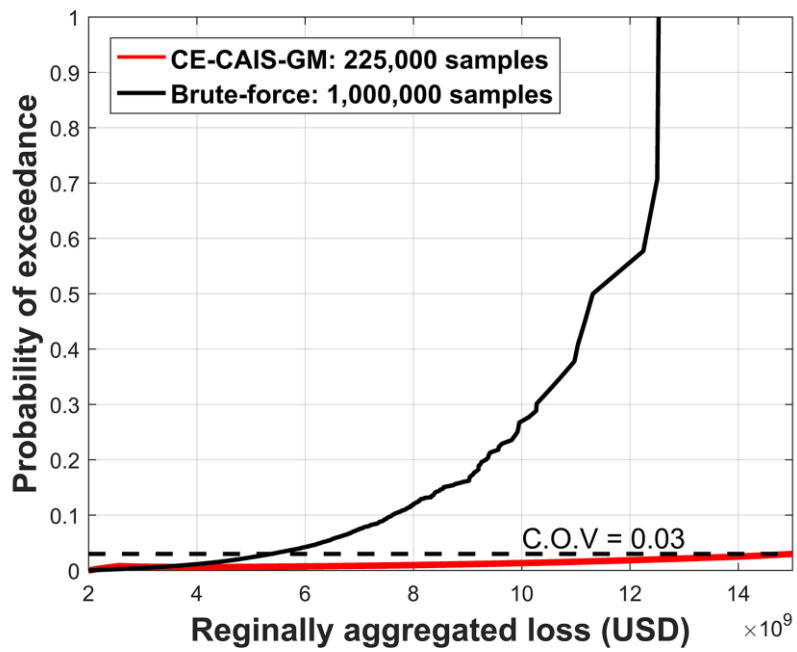

Figure 2.13. Coefficient of variation of exceedance probability

2.7. Conclusion

A new simulation-based method is proposed for community-level probabilistic seismic risk assessment (PRA) with two applications, (a) post-hazard traffic flow capacity and (b) aggregated regional monetary loss. The proposed approach, termed as cross-entropy-based concurrent adaptive importance sampling (CE-CAIS), improves the efficiency of the simulation-based method when multiple levels of probabilities need to be estimated. As a variation of CE-AIS, the proposed approach does not require prior knowledge, assumption or additional reliability analysis to identify important regions in selecting the importance sampling (IS) density. While CE-AIS needs to repeat pre-sampling to obtain a near-optimal IS density for each target event, CE-CAIS finds a near optimal IS density which provides fairly well and speedy convergence for all possible events, concurrently. This feature makes the proposed approach superior to the original CE-AIS approach for regional PRA because numerous post-hazard consequences often exist within probabilistic concepts.

Numerical examples demonstrated superb efficiency of CE-CAIS compared to brute-force MCS and original CE-AIS. In addition, several techniques were introduced to be plugged into a procedure to facilitate CE-CAIS in our applications: (a) matrix-based system reliability (MSR) method for the example of post-hazard traffic flow capacity and (b) principal component analysis (PCA) and central limit theorem (CLT) for the regionally aggregated monetary loss. Sampling-based MSR approach provides a way to obtain conditional system reliability of traffic network while PCA and CLT was used to reduce the dimensionality of random variables, so that CE-CAIS was successfully applied to numerical examples. With these additional techniques on main methodology, CE-CAIS found a near-optimal IS density to provide significant improvement in terms of computational cost not relying on any subjective judgement to select IS density model.

Appendix 2A. Simulation of Seismic Hazard on Uncorrelated Standard Normal Space

It is recommended to implement reliability analysis on uncorrelated standard normal space when exploiting cross-entropy based adaptive importance sampling (CE-AIS), especially for the cases to select Gaussian mixture as its importance

sampling (IS) density model. That is because 1) one can avoid scale-issue of individual variables and 2) the IS density model effectively mimics the optimal IS density when this reliability analysis is performed on uncorrelated standard normal distribution Kurtz and Song 2013. The same recommendation is also valid for cases of CE-AIS with using Von-mises Fisher mixture, which is originally developed to provide a method to deal with high-dimensional reliability problems Wang and Song 2016.

Consider a ground-motion prediction equation that was discussed in (2.7)

$$\ln Y_{ij} = f(M_j, R_{ij}, \lambda_i) + \sigma_{ij}\varepsilon_{ij} + \tau_j\eta_j \tag{2A.1}$$

The random variables in this equation are: (a) magnitude of earthquake, (b) source-to-site distance (c) intra-event residuals (d) inter-event residuals, and definitions are the same as those in (2.7). Since residuals are assumed to follow normal distribution, one can easily transform these residuals into uncorrelated standard normal variables by using Cholesky-decomposition. Thus, this section is going to introduce an effective way to realize ground-motion intensities from simple fault model with using uncorrelated standard normal variables only focusing on magnitude and source-to-site distance. However, we generate random numbers to describe rupture surface instead of source-to-site distance since source-to-site distance becomes deterministic for all sites with a realization of earthquake rupture.

2A.1. Magnitude-frequency relationship

There have been several models to describe magnitude-frequency distribution, which literally relates magnitude of earthquake with its occurrence rate for a specific source, e.g. Field et al. (2013), Cosentino et al. (1977). Since all of marginal cumulative density function (CDF) can be matched with CDF of other arbitrary random variable, it is natural to transform a non-normal random variable into standard normal variable according to its CDF value (Der Kiureghian 2005). Thus, the random magnitude of event on the seismic source, m, is represented by a single standard normal random variable u

$$m = {F_M}^{-1}(\Phi(u)). \tag{2A.1}$$

where ${F_M}^{-1}(\cdot)$ is inverse of CDF for random magnitude m while $\Phi(\cdot)$ is standard normal CDF. For the case of non-analytical CDF or discrete value of m, it is still valid to adopt table-look-up method to find the corresponding CDF values

by matching one with the other based on numerical evaluation.

2A.2. Realization of rupture surface

Simple fault assumes that earthquake rupture floats on the fault surface, and a rupture is also assumed to have maximum area up to that of active fault where the rupture is located. The relationship between magnitude of the event and area of rupture are described by magnitude-scaling relationship. In this section, we representatively introduce the suggestion by Field et al. 2013 in (2A.2), which is employed in main text, but other proposals are still applicable to the framework we suggested in this thesis.

$$M = \log_{10} A + \frac{2}{3}\log_{10}\frac{\max(1,\ \sqrt{\frac{A}{W^2}})}{(1+\max\left(1,\ \frac{A}{W^2\beta}\right))/2} + const \tag{2A.2}$$

Following Field et al. (2013), the rupture is assumed to have the same width W equals length L for small earthquakes, but the length L follows equation (2A.3) when width W saturates the downdip width W^* for large earthquakes.

$$L = \begin{cases} W & W < W^* \\ \dfrac{A}{W^*} & W = W^* \end{cases} \tag{2A.3}$$

By using magnitude-scaling relationship, the area of rupture becomes deterministic with a given magnitude value. Thus, the only remained part to realize earthquake rupture is its location on a source (active fault). Since a random magnitude defines the bounds of source, where the earthquake is able to occur, the location of the hypocenter should be conditional to the random magnitude on a given active fault.
Two random indicators are introduced to determine the location of upper-left corner instead of hypocenter for the simplistic application; but, the location of hypocenter is still ensured by that of upper-left corner with given width and length of rupture. A uniform random variable, $s_1 \subset [0,\ 1]$, is the ratio of (average) strike-directional coordinate of upper-left corner to the length of active fault to indicate horizontal location of the rupture while $s_2 \subset [0,\ 1]$ is also uniform random variable to describe the ratio of (average) dip-directional coordinate of upper-left corner to the width to indicate vertical location; s_1 and s_2 are statistically independent. In short, s_1 and s_2 are parameters to define relative coordinates of upper-left corner in terms of either strike- or dip- direction. Since these two independent indicators follow uniform distribution, these parameter is generated

from uncorrelated normal random variables by using (2A.4)

$$s_i = \Phi(u_i) \text{ for } i = 1, 2 \tag{2A.4}$$

Figure 2A.1 illustrates the relationship of s_1 and s_2 with a random magnitude m. Consider a typical simple fault model in Figure 2A.1 (a), and magnitude-area relationship in (2A.2) gives a restricted region where rupture surface may occur with a given magnitude m in Figure 2A.1 (b); in this figure $m = 5$. Then, random parameters, s_1 and s_2, indicate the location of the left-upper corner of rupture on the fault surface. For example, the central rupture event is realized with parameters, $s_1 = 0.5$ and $s_2 = 0.5$ as Figure 2A.1 (c). This parameterization helps to achieve a consistent sampling approach not changing the sampling domains to describe the location of upper-left corner. With the same parameters of $s_1 = 0.5$ and $s_2 = 0.5$, Figure 2A.1 (d) illustrates earthquake ruptures with different magnitudes from 5 to 8 with intervals of 0.25.

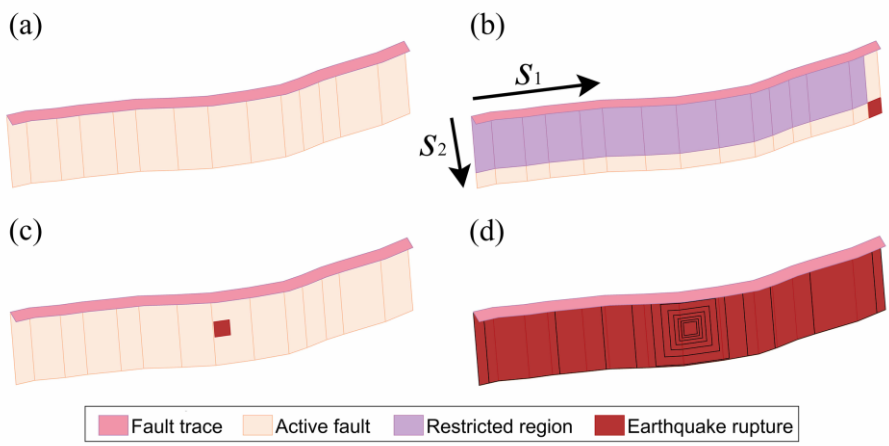

Figure 2A.1. Demonstration of sampling procedure for earthquake rupture

Appendix 2B. Updating Rules for CE−AIS−GM

This section demonstrates updating rule of cross-entropy based adaptive importance sampling with using Gaussian mixture (CE-AIS-GM), which was introduced by Kurtz and Song (2013). To discuss this updating rule, consider a numerical integration

$$I_t = \int H(\mathbf{x}) f(\mathbf{x};\mathbf{u}) d\mathbf{x} \cong \frac{1}{N}\sum_{i=1}^{N} H(\mathbf{x}_i) \tag{2B.1}$$

where $H(\mathbf{x})$ is a general function of variables $\mathbf{x}$, and $f(\mathbf{x};\mathbf{u})$ is the probability density function (PDF) of $\mathbf{x}$ with parameters $\mathbf{u}$.

2B.1. Optimal Importance Sampling Density

Revisit a derivation in Rubinstein and Kroese (2004) to identify optimal density to implement importance sampling. The integral (2B.1) becomes after introducing importance sampling (IS) density $h(\mathbf{x};\mathbf{w})$, which has the distribution parameter $\mathbf{w}$

$$I_t = \int \left[\frac{H(\mathbf{x}) f(\mathbf{x};\mathbf{v})}{h(\mathbf{x};\mathbf{w})}\right] h(\mathbf{x};\mathbf{w})\mathrm{dx} = E_{\mathbf{v}}\left[\frac{H(\mathbf{x}) f(\mathbf{x};\mathbf{v})}{h(\mathbf{x};\mathbf{w})}\right] \cong \frac{1}{N}\sum_{i=1}^{N} \frac{H(\mathbf{x_i}) f(\mathbf{x_i};\mathbf{v})}{h(\mathbf{x_i};\mathbf{w})} \tag{2B.2}$$

The variance of estimated expectation $q = H(\mathbf{x}) f(\mathbf{x};\mathbf{v}) / h(\mathbf{x};\mathbf{w})$ is

$$\mathrm{Var}(q) = \int_{\mathbf{x}} H(\mathbf{x})^2 \frac{f(\mathbf{x})^2}{h(\mathbf{x};\mathbf{w})^2} h(\mathbf{x};\mathbf{w}) d\mathbf{x} - I_t^2 \tag{2B.3}$$

The efficiency of simulation can be improved by selecting $\mathrm{h(x;w)}$ that provides lower $\mathrm{Var}(q)$ than that of $H(\cdot)$ in (2B.1). Cauchy and Schwarz inequality provides a lower bound of (2B.3)

$$\int_{\mathbf{x}} H(\mathbf{x})^2 \frac{f(\mathbf{x})^2}{h(\mathbf{x};\mathbf{w})} d\mathbf{x} \geq \left[\int_{\mathbf{x}} H(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}\right]^2 \tag{2B.4}$$

The equality in (2B.4) is satisfied when $h(\mathbf{x};\mathbf{v})$ is the "best" density function $p^*(\mathbf{x})$ which is express as

$$p^*(\mathbf{x}) = \frac{|H(\mathbf{x})| f(\mathbf{x};\mathbf{v})}{\int |H(\mathbf{x})| f(\mathbf{x};\mathbf{v}) d\mathbf{x}} \tag{2B.5}$$

2B.2. Cross-entropy-based Updating Rule with Using Gaussian Mixture

The best IS density in (2B.5) cannot be applicable in practice because the denominator is corresponding to I_t in most cases. Nevertheless, (2B.5) still implies that if one can choose an IS density of which shape is similar with $p^*(\mathbf{x})$, the numerical integral in (2B.1) becomes efficient by converging to I_t quickly.

