UNIVERSITY OF CALIFORNIA

Los Angeles

Probabilistic Evaluation of Seismic Levee Performance using Field Performance Data

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Civil Engineering

by

Dong Youp Kwak

2014

ABSTRACT OF THE DISSERTATION

Probabilistic Evaluation of Seismic Levee Performance using Field Performance Data

by

Dong Youp Kwak

Doctor of Philosophy in Civil Engineering

University of California, Los Angeles, 2014

Professor Jonathan P. Stewart, Chair

I characterize the seismic fragility of levees along the Shinano River system in Japan using field performance data from two **M** 6.6 shallow crustal earthquakes. I quantify levee damage using crack depth, crack width, and crest subsidence for 3318 levee segments each 50 m long. Variables considered for possible correlation to damage include peak ground velocity (*PGV*), geomorphology, groundwater elevation, and levee geometry. For site conditions beneath levees without geophysical measurements, a model for shear wave velocity is proposed considering soil type, penetration resistance, vertical effective stress, geomorphology, and spatial variation of boring-to-boring residuals. Seismic levee fragility is expressed as the probability of exceeding a damage level conditioned on *PGV* alone and *PGV* in combination with other variables. The probability of damage (at any level) monotonically increases from effectively zero for $PGV < 14$

cm/s to approximately 0.5 for $PGV \approx 80$ cm/s. Of the additional parameters considered, groundwater elevation relative to levee base most significantly affects fragility functions, increasing and decreasing failure probabilities (relative to the PGV-only function) for shallow and deep groundwater conditions, respectively.

I demonstrate applicability of the fragility models developed from data in the Shinano River region of Japan (SRJ) for geotechnical conditions along urban levees in the Central Valley region of California (CVC) by comparing penetration resistance data between regions for common soil types and geology. For Holocene flood plain deposits I find penetration resistance for coarse-grained soils in the SRJ and CVC study regions to be similar, whereas for fine-grained soils the CVC sediments are stiffer. For Holocene basin and Pleistocene deposits mostly appeared only in the CVC, both coarse- and fine-grained deposits are stiffer than Holocene floodplain deposits.

Spatial correlations of demand and damage are important to consider when evaluating the performance of a complete levee system. I develop a numerical methodology to evaluate the system fragility utilizing the spatial correlations of damage capacity and demands between segments. System level damage probabilities are found to decrease with the strength of these correlations. The damage demand exhibits positive correlation over larger distances than does the damage capacity.

The dissertation of Dong Youp Kwak is approved.

Robert E. Kayen

Hongquan Xu

Jian Zhang

Scott J. Brandenberg

Jonathan P. Stewart, Committee Chair

University of California, Los Angeles

2014

... To my beloved family.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This research pursued for my Ph.D. was supported by California Department of Water and Resource (CDWR) under contract number 4600008849. This support is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect those of the CDWR. I also acknowledge data providers, i.e., Ministry of Land, Infrastructure, Transportation and Tourism (MLIT), Shinano River Work Office (SWO), Niigata Prefectural Office (NPO), National Research Institute for Earth Science and Disaster Prevention (NIED), Geological Survey of Japan (GSJ), East Japan Railway Company, NEXCO East Japan, and Kashiwazaki City, which made this work feasible. I also thank my collaborators, Atsushi Mikami at the University of Tokushima in Japan, Raymond B. Seed at the University of California at Berkeley (UCB), Ariya Balakrishnan at CDWR, Leslie F. Harder, Jr. at HDR Inc., and Vlad G. Perlea at U.S. Army Corps of Engineers (USACE) for their valuable comments and suggestions.

I would like to show my gratitude to my advisor, Professor Jonathan P. Stewart, and co-advisor, Professor Scoot J. Brandenberg, who supported and guided me personally and academically not only for Ph.D. degree but also for living in Los Angeles during my stay in UCLA. I really appreciate their help for me. I also thank to my committee members and faculty members, Robert E. Kayen at United States Geological Survey (USGS), Mladen Vucetic, Christine Goulet, and Jian Zhang at Civil and Environmental Engineering, Hongquan Xu and Nicolas Christou at Statistics in UCLA, who were willing to share their knowledge and gave valuable comments for this work. I also thank to all my colleagues and friends, Kamil Afacan, Emel Seyhan, Eric Yee, and Lisa Star whom I shared great memory with in ULCA, Jinha Kang who was my roommate giving great tips always, and Jungin Choi and Minteak Yoo who were the Korean gang including me in our office. It was great pleasure to be with them in and out of the school.

Finally, I would like to thank to my family, my wife Junghwa Lee for being there as a friend and a supporter, and my son Dustin Kwak, celebrating 100 days old, for growing up well. From them I could learn what is the most important value in my life.

CURRICULUM VITAE

2007	B.S., Civil Engineering Hanyang University, Seoul, Korea
2007-2009	Graduate Research Assistant Department of Civil and Environmental Engineering Hanyang University, Seoul, Korea
2008	Teaching Assistant Department of Civil and Environmental Engineering Hanyang University, Seoul, Korea
2009	M.S., Civil Engineering Hanyang University, Seoul, Korea
2009	Research Assistant Hanyang University, Seoul, Korea
2010-2014	Graduate Research Assistant Department of Civil and Environmental Engineering University of California, Los Angeles
2010-2014	Teaching Assistant Department of Civil and Environmental Engineering University of California, Los Angeles

SELECTED PUBLICATIONS AND PRESENTATIONS

Journal Papers

Kwak, **D. Y.**, J. P. Stewart, S. J. Brandenberg, A. Mikami (2014). Characterization of seismic levee fragility using field performance data, submitted to *Earthquake Spectra*.

Boroschek, R. L., V. Contreras, **D. Y. Kwak**, and J. P. Stewart (2012). Strong Ground Motion Attributes of the 2010 MW 8.8 Maule, Chile, Earthquake, *Earthquake Spectra*, **28**(S1), S19-S38.

Park, D., **D. Y. Kwak**, C. G. Jeong, and T. Park (2012). Development of probabilistic seismic site coefficients of Korea, *Soil Dynamics and Earthquake Engineering*, **43**(9), 247-260.

Park, D., M. Sagong, **D. Y. Kwak**, and C. G. Jeong (2009). Simulation of Tunnel Response under Spatially Varying Ground Motion, *Soil Dynamics and Earthquake Engineering*, **29**(11), 1417-1424.

Conference Presentations

Kwak, **D. Y.**, S. J. Brandenberg, A. Mikami, A. Balakrishnan, and J. P. Stewart (2014). Applicability of levee fragility functions developed from Japanese data to California's Central Valley, *2014 Conference of United States Society on Dams*, San Francisco, California, Oral Presentation by Dong Youp Kwak.

Kwak, **D. Y.**, S. J. Brandenberg, J. P. Stewart, and A. Mikami (2012). Groundwater level evaluation for river flood control levees and its effect on seismic performance, *15th World Conference on Earthquake Engineering*, Lisbon, Portugal, Poster Presentation.

Kwak, **D. Y.**, A. Mikami, S. J. Brandenberg, and J. P. Stewart, (2012). Ground motion estimation for evaluation of levee performance in past earthquakes, *9th International Conference on Urban Earthquake Engineering / 4th Asia Conference on Earthquake Engineering*, Tokyo, Japan, Oral Presentation by Dong Youp Kwak.

Kwak, **D. Y.**, C. G. Jeong, D. Park, and S. S. Park (2008). Comparison of Frequency Dependent Equivalent Linear Analysis Methods, *14th World Conference on Earthquake Engineering*, Beijing, China, Poster Presentation.

Kwak, **D. Y.**, J. K. Ahn, D. Park, and Y. M. A. Hashash (2008). Influence of Rate Dependent of Soil Behavior on Propagated Ground Motion, *Conference of Geotechnical Earthquake Engineering & Soil Dynamics IV*, Sacramento, California, Poster Presentation.

Reports

Stewart, J. P., **D. Y. Kwak**, S. J. Brandenberg, and A. Mikami (2013). *Characterization of seismic fragility of levees using field performance data*, Project Report for California Department of Water Resources, UCLA, 142 pp.

1 STUDY OVERVIEW

1.1 INTRODUCTION

A levee is a natural or artificial embankment that provides flood protection adjacent to rivers or coastal areas. Most often flood control levees do not routinely retain water, serving that function only during flood events that are unlikely to coincide with a major earthquake. The objective of this research is to empirically characterize the seismic fragility of flood control levees from experience in a region where levee systems have been strongly shaken by multiple shallow crustal earthquakes.

Because levees are often constructed on soft soils, seismic hazards are generally driven by ground failure involving weak and potentially liquefiable soils in the foundations and in the levees themselves. Recently developed levee design standards consider seismic demands (USACE, 2011a; Sugita and Tamura, 2008; MLIT, 2012a), but the principal problem remains the substantial levee networks already in place that were not properly engineered at the time of their original construction.

Several prior studies have examined individual case histories of seismic levee failures, typically from liquefaction of embankment or foundation materials (Sasaki, 2009; Miller and Roycroft, 2004). The present work is fundamentally different in scope in two respects: (1) instead of looking at individual deformed sections, I systematically examine levee performance at a regular spacing interval along a river system, including segments with and without ground deformations; (2) I analyze damage relative to simple parameters representing seismic demand and levee/ground conditions in lieu of detailed, site-specific geotechnical analysis.

My results are expressed in terms of fragility functions that give the probability of damage as a function of ground motion intensity and other relevant factors. These are not the first fragility functions that have been developed for levees. Salah-Mars et al. (2008) estimated fragility for levees in the California Bay Delta region based on numerical analyses of seismic levee deformation potential combined with judgment-based relations for breach probability conditional on crest settlement. Rosidi (2007) evaluated levee fragility in a broadly similar manner for generic levee sections (not specific to a location). Moreover, procedures to estimate levee fragility for non-seismic hazards have been established from analytical simulations by Apel et al. (2004) and Vorogushyn et al. (2009) (instabilities from overtopping and piping, respectively, from river water rise) and from a combination of analysis and observation by Foster et al. (2009) (overtopping and seepage). My study is distinct from prior work in that seismic fragilities are estimated directly from analysis of field performance data, without an underlying numerical model of soil response. My results provide probabilities of various damage states, not of a binary failure or non-failure condition. As such, my work is similar in objective (if not in approach) to the first step of the fragility development process defined by Salah-Mars et al. (2008) and Rosidi (2007) (i.e., computation of deformation given ground motion level). My estimates of fragility are useful for preliminary seismic risk assessments of this critical infrastructure for regions having similar seismologic, hydrologic, and geologic conditions to those in the study region, particularly when detailed geotechnical data is not available.

1.2 SCOPE OF RESEARCH

I have collected information from the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) in Japan regarding levee performance during the 2004 **M** 6.6 Niigata-ken Chuetsu and the 2007 **M** 6.6 Niigata-ken Chuetsu-oki earthquakes. These events were selected because:

i. Levee performance was well documented by staff of the Shinano River Work Office (SWO) under MLIT and the Niigata Prefectural Office agencies (NPO) in Japan (whose staff manually inspected the full length of the levees in the effected regions),

ii. The level of ground shaking varied across the levee system such that some areas were strongly shaken on the surface projection of the fault ruptures (maximum recorded PGA ≈ 1.6g) and experienced damage, whereas other areas experienced more modest shaking and little damage (thereby bracketing a range of responses),

iii. Significant geotechnical data have been compiled for the region as part of engineering investigations, and

iv. The earthquake magnitudes were generally comparable with design-basis earthquakes in other regions where the results are needed for application, including much of California's Central Valley region.

The levees affected by those two earthquakes are along the Shinano River system in northwestern Japan, which has three components - Shinano River mid- and downstream (SH1), Shinano River upstream (SH2), and the tributary Uono River (UO). Figure 1.1 shows the study region including levees along SH1, SH2, and UO and finite fault solutions for the two considered reverse-slip events.

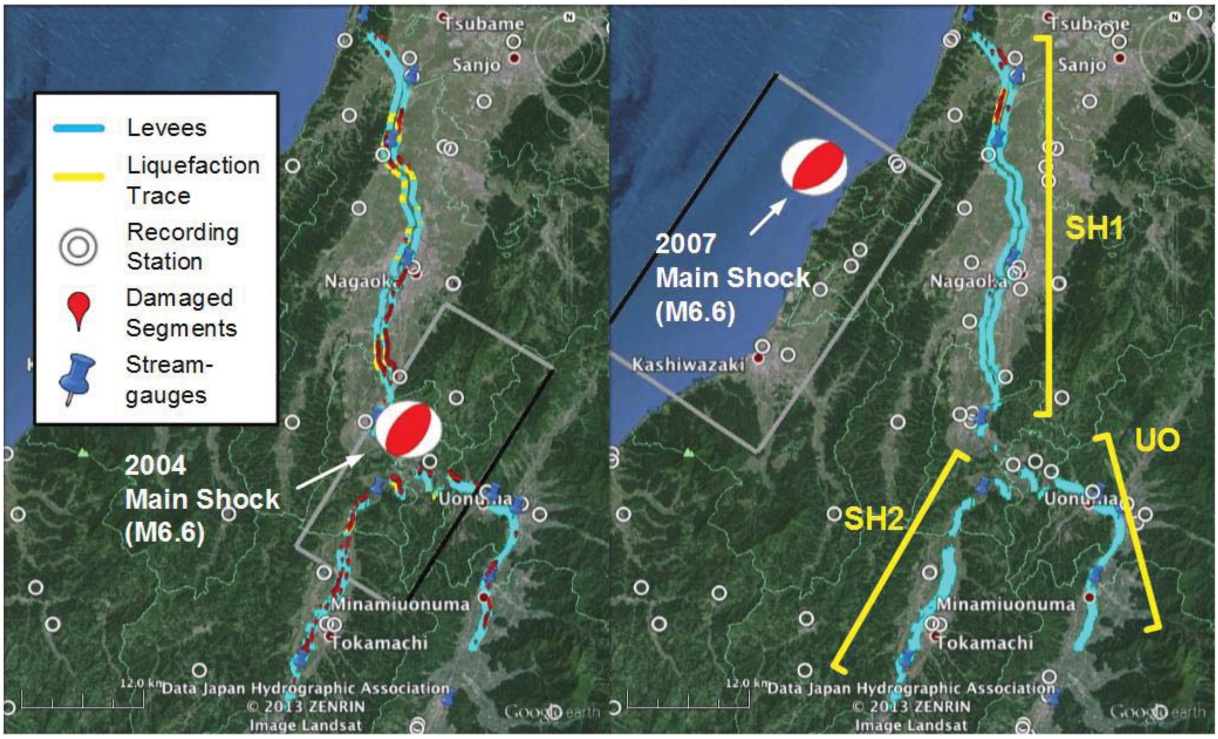

Figure 1.1 **Levees along the Shinano River system (SH1, SH2, and UO) on Google Earth map. Locations of levee damage, liquefaction trace, epicenters (beach balls) and finite fault planes (black line at top). Locations of recording stations and stream gauges. Finite fault solutions from Asano and Iwata (2009) and Miyake et al. (2010).**

In Chapter 2 I describe the collecting effort for damage description and information on levee geometry, geologic conditions, ground water and river water levels, and earthquake ground motion data from the region. I define the respective parameters that I anticipate may correlate to damage rates:

1. Intensity of shaking, as represented by *PGA* and *PGV* (geometric means),
2. Geomorphic conditions beneath the levee, e.g., Deltaic deposits, sand dunes, natural levee deposits, old river channel, etc.,
3. Water elevation in the levee relative to the elevation of the levee base, and
4. Levee cross-sectional shape (i.e., low vs. high aspect ratio).