The derivation in Rubinstein and Kroese (2004) provides a cross-entropy-based updating rule to find a near-optimal density. The distance between the

current IS density model and the best density function $p^*(\mathbf{x})$ can be measured by Kullback-Leibler CE

$$D\big(p^*(\mathbf{x}), h(\mathbf{x};\mathbf{w})\big) = \int p^*(\mathbf{x}) \ln p^*(\mathbf{x})\, d\mathbf{x} - \int p^*(\mathbf{x}) \ln h(\mathbf{x};\mathbf{w})\, d\mathbf{x} \tag{2B.6}$$

One can find $\mathbf{w}$ with given distribution type $h(\mathbf{x};\cdot)$ which gives a near-optimal IS density by minimizing Kullback-Leibler CE in (2B.5). Since $\int p^*(\mathbf{x}) \ln p^*(\mathbf{x})\, d\mathbf{x}$ in (2B.5) is already determined value regardless of IS density $h(\mathbf{x};\mathbf{w})$ and $p^*(\mathbf{x})$ is proportional to $|H(\mathbf{x})| f(\mathbf{x};\mathbf{v})$, an optimization problem is stated as the minimization of $D\big(p^*(\mathbf{x}), h(\mathbf{x};\mathbf{w})\big)$ to find a near-optimal density with parameter $\mathbf{w}^*$

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} D\big(p^*(\mathbf{x}), h(\mathbf{x};\mathbf{w})\big) = \arg\max_{\mathbf{w}} \int |H(\mathbf{x})| \ln h(\mathbf{x};\mathbf{w})\, f(\mathbf{x};\mathbf{v}) d\mathbf{x} \tag{2B.7}$$

To update IS density model adaptively, the previous IS density model $h\big(\mathbf{x};\mathbf{w}_{\mathrm{pre}}\big)$ with parameter $\mathbf{w}_{\mathrm{pre}}$ is used instead of $f(\mathbf{x};\mathbf{v})$ in (2B.7).

$$\begin{aligned} &\arg\max_{\mathbf{w}} \int |H(\mathbf{x})| \ln h(\mathbf{x};\mathbf{w})\, f(\mathbf{x};\mathbf{v}) d\mathbf{x} \\ &= \arg\max_{\mathbf{w}} \int |H(\mathbf{x})| \ln h(\mathbf{x};\mathbf{w}) \frac{f(\mathbf{x};\mathbf{v})}{h\big(\mathbf{x};\mathbf{w}_{\mathrm{pre}}\big)} h\big(\mathbf{x};\mathbf{w}_{\mathrm{pre}}\big) d\mathbf{x} \\ &= \arg\max_{\mathbf{w}} E_W\big[|H(\mathbf{x})| \ln h(\mathbf{x};\mathbf{w})\, W\big(\mathbf{x};\mathbf{v},\mathbf{w}_{\mathrm{pre}}\big)\big] \end{aligned} \tag{2B.8}$$

where $E_W[\cdot]$ denotes the mathematical expectation with respect to the density function $h\big(\mathbf{x};\mathbf{w}_{\mathrm{pre}}\big)$, and $W\big(\mathbf{x};\mathbf{u},\mathbf{w}_{\mathrm{pre}}\big)$ is the likelihood ratio, $f(\mathbf{x};\mathbf{v})/h\big(\mathbf{x};\mathbf{w}_{\mathrm{pre}}\big)$. With empirical expectation $E_W[\cdot]$, the optimization problem becomes

$$\begin{aligned} &\arg\max_{\mathbf{w}} E_W\big[|H(\mathbf{x})| \ln h(\mathbf{x};\mathbf{w})\, W\big(\mathbf{x};\mathbf{v},\mathbf{w}_{\mathrm{pre}}\big)\big] \\ &\cong \arg\max_{\mathbf{w}} \frac{1}{N} \sum_{i=1}^{N} |H(\mathbf{x}_i)| \ln h(\mathbf{x}_i;\mathbf{w})\, W(\mathbf{x}_i;\mathbf{v},\mathbf{w}) \end{aligned} \tag{2B.9}$$

The updating rule for near optimal IS density can be achieved by setting the gradient of $E_W[\cdot]$, $\nabla_{\mathbf{w}} E_W[\cdot] = 0$.

$$\frac{1}{N} \sum_{i=1}^{N} |H(\mathbf{x}_i)| W\big(\mathbf{x}_i;\mathbf{v},\mathbf{w}_{\mathrm{pre}}\big) \nabla_{\mathbf{v}} \ln h(\mathbf{x}_i;\mathbf{w}) = 0 \tag{2B.10}$$

Kurtz and Song (2013) and Wang and Song (2016) provided updating rules to implement CE-AIS by applying this condition to multimodal distributions, Gaussian mixture and Von-mises-Fisher mixture, respectively. In this paper, the

result of Kurtz and Song (2013) is visited.

Gaussian mixture (GM) is a linear combination of several Gaussian distributions with a weight for each distribution. The probability density function (PDF) of GM with O number of Gaussian distributions is

$$p(\mathbf{x}) = \sum_{o=1}^{O} \pi_o N(\mathbf{x}|\boldsymbol{\mu}_o, \boldsymbol{\Sigma}_o) \tag{2B.11}$$

where π_o denotes the relative weight of o-th Normal distribution, $0 \leq \pi_o \leq 1$, $\sum_{o=1}^{O} \pi_o = 1$, $o = 1, \ldots, O$, and $N(\mathbf{x}|\boldsymbol{\mu}_o, \boldsymbol{\Sigma}_o)$ is a unimodal Normal distribution with mean vector of $\boldsymbol{\mu}_o$ and covariance matrix of $\boldsymbol{\Sigma}_o$.

After substitute $p(\mathbf{x})$ into (2B.10), the updating rules proposed by Kurtz and Song (2013) are

$$\boldsymbol{\mu}_q = \frac{\sum_{i=1}^{N} |H(\mathbf{x}_i)| W(\mathbf{x}_i; \mathbf{u}, \mathbf{w}) \gamma_i(z_q) \mathbf{x}_i}{\sum_{i=1}^{N} |H(\mathbf{x}_i)| W(\mathbf{x}_i; \mathbf{u}, \mathbf{w}) \gamma_i(z_q)} \tag{2B.12}$$

$$\boldsymbol{\Sigma}_q = \frac{\sum_{i=1}^{N} |H(\mathbf{x}_i)| W(\mathbf{x}_i; \mathbf{u}, \mathbf{w}) \gamma_i(z_q) (\mathbf{x}_i - \boldsymbol{\mu}_q)(\mathbf{x}_i - \boldsymbol{\mu}_q)^{\mathrm{T}}}{\sum_{i=1}^{N} |H(\mathbf{x}_i)| W(\mathbf{x}_i; \mathbf{u}, \mathbf{w}) \gamma_i(z_q)} \tag{2B.13}$$

$$\pi_q = \frac{\sum_{i=1}^{N} |H(\mathbf{x}_i)| W(\mathbf{x}_i; \mathbf{u}, \mathbf{w}) \gamma_i(z_q)}{\sum_{i=1}^{N} |H(\mathbf{x}_i)| W(\mathbf{x}_i; \mathbf{u}, \mathbf{w})} \tag{2B.14}$$

where $\gamma_i(z_q) = \pi_q N(\mathbf{x}_i|\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q) / \sum_{o=1}^{O} \pi_o N(\mathbf{x}_i|\boldsymbol{\mu}_o, \boldsymbol{\Sigma}_o)$, the responsibility, which is referred as the probability that z_q takes 1 with an observation $\mathbf{x}_i$.

To fix upper/lower triangular elements of covariance matrix to zero, only diagonal elements are updated with the following rule

$$\sigma_{qr}^2 = \frac{\sum_{i=1}^{N} |H(\mathbf{x}_i)| W(\mathbf{x}_i; \mathbf{u}, \mathbf{w}) \gamma_i(z_q) (x_{ir} - \mu_{qr})^2}{\sum_{i=1}^{N} |H(\mathbf{x}_i)| W(\mathbf{x}_i; \mathbf{u}, \mathbf{w}) \gamma_i(z_q)} \tag{2B.15}$$

where σ_{qr}^2 is r-th diagonal component in updated q-th covariance matrix $\boldsymbol{\Sigma}_q$. x_{ir} and μ_{qr} are r-th component of $\mathbf{x}_i$ and $\boldsymbol{\mu}_q$, respectively. This strategy in (2B.15) is to increase the stability of CE-AIS-GM while sacrificing degree of freedom in covariance matrix in (2B.13); in other word, (2B.15) is alternative choice instead of (2B.13) for the purpose of stable execution. In this paper, no correlated shape GM of (2B.15) is adopted because it is likely that the optimal IS density in terms of conditional probability is circular enough to exploit diagonal covariance matrix. Of course, there are trade-off between stability and flexibility in CE-AIS-GM, but the stability has higher priority to flexibility in these applications.

2B.3. Initial parameters to implement CE-CAIS

Initial parameter has important role in successful implementation of CE-AIS. There are several recommendations to determine initial parameters of non-parametric distributions to implement CE-AIS as Kurtz and Song (2013) and Wang and Song (2016). However, determination for these initial points becomes more relaxed for the applications in this paper. That is because our estimation in sampling is conditional probability of multiple consequences after hazard events. This conditional probability-based approach has a great merit over non-conditional probability-based simulations. For example, the latter is relying on indicator function to estimate the probability of event; the estimator of probability is 1 for the case that the sampling point is in failure domain or 0 elsewise. This characteristic of non-conditional probability-based simulation implies that a good initial distribution for CE-AIS has to locate its center where is close to design points (or most probable points); if there is no failure point in previous round of pre-sampling, the cross-entropy based updating rule is going to blow up. On the other hand, it is not the case for conditional probability-based simulation since the conditional probability is always defined for all sampling points; for example, with given hazard realization, there is always conditional probability that loss exceeds a certain value. This property of conditional sampling relax the restriction on initial parameters.

For the applications in this paper, the initial parameters for mean-vector, $\boldsymbol{\mu}_o$, is all zeros $o = 1, \dots, O$, and the uniform elements for weight vector, $\boldsymbol{\pi}_o$. 8 Gaussian distributions were used for the flow-capacity example in Section 2.5, which is $o = 1, \dots, 8$ in (2B.11), and its initial parameters for 8 covariance matrices, $\boldsymbol{\Sigma}_o$, are diagonal matrices of which diagonal elements of each matrix are all the same; from 0.5 to 2 with 7 uniform intervals for 8 matrices. Similarly, 3 Gaussian distributions were used for the aggregated regional loss example in Section 2.6, which is $o = 1, \dots, 3$ in (2B.11), and initial covariance matrices, $\boldsymbol{\Sigma}_o$, are also diagonal matrices of which diagonal elements of each matrix are all the same as one of elements in {0.5, 1.25, 2} for 3 matrices. According to Kurtz and Song (2013), implementation of CE-AIS-GM is less sensitive to the number of unimodal Gaussian distributions in the mixture for an appropriate dimensional space of random variables. See Kurtz and Song (2013) for more discussions.

Chapter 3. Feature selection and clustering-based approach for complex lifeline network

It is necessary to evaluate seismic risk on urban infrastructures at network-level because spatially distributed infrastructures keep our community sustainable by transporting social supplies of the urban areas. The framework for this risk assessment is required to accept probabilistic concepts because uncertainties exist in both natural hazard and structural behavior. While Monte Carlo Simulation (MCS) is a straightforward solution to assess risk of network failure, MCS is not always effective especially when severe hazardous scenario should be included. That is because the simulation-based approach consumes a large computational cost to forecast rare event. To overcome this issue, a computationally efficient framework is introduced to obtain competently approximated annual exceedance probability curve for post-hazard performance of urban network. In the framework, a representative value is assigned to the each cluster to avoid undesirable computation after clustering samples according to the similarity. The Motivation is that generating samples is computationally affordable while repetitive network analyses for each sample are undesirable during the risk assessment. For this purpose, we illustrate features to capture the similarity between samples to develop a computationally efficient framework for probabilistic risk assessment for lifeline network. After that, our proposed approach is applied to two numerical examples, (a) a hypothetical road network consisting of 100 bridges and (b) the San Francisco Bay Area road network composed by 1,656 bridges. It is expected to expand our understanding on the impact of hazard and community-level resilience through the proposed framework.

3.1. Introduction

Urban networks are highly complex systems that are characterized by large numbers of infrastructures and their interdependencies (Der Kiureghian and Song, 2008). It is important to understand the risk of a lifeline network against natural or man-made hazards not only in terms of individual infrastructures but also at the levels of its network-level utility. The risk assessment of urban network should be also performed under probabilistic concepts since there are inherent randomness and uncertainties in potential hazards, structural behavior, post-hazard performance of infra-network, and socio-economic impacts. That is because that the resilience-

based decision making should be footing on such a probabilistic risk analysis, which is intend to keep our community being robust, redundant, resourceful, and rapid against urban disasters. (Bruneau and Reinhorn 2006).

Joint failures of infrastructures have a statistical dependence between themselves, so that its significant impact is a main challenge of the probabilistic risk assessment (PRA) for lifeline network (Song and Ok 2010, Lim and Song 2012). Thus, the PRA requires incorporating the joint failures into the framework to evaluate the post-hazard utility of network with its likelihood. Besides such a challenge, urban network has almost infinite number of post-hazard scenarios with difference level of probabilities, which makes identifying full mechanism of the post-hazard behavior of the network impossible. Thus, it is restricted not only to develop non-simulation-based risk assessment, but also to apply the framework into complex urban network (See examples of non-simulation approaches in Lim and Song 2012, Lee et al. 2011, Lim et al. 2015).

To overcome the challenge, Monte-Carlo Simulation (MCS) has been widely employed to perform PRA for lifeline network with simplicity (Bommer et al. 2002, Shiraki et al. 2007, Jayaram and Baker 2010, Miller and Baker 2015). However, a brute-force application of MCS has a limitation because it requires a huge computational time to identify low probability of post-hazard event. The low-probability event is still important to increase our communication with urban disasters in the context of risk mitigation and resilience-based decision making. As an effort to save computing resources, importance sampling (IS) approach has been employed by Jayaram and Baker (2010) to analyze post-hazard traffic flow for simplified (or aggregated) road network. The research also exploited *k*-means clustering on the ground-motion intensities to recognize similarities between realized hazard scenarios. Later, an optimization-based approach was developed to select the appropriate number of damaged scenarios (damage-maps) among the generated catalog (Miller and Baker 2015). The optimization is based on the ground-motion intensity profiles of several sites and the profile of an indirect indicator, as known as proxy-measure, for the target performance measure. Although these researches have shown sufficiently improved performance with the point of efficient PRA, the applications of such approaches have been still constraint. For example, Jayaram and Baker (2010) has limitations because (a) the post-hazard performance of the simplified network differs from that of real-scaled network, (b) heuristic IS densities in the research are not guaranteed to keep consistency not relying on the local properties of the network, and (c) the clustering, only focusing on the data about ground-motion intensities, cannot reflect the uncertainties in structural behavior of infrastructures. Also, (d) the clustering

algorithm, suggested in the paper, requires plentiful random-access memory (RAM) and computing time to insure sufficient work space in the computer since the algorithm needs to be implemented on the high dimensional space if the data is collected from amount of spatially distributed sites. On the other hand, Miller and Baker (2015) has another drawback because the optimization algorithm in the research employs tuned parameters to select damage-maps among the catalog. Such factors, which are potentially sensitive to the problem, are not desirable for general applications to complex and localized urban networks.

In this paper, a computationally efficient framework of PRA is introduced for lifeline network by succeeding two benchmark studies (Jayaram and Baker 2010, Miller and Baker 2015). Our main difference from the benchmark studies is that we identify the similarity between samples only through collecting the data about failures of infrastructures rather than exploiting data about hazard intensities. That is because the damage patterns on infrastructures can be differently realized even after the same outcomes of ground-motion intensities while collecting the data on the structural behavior does not require particularly large amount of computational cost. For the purpose of convenient data analysis, indirect (but computationally reasonable) measures, as known as proxy measures, are employed again to evaluate similarity between damage-maps; it is necessary to reduce the dimensionality of data to facilitate a computationally fast clustering algorithm. In final, the *k*-means clustering is executed to make groups and to assign a representative post-hazard performance value to each cluster. The framework is less sensitive to the localized network because there is no heuristic IS density or parameter, which should be tuned according to the characteristics of problems; the suggestion can be comprehended on the extension of unsupervised learning with extracted features and clustering approach to develop computationally efficient framework of PRA for urban network.

To achieve our research goal, we first review the framework of PRA for spatially distributed network under seismic hazard; the review provides investigations on uncertainties in ground-motion intensities, fragility of infrastructures, and network analysis for its functionality (Section 3.2). In Section 3.3, we extract features of realized damage-scenarios (damage-maps) to evaluate the similarity between samples, and *k*-means clustering algorithm is demonstrated to develop the computationally efficient framework. After instructing methods for feature extraction and clustering algorithm, the proposal is applied to two numerical examples in Section 3.4 (a hypothetical road network consisting of 100 bridges) and 3.5 (the San Francisco Bay Area road network composed by 1,656 bridges in morning trip time). Finally, Section 3.6 presents a summary and

discussions on findings and future research needs.

3.2. Probabilistic Seismic Risk Assessment for Lifeline Network

This section provides an overview on the framework of PRA for lifeline network, and demonstrates uncertainties in the risk assessment. In this paper, two types of uncertainties are integrated into the framework, the randomness in potential hazards and the uncertainties in structural behavior. Other sources of uncertainties are neglected since there is no consensus on the post-hazard randomness in the infra-network or socio-economic impacts. Although the seismic hazard is mainly discussed in this paper, the same framework is expandable to other natural or man-made disaster if there are appropriate probabilistic models for the disaster.

3.2.1. Randomness in ground-motion prediction

Ground-motion prediction equation (GMPE) is an empirical equation that is derived from observed earthquake events to forecast ground motion intensity at the site with given outcomes of characteristic of the event. For example, spectral acceleration at the site is estimated by the magnitude, source-to-site distance and other explanatory variables in (3.1). A general form of GMPE is

$$\ln Y_{ij} = f\left(M_j, R_{ij}, \theta_i\right) + \sigma_{ij}\varepsilon_{ij} + \tau_j\eta_j \tag{3.1}$$

where Y_{ij} denotes the interesting ground-motion intensity measure, -e.g. peak ground acceleration (PGA) or spectral acceleration (SA) at site i for the given event j-, $f\left(M_j, R_{ij}, \lambda_i\right)$ is the formula to represent attenuated mean value of Y_{ij} with M_j (magnitude of event j), R_{ij} (source-to-site distance to site i in earthquake j), and θ_i (a set of other explanatory at site i). ε_{ij} is the intra-event residual to describe site-to-site randomness within the same event j while η_j is the inter-event residual to describe event-to-event randomness that is the same value at all sites, and ε_{ij} and η_j are standard normal random variables, i.e. zero mean and unit standard deviation. σ_{ij} and τ_j are the standard deviations of the intra-event and inter-event residuals, respectively. Here, ε_{ij} (intra-event residual at site i) is spatially correlated with those of other sites, e.g. ε_{kj} (intra-event residual at site k). Such correlations are often described by a auto-correlation model, which

is a given function of the distance between two sites. To implement PRA, there are random inputs of GMPE: magnitude of event, earthquake rupture (the seismological distance R_{ij} is automatically determined by the given rupture), intra-event residuals, inter-event residuals, and uncertainties in explanatory variables. In this paper, uncertainties in explanatory variables are neglected for the simplicity, but the sampling procedures for the remained variables are provided to realize seismic hazard from the probabilistic model.

3.2.1.1. Realization of magnitude

In PRA framework, the earthquake event is often assumed to follow Poisson process (memoryless) and to have statistical independence with other events. Accepting such assumptions, the probability that earthquake event occurs on the seismic source s is evaluated as (3.2) using total probability theorem

$$P_s = \frac{\lambda_s}{(\sum_{w=1}^{S} \lambda_w)} \tag{3.2}$$

where λ_s, is the annual rate on source s that magnitude of event exceeds a threshold. The minimum value has been introduced to prevent wasting computing resource to calculate outcomes of trivial events. For example, events of which magnitudes are less than 5 are discarded in this paper. Using (3.2), the source of the event can be randomly selected according to its relative probability of occurrence.

After choosing the source of event, a random magnitude is realized based on its marginal density. Since it is difficult to directly exploit joint distribution of magnitude and source-to-site distance to realize the seismic hazard, it is more convenient to employ the marginal distribution of magnitude and the conditional distribution of source-to-site distance with a given magnitude. Using Bayes theorem, this relationship is represented as

$$f_{M,\boldsymbol{R}|S}(m, \boldsymbol{r}|s) = f_{\boldsymbol{R}|M,S}(\boldsymbol{r}|m, s) f_{M|s}(m|s). \tag{3.3}$$

There have been trials to provide a marginal density of the magnitude, which is so called magnitude-frequency distribution at certain seismic source s. If the magnitude-frequency distribution is provided in a closed form of distribution, i.e. truncated exponential distribution (Cosentino et al. 1977), there would be no difficulty to generate random magnitude. Similarly, the random magnitude would be realized without difficulty if the distribution is provided in the form of discrete PMF (Field et al. 2013). For the case of a nontrivial or complex distribution, see and refer the Der Kiureghian (2005).

3.2.2.2. Realization of earthquake rupture

The relationship between magnitude of the event and area of the rupture are described by magnitude-scaling relationship. Representatively, the suggestion by Field et al. (2013) is introduced in (3.4). Although we accepted the proposal by Field et al. (2013) in research, but suggestions by others may also be applicable to the framework.

$$M = \log_{10} A + \frac{2}{3}\log_{10}\frac{\max(1,\ \sqrt{\frac{A}{W^2}})}{(1+\max\left(1,\ \frac{A}{W^2\beta}\right))/2} + const \tag{3.4}$$

Simple fault model assumes the rupture, which triggers earthquake, floats on the fault surface (Figure 3.1 (c)). Also, the rupture area is restricted to have the maximum value which is the same with that of the active fault where the rupture is located. According to Field et al. (2013), the rupture is assumed to have squared shape, which means width W equals to length L, if the width W has not saturate the down-dip width W^* yet for small magnitude events; however, the length L follows equation (3.5) after width W saturates the down-dip width W^* for large earthquake events.

$$L = \begin{cases} W & W < W^* \\ \dfrac{A}{W^*} & W = W^* \end{cases} \tag{3}$$

Neglecting the randomness in magnitude-scaling relationship, the area of rupture becomes deterministic to the realized magnitude. Thus, the only remained random variable is the center (or other representative point) of rupture on the source (active fault). Figure 3.1 illustrates the procedure to realize the rupture on the source. First, a random magnitude restricts the bounds of source by using magnitude-scaling relationship, where the earthquake is able to occur (Figure 3.1 (a)); in other word, the hypocenter is realized conditionally with a given magnitude value. Next, two random parameters are introduced to indicate the location of upper-left corner; remind that location of the corner still guarantees a single realization of earthquake rupture. The parameters, $s_1 \subset [0,\ 1]$ and $s_2 \subset [0,\ 1]$, are statistically independent and uniformly random variables. Such indicators are to pinpoint relative coordinates of upper-left corner in terms of strike- and dip-direction, respectively. For more details, s_1 is the ratio of strike-directional coordinate of the corner to the active fault length while $s_2 \subset [0,\ 1]$ is the ratio of dip-directional coordinate of the corner to the width. (Figure 3.1 (b)). For example, Figure 3.1 (c) illustrates the central rupture event with $s_1 = 0.5$, $s_2 = 0.5$, and

$m = 5$. For the same s_1 and s_2 values of 0.5, Figure 3.1 (d) shows earthquake ruptures for different magnitudes from 5 to 8 with intervals of 0.25. Finally, the vector of source-to-site distance $\boldsymbol{r}$ in (3.3) is only described by two random variables s_1 and s_2. Although simple fault model are mainly employed to describe seismicity of the region, a similar approach is possible to indicate the location of rupture for other source types, e.g. point or area source.

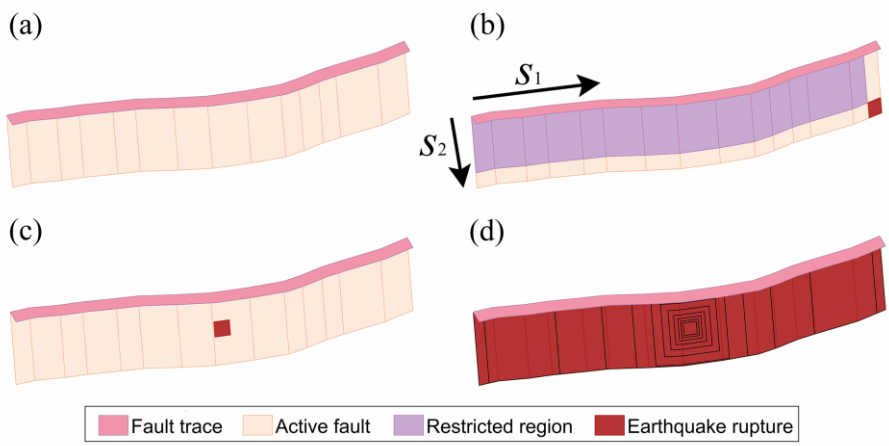

Figure 3.1. Demonstration of sampling procedure for earthquake rupture

3.2.1.3. Realization of residuals

Since normally distributed residuals are assumed in GMPE, uncorrelated standard normal variables are easily transformed into intra/inter-event residuals by using Cholesky-decomposition (or other linear transformation). It is widely adopted to use homogeneous auto-correlation model to describe the spatial correlation, e.g. Jayaram and Baker (2009), so that the correlation matrix for intra-event residuals can be constructed without difficulty.

3.2.2. Uncertainties in structural behavior

Fragility model is a conditional probability that post-hazard status of an asset exceeds k-th damage state (ds_k) against given hazard-intensity Y. The resisting capacity of the asset is denoted as A_m. It is common to assume the resisting capacity follows log-normal distribution for fragility model of point (or node) type structures –e.g. building, bridge, or station. Thus, a representative

fragility for point type assets is

$$P(DS > ds_k|Y) = \Phi\left(\frac{\ln Y - \ln A_m}{\beta_{DS}}\right) \tag{3.6}$$

where $\Phi(\cdot)$ is CDF of standard normal distribution while β_{DS} is randomness of structural capacity. Especially for risk assessment for seismic hazard, the hazard intensities Y in (3.6) become the ground-motion intensities like PGA or SA. These intensities are forecasted by GMPE and realized its affiliated random variables in (3.1). In this paper, the numerical examples constrain fragile assets in the road network to only bridges, so that other fragility models for line (or link) type assets are not demonstrated in this section.

3.2.3. Network risk assessment in the context of utility

For PRA of the lifeline network, the annual probability that the loss of the utility exceeds a certain value l_i is formulated as (3.7) by total probability theorem.

$$P(L \geq l_i) = \int P(L \geq l_i|\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} \tag{3.7}$$

where $\mathbf{y}$ denotes the ground-motion intensities at the sites of individual infrastructures, and $f_{\mathbf{Y}}(\mathbf{y})$ represents the joint probability density function (PDF) of ground-motion intensities. This joint PDF of ground-motion intensities, $f_{\mathbf{Y}}(\mathbf{y})$, is rewritten by seismic source and affiliated random variables in GMPE (magnitude, source-to-site distance, intra/inter-event residuals).

$$P(L \geq l_i) = \left(\sum_{s=1}^{S} \lambda_s\right) \cdot \int P(L \geq l_i|m, \boldsymbol{r}, \eta, \boldsymbol{\varepsilon}) f_{\boldsymbol{X}}(m, \boldsymbol{r}, \eta, \boldsymbol{\varepsilon}) dm d\boldsymbol{r} d\eta d\boldsymbol{\varepsilon} \tag{3.8}$$

where λ_s is the annual occurrence rate of the event at source s, for $s = 1, \ldots, S$, and $f_{\boldsymbol{X}}(\cdot)$ is corresponding joint PDF of magnitude m, source-to-site distance r, inter-event residual η, and intra-event residuals $\boldsymbol{\varepsilon}$. However, $\sum_{s=1}^{S} \lambda_s$ can exceed 1 with increasing number of seismic sources, which violates the fundamental axiom of probability that "probability cannot exceed 1". Since λ_s in the equation (3.8) assumes significantly small annual rates, λ_s's, the annual occurrence rate at reginal scale, Λ, is newly defined not to violate axiom of probability.

$$\Lambda = 1 - \exp\left(-\sum_{w=1}^{S} \lambda_w\right) \tag{3.9}$$

This aggregated annual rate in the target region is derived by the fact that the sum of independent random variables, which follow Poisson distribution, also follows Poisson distribution of which mean occurrence rate is the same with sum of individual mean occurrence rates. Thus, (3.8) becomes

$$P(L \geq l_i) = \Lambda \cdot \int P(L \geq l_i | m, \boldsymbol{r}, \eta, \boldsymbol{\varepsilon}) f_{\boldsymbol{X}}(m, \boldsymbol{r}, \eta, \boldsymbol{\varepsilon}) dm d\boldsymbol{r} d\eta d\boldsymbol{\varepsilon} \tag{3.10}$$

Assuming simple fault models and accepting the parameters s_1 and s_2 in Sub-section *3.2.1*, the (3.10) is

$$P(L \geq l_i) = \Lambda \cdot \int P(L \geq l_i | m, s_1, s_2, \eta, \boldsymbol{\varepsilon}) f_{\boldsymbol{X_s}}(m, s_1, s_2, \eta, \boldsymbol{\varepsilon}) dm ds_1 ds_2 d\eta d\boldsymbol{\varepsilon} \tag{3.11}$$

Considering the dependence between random variables in (3.11), joint PDF $f_{\boldsymbol{X_s}}(\cdot)$ is rewritten by Bayes theorem

$$\begin{aligned} f_{\boldsymbol{X_s}}(m, s_1, s_2, \eta, \boldsymbol{\varepsilon}) &= \left[\sum_{s=1}^{S} \frac{\lambda_s}{(\sum_{w=1}^{S} \lambda_w)} \right] \cdot f_{M|S}(m|s) \cdot f_{s_1}(s_1) \cdot f_{s_2}(s_2) \cdot f_{H,E}(\eta, \boldsymbol{\varepsilon}) \\ &= P_S \cdot f_{M|S}(m|s) \cdot f_{s_1}(s_1) \cdot f_{s_2}(s_2) \cdot f_{H,E}(\eta, \boldsymbol{\varepsilon}) \end{aligned} \tag{3.12}$$

Finally, the annual probability of exceedance can be estimated by brute-force MCS

$$P(L \geq l_i) \cong \Lambda \cdot \frac{1}{N} \sum_{n=1}^{N} I(L \geq l_i), \tag{3.13}$$

where $I(L \geq l_i)$ is the indicator function which is 1 if the decreased utility of the network (loss) exceeds a certain value l_i with realized random variables, otherwise zero. Such probabilistic simulation necessarily realizes structural damage of individual infrastructures because the structural outcomes of the hazard can be different even after the same event as long as the probabilistic concept is adopted. To realize the damage status, an uniform random variables or standard normal random variables is generated and compared with the conditional failure probability, $P(DS > ds_k|Y)$, or conditional reliability index, $-(\ln Y - \ln A_m)/\beta_{DS}$, in (3.6), respectively.