Relatively straightforward parameterization of geomorphic condition and levee cross-sectional shape are described, whereas intensity of shaking and water elevation that need more detailed analyses are further discussed in Chapter 5. Each of these four factors potentially affecting levee performance are assigned to each 50 m levee segment along with damage level assigned from damage description in the database.

In chapter 3, site conditions between the study region in Japan and the Central Valley region of California are compared. I have collected from California Department of Water Resources (CDWR) geotechnical borehole data from urban levees in the Central Valley region to facilitate comparisons of geotechnical conditions. This comparison gives the idea that the similar level of seismic performance in the study region levees in Japan would be expected for levees in California's Central Valley.

In Chapter 4, a correlation model for shear wave velocity (V_s) is developed. V_s is regressed with respect to soil type, penetrations resistance, and vertical effective stress considering geomorphology and spatial variation of boring-to-boring residual (data minus model). This model is utilized to evaluate site conditions in seismic stations and levee foundations where geotechnical information is available without geophysical measurements.

In Chapter 5, I describe estimation of intensity of shaking and hydrological conditions at the time of earthquakes, which are challenging parameters in terms of spatial and temporal variations. Intensity of shaking, as represented by *PGA* and *PGV* (geometric means), is spatially evaluated in the study region using Kriging methods. Levee ground water elevations at the time of earthquakes are evaluated utilizing river water elevation considering irrigation effect. Hydrological condition is parameterized as levee ground water elevation relative to the levee base.

My database comprises 3318 levee segments of 50 m length for the Shinano River system for both 2004 and 2007 earthquakes. Each 50 m segment has post-earthquake observations of performance following each event from which a damage assignment is made. By combining the damage information with the potentially contributing factors, I have developed models for levee fragility using this data set (presented in Chapter 6).

The analysis of spatial correlations for demand and damage is significant in the performance of a levee system. This issue is particularly important to the application of the present results to a distributed system of levees as would be encountered in practical applications. I develop a methodology properly estimating system fragility incorporating correlations of damage capacities and demands. This work is addressed in Chapter 7 with example of Shinano River levee system.

All findings are summarized in Chapter 8 with recommendations and future research.

2 DATA RESOURCES

2.1 INTRODUCTION

Empirical analysis of levee fragility requires information on the spatial extent of levee damage and its severity, along with various metrics that may correlate to damage. The correlating metrics considered here include ground shaking intensity, geotechnical/geological conditions in the levee and its foundation, hydrological conditions (water level), and geometric parameters related to the levee cross section.

This chapter describes the full suite of data collected for the Niigata region of Japan. The damage status on levees is documented after field reconnaissance along the study region. The specific geotechnical data that was collected consists of geomorphologic and geologic classification from Japanese national maps and borehole data that includes information on soil type and penetration resistance. The hydrological conditions (i.e., ground water level on levee foundation and river water elevation) are collected from boreholes and stream gauges. The intensity measures (*IM*s) are obtained from seismic networks recorded ground motion of the 2004 and 2007 Niigata earthquakes.

2.2 DAMAGE DESCRIPTION

As shown in Figure 1.1, the 2004 earthquake fault plane was located beneath the Shinano River system and produced broadly distributed damage. Many segments experienced strong shaking (up to 1.6g *PGA*). The 2007 earthquake fault plane was located off-shore and produced modest shaking intensity in the study region (0.1~0.4g *PGA*). Damage was concentrated in downstream portions of the levee system. Figure 1.1 also shows locations of surface manifestation of liquefaction, some (but not all) of which are co-located with areas of levee damage.

The locations and severity of damage are based on post-earthquake damage reports by the Shinano River Work Office (SWO, 2007, 2008) and OYO (2008), which measured at regular intervals crack depth and width, vertical slip across cracks, and relative settlement between damaged and undamaged levee sections. The SWO reports also provide a photographic record of the levee performance at regular intervals. Segments without measured damage quantities did not suffer damage beyond a visually apparent level, and are confirmed cases of no damage rather than levee segments that were not inspected. Figure 2.1 shows examples of various damage states.

Figure 2.1 Example of damage states on levee. (a) Damage level 2: crack ~ 7 km inland from ocean at the Shinano River during 2007 earthquake (from OYO, 2008) and (b) damage level 4: lateral spreading ~ 40 km inland from ocean during 2004 earthquake (from SWO, 2008).

I classify damage severity in five levels as shown in Table 2.1 for 50 m (in length) levee segments throughout the Shinano River system (3318 segments up to 80 km from river mouth). To place the subsidence numbers in perspective, average levee heights range from 5.7 to 4.5 m in downstream and upstream areas, respectively, so the subsidence associated with damage level 4 (i.e., > 100 cm) corresponds to at least 20% of the levee height. When the available damage metrics produce different damage classifications for a given levee segment, I select the most severe classification. Of the 3318 segments, damage levels of one or greater were found for 652 segments in the 2004 event and 78 segments in the 2007 event (damage rates of 19.7% and 2.4%, respectively). Figure 2.2 shows percentages of segments in each damage level (*DL*) for the two events.

Table 2.1 Damage levels assigned to levee segments.

Damage Level	Crack depth (cm)	Crack width (cm)	Subsidence (cm)	Description
0	0	0	0	No damage reported
1	0~100	0~10	0~10	Slight damage, small cracks
2	100~200	10~50	10~30	Moderate damage, cracks or small lateral spreading
3	200~300	50~100	30~100	Severe damage, lateral spreading
4	> 300	> 100	> 100	Levee collapse

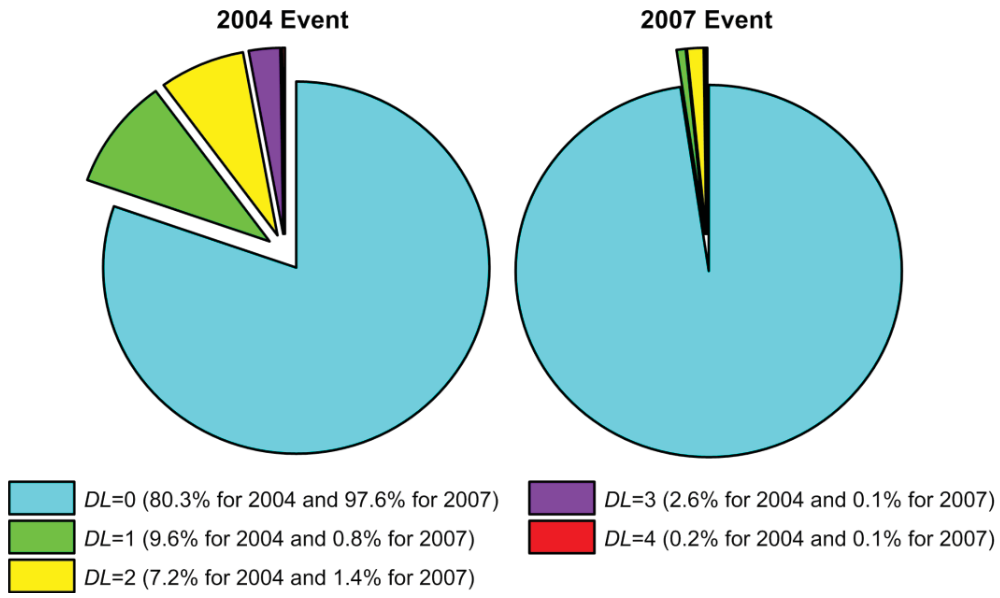

Figure 2.2 Percent of segments for damage level (*DL*) 0-4 from the 2004 and 2007 earthquakes.

Figure 2.3 shows rates of surface manifestation of liquefaction conditional on damage level. Levee segments with no or minor damage levels (*DL* = 0 and 1) have low rates of liquefaction manifestation, whereas levees with moderate to severe damage levels (*DL* > 1) have surface manifestation rates of 50-80%. This indicates that damaged levee segments were often, but not always, accompanied by the surface manifestation of liquefaction.

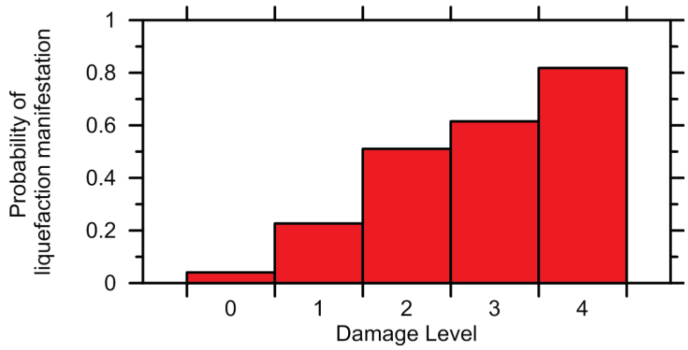

Figure 2.3 Probability of surface manifestation of liquefaction for each damage level.

2.3 EARTHQUAKE RECORDINGS

Ground motion recordings were gathered from four data providers: Japan Society of Civil Engineers (JSCE), National Research Institute for Earth Science and Disaster Prevention (NIED), Japan Meteorological Agency (JMA), and National Institute for Land and Infrastructure Management (NILIM). JSCE provided earthquake strong motion data at a web site (JSCE, 2011) where recordings of the 2007 earthquake were available along with boring logs. East Japan Railway Company, East Nippon Expressway, and Kashiwazaki City Office maintain the networks that provided the ground motion data and boring logs for the JSCE web site, which is no longer accessible as of this writing. I utilize 15 JSCE records from stations located within 100 km of the 2007 event fault plane. NIED maintains two seismic networks known as the Kyoshin Network (K-NET) and the Kiban Kyoshin Network (KiK-net) (NIED, 2012b). I select stations within 100 km of the fault plane, which comprises 55 K-NET stations and 38 KiK-net stations for the 2004 event, and 49 K-NET stations and 30 KiK-net stations for the 2007 event. Each station has a three-component digital strong-motion accelerograph as well as geophysical logs of P and S-wave velocities from downhole measurements. JMA maintains a web site from which data was obtained for this study (JMA, 2012). The seismic stations for which data is distributed on the JMA site are operated both by JMA and local governments. I select 26 JMA stations for the 2004 event and 52 stations for the 2007 event. Due to lack of geophysical measurements for site condition (i.e., time-averaged shear wave velocity up to 30 m depth, V_{s30}) in JMA stations, I utilize the Japan Seismic Hazard Information Station (J-SHIS) database providing V_{s30} based on geomorphology, slope, elevation, and distance from mountain or hill (Matsuoka et al., 2006). NILIM maintains a seismic network comprised of stations near road and river facilities. Data from this network is distributed via a web site (NILIM, 2013) in the form of tabulated intensity measures (maximum of the three components only) and intensities, but digital acceleration histories and component-specific intensity measures are not available. Two seismic stations from this network are located near levees within the study region (9.2 and 40.7 km from river mouth at right-side). Component-specific *IM*s from those stations, as documented by OYO (2008), were used in my study. Locations of the utilized ground motion stations near the Shinano River system are shown in Figure 1.1, and coordinates and station codes are in Table 2.2.

Figure 2.4 shows contour maps of *IM* (i.e., *PGV*) for 2004 and 2007 earthquakes calculated by linear interpolation. However, the use of *IM*s from linear interpolation is not recommended to levees because it does not capture the site effect possibly strong in soft foundations. Section 5.2 describes how the site effect is considered to estimate appropriate ground motions for levee segments.

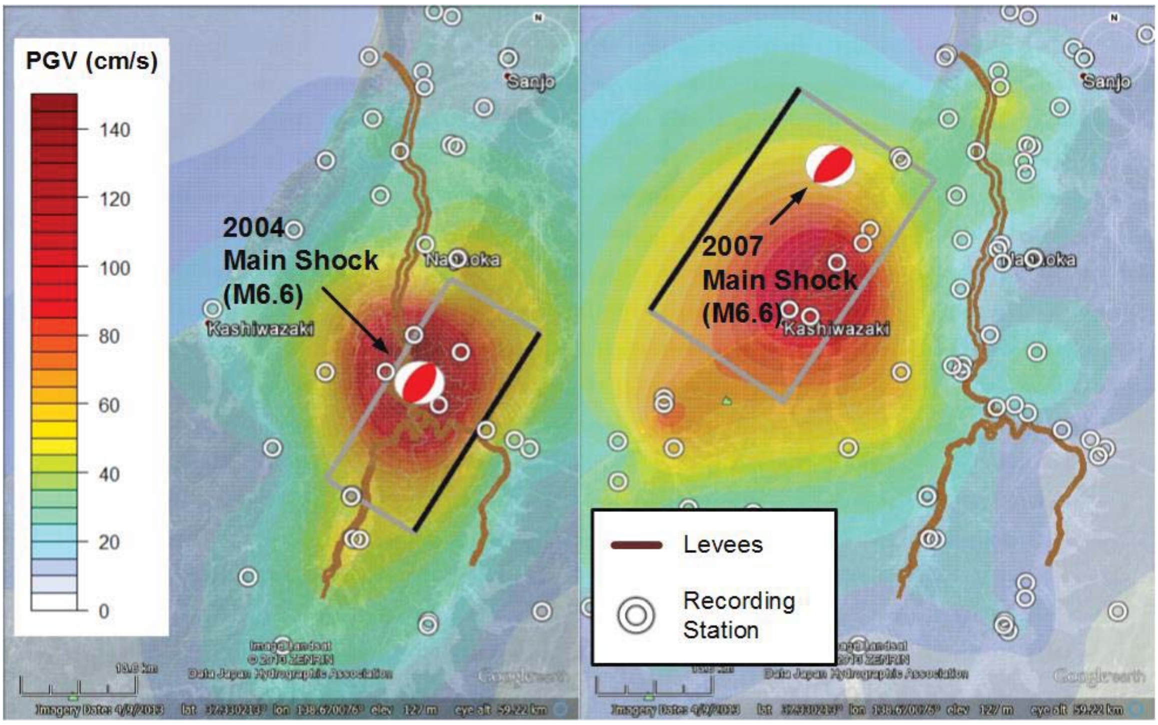

Figure 2.4 **Contour maps *of PGV*s linearly interpolated for 2004 and 2007 earthquakes.**

Table 2.2 **List of seismic stations from various networks around the study region. Joyner-Boore rupture distance (*R_{JB}*; Joyner and Boore, 1981), *PGA*, and *PGV* are indicated.**

		Location			2004 Earthquake			2007 Earthquake		
Site Code	Network	Longitude (deg)	Latitude (deg)	V_{s30} (m/s)	R_{JB} (km)	*PGA* (g)	*PGV* (cm/s)	R_{JB} (km)	*PGA* (g)	*PGV* (cm/s)
FKS021	NIED	139.8633	37.6535	372	83	0.123	5.4	98	0.027	2.2
FKS022	NIED	139.6467	37.6002	212	63	0.144	7.8	79	0.048	3.0
FKS023	NIED	139.9294	37.4774	183	83	0.057	4.5	-	-	-
FKS025	NIED	139.9000	37.3077	608	80	0.055	2.2	-	-	-
FKS026	NIED	139.5386	37.2660	332	49	0.113	4.2	75	0.056	2.3
FKS027	NIED	139.6808	37.0700	690	68	0.073	1.5	96	0.043	0.8
FKS028	NIED	139.3144	37.3491	306	28	0.165	12.1	53	0.057	5.4
FKS029	NIED	139.3802	37.0159	433	49	0.198	3.5	77	0.090	1.6
FKS030	NIED	139.5133	37.4530	496	46	0.120	4.5	67	0.031	1.7
FKSH01	NIED	139.7150	37.7565	704	77	0.050	1.4	88	0.040	0.9
FKSH03	NIED	139.7533	37.6078	350	72	0.085	5.1	-	-	-
FKSH04	NIED	139.8126	37.4508	246	72	0.073	2.8	93	0.027	1.8
FKSH05	NIED	139.8725	37.2544	596	78	0.064	2.7	-	-	-
FKSH06	NIED	139.5198	37.1723	680	50	0.139	4.2	78	0.044	1.8
FKSH07	NIED	139.3756	37.0103	829	49	0.110	2.5	77	0.040	1.2
FKSH21	NIED	139.3146	37.3422	365	28	0.329	15.1	53	0.101	4.1
GNM001	NIED	139.2248	36.7722	500	-	-	-	82	0.096	2.0
GNM002	NIED	138.9695	36.7819	444	43	0.313	6.8	67	0.060	1.5
GNM003	NIED	139.0784	36.6578	374	59	0.318	8.9	84	0.067	5.2
GNM004	NIED	138.5918	36.6172	315	63	0.028	2.1	75	0.043	2.8
GNM005	NIED	138.5177	36.5133	343	76	0.011	2.8	86	0.011	2.1
GNM006	NIED	138.7523	36.5103	370	71	0.033	1.2	88	0.017	1.1
GNM007	NIED	139.0101	36.4619	288	78	0.070	3.1	-	-	-
GNM008	NIED	139.1368	36.3478	371	93	0.033	3.7	-	-	-
GNM009	NIED	139.3251	36.4106	393	93	0.116	3.2	-	-	-
GNM013	NIED	139.0176	36.3181	259	94	0.087	4.9	-	-	-
GNM014	NIED	138.7179	36.3500	580	90	0.018	1.2	-	-	-
GNMH07	NIED	139.2104	36.6998	648	60	0.095	2.4	87	0.038	1.2
GNMH08	NIED	138.5244	36.4917	339	78	0.015	3.4	89	0.021	3.5
GNMH09	NIED	138.9068	36.6212	624	59	0.021	1.8	80	0.013	1.8
GNMH13	NIED	139.0627	36.8620	323	-	-	-	65	0.080	3.4
ISK002	NIED	137.2877	37.4443	167	-	-	-	96	0.093	11.9
ISKH01	NIED	137.2844	37.5266	345	-	-	-	98	0.051	4.9
NGN001	NIED	138.3660	36.8514	305	50	0.076	7.2	52	0.164	19.1
NGN002	NIED	138.2069	36.8069	388	63	0.106	5.9	62	0.211	14.4
NGN003	NIED	138.4130	36.7403	521	58	0.078	3.1	62	0.048	2.4
NGN004	NIED	138.1938	36.6486	465	78	0.029	1.8	78	0.033	3.9
NGN005	NIED	137.8544	36.6981	513	95	0.013	1.4	89	0.017	2.2
NGN007	NIED	138.1191	36.5339	320	92	0.024	1.3	93	0.033	2.7
NGNH07	NIED	138.3761	36.7434	393	60	0.050	3.2	63	0.048	4.5
NGNH27	NIED	138.0479	36.5770	381	92	0.020	2.6	91	0.038	4.9
NGNH28	NIED	138.0964	36.7075	587	78	0.030	4.0	77	0.037	4.5
NGNH29	NIED	138.4407	36.9103	465	41	0.122	6.7	43	0.104	5.7
NGNH36	NIED	137.8482	36.6984	489	96	0.014	1.5	-	-	-
NIG001	NIED	138.4337	38.2584	641	-	-	-	72	0.062	3.6
NIG002	NIED	138.4398	38.0745	249	83	0.035	2.4	52	0.041	4.8
NIG003	NIED	138.3226	37.9976	188	81	0.083	3.9	47	0.164	7.1

Table 2.2 (continued)

Site Code	Network	Location		Vs30	2004 Earthquake			2007 Earthquake		
		Longitude (deg)	Latitude (deg)	V_{s30} (m/s)	R_{JB} (km)	*PGA* (g)	*PGV* (cm/s)	R_{JB} (km)	*PGA* (g)	*PGV* (cm/s)
NIG004	NIED	138.2795	37.8204	320	-	-	-	34	0.101	4.0
NIG005	NIED	138.4981	37.9204	290	65	0.028	1.4	34	0.145	5.9
NIG008	NIED	139.4055	38.0530	221	84	0.052	2.2	82	0.039	2.1
NIG009	NIED	139.3375	37.9507	180	71	0.031	2.6	69	0.035	2.3
NIG010	NIED	139.0108	37.9146	174	56	0.090	7.6	48	0.048	8.4
NIG011	NIED	139.1442	37.8013	150	48	0.059	6.5	46	0.053	4.6
NIG012	NIED	139.4772	37.6863	235	55	0.294	14.5	66	0.129	6.9
NIG013	NIED	138.8834	37.7638	175	38	0.116	12.9	28	0.139	14.4
NIG014	NIED	138.9559	37.6410	128	25	0.113	14.4	21	0.106	15.7
NIG015	NIED	139.1886	37.6935	490	40	0.067	4.5	42	0.032	4.5
NIG016	NIED	138.7676	37.6421	374	27	0.095	6.0	11	0.298	11.4
NIG017	NIED	138.8431	37.4416	275	4	0.433	41.1	11	0.223	15.8
NIG018	NIED	138.5579	37.3724	199	22	0.128	22.3	0	0.625	97.7
NIG019	NIED	138.7898	37.3057	307	0	1.255	109.0	16	0.412	34.1
NIG020	NIED	138.9620	37.2332	333	5	0.475	28.1	33	0.132	6.4
NIG021	NIED	138.7468	37.1281	419	6	1.476	51.8	24	0.258	16.7
NIG022	NIED	138.8462	37.0364	194	13	0.339	20.4	38	0.133	10.2
NIG023	NIED	138.6529	37.0147	654	21	0.390	25.4	32	0.049	4.7
NIG024	NIED	138.4440	37.1268	379	26	0.250	11.2	21	0.179	16.8
NIG025	NIED	138.2235	37.1607	135	44	0.198	17.2	28	0.208	23.9
NIG026	NIED	138.2507	37.0227	282	46	0.077	4.4	40	0.118	8.8
NIG027	NIED	137.8624	37.0235	369	79	0.059	2.1	61	0.087	4.8
NIG028	NIED	138.8862	37.4261	388	1	0.648	35.7	15	0.141	10.8
NIGH01	NIED	138.8876	37.4272	388	1	0.741	61.3	15	0.136	11.2
NIGH03	NIED	139.4289	38.1326	251	93	0.033	2.3	89	0.027	2.4
NIGH04	NIED	139.5428	38.1312	392	98	0.055	2.5	96	0.062	3.2
NIGH05	NIED	139.2788	37.9759	245	71	0.096	5.0	68	0.077	5.6
NIGH06	NIED	139.0676	37.6527	336	31	0.378	25.3	30	0.150	8.5
NIGH07	NIED	139.2610	37.6658	528	41	0.117	3.2	47	0.074	2.2
NIGH08	NIED	139.4648	37.6708	327	53	0.133	9.7	-	-	-
NIGH09	NIED	139.1279	37.5386	463	23	0.384	14.3	32	0.119	4.6
NIGH10	NIED	139.3648	37.5438	653	38	0.168	8.7	53	0.074	3.8
NIGH11	NIED	138.7440	37.1727	375	2	0.508	48.2	21	0.144	19.0
NIGH12	NIED	138.9821	37.2238	553	7	0.401	22.2	35	0.154	5.8
NIGH13	NIED	138.3965	37.0544	461	33	0.076	5.7	30	0.254	23.5
NIGH15	NIED	138.9951	37.0534	686	18	0.199	6.9	47	0.051	2.6
NIGH16	NIED	137.8480	36.9377	525	83	0.031	1.2	68	0.070	4.1
NIGH17	NIED	138.0966	36.8570	383	67	0.060	5.6	63	0.025	3.1
NIGH18	NIED	138.2594	36.9425	311	50	0.103	7.2	47	0.100	8.4
NIGH19	NIED	138.7849	36.8114	625	38	0.072	3.1	57	0.107	5.0
TCG002	NIED	139.8033	36.9900	609	82	0.035	1.6	-	-	-
TCG003	NIED	139.7153	36.8144	362	86	0.048	1.9	-	-	-
TCG004	NIED	139.4204	36.8061	431	65	0.049	2.6	93	0.040	2.1
TCG009	NIED	139.7155	36.7258	226	91	0.093	4.4	-	-	-
TCG010	NIED	139.4376	36.6347	520	79	0.045	1.8	-	-	-
TCGH07	NIED	139.4535	36.8817	420	62	0.123	3.5	-	-	-
TCGH08	NIED	139.6459	36.8828	723	77	0.047	2.0	-	-	-
TCGH09	NIED	139.8364	36.8625	468	92	0.030	1.6	-	-	-

Table 2.2 (continued)

Site Code	Network	Location		V_{s30}	2004 Earthquake			2007 Earthquake		
		Longitude (deg)	Latitude (deg)	(m/s)	R_{JB} (km)	*PGA* (g)	*PGV* (cm/s)	R_{JB} (km)	*PGA* (g)	*PGV* (cm/s)
TCGH11	NIED	139.7694	36.7084	329	97	0.054	1.8	-	-	-
TCGH17	NIED	139.6922	36.9854	1433	74	0.067	1.4	-	-	-
TYM001	NIED	137.6257	36.9760	359	-	-	-	81	0.025	1.7
TYM004	NIED	137.5274	36.8624	484	-	-	-	95	0.033	2.4
YMT014	NIED	139.8697	37.9207	361	98	0.075	2.6	-	-	-
YMTH05	NIED	139.8017	37.9883	533	99	0.019	1.4	-	-	-
65003	JMA	138.2363	37.1480	347	-	-	-	29	0.134	19.7
65004	JMA	138.4454	37.1272	379	26	0.296	15.2	21	0.199	19.7
65005	JMA	138.4254	37.1591	641	26	0.175	6.8	19	0.464	23.6
65006	JMA	138.6082	37.1327	518	-	-	-	18	0.192	17.1
65007	JMA	138.6068	37.0877	400	16	0.194	17.9	23	0.108	17.2
65008	JMA	138.5118	37.1547	409	-	-	-	15	0.620	21.7
65009	JMA	138.3829	37.0794	400	-	-	-	28	0.310	25.9
65010	JMA	138.3912	37.2766	195	30	0.112	10.7	9	0.424	70.1
65011	JMA	138.3293	37.2302	195	-	-	-	17	0.267	28.2
65012	JMA	138.3301	37.1872	201	-	-	-	21	0.228	41.4
65013	JMA	138.4049	37.2233	376	28	0.129	12.3	14	0.356	49.8
65017	JMA	138.2941	37.0524	313	-	-	-	35	0.079	12.7
65018	JMA	138.3344	37.0816	340	-	-	-	30	0.232	15.8
65019	JMA	138.3460	37.1249	370	34	0.088	13.4	26	0.260	37.9
65020	JMA	138.0913	37.1580	454	-	-	-	36	0.151	13.3
65024	JMA	138.9557	37.6411	128	25	0.142	15.0	21	0.114	16.2
65025	JMA	138.5587	37.3716	199	-	-	-	0	0.574	95.1
65027	JMA	139.0401	37.6663	162	-	-	-	29	0.077	13.4
65028	JMA	138.7554	37.1272	419	6	1.092	44.4	25	0.216	20.7
65029	JMA	138.8854	37.5461	210	14	0.209	30.7	11	0.127	20.1
65033	JMA	138.9226	37.5880	165	19	0.142	19.5	16	0.286	25.0
65034	JMA	138.8779	37.5480	160	14	0.229	30.9	11	0.332	33.3
65035	JMA	138.7885	37.3938	167	-	-	-	10	0.221	22.4
65036	JMA	138.7829	37.4936	174	12	0.336	28.7	4	0.311	32.7
65037	JMA	138.8096	37.5408	430	15	0.322	24.0	5	0.337	27.6
65038	JMA	138.7724	37.5758	292	20	0.248	29.3	5	0.248	33.2
65039	JMA	138.7093	37.5308	367	20	0.177	15.1	0	0.362	36.1
65040	JMA	138.7704	37.6447	374	-	-	-	11	0.336	11.2
65041	JMA	138.8899	37.3266	400	0	0.673	106.4	22	0.337	34.8
65042	JMA	138.8612	37.2705	407	0	1.451	113.4	23	0.159	12.5
65043	JMA	138.9240	37.2438	338	1	0.409	47.8	30	0.197	8.3
65050	JMA	138.8474	37.0398	194	12	0.414	23.2	38	0.118	11.8
65053	JMA	138.7451	37.1741	375	2	0.636	42.7	21	0.138	19.8
65056	JMA	138.6387	37.2244	377	7	0.202	12.6	10	0.555	39.4
65057	JMA	138.7090	37.3052	373	6	0.561	49.6	10	0.566	61.3
65058	JMA	138.6226	37.4219	164	-	-	-	0	0.400	128.4
65059	JMA	138.6668	37.4569	329	18	0.202	17.9	0	0.863	70.7
65065	JMA	139.1825	37.7444	161	-	-	-	44	0.079	9.2
65083	JMA	138.8554	37.6911	162	30	0.153	17.6	20	0.145	14.4
65084	JMA	138.8396	37.6263	203	23	0.147	20.1	13	0.203	34.6
65085	JMA	138.8796	37.6858	159	30	0.114	16.9	21	0.118	14.5
65301	JMA	138.8385	37.4363	275	-	-	-	11	0.186	12.9