For risk assessment of lifeline network, it is necessary to define the utility of network. Depending on the purpose of the system, the utility can be defined as connectivity from source-node to sink-node, flow capacity (maximum flow) from origin to destination, or total travel cost (time) on the transportation network. In this paper, the increased travel time with fixed demand is selected as the post-hazard utility of road network under user-equilibrium (UE) assumption. Although the UE with fixed demand may be not realistic, such a hypothesis can be helpful to increase our communication with urban disaster, and its computational cost is still

challengeable. Also, the assumption can be reasonable to describe the situation when commuters in the urban community get used to the damaged network while the network has not been fully recovered after the hazardous event (Miller and Baker 2015).

The UE traffic assignment is an optimization assuming that each user on the network follows the route which would minimize his or her travel time. Adopting this UE principle, the optimization problem is

$$\min_{\mathbf{q}} \sum_{l\in\{\text{links}\}} \int_0^{q_l} t_a^l(x_l)dx_l$$

Subject to:

$$\begin{aligned}
&\text{(i)} \sum_{m\in\{\text{paths}\}} f_m^{od} = Q^{od}, \qquad \forall o \in \{\text{org}\}, \forall d \in \{\text{dest}\},\\
&\text{(ii)}\ x_i = \sum_{o\in\{\text{org}\}} \sum_{d\in\{\text{dest}\}} \sum_{m\in\{\text{paths}\}} f_m^{od}\,\delta_{ml}^{od}, \quad \forall l \in \{\text{links}\},\\
&\text{(iii)}\ f_m^{od} \geq 0, \quad \forall o \in \{\text{org}\}, \forall d \in \{\text{dest}\}, \forall m \in \{\text{paths}\}.
\end{aligned} \tag{3.14}$$

where the solution, $\mathbf{q}$, is a vector of which element, q_l, is the optimal link-flow of the link l under UE principle, and $t_a^l(\cdot)$ is travel cost (time) of the link l with a feasible (not be optimal yet) flow x_l. Sets $\{\text{org}\}$ and $\{\text{dest}\}$ are compounded by all origins and destinations of demands, respectively, while set $\{\text{paths}\}$ consists of all paths (subsets of nodes or links) which are connecting all pairs of origin node o and destination node d (o-d pair). Q^{od} is the demand on the travel from origin node o to destination node d, and f_m^{od} is a path-flow on the path m that is connecting the same o-d pair. δ_{ml}^{od} is a Dirac-delta function that is 1 if link l is included in the path m for a given o-d pair, else 0.

Bureau of Public Roads (1964) provides the travel time on link l

$$t_a^l(x_l) = t_0^l\left[1 + \alpha\left(\frac{x_l}{c_l}\right)^{\beta}\right] \tag{3.15}$$

where t_0^l denotes the freeway travel time for link l assuming the link is empty, and c_l is the capacity of link l while α and β are calibrating parameters, 0.15

and 4, respectively. In final, the total travel time of the network (TT) can be obtained as (3.16) by solving the optimization problem in (3.14)

$$TT = \sum_{l \in \{\text{links}\}} t_a^l(q_l) \cdot q_l. \tag{3.16}$$

The increased travel time of the network is evaluated in the percentage unit to compare post-hazard performance of the network with the performance under the normal operation. The increased travel time in percentage (ITT) is defined as

$$ITT = \frac{TT_d - TT_0}{TT_0} \times 100\ (\%). \tag{3.17}$$

where TT_d is the total travel time of the network under damaged situation after the hazard, and TT_0 is the total travel time of the network if there is no damaged infrastructure.

3.3. Feature Selection and Clustering-based Approach

In this section, a computationally efficient framework is introduced for PRA of lifeline network based on two benchmark studies (Jayaram and Baker 2010, Miller and Baker 2015). Jayaram and Baker (2010) suggested a computationally efficient framework by using importance sampling (IS) and data reduction technique (*k*-means clustering), but the research has a limitation in that a heuristic IS density is employed, which cannot guarantee its effectiveness if the suggested heuristic strategy is applied to other localized network. Also, the *k*-means clustering in that research has been implemented only focusing on ground-motion intensities, so that the suggestion can be inefficient if the clustering is executed on the high dimensional data; for data of *N* spatially distributed assets, a huge computational cost of the algorithm is required to secure the work space of the computer with the large *N* value. Later, Miller and Baker (2015) proposed an optimization-based algorithm to select representative damage scenarios (damage-maps) by using proxy measures. The proxy measure is an indirect estimators for the target network performance measure (computationally challenging), but has a merit in the point of computational cost. In the optimization algorithm, an approximated annual occurrence profile is obtained for post-hazard target

performance measure (increased total travel time) by proxy measures (spectral acceleration and percentage of functional bridges). However, a limitation exists in Miller and Baker (2015) because the optimization requires parameters which should be tuned with heuristic method, so that the optimization can be sensitive to characteristics of the network. Also, the method for traffic assignment of the research should be modified since such a heuristic algorithm, i.e. incremental traffic assignment, produces quite different results depending on the order of assignment; the solution is changed if the different order of o-d pairs is adopted.

Therefore, our research goal is to develop (a) computationally efficient, (b) less problem-sensitive framework of PRA for urban network that is (c) applicable to full-scaled network. To achieve our goals, a framework is suggested to identify the similarity between damage-maps. Main idea of the approach is to decrease the undesirable repetitive network analyses, which are computationally challenging in MCS procedure. It is inevitable to generate random hazard and failure of structures like brute-force MCS, but this step is relatively affordable with point of computational cost. After realization, the damage-maps are grouped into a few clusters according to their similarities, so that a representative network-level utility is assigned to all elements of each cluster. In this approach, features to check the similarity play a key-role for successful implementation. The following sub-sections provide a guide to select features of damage-maps and to obtain approximated profile of target performance measure.

3.3.1. Feature selection: proxy measures

In our framework, the similarity is identified by proxy measures based on failures of infrastructures, e.g. number of damaged bridges, instead of exploiting data on both hazard intensities and structural damage. Although this approach may miss the information about the hazard, there are two supporting points of our proposal. First, a clustering algorithm can be implemented on the computationally comfortable environment since additional computational cost will be required to execute the algorithm if the dimensionality of data is significantly increased by spatially distributed hazard intensities; it is difficult to find a representative explanatory variable which covers overall pattern of the hazard intensities. Second, the post-hazard performance of urban network can be different even after the same ground-motion intensities. In other word, information about structural failures is more important than hazard characteristics while the structural damage of infrastructures is easily realized without much additional effort. That is because the information on ground-motion intensities is already embedded in the failures of

structures.

Proxy measures are indirect estimators to predict target performance measure. For example, total travel time on the traffic network tends to be increased if more bridges are under non-functional status after the earthquake event; i.e. the number of damaged bridges (NDB) is a proxy measure for increased travel time (ITT). However, the proxy measure is not omnipotent because infrastructures on urban community contribute on the network at different level; an example can be a bottle-neck effect induced by failures of key-elements. Figure 3.2 illustrates such situations in the urban network. Although the same numbers of bridges are damaged for both situations, the increased travel time on the left is almost a half of the realization on the right due to topology of the network. To mitigate this effect, additional features are introduced to mark samples (damage-maps) in terms of characteristics for network performance. By adding appropriate features, strength for classification is reinforced to evaluate the similarity between samples. For this purpose, such features would be still another proxy measures since a proxy measure is able to indicate overall tendency of target performance measure, so they are attractive candidates for these features.

As a proxy measure for ITT of road network, the number of damaged bridges (NDB) is selected feature to identify the similarity of damage-maps.

$$\mathrm{NDB} = \sum_{i=1}^{N} I_i. \tag{3.18}$$

where I_i is indicator function that is 1 for non-functional state of i-th bridge, else 0, for $i = 1, \dots, N$. For the purpose of effective clustering, two additional features are introduced as proxy measures of the target performance measure. These proxy measures share a common property in that both measures require reference status of the network. To evaluate both features, a network analysis is required for the non-damaged case as the reference.

Lost travel time with traffic flow (LTT) is defined as

$$LTT = \sum_{\iota \in \{lost\}} t_{a0}^{\iota} \cdot q_{0\iota} \tag{3.19}$$

where ι denotes a link that loses its functionality after the hazardous event while t_{a0}^{ι} is a travel time of the link ι if the network are fully operated (non-damaged case). Similarly, $q_{0\iota}$ represents an originally intended flow of the link ι under the non-damaged status. A set $\{lost\}$ consists of all links that are not operated after the event due to structural failure. LTT represents the sum of demands on lost links

that are weighted by corresponding travel times. Lost capacity with traffic flow (LCT) is similarly defined as (3.20), but the main difference is that LCT adopts link capacity c_ι instead of travel time t_{a0}^{ι} for the lost link ι.

$$LCT = \sum_{\iota \in \{lost\}} c_\iota \cdot q_{0\iota}. \tag{3.20}$$

Above proxy measures are evaluated by prosaic calculation within very short time, so that the generated damage-maps are arranged into feature space to evaluate the similarities without consuming amount of computational cost. From Figure 3.2, one can check that two damage-maps, which cannot be characterized only by NDB, are well-distinguished by the additional features (LTT an LCT) with respect to target performance measure. Although NDB values of both damage-maps are equal to each other, different ITT values can be anticipated by two additional features.

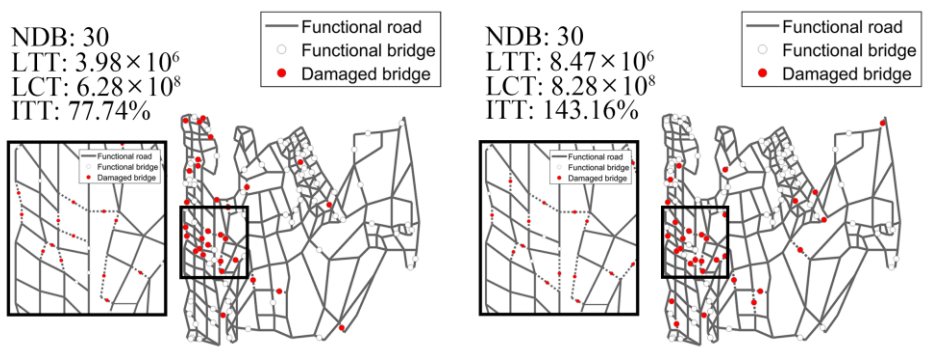

Figure 3.2. Realizations of the network to help understand proxy measures

3.3.2. Clustering-based approach for PRA

The *k*-means clustering is popular for clustering analysis of data. Through the clustering analysis, each observation in the data belongs to the cluster of which center (mean) is the nearest to the observation according to defined distance function, e.g. L_2-norm, so that N observations are partitioned into k clusters. Such clustering analysis is an optimization process to minimize sum of distances from each point within the cluster to the center. (3.21) represents the objective function of the optimization accepting L_2-norm as the distance function to evaluate distances from within-clusters (observations in the cluster) to the their center.

$$\underset{\mathbf{S}}{\operatorname{argmin}} \sum_{c=1}^{k} \sum_{\mathbf{x} \in \{S_c\}} \|\mathbf{x} - \boldsymbol{\mu}_c\|^2 \tag{3.21}$$

where $\{S_c\}$ is the set of which elements are positioned within c-th cluster while $\boldsymbol{\mu}_c$ is the center of c-th cluster (mean). $\mathbf{x}$ is an observation (element) in the set S_c for $c = 1, \ldots, k$, and $\|\cdot\|^2$ is an L_2-norm operator in Euclidian space.

The *k*-means clustering is implemented on the feature space to establish the computationally efficient framework for PRA of urban network. In our suggestion, the observation $\mathbf{x}$ in (3.21) is corresponding to extracted features (proxy measures) for each damage-map. To eliminate the scale-effect of features, all proxy measures are normalized to have zero means and unit standard deviations. After clustering the damage-maps, a representative value is assigned to each cluster to obtain approximated risk profile instead of evaluating target performance measure for every sample. An approximated profile is obtained by (3.22) after clustering samples according to their proxy measures

$$P(L \geq l_c) \cong \Lambda \cdot \sum_{c=1}^{k} w_c \cdot I_c(L \geq l_i), \tag{3.22}$$

where $I_c(L \geq l_i)$ is an indicator function for c-th cluster which is 1 if representative value of the cluster exceeds certain value l_i, else 0. The weight of cluster, w_c, is the ratio of the number of observations within the cluster, N_c, to the total number of observations (damage-maps) N; in short, $w_c = N_c/N$. In this clustering-based approach, a representative network analysis is performed for the randomly selected damage-map within the cluster. Appendix B in Jayaram and Baker (2010) provides a proof that the estimator in (3.22) is unbiased.

The procedure for computationally efficient PRA of lifeline network is summarized as follows:

Step 1: Generate N damage scenarios considering sources of uncertainties.
Step 2: Compute proxy measures (features) and normalize the measures
Step 3: Execute *k*-means clustering with proxy measures to partition the data.
Step 4: Perform a representative network analysis for each cluster.
Step 5: Obtain approximated risk profile by (3.22).

3.4. Post-hazard Traffic Flow of Hypothetical Road Network

Figure 3.3 illustrates a hypothetical road network with 100 road-bridges,

which is hand-drawn by authors, surrounded by 9 active faults. Geometric and seismic properties of active faults are from Field et al. (2013) and modified to facilitate numerical tests in this example. Figure 3.3 (a) is a visualization of the network with its bridge-components. Failures of bridges are assumed to be the only source of decreased network utility after the hazard. The fragility model of each bridge follows definition FEMA-443 (2003), and the bridge types were randomly selected from the same reference. The damage status that exceeds 'extensive' is set to the failure of bridge by Miller and Baker (2015), so the same criterion is accepted to define the failure of bridge. Magnitude-scaling relationship in Shaw (2009) with parameters by appendix E of Field et al. (2013), and GMPE in Boore and Atkinson (2008) are adopted in this application. Spectral acceleration at 1.0 sec ($SA_{1.0sec}$) was required as the input of PRA, and the model by Jayaram and Baker (2009) is used to consider the spatial correlation model for intra-event residuals since fragility model in FEMA-443 (2003) is based on the $SA_{1.0sec}$ for the bridge assets. Figure 3.3 (b) details demands of the traffic network with the origins and destinations; total 90 o-d pairs were introduced. To solve UE traffic assignment problem, the example employs a Frank-Wolfe based algorithm (FW) suggested by D. van Vliet (1986). Although there are other path-based algorithms that has been developed to improve the convergence speed of the optimization, FW algorithm is applied in this example because (a) this link-based solution has eminent stability, (b) computing code can be built with a slight effort, and (c) its convergence speed is still tolerable in this problem. Next, relative gap (RGAP) defined as (3.23) is adopted as a criterion to terminate the algorithm.

$$RGAP = \frac{(\text{total travel time}) \ - (\text{total travel time on shortest paths})}{(\text{total travel time})} \tag{3.23}$$

Since UE assumes that user of the network tires to follow the route that would minimize his or her travel time, any activated route becomes the shortest path for all o-d pairs of the network, so that RGAP converges to 0 with iterations of the algorithm. In this application, the stop criterion is set to RGAP of 0.01.

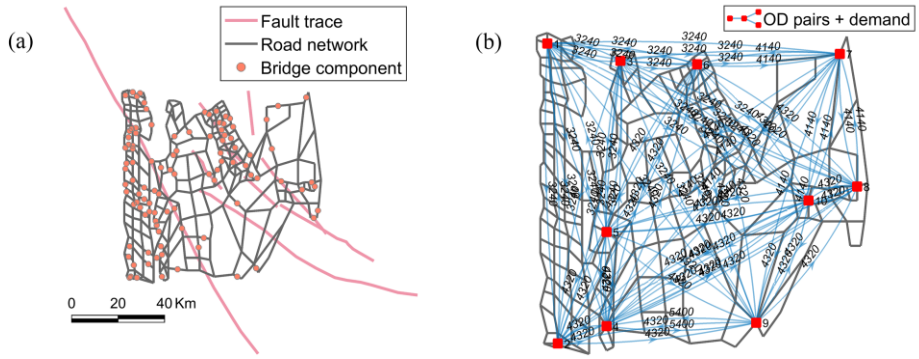

Figure 3.3. Illerstration of the hypothetical road network

A million number of damage scenarios (damage-maps) are generated to adopt the simulation-based framework of PRA while reflecting the uncertainties demonstrated in Section 3.2. Since there are 62,864 unique realizations among a million realizations, at least 62,864 network analyses are required to implement brute-force MCS in (13). Although each realization only requires 6.55 seconds on an average to assign the traffic flow on the network according to UE principle, the computational cost is severely cumulated by repetitive network analysis of MCS, so that sufficient computing resources are required to obtain the seismic risk profile of the network. In this application, 3,627 seconds is consumed to realize 1,000,000 damage scenarios for 100 bridges, which is the essentially required for both brute-force MCS and the suggested clustering-based approach. For brute-force MCS, network analyses are implemented for 62,864 unique realizations, and 4,037 seconds are required to obtain the result of network analyses with 102 CPUs of Intel Xeon E5-2407 2.20GHz and 312 GB of 6 nodes RAM; this step is executed by parallel computing.

To validate the suggestion in Section 3.3, two results are provided according to the number of clusters, 400 and 1,000, which are defined values (k) in k-means clustering. Since the clustering-based approach requires extracting features before implementing the clustering, the proxy measures for all damage-maps are evaluated with computational time of 14 seconds except for the non-damaged cases; this computation cost can be different for each trial, but does not show a remarkable variability. Non-damaged scenario is skipped to calculate its proxy measures because its target performance measure, i.e. increased total travel time (ITT), is necessary evaluated to obtain proxy measures of other damage maps, and it has obvious proxy measures of zeros. The clustering is executed after eliminating a bunch of non-damaged cases, but not on the unique data set. The

duplicated data are allowed during the clustering because such overlapping reflects the probabilistic density of events. The computational time for this step depends on the k value of the clustering; 73 seconds for 400-means clustering and 1,099 seconds for 1,000-means clustering. Although the lower k value requires a short computational time for the step, the resolution of the approximation in (3.22) becomes also coarse with the lower k value; this is vice versa for the larger k value, so there is a trade-off between resolution of the approximation and computing speed of the clustering. As the final step, the network analyses are implemented by representative damage-maps for each cluster, so 400 network analyses are performed for 400-means clustering while 1,000 analyses are performed for 1,000-means clustering; total (k+1) traffic assignment is executed because there are 1 execution for the non-damaged case in Step 2 and k excitations for k clusters in Step 4. This network analysis step is also executed by parallel computing with 102 CPUs of Intel Xeon E5-2407 2.20GHz and 312 GB of 6 nodes RAM, which is the same resource with the application by brute-force MCS. Table 3.1 summarizes the computational cost for each step in the risk assessment, and the approximated curves are illustrated for annual exceedance probability of the event in Figure 3.4. Although the suggestion requires additional steps to compute proxy measures and to execute clustering algorithm, the reduced computation on repetitive traffic assignments improves the efficiency of the framework effectively.

Table 3.1. Computational time of each step for risk assessement
(Hypothetical network)

Step	**Brute-force MCS (62,864 analyses)**	**Clustering-based approach (1,000 means)**	**Clustering-based approach (400 means)**	**Computer spec.**
Generating samples (10^6)	3,627 sec			CPU: Intel i5-4570 3.2GHz (1 core) RAM: 16.0 GB
Evaluating proxy measures	-	14 sec	13 sec	CPU: Intel(Xeon) E5-2407 2.20GHz (1 core) RAM: 16.0 GB
k-means clustering	-	1,099 sec	73 sec	CPU: Intel(Xeon) E5-2407 2.20GHz (1 core) RAM: 16.0 GB
Network analysis	4,037 sec	124 sec	77 sec	CPU: Intel(Xeon) E5-2407 2.20GHz (102 cores) RAM: 312 GB x 6 nodes
Total	7,664 sec (2.21 hr)	4,864 sec (1.35 hr)	3,790 sec (1.05 hr)	

To discuss about the accuracy of approximation, the error of approximated curve is defined as (3.24) where the measurement is the average of normalized absolute gaps between annual exceedance probabilities by two approaches, the brute force MCS and the clustering-based approach.

$$\frac{1}{E}\sum_{i=1}^{E}\left|\frac{\hat{p}_i^{mcs}-\hat{p}_i^{clu}}{\hat{p}_i^{mcs}}\right| \tag{3.24}$$

where $\hat{p}_i^{mcs}$ is the probability that target performance measure exceeds i-th level estimated by brute-force MCS while $\hat{p}_i^{clu}$ is that by the clustering-based approach. Since the accuracy of approximated profile differs depending on target event-ranges in (3.24), the target range is defined as 'from 0 to $\mathrm{ITT}_{\max}$ with 0.001 intervals' where $\mathrm{ITT}_{\max}$ is the maximum value of target performance measure for (3.24). For example, when $\mathrm{ITT}_{\max}$ is set to 50, first damage level ($i = 1$) in (3.24) becomes 0, the second level ($i = 2$) is 0.001, the third ($i = 3$) is 0.002, and so on until the maximum value $\mathrm{ITT}_{\max}$ becomes 50 ($i = E$). Under such a definition, the interpolated risk profiles are used in this error measurement for both brute-force MCS and the clustering-based approach. The error measurement is summarized for this application in Table 3.2. In the Table 3.2, the probabilities are not multiplied by the annual rate Λ yet. Because this parameter is not improved by further progress of the simulation, it is appropriate to compare the error measurement only by the simulation-driven parts, $\hat{p}_E^{mcs}$ and $\hat{p}_i^{clu}$. From the Table 3.2, one can check that the clustering-based approach gives a quite acceptable approximation even for the cases that corresponding probabilities are less than 0.0025 or 0.001, which is the minimum probability that can be estimated by brute-force MCS with 400 or 1000 samples, respectively; brute-force MCS with *N* samples can estimate the probability of event up to 1/*N* apart from whether the forecast is acceptable or not. In summary, a result is extracted that the clustering-based approach provides a quite good approximation with almost a half computational cost if the same resource is exploited.

Table 3.2. Error measurement for each *k* value (Hypothetical network)

$\mathrm{ITT}_{\max}$ ($\hat{p}_E^{mcs}$)	50% (2.9×10^{-3})	100% (9.39×10^{-4})	150% (1.48×10^{-4})	200% (1.72×10^{-5})	250% (3.06×10^{-6})	300% (1.02×10^{-6})
Clustering-based approach (1,000)	0.035	0.041	0.0824	0.1394	0.2138	0.2857
Clustering-based approach (400)	0.091	0.129	0.1753	0.2281	0.2685	0.3313

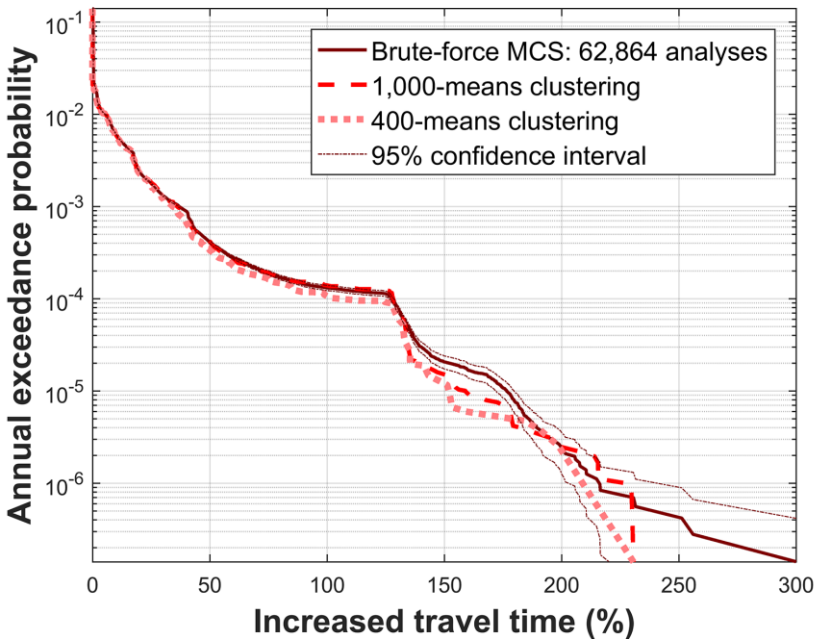

Figure 3.4. Annual exceedance probability curves (Hypothetical network)

Figure 3.5 plots the data (damaged scenarios) on the normalized feature space allowing duplications. In Figure 3.5 (a), data are scattered on the space according to their normalized proxy measures while the points are coloured with their corresponding target performance measures. Figure 3.5 (b) provides the same scatter plot with Figure 3.5 (a), but the data is coloured with the approximated target performance measures by the clustering-based approach with 1,000-means. To discuss the suggested approximation with the more comprehensive view, the 3D scatter plots are projected into the plane of which x-axis is the first principal component of data while y-axis is the second principal component. First principal component is the variable which shows the greatest variability of data, and second principal component is orthogonal to first principal component which shows the secondly greatest variability; next principal components are similarly defined with their orthogonality and variability. An orthogonal transformation, as known as principal components analysis, is implemented to configure these principal components. In Figure 3.5 (c) and (d), the data are scattered in 2D of first and second principal components. Although there are minor differences between both figures, the overall tendency of network-level damage can be captured by the clustering-based approach, so that the approximated risk profile is acceptable in spite of decreased computational time.

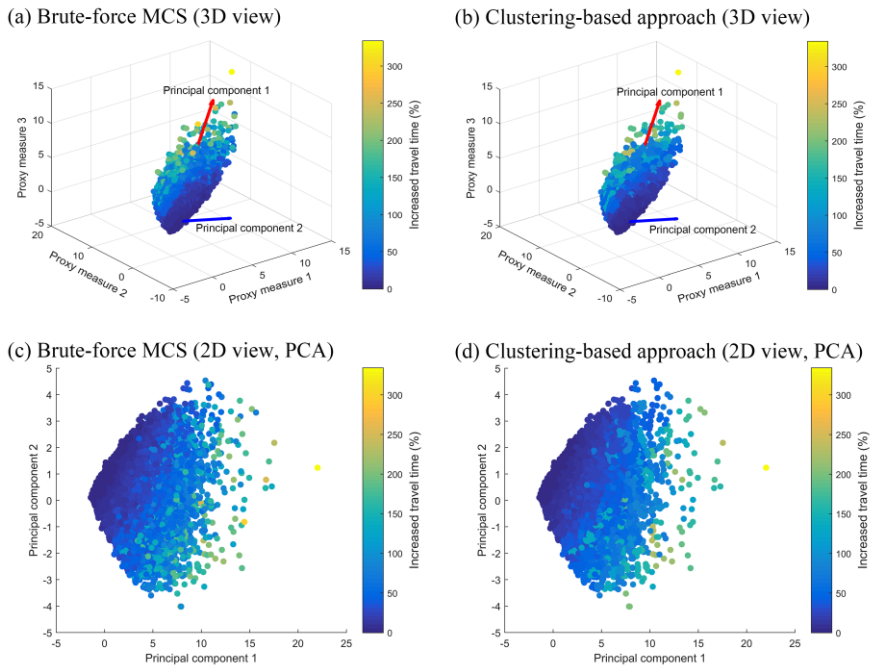

Figure 3.5. Scatter plots of data on the feature space (Hypothetical network)

3.5. Traffic Network on Bay Area, San Francisco

In addition to Section 3.4, another application is represented to show applicability of our framework. The seismic risk on San Francisco Bay Area road network is investigated in this Section. The data of this network is downloaded from database of PhD thesis of Dr. Mahalia Miller, Stanford University (currently at Google) [https://purl.stanford.edu/mh896js1648]; the author is really appreciated for her efforts and kindness in sharing data. This complex network consists of total 11,921 nodes and 32,858 links including 9,635 physical nodes and 24,404 physical while the other components are centroid nodes and links, which has no physical meaning but has meant to depict boundary flows or to aggregate minor elements. The capacity of the non-physical link is set to 10^{10}, which is large enough that travel time of the link is not influenced by the increased flow. Also, 34 artificial nodes are introduced in this network. These additional non-physical nodes are named as super-district nodes that are dedicated to describe traffic demands of commuters who move from districts to other districts at the morning trip time (Miller 2014); internal demands on the district are neglected since there is no data of the origin and destination at inner district scale. The links, which connect super-district nodes and other nodes, are named as super-district links, and the capacities

of those links are set to 10^{10} due to the same reason with pre-existing non-physical links. As the numerical example in Section 3.4, the network is assumed to show decreased utility only by the damage of its bridge components. Since only 1,656 bridges contribute on the network (directly and indirectly) among total 1,742 bridges in the data set, we realize ground-motion intensities at 1,656 sites. The seismicity around the urban network is forecasted by Field et al. (2013), and the same parameter is adopted from appendix E of Field et al. (2013) to exploit Magnitude-scaling relationship of Shaw (2009). Total 67 active faults are included in the application (UCERF ver. 3.1 of Field et al. 2013). Next, GMPE of Boore and Atkinson (2008) are employed in this application while the model by Jayaram and Baker (2009) is used to consider the spatial correlation of intra-event residuals at $SA_{1.0sec}$. The fragility model is also adopted from the original data set by the benchmark study (Miller 2014, Miller and Baker 2015), and the same damage status that exceeds 'extensive' is set to the failure of bridge to keep consistency with the study. The extensive damage of the bridge yields non-functional status of links which are influenced by the failure of bridge either directly or indirectly. In the application, we also eliminated all possibility that non-physical links are damaged from the hazardous event because such an effect has not been investigated yet. Figure 3.6 illustrates the road network and 69 active faults surrounding the network. Figure 3.7 shows the super-district nodes of the network as its origins and destinations.