Table 2.2 (continued)

Site Code	Network	Location		Vs30	2004 Earthquake			2007 Earthquake		
		Longitude (deg)	Latitude (deg)	V_{s30} (m/s)	R_{JB} (km)	PGA (g)	PGV (cm/s)	R_{JB} (km)	PGA (g)	PGV (cm/s)
65321	JMA	138.7918	37.3130	307	-	-	-	16	0.290	23.3
690E1	JMA	138.8885	37.7613	175	-	-	-	28	0.137	13.5
690F1	JMA	138.8751	37.0648	353	-	-	-	38	0.107	19.7
6CB51	JMA	138.7063	37.5347	375	-	-	-	0	0.517	44.9
6CB61	JMA	138.1599	37.1060	641	-	-	-	36	0.150	11.9
6E1C1	JMA	138.2469	37.1080	358	-	-	-	32	0.262	24.8
70025	JMA	138.2358	36.7545	484	-	-	-	66	0.219	21.6
70026	JMA	138.2519	36.7673	482	-	-	-	64	0.565	58.2
70027	JMA	138.0894	36.7023	587	-	-	-	77	0.108	8.8
70031	JMA	138.3245	36.7731	525	-	-	-	61	0.230	22.2
Egkg	JSCE	138.8365	37.2677	431	-	-	-	22	0.219	17.4
Jts	JSCE	138.2627	37.1483	291	-	-	-	28	0.123	23.2
Kzk	JSCE	138.3910	37.2702	163	-	-	-	10	0.276	79.5
Kd	JSCE	138.9729	37.2161	492	-	-	-	35	0.189	8.3
Mkm	JSCE	138.8760	37.0812	366	-	-	-	36	0.113	13.4
Nok	JSCE	138.7907	37.4458	334	-	-	-	7	0.155	21.2
Nshmtsu	JSCE	138.8735	37.5302	161	-	-	-	10	0.245	24.0
Nshya	JSCE	138.6571	37.4427	419	-	-	-	0	0.850	81.4
Sjotsuba	JSCE	138.9415	37.6511	158	-	-	-	21	0.131	16.9
Kwz	JSCE	138.5869	37.3643	281	-	-	-	0	0.403	93.0
Ntnh	JSCE	138.1077	37.1713	479	-	-	-	34	0.187	14.5
Ojy	JSCE	138.7776	37.3120	465	-	-	-	15	0.213	27.7
Skw	JSCE	138.8787	37.2616	384	-	-	-	25	0.207	14.6
Sno	JSCE	138.8449	37.4222	298	-	-	-	13	0.240	19.5
Sok	JSCE	138.8765	37.5177	162	-	-	-	10	0.250	31.3
Ookawatsu	NILIM	138.8421	37.6099	160	21	0.140	17.0	12	0.432	47.6
Myoken	NILIM	138.8279	37.3450	448	0	1.513	145.0	16	0.260	17.5

2.4 GEOMORPHOLOGY MAPS

After searching many alternate sources for geologic and geomorphic data, I selected the 1:25,000 geomorphic maps prepared by the Geospatial Information Authority of Japan (GSI, 1977), which is shown in Figure 2.5. These maps are made for flood control use in the vicinity of rivers, and show relatively precise boundaries of geomorphic categories.

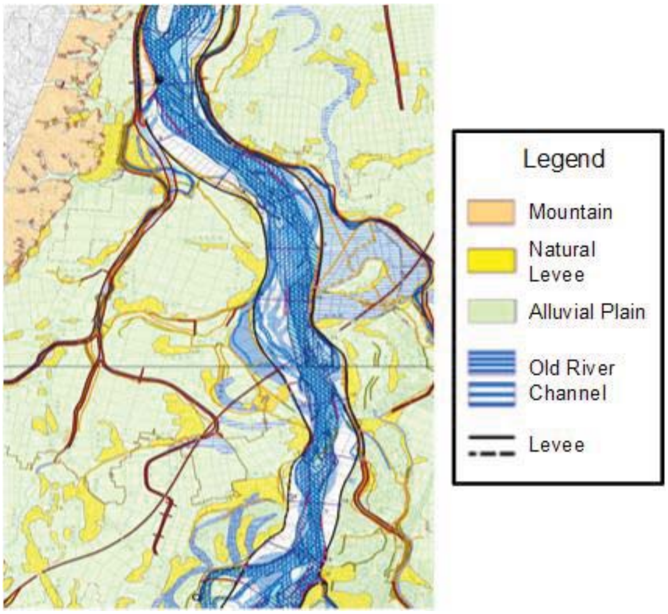

Figure 2.5 1:25,000 scale geomorphology map for the study area from GSI (2011).

Categories in these maps along the levees include mountain, terrace, alluvial fan, natural levee, alluvial plain, old river highland, old river channel, and back marsh. These geomorphic categories correlate with hydrologic conditions, and I adopt a grouping strategy proposed by Wakamatsu and Matsuoka (2011) and used by MLIT (2012) for liquefaction applications: (1) mountain and gravelly terrace, typically having deep groundwater, (i.e., groundwater depth > 3 m below ground surface); (2) alluvial fan, natural levees, alluvial plain, and old river highland, typically having shallow groundwater, (depth < 3 m); and (3) old river channel and back marsh, typically having very shallow groundwater, (depth < 1 m). I denote the grouped categories as G_N. The numbers of 50 m segments for each group are 264, 2485, and 312 for G_N = 1, 2, and 3, respectively.

Figure 2.6 shows the percentages of levee segments having G_N = 1– 3 within 2 km bins along the river length. All areas are dominated by G_N 2 (mostly alluvial plain), which is expected given the location of levees adjacent to rivers. Downstream areas (< 5 km from river mouth) and upstream areas of the Shinano River (> 40 km from river mouth) include a moderate portion of G_N 1 (more competent geologic materials) whereas midstream areas (20-40 km from river mouth) and upstream areas (> 60 km from river mouth) include significant fractions of G_N 3 materials. Group numbers are evaluated as a possible predictor of levee damage in Chapter 6.

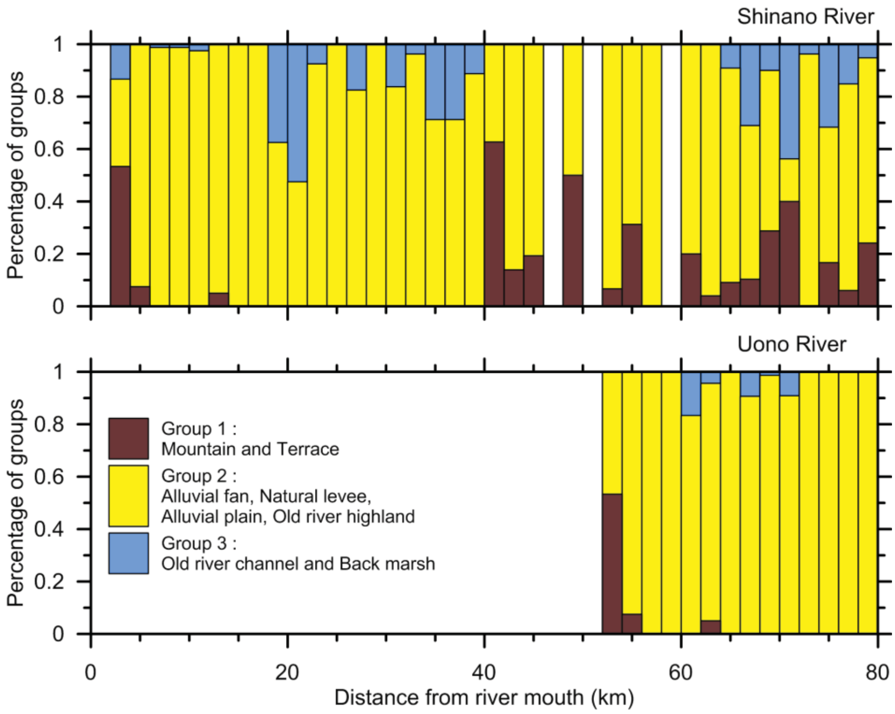

Figure 2.6 **Percentage of geomorphology groups within 2 km bins along the levees. Note that no levees are present along the Shinano River from 0-2 km, 46-48 km, 50-52 km, and 58-60 km.**

2.5 SOIL CONDITIONS AND LEVEE GEOMETRY

Starting before the 2004 earthquake, borings have been drilled by vendors contracted with SWO (YE, 2004, 2009; CD, 2004; Kittaku, 2004, 2007a, 2007b, 2007c; INA, 2008; SC, 2007; DET, 2007; NCG, 2007; OYO, 2009; NK, 2009; FC, 2009) along the Shinano River levees, up to 80 km from the river mouth, for the purpose of seepage and slope stability analyses. As shown in Figure 2.7, a given levee section typically has three borings − near the crest, river-side, and land-side slope or berm. Cross sections drawn from these borehole data show borehole water table and subsurface soil conditions as well as levee height and width. A total of 157 cross sections were collected. Levee geometry and soil conditions are evaluated utilizing these cross sections and subsurface information.

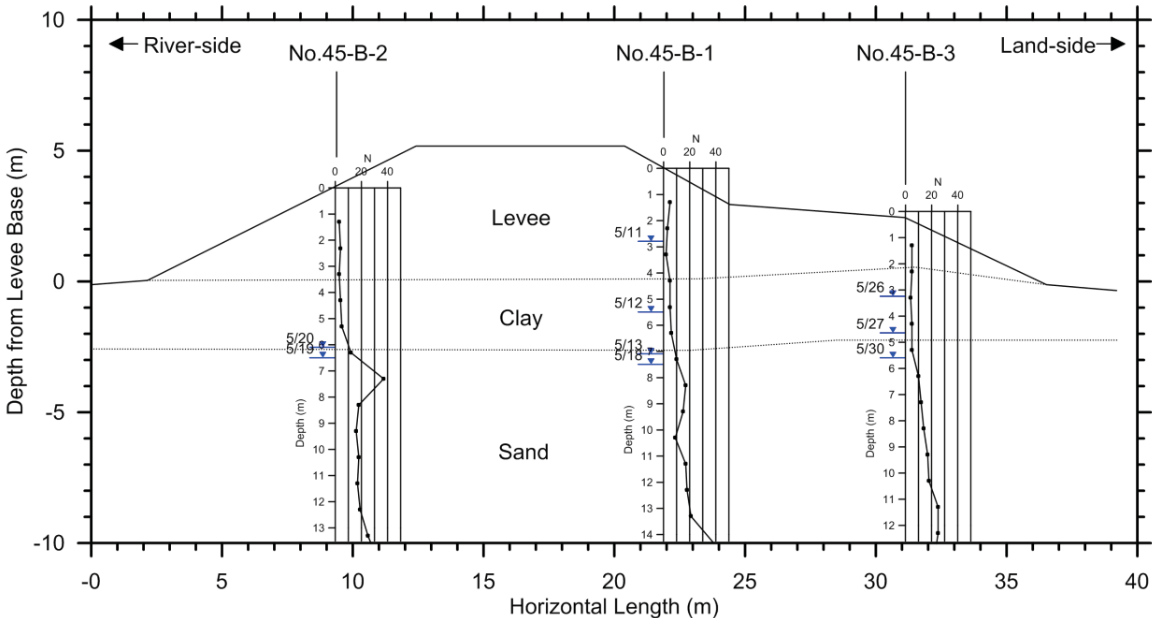

Figure 2.7 **Example of cross sections through levee showing levee base and levee ground water elevations from boreholes on various dates (from OYO, 2008).**

2.5.1 Embankment and Foundation Soil Conditions

The soil conditions within and beneath the levees are evaluated from boring logs that include standard penetration tests and other types of sampling. I have obtained both graphical representations of the boring logs and digitized versions of the boring logs, including penetration resistance data. This data includes layer descriptions according to a Japanese soil classification system, layer boundary depths, groundwater depths at various times, SPT penetration resistance (without energy or overburden corrections), and depths of SPT measurements.

The soil classifications in the boring logs follow a system that is similar to the Unified Soil Classification System (USCS), in that it provides the soil type and the soil fines content in an approximate manner. The system uses three letters (JHPC, 2005):

1) First letter represents the major soil type comprising > 50% of the soil mixture by dry weight (G for gravel, S for sand, M for silt, C for clay, Pt for peat);
2) Second letter, as applicable, represents a minor soil type comprising 15-50% of the soil mixture by dry weight;
3) Third letter, as applicable, represents a sub-minor soil type comprising 5-15% of the soil mixture by dry weight.

For example, a material consisting of 67.5% gravel (G) and 22.5% sand (S), and 10% silt (M) would classify as GS-M.

I compile the available borehole data for the study region for the levee fill materials and within the foundation to a depth of twice the levee height. As shown in Figure 2.8, the major soil types are sands and gravels with relatively low fines content (essentially clean to < 15%). Fine-grained plastic materials (silts and clays) comprise less than 35% of the levee embankment materials and less than 20% of the foundation materials. Peats are rarely encountered in the study area (0.3% in the foundation).