With the point of network analysis, the main difference of our research with the benchmark study is that we adopted user-equilibrium (UE) assignment instead of iterative traffic assignment (or incremental traffic assignment, ITA). Since ITA is a heuristic algorithm, not based on the optimization, its inherent limitation is that the result can be different by assigning traffic flow with different orders; as an optimization-based algorithm, UE does not suffer from such a problem. To solve UE traffic assignment, a Frank-Wolfe based algorithm (FW), suggested by D. van Vliet (1986), is employed again for damaged network cases while gradient projection algorithm (GP) by Michael Florian et al. (2009) is adopted for non-damaged network cases. The GP is initialized with the result of a few iterations by FW. The GP showed relatively fast convergence over FW, but the lack of stability was revealed in several damage-maps. That is why only FW approach is only employed to assign traffic flows to the damage-maps except for the non-damaged cases. However, in most damaged-maps, even FW do not show perfect convergence within an affordable computing time, so that the stopping criteria are set to (a) relative gap (RGAP) in (3.23) less than 0.01 and (b) maximum iteration of 30.

After generating 100,000 damage-maps, there are only 10,507 damaged cases; due to resisting capacity of bridges, there is no damaged bridge in 89,493 cases. The network has probability of 0.0109 that at least a destination node is disconnected from its origin node; in other word, at least 1 o-d pair is cut. Such disconnections were neglected in Miller and Baker (2015), but it is revealed that there are 1,089 connection failures among 100,000 realizations. Through investigations on such cases, the disconnection is mainly contributed from the fact that several super-district nodes are connected directly by physically damageable links. Since there is no agreed research on the effect of disconnection, in this paper, the risk assessment is only focusing on the increased travel time (ITT) for the cases that all o-d pairs are linked. After eliminating all failures in o-d pairs, only 9,418 realizations are collected, while 7,106 damaged samples are unique. The computational cost is still affordable to check connectivity for all O-D pairs when it is compared with that of flow analysis; using parallel computing, only 75 seconds is required to check the connectivity with 102 CPUs of Intel Xeon E5-2407 2.20GHz and 312 GB of 6 nodes RAM while 23,031 seconds is required to analyze post-hazard performance of the network with the same resources (See third and sixth rows of Table 3.4).

Table 3.3. Error measurement for each k value (Bay Area road network)

ITT_{max} ($\hat{P}_E^{mcs}$)	50 (3.8×10^{-3})	100 (1.5×10^{-3})	150 (8.99×10^{-4})	300 (2.99×10^{-4})	600 (1.33×10^{-4})	900 (5.12×10^{-5})
Clustering-based approach (1,500)	0.0294	0.0528	0.0998	0.1328	0.1270	0.0901
Clustering-based approach (1,000)	0.0422	0.0681	0.1477	0.1772	0.1279	0.2677
Clustering-based approach (500)	0.1018	0.2228	0.3950	0.5708	0.5772	0.7682

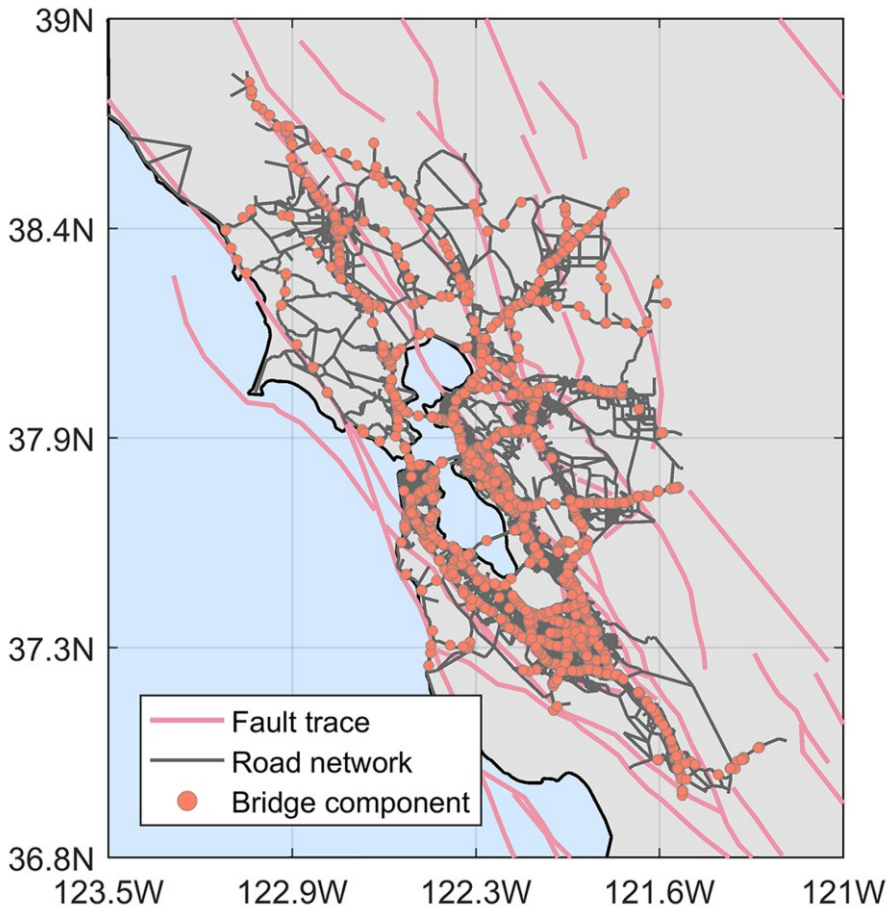

Figure 3.6. Illustration of San Francisco Bay Area road network

Table 3.4. Computational time of each step for risk assessement (Bay Area road network)

Step	**Brute-force MCS**	**Clustering-based approach (1,500 means)**	**Computer spec.**
Generating samples (10^5)	1,603 sec		CPU: Intel i5-4570 3.2GHz (1 core) RAM: 16.0 GB
Connectivity check	75 sec		CPU: Intel(Xeon) E5-2407 2.20GHz (102 cores) RAM: 312 GB x 6 nodes
Evaluating proxy measures	-	313 sec	CPU: Intel(Xeon) E5-2407 2.20GHz (1 core) RAM: 16.0 GB
k-means clustering	-	84 sec	CPU: Intel(Xeon) E5-2407 2.20GHz (1 core) RAM: 16.0 GB
Network analysis	23,031 sec (6.40 hr)	4,782 sec (1.32 hr)	CPU: Intel(Xeon) E5-2407 2.20GHz (102 cores) RAM: 312 GB x 6 nodes
Total	24,709 sec (6.86 hr)	6,857 sec (1.90 hr)	

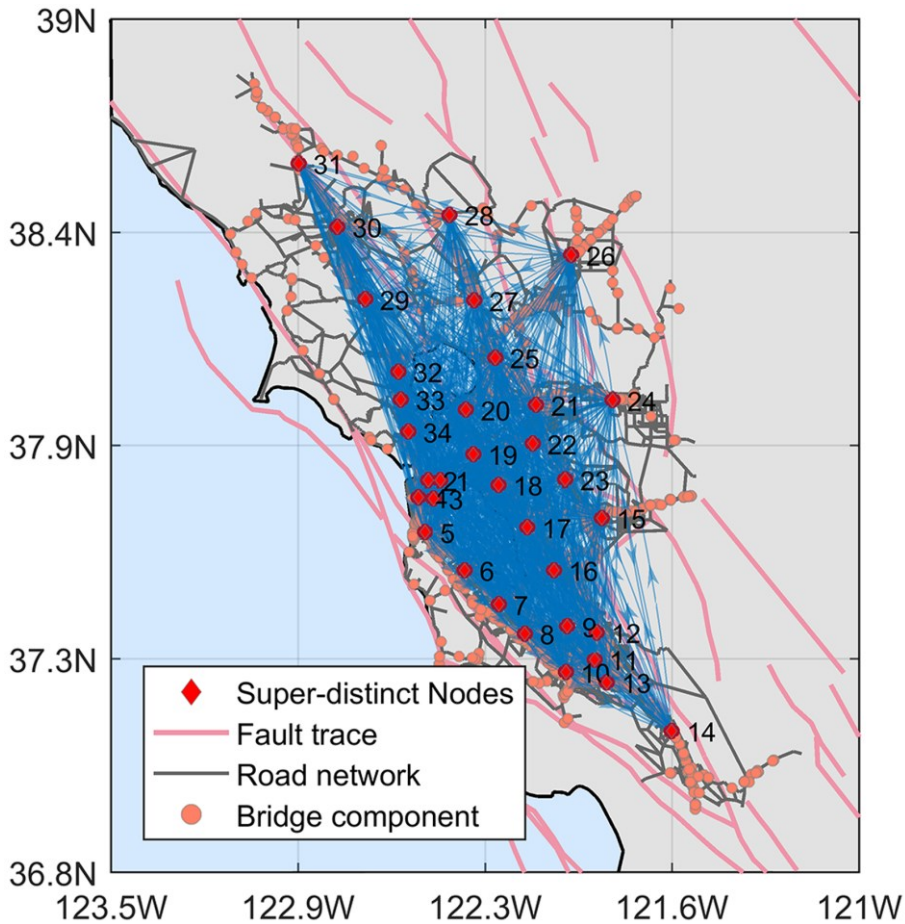

Figure 3.7. Super-distinct nodes and demands of the network

For brute-force Monte-Carlo simulation (MCS), total 7,107 traffic assignments were implemented, which requires 6.29 hours of computational time with 102 CPUs of Intel Xeon E5-2407; the unique damaged samples include 7,106 unique damage-maps and 1 non-damaged case. In the clustering-based approaches, the data is grouped into 500, 1,000, and 1,500 clusters, respectively. Since *k*-means clustering algorithm results different outcomes depending on its initializations, option for replicates is set to 20 to obtain the reproducible result. In other word, 20 independent clustering's are implemented on the same data with random initializations for each *k* value, but a representative outcome is selected as the global solution when its objective function shows the minimum value among the independent executions.

Figure 3.8 represents the approximated profiles that are obtained by the clustering-based approach; the dark red line represents the result by brute-force

MCS with 7,107 network analyses while the each green/blue/red line represents the result by 500/1,000/1,500-means clustering, respectively. The same measurement in (3.24) is employed to evaluate the error of the approximations with the base-line (brute-force MCS) in Table 3.3. With the increasing number of clusters, the approximation show higher accuracy, but there is a trade-off between the accuracy and rapidity of the approximation. The computation time for 1,500 clusters is provided in Table 3.4 since the approximation with 1,500 clusters showed a quite acceptable accuracy with significant evaluation points. The about only 2 hours were required for whole procedure of the proposed clustering-based approach while almost 7 hours were required by brute-force MCS. From this observation, it seems like that our proposed approach has a great merit with the point of computational cost. Since we did not allowed parallel computing in Step 1 (Generating samples) during the application, there is room for additional improvement if parallel computing-based algorithm is implemented in realizing damage-maps. Similarly with Figure 3.5, Figure 3.9 is provided to compare the data on the normalized feature space; however, the color-map of data in Figure 3.9 is provided in logarithmic scale due to the wide variability of the target performance measure. The same result with Figure 3.5 is extracted from Figure 3.9; in spite of the minor differences, the overall tendency of target performance measure (ITT) can be captured by clustering data according to their proxy measures.

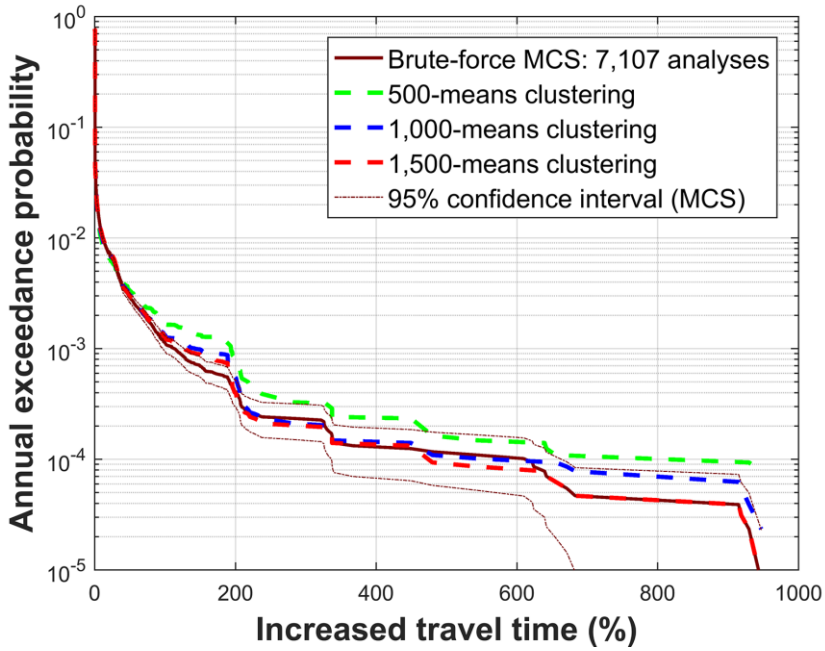

Figure 3.8. Annual exceedance probability curves

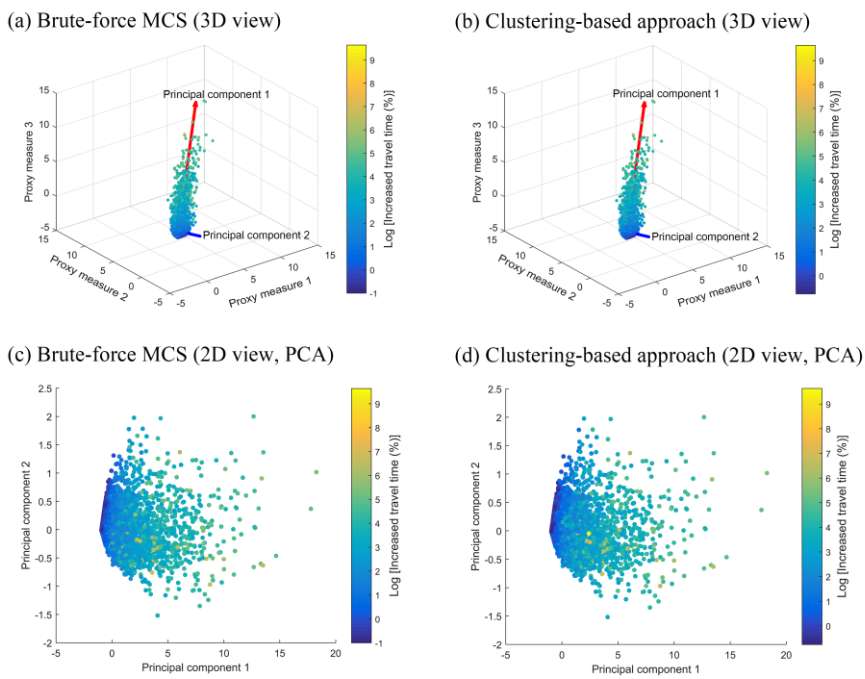

Figure 3.9. Scatter plots of data on the feature space

3.6. Conclusion

A computationally efficient framework is introduced for probabilistic risk assessment (PRA) on spatially distributed urban network. By succeeding two benchmark studies (Jayaram and Baker 2010, Miller and Baker 2015), the proposed method has a merit on computational cost while not requiring any problem-specific parameter. We identify the similarity between samples (damage-maps) only through the realized damage of infrastructures rather than using additional information on the hazard. That is because (a) the outcome of structural damage would be different even after the same realized ground-motion intensities in the PRA framework and (b) the spatial patterns of ground-motion intensities are already included in the damage patterns of infrastructures while realizing structural damage does not require particularly large computation.

In the proposed framework, the damage-maps are clustered by the indirect post-hazard performance indicators, as known as proxy measures. Such features are employed to compare the similarity of samples with their convenient computation. Except for their fast evaluation, another merit of proxy measures is that the reduced dimensionality of data provides comfortable environment for the clustering algorithm. The damage-maps are clustered into a few groups according to their similarities by using the *k*-means clustering. In final, representative network analyses are performed for each cluster, and the representative value is assigned to all damage-maps within the cluster while effectively decreasing the number of network analyses.

The proposed approach is applied into the hypothetical road network and the San Francisco Bay Area road network to check its validity. Through the applications, quite acceptable approximated profiles are obtained for annual probability of post-hazard performance of the network with reduced usage on computing resources. Such an approach is expected to contribute not only to establish an efficient framework for PRA at community-level with constraint computing resource, but also to increase the communication of our community with urban disaster and resilience. On the extended discussion, further researches are required on features of infra-network to establish a more effective and efficient framework for risk identification and appropriate response against urban disaster.

Appendix 3A. Scenario studies on Bay Area road network example

In this section, studies on three representatively selected scenarios are demonstrated to increase understanding on the mechanism of urban network after hazardous events. Figure 3A.1 represents a set of sequential realizations of the earthquake rupture, ground-motion intensities (spectral accelerations at 1.0 sec, $SA_{1.0\ sec}$), structural damage on bridges, and increased travel time (ITT) on road links where the earthquake is assumed to be occurred on the Central Calaveras with moment magnitude of 6.9 (Scenario 1). Similarly, Figure 3A.2 and 3A.3 illustrate the outcomes of events on Green Valley with moment magnitude 6.9 (Scenario 2) and San Andreas (Peninsula) with moment magnitude 7.5 (Scenario 3), respectively.

According to the main text, the proxy measures and target performance measure are summarized as:

- NDB: Number of damaged bridges
- LTT: Lost {travel time}×{traffic flow}
- LCT: Lost {capacity}×{traffic flow}
- ITT: Increased travel time (in percentage)

where the formulaic definitions follows (3.18), (3.19), and (3.20), respectively. Table 3A.1 represents the evaluated proxy measures and target performance measure of three damage-maps, and it is expected that one can increase his or her insight on post-hazard behavior of urban network by comparing these values. From Table 3A.1, one can check that the Scenario 1 is selected because its increased travel time (ITT) is corresponding to one of jumping points on the risk profile in the Figure 3.8 while Scenario 2 and 3 have the maximum values in ITT and NDB, respectively; since the Scenario 2 represents an extreme ITT value in Figure 3.8, the data point has been omitted on the plot.

Table 3A.1. Evaluated proxy measures and target performance measure of the three scenarios

	Scenario 1	Scenario 2	Scenario 3
NDB	103	29	252
LTT	2.620×10^7	7.894×10^6	4.057×10^7
LCT	2.543×10^9	9.298×10^8	5.349×10^9
ITT (%)	649.703	1519.378	224.737

From the Table 3A.1 and Figure 3A.1, Scenario 1 is interpreted as the situation

that a locally focused traffic jam is triggered severely by damage of bridges on that area, where road components are densely located and higher traffic demands exist. The scenario implies that the functionally important components are concentrated on the specific region, which makes the urban community vulnerable to specific external impacts that can be induced by the natural or man-made hazards. Next, Scenario 2 and 3 are compared to discuss about the existence of key-elements in the lifeline network. For example, Scenario 2 shows intense ITT value in spite of the small values of proxy measures; such a situation is activated by only 29 failures among total 1,656 bridges. It looks like that the severe damage on the network is mainly contributed by two reasons: (a) direct damage of important elements and (b) poor utility of back-up systems (alternative route). On the other hand, Scenario 3 indicates relatively affordable ITT value in spite of the extreme NDB value. Although there are amount of non-functional bridges in Scenario 3, the situation is comprehended that a quite reasonable alternative routes are provided for that event after severe damage on the components.

As a result, a common result is extracted from the case studies that risk assessment on infrastructures should be implemented not only at the component-level fragility but also in the context of network-level performance to prepare and response moderately against the urban disaster. However, it is really hard to fully understand the post-hazard mechanism of urban network, so that demands on further researches exist to develop and extract key-features of the network. The proposed framework can be comprehended as a part of extension of such researches, especially to improve the efficiency of the framework for risk identification. As the outcomes of further efforts, other valuable products are expected to be achievable while increasing the communication with urban disaster and resilience.

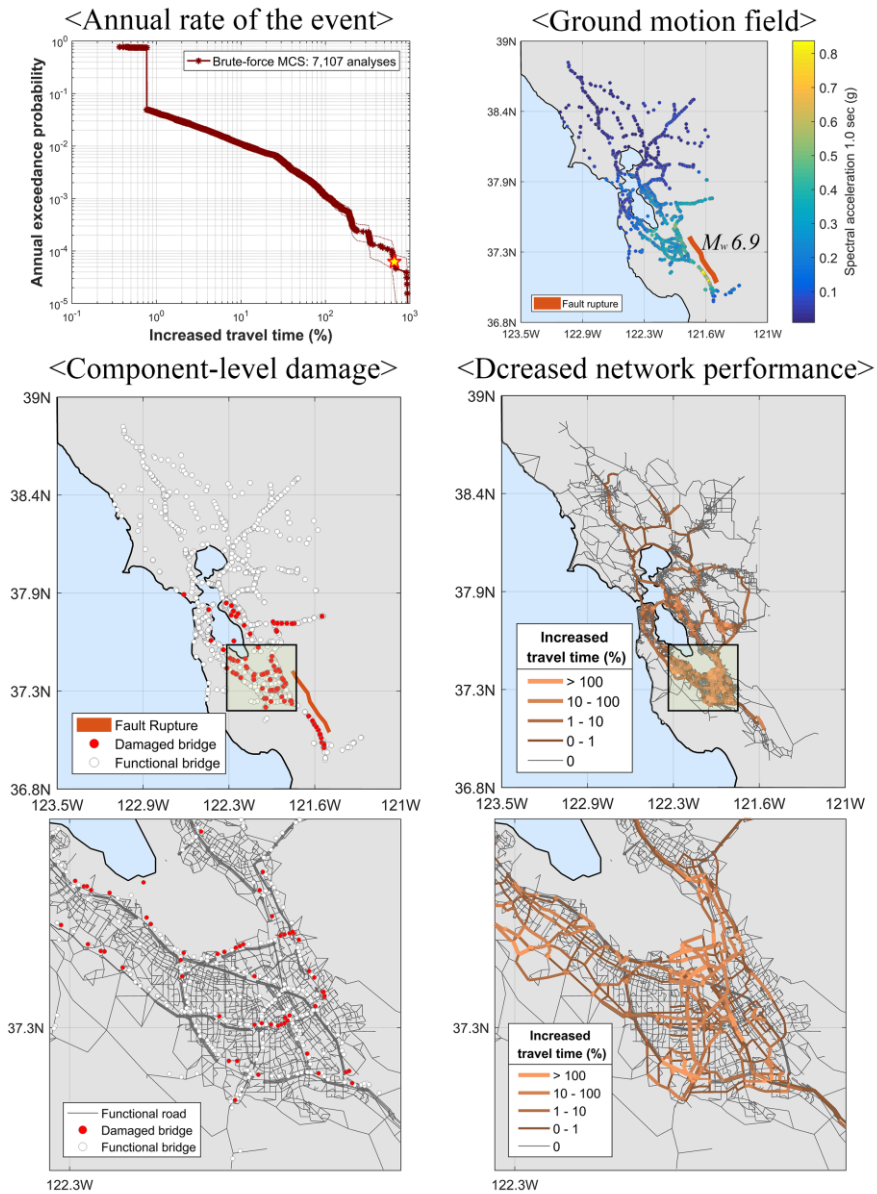

Figure 3A.1. Scenario study on earthquake event with moment magnitude of 6.9 on Calaveras (Central) 2011

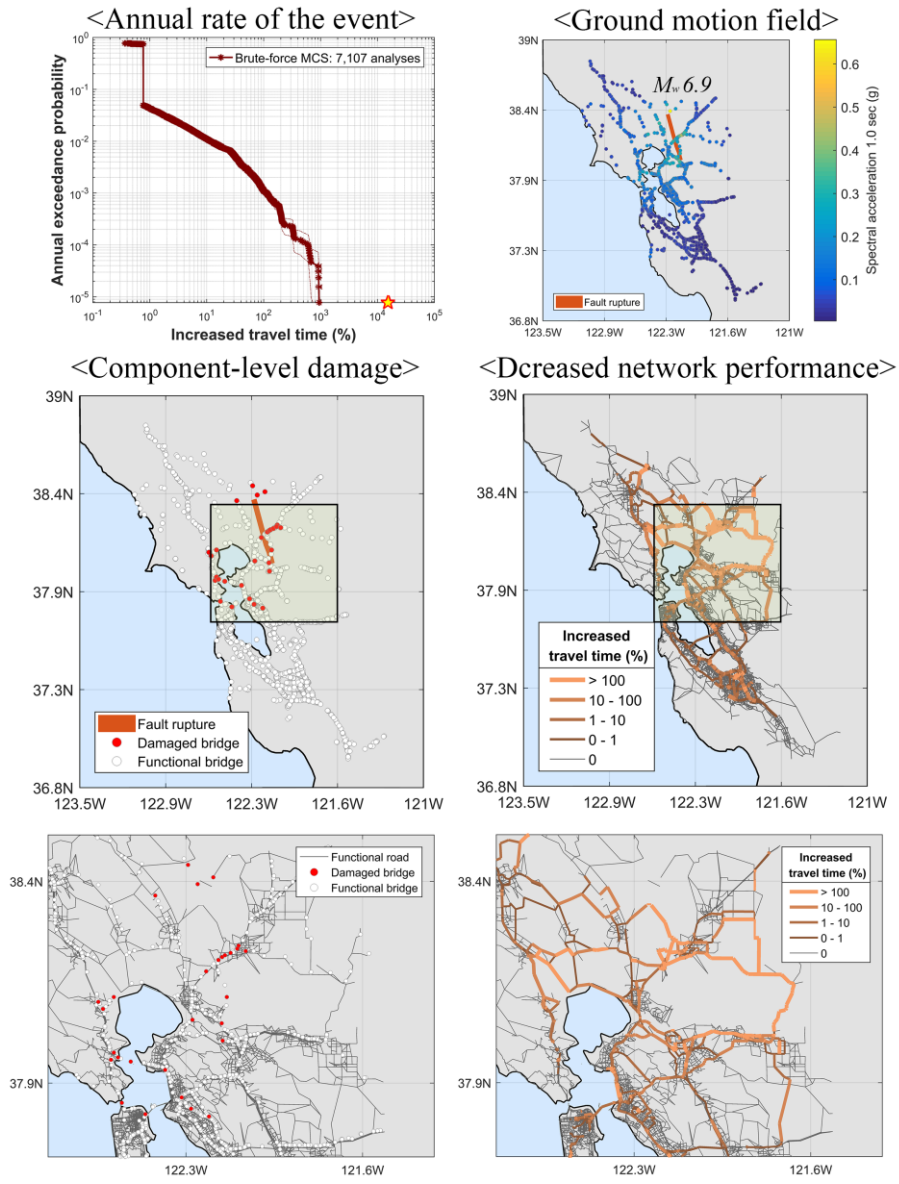

Figure 3A.2. Scenario study on earthquake event with moment magnitude of 6.9 on Green Valley 2011