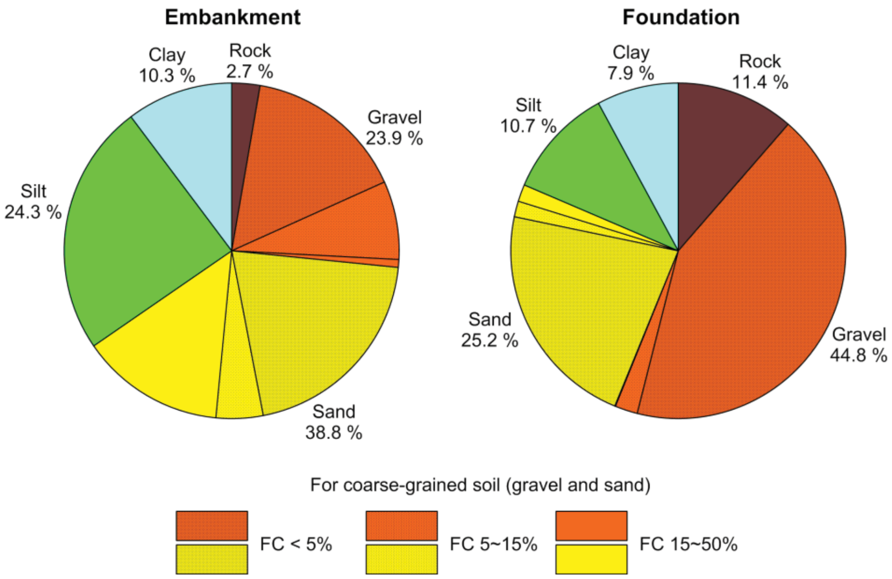

Figure 2.8 **Soil composition for levee embankment and foundation along the Shinano River levees.**

2.5.2 Levee Geometry

I quantify levee shape as shape factor (S_F), computed as average levee height over average levee width. The average levee height is the mean of the left- and right-side heights from crest to toe, and the average levee width is the mean of the crest and levee base widths, including berms. These dimensions are evaluated from 157 cross sections (similar to Figure 2.7) and intermediate locations are spatially interpolated. Values of S_F are high for tall slender levees (expected to have higher static shear stresses, but presumably also constructed with more competent soil materials) and low for short, broad levees. For example, the levee cross section shown in Figure 2.7 has S_F = 0.24. Figure 2.9 shows the range of S_F, which is 0.2-0.3 for downstream levees (relatively short and broad) and 0.25-0.35 for upstream levees (relatively slender). The gaps from 40 to 60 km occur because the river has carved a natural deep channel in stiff soil/rock, so no levees are present. The S_F is assigned for each 50 m levee segment as a possible factor contributing to levee damage.

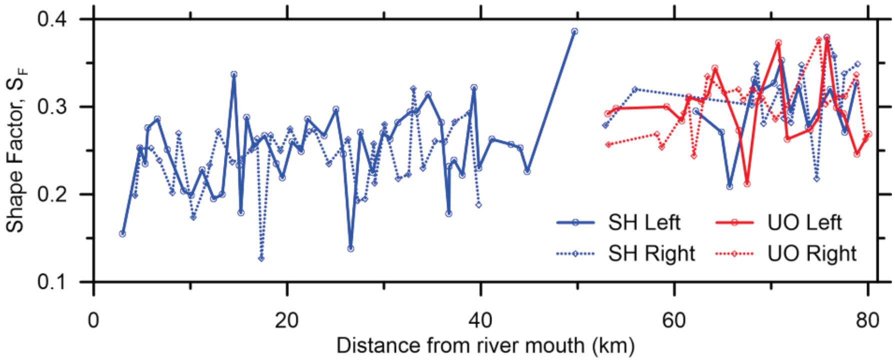

Figure 2.9 ***S_F* along river length. The range of *S_F* for downstream areas is 0.2~0.3 and 0.25~0.35 for upstream areas. Gaps in the figure correspond to locations where levees are not present.**

2.6 HYDROLOGY CONDITIONS

2.6.1 Ground Water Level

As shown in Figure 2.7, levee ground water levels (*LGWE*) were measured in boreholes. The measurement of water levels is sensitive to the method of drilling and in some cases was affected by in-situ permeability tests that were performed. In the case of auger methods, water levels typically rise with time as the boreholes fill to the water table elevation. In the case of the rotary wash drilling method, water levels typically drop with time as the drilling fluid flows from the borehole until the water table elevation is reached. The in-situ permeability tests were performed as a part of seepage analysis within most boreholes on the river- and land-sides of levees within the study region. Those tests consisted of filling boreholes with water or pumping water from boreholes (permeability then being measured from water table drop or rise, respectively). In either case, I seek stabilized water levels at some amount of time following testing.

The reports by above agencies indicate water table elevations in three ways: the water elevation that is encountered using drilling methods other than rotary wash, stabilized water elevations following in-situ permeability tests, and water elevations for which the method of drilling is unknown and the possible occurrence of a permeability test is also unknown.

I prioritize the selection of water elevations from boreholes as follows:

1. Water levels are taken from boreholes that are advanced without rotary wash drilling, without in-situ permeability tests, and located at levee crests.
2. Stabilized water elevations following in-situ permeability tests, irrespective of the drilling method.
3. When the method of drilling is unknown, select the last (in time) water elevation among the available measurements within the borehole at the levee crest.

The objective of these screening criteria is to obtain stable water elevations, which may include perched ground water.

2.6.2 River Water Elevations

As shown in Figure 1.1, river water elevations (*RWE*) are measured from seven streamgauge stations at the Shinano River (SH1 and SH2) and four at the Uono River (UO) hourly and daily; I utilize the day-based database. Figure 2.10 shows *RWE*s at streamgauges at the month of 2004 and 2007 earthquake (October for 2004 and July for 2007). The 2004 earthquake occurred two days after a flood event, at which *RWE*s are higher than those at the 2007 earthquake. Nevertheless the difference of river elevations between 2004 and 2007 earthquake is less than 1 m. *RWE*s in SH1 and SH2 fluctuate more than those in UO.

I sample *RWE*s not only on the earthquake date but also the date of subsurface exploration. The total eleven stations are too sparse spatially (average distance between adjacent stations is 13 km) to provide accurate *RWE*s for each 50 m levee segment considered in this study. For this reason, I also utilize *RWE* data from relatively detailed surveys performed following a flood (Oct 21 2004) and for maintenance purposes (Oct 2009 ~ Feb 2010). These detailed surveys are used to improve the knowledge of the spatial variation of *RWE*. I describe in detail the estimation of *RWE* utilizing streamgauge data and detailed surveys in Section 5.3.

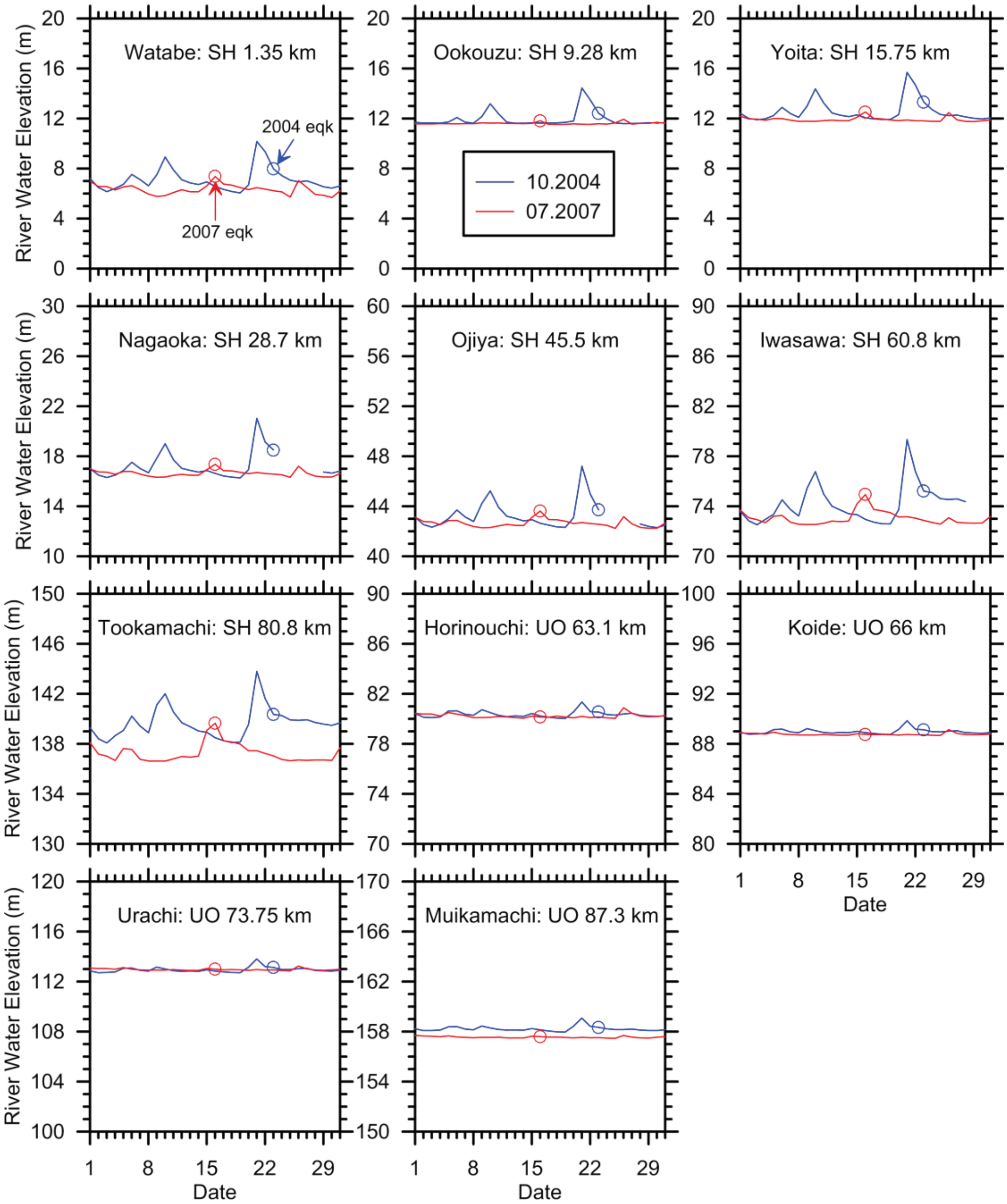

Figure 2.10 **River water elevations at the month of 2004 and 2007 earthquakes on seven streamgauges along the Shinano River and four along the Uono River. The dates of earthquakes are pointed as circles. Name of streamgauge, river, and distance from the river mouth are marked.**

3 COMPARISON OF SITE CONDITIONS BETWEEN STUDY REGION AND CALIFORNIA'S CENTRAL VALLEY

3.1 INTRODUCTION

Geotechnical conditions are compared beneath levees in the Shinano River system in Niigata region of Japan (SRJ) with those for urban levees in the Central Valley region of California (CVC). The purpose of this comparison is to evaluate the degree of similarity of the soil conditions for these two levee systems conditioned on soil type and surface geology, which is needed to judge whether the fragility models developed for Japan (presented in Chapter 6) can be applied in the CVC. Some Central Valley levees are on the eastern margin of the Delta, but Delta levees resting atop peaty organic soil are not considered both because my focus is on urban levees and soil conditions comparable to those in the Delta are not present in SRJ.

3.2 DATA RESOURCE

3.2.1 Boring Logs

Available data include penetration test measurements from 410 borings along the Shinano River and Uono River levee systems in Niigata, and 643 borings along rivers, creeks, bypasses, canals, and sloughs in the Central Valley. The CVC boring logs were divided among the North, Central, and South regions as shown in Figure 3.1. Table 3.1 indicates the name, length, and number of boreholes within each sub-region. The total length of levees considered for the SRJ levee system is 166 km (Figure 1.1), whereas the total length for CVC levee system is 358 km.

Table 3.1 List of locations where levees present in this study. Region, river (or bypass, canal, or creek, etc.) name, length, and number of boreholes are indicated

Region		River (or bypass, canal, or creek, etc.)	Length (km)	Number of boreholes
Niigata, Japan	Shinano River system	Shinano River – Left-side	59	177
		Shinano River – Right-side	60	153
		Uono River – Left-side	23	40
		Uono River – Right-side	24	40
		Total	166	410
Central Valley	North	Feather River	72	127
		Yuba River	21	44
		Sutter Bypass	28	34
		Wadsworth Canal	7	10
		Mud Creek	4	6
		Sycamore Creek	7	6
		WPIC	10	11
		Jack Slough	4	23
		Total	153	261
	Central	Sacramento River	27	51
		American River	19	11
		Willow Slough Bypass	12	17
		Yolo Bypass	21	51
		Natomas East Main Drainage Canal	27	36
		South Fork Putah Creek	13	17
		Cache Creek Settling Basin	9	5
		Total	128	188
	South	San Joaquin River	27	84
		Calaveras River	10	35
		Bear Creek	25	43
		French Camp Slough	3	15
		Mormon Slough	12	17
		Total	77	194

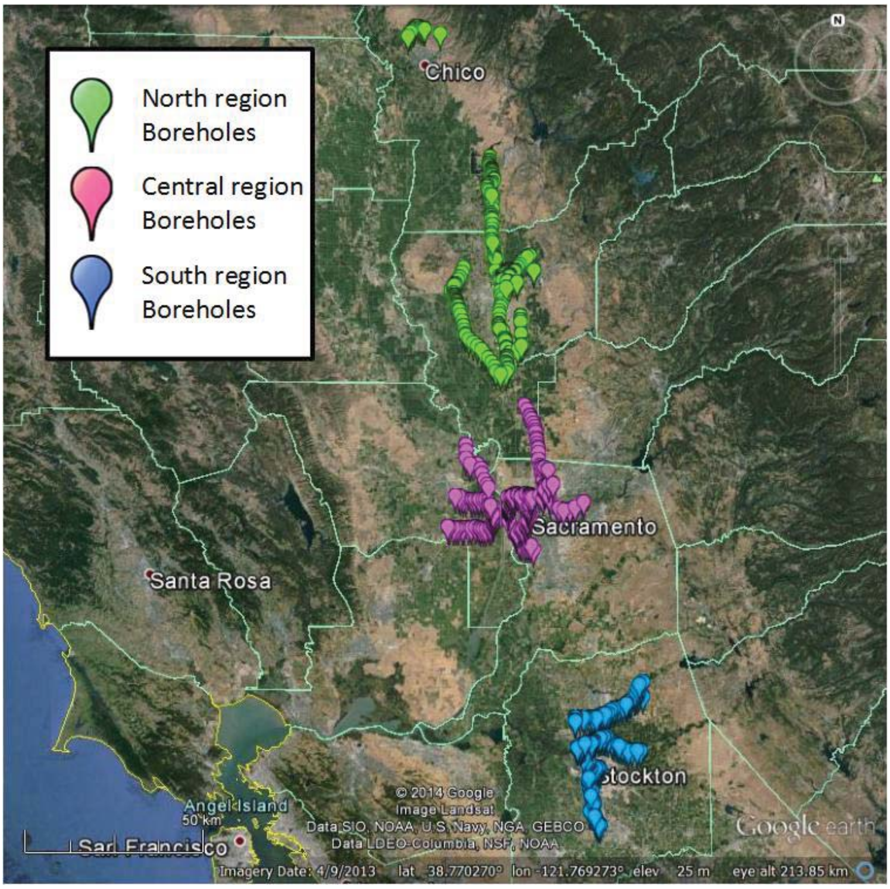

Figure 3.1 Google Earth maps showing borehole locations of North, Central, and South region levees in California's Central Valley.