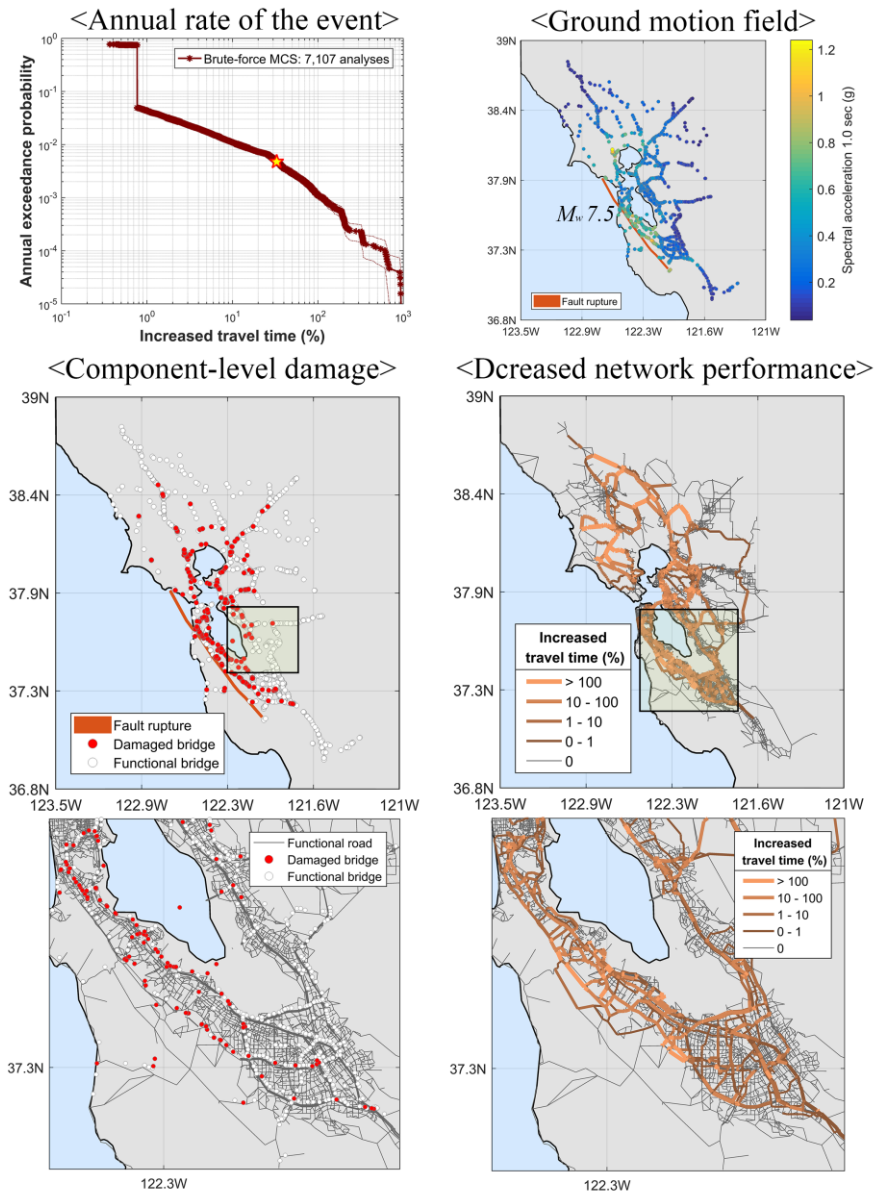

Figure 3A.3. Scenario study on earthquake event with moment magnitude of 7.5 on San Andreas (Peninsula) 2011

Chapter 4. Conclusion

New simulation-based methods are proposed for community-level probabilistic seismic risk assessment (PRA) in Chapter 2 and 3.

In Chapter 2, a new importance sampling (IS) based framework is established for community-level probabilistic seismic risk assessment (PRA) with two applications, (a) post-hazard traffic flow capacity and (b) aggregated regional monetary loss. The proposed approach, termed as cross-entropy-based concurrent adaptive importance sampling (CE-CAIS), improves the efficiency of the simulation-based method when multiple levels of probabilities need to be estimated. As a variation of CE-AIS, the proposed approach does not require prior knowledge, assumption or additional reliability analysis to identify important regions in selecting the importance sampling (IS) density. While CE-AIS needs to repeat pre-sampling to obtain a near-optimal IS density for each target event, CE-CAIS finds a near optimal IS density which provides fairly well and speedy convergence for all possible events, concurrently. This feature makes the proposed approach superior to the original CE-AIS approach for regional PRA because numerous post-hazard consequences often exist within probabilistic concepts. Numerical examples demonstrated superb efficiency of CE-CAIS compared to brute-force MCS and original CE-AIS. In addition, several techniques were introduced to be plugged into a procedure to facilitate CE-CAIS in our applications: (a) matrix-based system reliability (MSR) method for the example of post-hazard traffic flow capacity and (b) principal component analysis (PCA) and central limit theorem (CLT) for the regionally aggregated monetary loss. Sampling-based MSR approach provides a way to obtain conditional system reliability of traffic network while PCA and CLT was used to reduce the dimensionality of random variables, so that CE-CAIS was successfully applied to numerical examples. With these additional techniques on main methodology, CE-CAIS found a near-optimal IS density to provide significant improvement in terms of computational cost not relying on any subjective judgement to select IS density model.

In Chapter 3, a computationally efficient framework is introduced for probabilistic risk assessment (PRA) on spatially distributed urban network. By succeeding two benchmark studies (Jayaram and Baker 2010, Miller and Baker 2015), the proposed method has a merit on computational cost while not requiring any problem-specific parameter. We identify the similarity between samples (damage-maps) only through the realized damage of infrastructures rather than using additional information on the hazard. That is because (a) the outcome of structural damage would be different even after the same realized ground-motion

intensities in the PRA framework and (b) the spatial patterns of ground-motion intensities are already included in the damage patterns of infrastructures while realizing structural damage does not require particularly large computation. In the proposed framework of Chapter 3, the damage-maps are clustered by the indirect post-hazard performance indicators, as known as proxy measures. Such features are employed to compare the similarity of samples with their convenient computation. Except for their fast evaluation, another merit of proxy measures is that the reduced dimensionality of data provides comfortable environment for the clustering algorithm. The damage-maps are clustered into a few groups according to their similarities by using the *k*-means clustering. In final, representative network analyses are performed for each cluster, and the representative value is assigned to all damage-maps within the cluster while effectively decreasing the number of network analyses. The proposed approach is applied into the hypothetical road network and the San Francisco Bay Area road network to check its validity. Through the applications, quite acceptable approximated profiles are obtained for annual probability of post-hazard performance of the network with reduced usage on computing resources. Such an approach is expected to contribute not only to establish an efficient framework for PRA at community-level with constraint computing resource, but also to increase the communication of our community with urban disaster and resilience. On the extended discussion, further researches are required on features of infra-network to establish a more effective and efficient framework for risk identification and appropriate response against urban disaster.

Bibliography

1. Atkinson, G.M. and D. M. Boore. Ground Motion Relations for Eastern North America, Bulletin of the Seismological Society of America 1995, Vol. 85, pp. 17-30.
2. Bai J. W, Hueste M.B.D, Gardoni P, Probabilistic assessment of structural damage due to earthquakes for buildings in Mid-America. Journal of structural engineering 2009, 135(10), 1155-1163.
3. Bommer J, Spence R, Erdik M, Tabuchi S, Aydinoglu N, Booth E, del Re D, Peterken O. Development of an earthquake loss model for Turkish catastrophe insurance. Journal of Seismology 2002; 6(3):431–446.
4. Boore D.M, Atkinson G.M, Ground-motion prediction equations for the average horizontal component of PGA: PGV and 5% damped SA at spectral periods between 0.01 s and 10.0 s. Earthquake Spectra 2008; 24(1):99–138.
5. Bruneau M, Reinhorn A. Overview of resilience concept, Proc. 8th US national Conference on Earthquake Engineering, April 18-22, 2006, San Francisco, CA, USA.
6. Cosentino P, Ficarra V, Luzio D. Truncated exponential frequency-magnitude relationship in earthquake statistics, Bulletin of the Seismological Society of America 67.6 1977, 1615-1623.
7. Der Kiureghian A, First- and Second-Order Reliability Method, Chapter 14: Engineering Desing Reliability Handbook, E. Nikolaidis, D.M. Choicel & S. Singhal (eds), CRC Press, Boca Rapton, FL, 2005.
8. Der Kiureghian A, Song J. Multi-scale reliability analysis and updating of complex systems by use of linear programming, Reliability Engineering & System Safety 2008, 93, 288-297.
9. FEMA-443, HAZUS-MH technical manual, Federal Emergency Management Agency, Washing DC, USA, 2003.
10. Field E.H, Biasi G.P, Bird P, Dawson T.E, Felzer K.R, Jackson D.D, Johnson K.M, Jordan T.H, Madden C, Michael A.J, Milner K.R, Page M.T, Parsons T, Powers P.M, Shaw B.E, Thatcher W.R, Weldon. R.J. II, Zeng Y, Uniform California earthquake rupture forecast, version 3 (UCERF3)—The time-independent model: U.S. Geological Survey Open-File Report 2013–1165, 97 p., California Geological Survey Special Report 228, and Southern California Earthquake Center Publication 1792, 2013, http://pubs.usgs.gov/of/2013/1165/.
11. GEM, OpenQuake Engine User Manual v2.1, Global Earthquake Model, 2016, https://www.globalquakemodel.org/media/cms_page_media/432/oq-manual-19_1.pdf
12. Jayaram, N, Baker, J. W, Correlation model for spatially distributed ground-motion intensities. Earthquake Engineering & Structural Dynamics 2009; 38(15), 1687-1708.
13. Jayaram N, Baker J.W, Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment, Earthquake Engineering & Structural Dynamics 2010; 38(15):1109-1131.
14. Kang W.-H, Lee Y.-J, Song J, Gencturk B, Further development of matrix-based system reliability method and applications to structural systems, Structure and Infrastructure Engineering: Maintenance, Management, Life-

cycle Design and Performance 2012, Vol. 8(5):441-457.
15. Kang W.-H, Song J, Gardoni P, Matrix-based system reliability method and applications to bridge networks. Reliability Engineering & System Safety 2008, Vol. 93:1584-1593.
16. Kim Y.S, Seismic loss assessment and mitigation of critical urban infrastructure systems. Thesis (PhD). University of Illinois at Urbana-Champaign, 2007.
17. Kurtz N, Song J, Cross-entropy-based adaptive importance sampling using Gaussian mixture, Structural Safety 2013, Vol. 42:35-44.
18. Lee Y.-J, Song J, Gardoni P, Lim H.-W, Post-hazard flow capacity of bridge transportation network considering structural deterioration of bridges, Structure and Infrastructure Engineering: Maintenance, Management, Life-cycle Design and Performance 2011; Vol. 7(7):509-521.
19. Lim H.-W, Song J, Efficient risk assessment of lifeline networks under spatially correlated ground motions using selective recursive decomposition algorithm, Earthquake Engineering and Structural Dynamics 2012; Vol. 41(13):1861-1882.
20. Lim H.-W, Song J, Kurtz N, Seismic reliability assessment of lifeline networks using clustering-based multi-scale approach. Earthquake Engineering and Structural Dynamics 2015; Vol. 44(3):355-369.
21. Loth C, Baker J.W, Spatial cross-correlation of spectral accelerations at multiple periods: model development and risk assessments considering secondary earthquake effects, Technical report; USGS Award G10AP00046, 2011.
22. Miller M, Seismic risk assessment of complex transportation networks. Thesis (PhD). Stanford University, 2014.
23. Miller M, Baker J.W, Ground-motion intensity and damage map selection for probabilistic infrastructure network risk assessment using optimization, Earthquake Engineering & Structural Dynamics 2015; 44(7):1139–1156.
24. Rubinstein R, Simulation and the Monte Carlo methods. Hoboken (NJ): John Wiley and Sons, Inc.; 1981.
25. Rubinstein R, Kroese D, The cross entropy method: a unified approach to combinatorial optimization, Monte-Carlo Simulation, and machine learning, Springer. Inc, New York (NY), 2004.
26. Shaw B.E, Constant stress drop from small to great earthquakes in magnitude-area scaling, Bulletin of the Seismological Society of America 2009; v. 99; no. 2A:871–875.
27. Song J, Kang W.-H, System reliability and sensitivity under statistical dependence by matrix-based system reliability method. Structural Safety 2009, Vol. 31(2):148-156.
28. Song J, Ok S.-Y. Multi-scale system reliability analysis of lifeline networks under earthquake hazards. Earthquake Engineering & Structural Dynamics 2010, 39(3), 259-279.
29. Steelman J, Hajjar J.F, Capstone Scenario Applications of Consequence-Based Risk Management for the Memphis Testbed; A Report of the Mid-America Earthquake Center, Mid-America Earthquake Center, 2008.
30. Steelman J, Song J, and Hajjar J.F, Integrated data flow and risk aggregation

for consequence-based risk management of seismic regional losses; A Report of the Mid-America Earthquake Center, Mid-America Earthquake Center, University of Illinois, 2007.
31. Ukkusuri S, Accounting for uncertainty, robustness and online information in transportation networks. Thesis (PhD). University of Texas at Austin, 2005.
32. Shiraki N, Shinozuka M, Moore JE, Chang SE, Kameda H, Tanaka S. System risk curves: Probabilistic performance scenarios for highway networks subject to earthquake damage. Journal of Infrastructure Systems 2007; 13(1):4354.
33. Wang Z, Song J, Cross-entropy-based adaptive importance sampling using von Mises–Fisher mixture for high dimensional reliability analysis, Structural Safety 2016, Vol. 59: 42-52.

Abstract

공간적으로 분포한 건축물과 인프라 시설물이 도시 지역에서 중요한 가치를 지닌다는 점에서 지진 재난 리스크 분석은 필수적이다. 그리고 지진과 같은 자연 재해로 인한 도시지역의 재난 리스크를 분석하는 데 있어서 재해의 특성과 구조물의 거동 등에서 다양한 불확실성이 존재하므로 지역단위에서의 재난 리스크는 필연적으로 확률론적 해석에 기반하여 평가되어야 한다. 지역단위 리스크 분석을 수행되는 데 있어서 몬테-카를로 시뮬레이션 (Monte-Carlo Simulation, MCS)은 그 간편한 시행 방법을 이유로 가장 근본적인 확률해석기법으로 사용되어 왔으나, 몬테-카를로 시뮬레이션은 낮은 수준의 확률을 갖는 사건을 규명하는 데 많은 계산소요를 요구한다는 점에서 근본적인 결함을 가지고 있다. 이러한 낮은 확률의 사건은 도시지역과 지역사회에 심각한 영향을 끼치는 경우가 많으므로, 올바른 재해에 대한 대비와 대응을 위해 필연적으로 규명되어야 할 필요가 있다. 이러한 문제점을 극복하기 위해 본 연구는 지역단위를 대상으로 한 확률론적인 지진 재난 리스크 분석을 위하여 계산상으로 효율적인 프레임워크를 구축하는 것을 목표로 한다. 이에 대한 첫번째 접근법으로서 Chapter 2에서는 cross-entropy을 기반으로 한 적응형 중요 샘플링 기법(adaptive importance sampling)을 사용하여 효율적인 확률 시뮬레이션을 제시한다. 그리고 복잡한 라이프라인 네트워크에 대한 보다 효율적인 확률기반의 리스크 분석 기법을 제시하기 위해 군집화(clustering)기법을 바탕으로 한 리스크 분석 방법이 Chapter 3에서 제시된다. Chapter 2와 3에서 제시된 기법마다 그 적용성을 검토하기 위해서 몇 가지의 수치적 예제가 제공되며, 다양한 후속 연구를 통해서 지역 사회에 대한 재난과 복원력에 대한 이해를 증진시킬 수 있는 결과물을 얻어낼 수 있을 것으로 예상된다.