3.2.2 Soil Compositions

The soil conditions beneath the levees are evaluated from boring logs that include standard penetration tests and other types of sampling. For CVC, boring logs were either images or digitized versions obtained from the California Department of Water Resources (CDWR). The CVC soil types given in the boring logs according to the USCS (classified from formal index testing or visual inspection). Soil conditions for SRJ are previously described in Section 2.5.1.

I utilize data only within the foundation to 10 m depth because the potential for earthquake-induced levee damage is mostly contained within the near surface soils. The levees themselves do not contribute significantly to seismic instabilities because they are most often unsaturated (i.e., these flood control levees are most often not retaining water). Furthermore, I exclude refusal (i.e., $N \geq 50$) and zero blow count ($N = 0$) cases from consideration to focus on soil conditions for which SPT blow count is a reasonable indicator of soil strength. Figure 3.2 shows the percentage of soil type (i.e., rock, gravel, sand, silt, and clay) present in the SRJ and CVC systems. For SRJ, the major soil types are coarse-grained sands and gravels. Fine-grained soils comprise less than 20% of soil types. Peats are rarely encountered in the study area (0.3%). For CVC, approximately 60% of the foundation materials are fine-grained soils (silts and clays). Higher proportions of fine-grained soils are present in the CVC region than in the SRJ region.

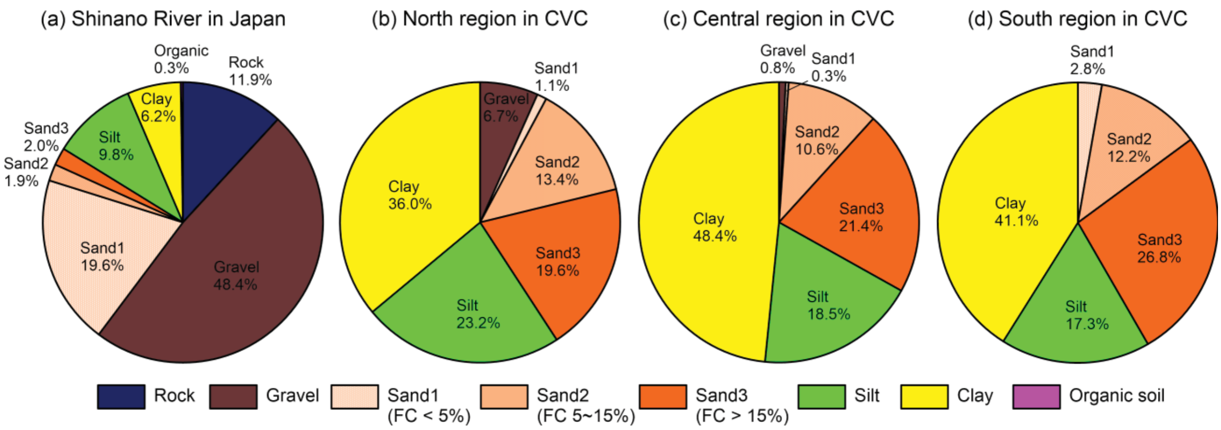

Figure 3.2 **Percentage of foundation materials to 10 m depth for primary soil types in (a) Shinano River system in Japan and (b) North, (c) Central, and (d) South regions in Central Valley of California. Sand is separated further according to fines content (< 5%, 5 ~ 15%, and > 15%).**

3.2.3 Geologic Conditions

High resolution geologic maps (i.e., 1:24,000 to 1:62,500) were utilized to assign surface geology conditions at the location of each boring log. I found that low-resolution maps (i.e., 1:200,000) are inadequate to capture the fluvial deposits adjacent to major rivers, and incorrect characterization arises from the low-resolution maps in many cases.

For the CVC system, I use surface geology maps by the U.S. Geological Survey (USGS). Table 3.2 shows map resources, geologic units and number of boreholes for each group for CVC. Based on the understanding of unit descriptions available from high resolution USGS maps, I group geologic units as (1) Holocene floodplain deposit, (2) Holocene basin-deposit, and (3) Pleistocene deposit. The borings often lie within Holocene floodplain deposits for levees along rivers, but also lie within Holocene basin-deposits and Pleistocene deposits adjacent to creeks, bypasses, canals, and sloughs.

For the SRJ system, I use high resolution geologic maps (1:50,000) from the Geospatial Information Authority of Japan (GSI, 2013) under MLIT. Most of the SRJ borings (97%) lie within Holocene floodplain deposits of the Shinano and Uono Rivers.

Table 3.2 List of high resolution geologic maps for study regions in CVC. Geologic units for CVC assigned to each group are provided along with numbers of boreholes.

Region	Resolution	Source	Geologic groups	Number of boreholes
North in CVC	1:62,500	Helley and Harwood (1985)	1. Holocene floodplain deposit (Qa) 2. Holocene basin deposit (Qb) 3. Pleistocene deposits (Qml, Qmu, Qrl, Qru, Qrb)	63 43 155
Central in CVC	1:62,500	Helley (1979)	1. Holocene floodplain deposit (Qha) 2. Holocene basin deposit (Qhb) 3. Pleistocene deposits (Qml, Qmu, Qrl)	79 88 21
South in CVC	1:24,000	Atwater (1982)	1. Holocene deposits (Qfp, Qpm) 2. Holocene and upper Pleistocene (Qcr) 3. Pleistocene deposit (Qm)	102 89 3

3.3 COMPARISON OF PENETRATION RESISTANCE

3.3.1 Correction on SPT Blow Counts

The practices and tools to carry out SPT penetration tests differ between Japan and California. To compare the same metrics between these regions, I correct blow to a standard energy efficiency level of 60% and to an effective overburden pressure of 1.0 atm. I denoted this as $(N_1)_{60}$. The calculation of $(N_1)_{60}$ requires hammer efficiency and vertical effective stress. Furthermore, to consider the effect of fines content for coarse-grained soil, I correct $(N_1)_{60}$ to clean sand equivalent blow counts, denoted as $(N_1)_{60\text{-}CS}$, which requires fines content.

The borehole data of SRJ includes SPT penetration resistance data at approximately 1 m depth intervals. Hammer efficiency is assigned for each boring based on hammer type and drop mechanism, which is described in the electronic version of the boring logs as having four categories: automatic drop hammer, semi-automatic drop hammer, mechanical trip device (referred to as tonbi) that is nearly free-fall, and the rope-pulley method (JHPC, 2005). I assigned energy efficiencies of 78% to automatic, semi-automatic, and tonbi methods and 67% to the rope-pulley method, which are average values for those methods (Seed et al., 1985). Effective stresses at SPT locations are calculated as total stress minus pore water pressure. Unit weights for total stresses were computed assuming typical values in engineering practice, e.g., moist unit weights γ_t of rock, gravel, sand, silt, and clay above the water table are taken as 24, 21, 19, 16, and 14 kN/m^2, and saturated unit weights γ_{sat} of rock, gravel, sand, silt, and clay are taken as 25, 22, 20, 18, and 16 kN/m^2, respectively (NAVFAC, 1986). Water pressures were computed from the water table depths provided in the cross sections assuming hydrostatic conditions.

For CVC, hammer efficiencies for boreholes are directly provided in electronic versions of boring logs (68% ~ 88%). Effective stresses at SPT locations are calculated as total stress minus pore water pressure calculated from groundwater elevations reported in boring logs. For case that groundwater elevations were not reported, I assume that the water table is 3, 4, and 2 m below the levee base for North, Central, and South regions, respectively, for Holocene deposits (i.e., Holocene floodplain and basin deposits). For Pleistocene deposits, 8, 10, and 6 m are used. Those numbers are medians of measured ground water depths for each case.

As shown in Figure 3.2, the coarse-grained soil in CVC contains more fines content than SRJ, which affects liquefaction resistance. To account for this fines content difference on the comparison of penetration resistance for coarse-grained soils, I estimate the clean sand equivalent $(N_1)_{60\text{-}CS}$ for both regions using the following equation (Idriss and Boulanger, 2008):

$$\left(N_1\right)_{60\ CS} = \left(N_1\right)_{60} + \exp\left(1.63 + \frac{9.7}{FC + 0.01} \quad \left(\frac{15.7}{FC + 0.01}\right)^2\right) \tag{3.1}$$

where FC is fines content in percent. For SRJ, the fines correction was insignificant, whereas for CVC, the blow counts increased by 4 to 5 on average.

3.3.2 Distribution of Penetration Resistance

Histograms of $(N_1)_{60}$ or $(N_1)_{60\text{-}CS}$ between SRJ and CVC for major soil types (i.e., sand, silt, and clay) are compared conditioned on geologic groups. Figure 3.3 shows histograms of $(N_1)_{60\text{-}CS}$ values for Holocene floodplain deposits comprised of sand. I show these data first because this combination of soil type and geologic condition are expected to be the most susceptible to liquefaction, which is a significant driver of earthquake-induced levee damage. In this figure, *Num* indicates the number of blow counts that correspond to this condition, X_m is the median value, and σ_{ln} is the standard deviation in natural log units. The distributions are similar, with the SRJ region exhibiting slightly lower median values than the CVC region.

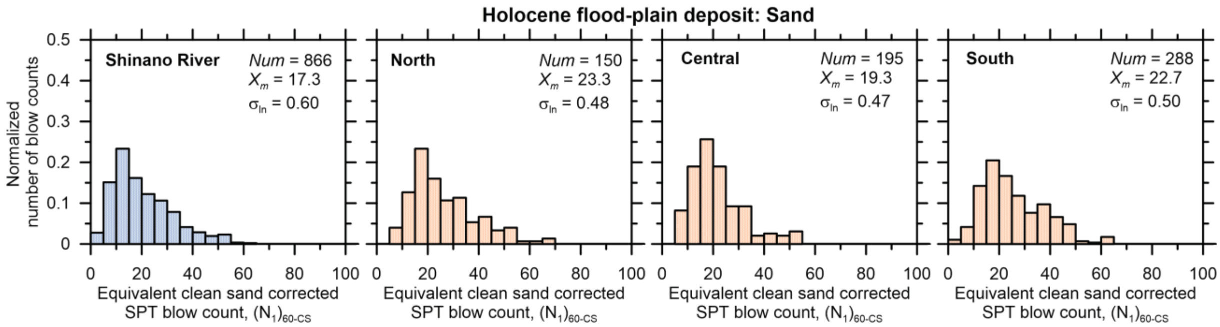

Figure 3.3 Histograms of energy- and overburden- corrected SPT blow counts for sandy soil modified by equivalent clean sand condition [i.e., $(N_1)_{60\text{-}CS}$; Idriss and Boulanger, 2008] in Holocene floodplain deposit. Median (X_m) and standard deviation (σ_{ln}) in natural log unit are presented.

Figure 3.4 shows histograms of $(N_1)_{60}$ values for fine-grained (silt and clay) Holocene floodplain deposits. Fine-grained soils for the CVC regions are further separated according to the liquid limits (ML and CL vs. MH and CH), where the majority is low plasticity material (i.e., ML and CL). The plasticity information is unknown for the SRJ region. In Figure 3.4, histograms are shown as the same manner for sand, but use $(N_1)_{60}$ rather than $(N_1)_{60\text{-}CS}$. SPT blow count is known to be a poor indicator of the strength of plastic fine-grained soils, and I don't recommend correlating strength with blow count for these materials in design applications. Nevertheless, blow count provides a reasonable point of comparison for the different systems, and other more relevant data (e.g., vane shear, CPT) are unavailable for this comparison. For silt, the SRJ and CVC-Central distributions are the most similar, whereas the CVC-North and South regions exhibit a significantly higher median. For clay soil type, all CVC regions are significantly stiffer than SRJ.

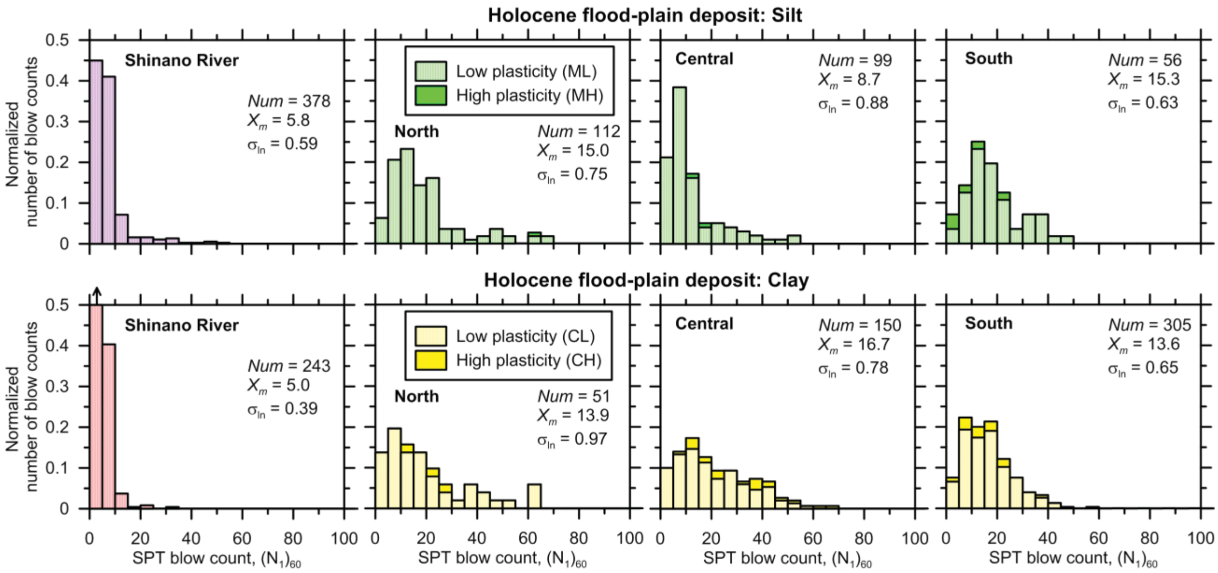

Figure 3.4 Histograms of energy- and overburden- corrected SPT blow counts conditional on geologic groups for silty and clayey soil. Median (X_m) and standard deviation (σ_{ln}) in natural log unit are presented.

Figure 3.5 shows histograms of $(N_1)_{60\text{-}CS}$ values for sandy soil type within Holocene basin deposits and Pleistocene deposits for the CVC region. Histograms for the SRJ region are not shown because so few borings were advanced in geological units other than Holocene floodplain deposits. The median values are significantly higher for these types of surface geology than for the Holocene floodplain deposits except South-Pleistocene deposit for which the number of data points is inadequate for comparison (*Num* = 9).

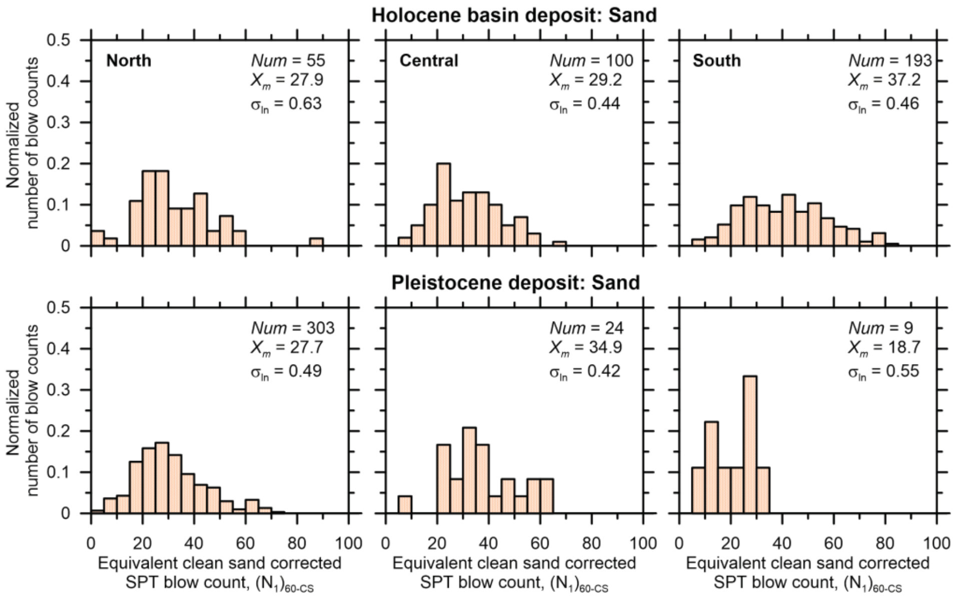

Figure 3.5 **Histograms of energy- and overburden- corrected SPT blow counts for sandy soil modified by equivalent clean sand condition in Holocene basin deposits and Pleistocene deposits in the CVC region.**

Figure 3.6 shows histograms of $(N_1)_{60}$ values for fine-grained soils having surface geology classifications of Holocene basin or Pleistocene. Again, the median values are significantly higher for these types of surface geology than for the Holocene floodplain deposits. Fine-grained materials for CVC with this geology type are again generally stiffer than fine-grained materials at SRJ.

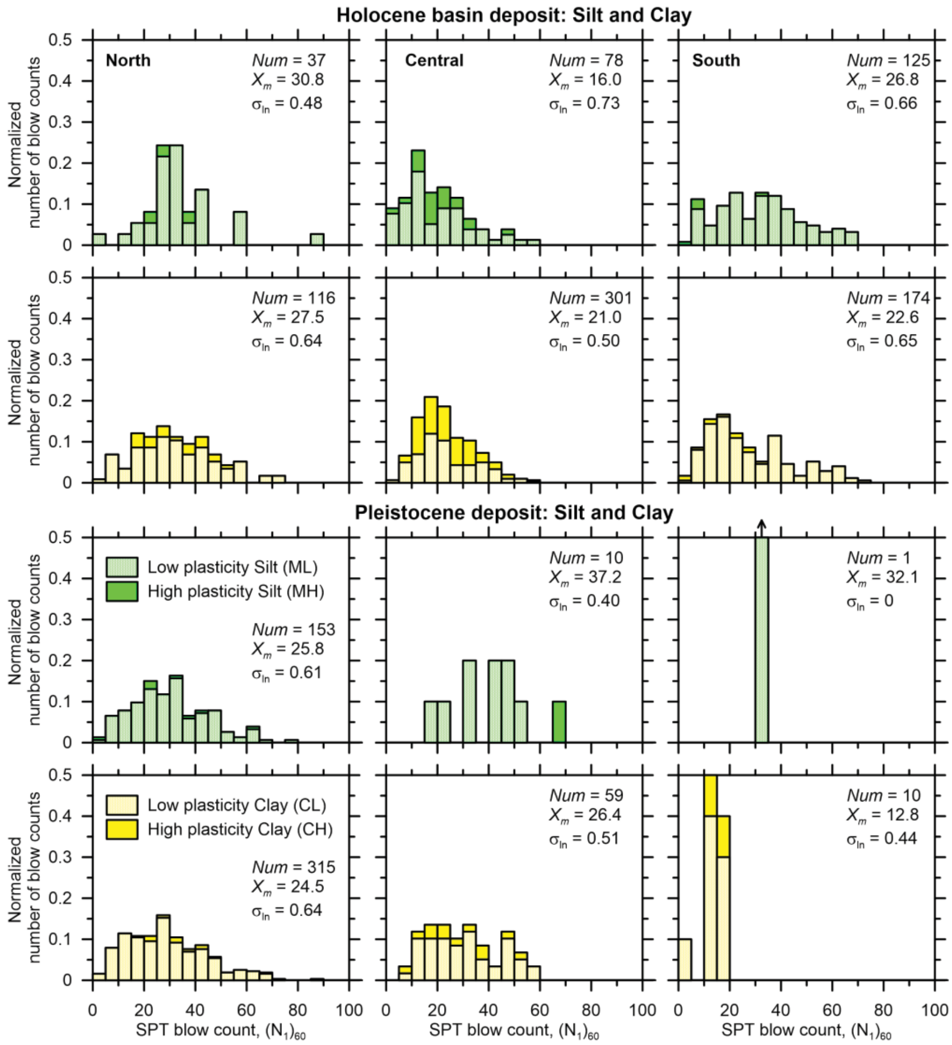

Figure 3.6 **Histograms of energy- and overburden- corrected SPT blow counts for silty and clayey soil in Holocene basin deposits and Pleistocene deposits in the CVC region.**

3.4 SUMMARY AND DISCUSSION

I compared penetration resistance measurements for borings along the Shinano River system in Niigata, Japan with borings along urban levees in the Central Valley region of California. Sands in Holocene floodplain deposits were found to exhibit similar values of $(N_1)_{60\text{-}CS}$ in the two study regions. Holocene basin deposits and Pleistocene deposits in the Central Valley region exhibited higher median blow counts (such geology types were scarce in the Japanese borings).

The purpose of comparing penetration resistance in these two regions was to ascertain whether fragility functions developed from the Japanese dataset could be applicable to the geological conditions in the Central Valley. Liquefaction was responsible for most of the heavily damaged levees in Japan, particularly for low to moderate ground shaking levels that characterize Central Valley seismic hazard. Therefore I conclude that the fragility functions are applicable for coarse-grained Holocene floodplain deposits because the distributions of blow counts between the two regions were similar. Furthermore, this combination of soil type and geology is anticipated to significantly influence levee fragility. The fragility functions would be expected to over-predict seismic levee damage for fine-grained Holocene floodplain deposits, as well as Holocene Basin deposits and Pleistocene deposits encountered for Central Valley urban levees along creeks, bypasses, canals, and sloughs.

These conclusions are applicable only to Central Valley urban levees, and not to Delta levees that constantly impound water. Delta levees are anticipated to be much more susceptible to earthquake damage since the unengineered levee fills are saturated and often susceptible to liquefaction, and high groundwater elevation is associated with higher rates of levee damage (Chapter 6). Furthermore, the peat soils that underlie Delta levees are very scarce in the Japanese dataset, but may contribute to levee damage.

4 CORRELATION OF SHEAR WAVE VELOCITY WITH PENETRATION RESISTANCE AND VERTICAL EFFECTIVE STRESS

4.1 INTRODUCTION

Shear wave velocity (V_s) profiles or averaged 30 m shear wave velocity (V_{s30}) is essential to evaluate ground motion amplification by local site effect. Ground motion can be amplified or de-amplified depending on site conditions, shaking amplitude, and frequency content. Amplification is expected for soft sites shaken weakly, while de-amplification may be observed with strong motions due to non-linear site effects associated with mobilization of large shear strains and associated modulus reduction and damping behavior. The stiffer sites have the less site effects because the soil conditions are more similar to the reference rock condition from which the site amplification factors are referenced. Site conditions are often assessed by V_s profiles. When a strong impedance contrast exists at the contact between a soil layer and the underlying stiff material, site response analysis would be beneficial to model site effects. However, in practice, site response analysis is seldom used, but rather a single indicator, V_{s30}, is used to define site amplification factors. This parameter does not perfectly capture site amplification for vertically propagating shear waves, but enormous data and empirical studies support the use of V_{s30}.

V_s profiles are measured in field using invasive methods that require a borehole (e.g., down-hole, cross-hole, suspension logging) and non-invasive surface methods [e.g., spectral analysis of surface waves (SASW), multi-channel analysis of surface waves (MASW), and refraction microtremor (ReMi)]. Each method has benefits and drawbacks, and involves different sample sizes. For example, surface wave methods provide good quality measurements at shallow depths while borehole methods may be difficult to interpret, but surface wave methods provide relatively poor resolution deeper in the profile where borehole methods tend to provide better accuracy. Moss (2008) showed that uncertainty in V_{s30} measurements are on the order of 1% to 3% coefficient of variation (standard deviation / mean) for down-hole and cross-hole methods, and about 5% to 6% for SASW. A geophysical survey, or better yet a combination of different geophysical surveys, is the best method for measuring a shear wave velocity profile and for ultimately measuring V_{s30}. However, V_{s30} is often estimated at sites where geophysical surveys are not available and cannot reasonably be obtained. For this purpose practitioners turn to more approximate correlations.

Methods for estimating V_{s30} based on geography and geomorphology parameters have been developed, and are perhaps the most crude approximation of V_{s30}. Wald and Allen (2007) and Yong et al. (2012) provide slope-base and terrain-base V_{s30} correlations, respectively, and Matsuoka et al (2006) provide V_{s30} correlation using Japan Engineering Geomorphological Map (JEGM). Moss (2008) indicates that such estimates result in coefficients of variation ranging from 10% to 50%, with a mean of about 30%, and Seyhan et al. (2014) indicate that geographic- or geomorphic-based V_{s30} estimation has standard deviation higher than 0.3 for California and Taiwan regions, and > 0.4 for Japan region.

Methods have also been proposed to correlate shear wave velocity with geotechnical site investigation data (i.e., SPT blow counts, N, and effective stresses, σ_v'). These methods provide a level of accuracy that is intermediate between geophysical measurements and correlations with surface geology (Seyhan et al., 2014). V_s models correlating with energy-corrected blow counts, N_{60} and overburden pressure-corrected blow counts, $(N_1)_{60}$ have been studied (Hasancebi and Ulusay, 2006; Dikmen, 2009; Andrus et al., 2004). Brandenberg et al. (2010) recommend a model correlating V_s with N_{60} and σ_v' because shear wave velocity and penetration resistance normalize differently with overburden pressure. Therefore direct correlation of N_{60} to V_s is likely to be biased with respect to σ_v'. Furthermore, independent knowledge of overburden scaling factors is rarely available.

In this study, I develop V_s models correlated with geotechnical metrics (i.e., N_{60} and σ_v') in addition to geographical (i.e., slope and elevation) or geomorphological (i.e., JEGM) proxies utilizing K-NET site database, global digital elevation map, and geomorphology map. I reduce the systematic errors in local regions by using spatial interpolation method (i.e., Kriging) on boring-to-boring residuals. The developed model is applied to evaluate V_{s30} on levees of the Shinano River system, which is utilized for ground motion predictions in Chapter 5.

4.2 DATA RESOURCES

4.2.1 Shear Wave Velocity and Penetration Resistance

NIED maintains the K-NET network with approximately 2 km spatial distance between sites in Japan, where SPT blow counts, V_p, V_s, bulk density, and soil types are provided up to 10 to 20 m depth for each layer with 1 m depth interval as shown in Figure 4.1. I utilize 1102 sites from K-NET database having total 16845 sets of shear wave velocity, blow count, and effective stress measurements as the main data set in this study. Selected sites include 92 sites in which seismometers have been relocated and are not currently operated at the original site, while sites with incomplete data profiles are excluded.

Twelve soil types are indicated in the boring logs: Surface soil, Fill soil, Gravel, Gravelly soil, Sand, Sandy soil, Silt, Clay, Organic soil, Volcanic ash clay, Peat, and Rock. For the purpose of V_s model development, I group similar soil types into seven categories: 1) Rock; 2) Gravel and Gravelly soil; 3) Sand and Sandy soil; 4) Silt; 5) Clay and Volcanic ash clay; 6) Organic soil and Peat; 7) Surface soil and Fill soil.

Typical use of energy ratio for SPT in Japanese practice is 67% for 'corn-pulley' method (i.e., pull and release the hammer connected with rope rolled in a drum; Japanese Industrial Standard JIS A 1219) and 78% for automatic-hammer method (Seed et al., 1987). Most of K-NET sites were installed in later 90's, when the corn-pulley method is common. This hammer dropping method is indicated in the boring profiles of K-NET sites provided in Geo-Station (NIED, 2014), where most of them used the corn-pulley method. I suppose that the energy ratio is 67%.

Effective stress is calculated by total stress minus water pressure. Location of ground water elevation is defined when P-wave velocity becomes higher than 1450 m/s (i.e., typical P-wave velocity of water), and total density is calculated assuming 20% saturation for soil above water table and the specific gravity as 2.7.

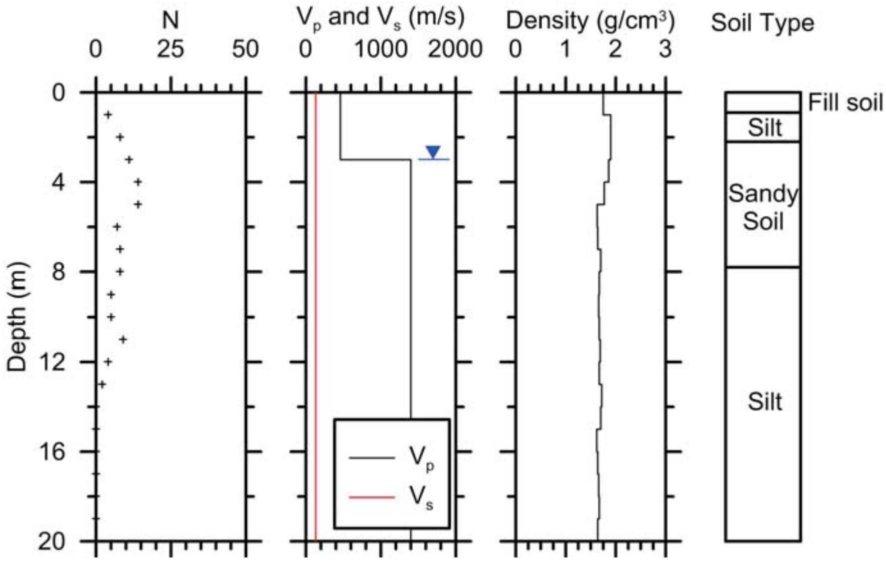

Figure 4.1 **Example of a K-NET station profile (AIC003) showing SPT N value, P- and S-wave velocity (V_p and V_s), density, and soil type (from NIED, 2012b). Location of water table is assumed where V_p is higher than 1450 m/s.**

The Port and Airport Research Institute (PARI) in Japan maintains a seismic network on around 61 ports with site condition data (PARI, 2014). The format of site condition data is analogous to one of K-NET, while PARI site often has penetrations resistance deeper than 20 m. The 29 complete site conditions set are compiled, and this data set is utilized for validation of V_s prediction later, but these data points are not included in model development. The location of PARI sites are not evenly spread throughout Japan like K-NET, but mostly located near oceans. Nevertheless, this data set is useful for validation purpose.

4.2.2 Geomorphic Conditions

The Japan Engineering Geomorphology Map (JEGM) provides 24 geomorphology categories for each approximately 250 m × 250 m grid for the entire land area of Japan (Wakamatsu and Matsuoka, 2011). Matsuoka et al. (2006) regressed V_{s30} using the JEGM categories along with elevation, ground slope, and distance from Tertiary mountain or hill. The JEGM categories and regressed V_{s30} values for each cell of Japan (6.2 million cells) are given through the J-SHIS

(Japan Seismic Hazard Information Station; NIED, 2013) website. Note that the database provided in J-SHIS use the Tokyo datum. I group the 24 JEGM categories (shown in Table 4.1) as seven groups according to the surface soil descriptions (Wakamatsu and Matsuoka, 2011) because they are similar in terms of surface soil types, and to provide enough data within a particular group to make a statistically meaningful regression for the V_s model development.

Table 4.1 Grouped JEGM categories and surface soil description along with number of sites for each group.

Group	JEGM categories	Surface soil descriptions	Number of borings
SG1	Mountain; Mountain footslope; Hill; Volcano; Volcanic footslope; Volcanic hill; Rocky strath terrace	Hard to soft rock and volcanic deposit	268
SG2	Gravelly terrace; Alluvial fan; Marine sand and gravel bars	Dense gravel and sand	333
SG3	Terrace covered with volcanic ash soil	Stiff volcanic ash	72
SG4	Valley bottom lowland	Dense gravel to soft cohesive soil	217
SG5	Natural levee; Sand dune; Dry riverbed; River bed	Loose sand	31
SG6	Back marsh; Lowland between coastal dunes and/or bars	Soft cohesive soil	68
SG7	Abandoned river channel; Delta and coastal lowland; Reclaimed land; Filled land	Loose sand overlying soft cohesive soil	113

4.2.3 Topographic Conditions

Two topographic parameters (elevation and ground slope) were utilized as conditioning variables. The U.S. Geological Survey (USGS) and the National Geospatial-Intelligence Agency (NGA) have developed a global and continental scale digital elevation model (called Global Multi-resolution Terrain Elevation Data 2010, GMTED2010; USGS, 2010). Three resolution models are provided: 7.5-, 15-, and 30-arc-second spatial resolutions. I use 30-arc-second dataset because the smoother elevations provide the more representative depositional conditions, and are less influenced by man-made sources of topographic relief such as levees (Allen and Wald, 2011). The 30-arc seconds indicates approximately 1 km horizontal grid spacing. Ground slopes are obtained utilizing elevations of the eight neighboring cells. Detailed technical procedures obtaining elevation and ground slope are described as follows:

- Elevation: Elevation is directly taken from GMTED2010, where **R** (a computer language for statistical computing and graphics; RCT, 2013) is used to pull out the elevation (v). A function "*raster*" in **R** reads the GIS elevation database from GMTED2010 and forms a matrix (i.e., raster layer) where row and column indices and cell values are analogous to longitude (x), latitude (y), and v. This raster layer is converted to a three-column matrix of x, y, and v by a function "*rasterToPoints*." I then find v for target coordinates from the matrix.
- Ground slope: The ground slopes are calculated using the "*terrain*" function (Hijmans, 2013) in **R**, which was originally utilized by Horn (1981). Ground slope in a grid is calculated utilizing elevations of 9 grids including the grid itself and adjacent 8 grids. Inputting the elevation raster layer from above and setting an option as slope [i.e., opt = c("slope")], the function *terrain* outputs a raster layer of ground slope. Again, I use *rasterToPoints* to find ground slopes at target points.

Table 4.2 summarizes data utilized in this study for development of shear wave velocity – penetration resistance – effective stress correlation considering topographic parameter effect. All data is web-based.

Table 4.2 Summary of database used in this study

Database	Provider	Data utilized	Description
K-NET site conditions	NIED	• Lat/Long • V_s • SPT N • Density • Soil type	V_p, V_s, N, bulk density, and soil type are provided for each 1 m interval soil layer up to 10 m to 20 m depth. http://www.kyoshin.bosai.go.jp/
PARI site conditoins	PARI	• Lat/Long • V_s • SPT N • Density • Soil type	V_p, V_s, N, bulk density, and soil type are provided for each soil layer with 1 m depth interval. Penetration depth is varied. http://www.eq.pari.go.jp/kyosin/
JEGM	NIED	• Lat/Long • JEGM category • V_{s30}	JEGM category (Wakamatsu and Matsuoka, 2011) and V_{s30} (Matsuoka et al, 2006) are provided for 7.5-arc-second longitude and 5-arc-second latitude grid (approximately 250 m × 250 m) for Japan. http://www.j-shis.bosai.go.jp/en/
GMTED2010	USGS NGA	• Lat/Long • Elevation • Ground slope	Elevations with various resolutions (30, 15, and 7.5 arc second) are provided. The 30-arc-second (approximately 1 km × 1 km) is utilized. Ground slope is obtained from elevations around 8 cells (Horn, 1981). http://topotools.cr.usgs.gov/gmted_viewer/

4.3 SHEAR WAVE VELOCITY MODEL DEVELOPMENT

The data compiled is classified as two types; 1) layered data within a borehole such as V_s, soil type (ST), energy-corrected SPT blow count (N_{60}), and vertical effective stress (σ_v'); 2) site term such as geomorphology and topography (i.e., JEGM, elevation, ground slope). I first develop a basic model considering only layered data ignoring site terms and then observe whether this approach causes bias with respect to site terms, thereafter making any necessary corrections. K-NET data, which are evenly scattered over Japan, are utilized for the basic model development, whereas PARI data, which are intentionally located at ports, are used for the model validation in Section 4.5.

4.3.1 Basic Model

The general form of the mixed-effect linear regression is used for the functional form of the model as follows:

$$\ln\left(V_s\right)_{ij} = \beta_0 + \beta_1 \ln\left(N_{60}\right)_{ij} + \beta_2 \ln\left(\sigma'_v\right)_{ij} + \eta_i + \varepsilon_{ij} \tag{4.1}$$

where βs are regression coefficient, η_i is boring-to-boring residual, and ε_{ij} is within-boring residual. i and j are site index and layer index, respectively. I regress the model to have zero mean of η_i and zero mean of ε_{ij} rather than to have zero mean of overall residual.

Note that Eqn. (4.1) is undefined for soft soils with zero blow count (i.e., the sampler penetrates into the soil under its own weight, also called "push"). Furthermore, the SPT test is also not physically meaningful when blow counts exceed about 50 because this corresponds to a refusal condition. In these cases, the specific values of blow count are not particularly meaningful, but knowing that a push or refusal condition exists is meaningful. For this reason, I grouped the data into three sets based on the range of field SPT blow count (i.e., N) values: i) push (N = 0), ii) $1 \leq N < 50$, and iii) refusal ($N \geq 50$). For push and refusal cases, I exclude the N values and use only σ_v' for model development. Push was only encountered within a few borings, so I developed regression constants for this condition regardless of soil group. On the other hand, I separated refusal cases based on soil group because a large enough sample within each group was obtained for a meaningful regression.

Figure 4.2 shows V_s for N = 0~50 case against σ_v' along with fit lines using median and plus and minus standard deviations of N_{60}, and against N_{60} with fit lines of σ_v' using the same format. V_s of coarse-grained soil groups (group 1 to 3) is sensitive on σ_v', whereas V_s of fine-grained soil groups (group 4 to 6) is relatively more sensitive on N_{60}. V_s in the range of $\sigma_v' < 20$ kPa and $N_{60} < 10$, and V_s in the entire range of σ_v' for SG6 may be biased from fit lines by comparing with the median of binned V_s. Nevertheless, fit lines for all soil groups generally follow the median points well indicating that Eqn. (4.1) is an agreeable function for V_s prediction. Figure 4.3 shows V_s for push and refusal cases. Refusal case shows generally higher V_s than push and N = 0~50 cases within the same range of σ_v'. The correspondence between fit lines and median points are worse than the N = 0~50 case, but again median points are not significantly biased from fit lines. Table 4.3 and 4.4 indicates regression coefficients (βs) and standard deviation in natural log unit for all seven soil groups and three blow count ranges.

Table 4.3 Regression coefficients according to soil group and penetration resistance for *N* = 0 ~ 50 and refusal cases.

Soil Group	$N = 0 \sim 50$				$N \geq 50$			
	β_0	β_1	β_2	σ_{ln}	β_0	β_1	β_2	σ_{ln}
SG1	3.727	0.185	0.293	0.371	4.892	0	0.294	0.456
SG2	3.84	0.154	0.285	0.369	4.557	0	0.302	0.36
SG3	3.913	0.167	0.216	0.328	4.51	0	0.274	0.338
SG4	3.879	0.255	0.168	0.349	2.742	0	0.643	0.352
SG5	4.119	0.209	0.165	0.369	4.549	0	0.257	0.456
SG6	3.98	0.275	0.108	0.416	3.208	0	0.49	0.245
SG7	4.089	0.153	0.208	0.384	4.274	0	0.304	0.462

Table 4.4 Regression coefficients for push case.

Soil Group	$N = 0$			
	β_0	β_1	β_2	σ_{ln}
SG1~7	4.537	0	0.033	0.314

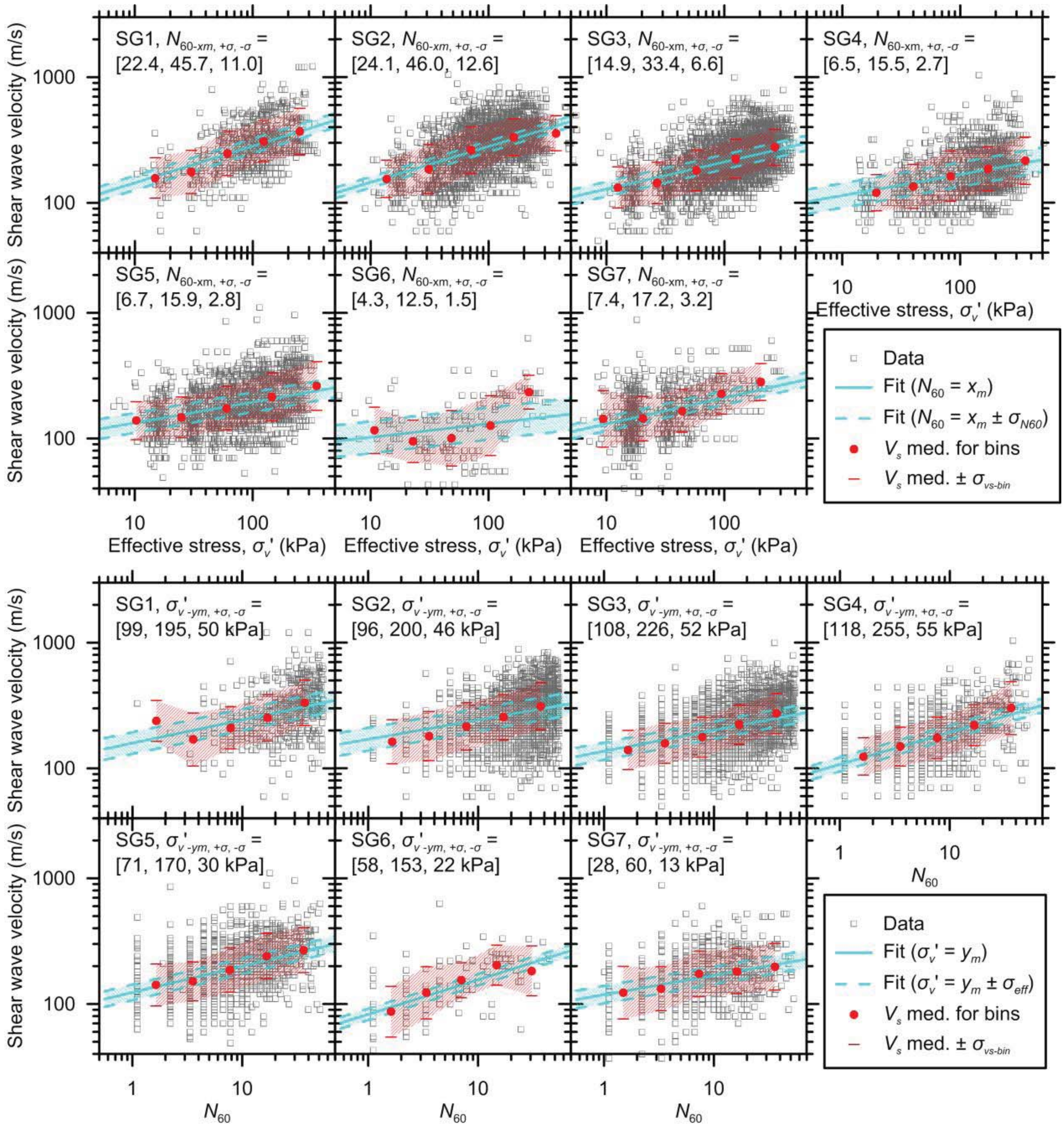

Figure 4.2 **V_s along with σ_v' and N_{60} for N = 0~50 case for each soil group (SG). Fits using three N_{60}s (median of N_{60} denoted as x_m, median plus and minus one standard deviation of natural log unit) for σ_v' abscissa and three σ_v's (median of σ_v' denoted as y_m, median plus and minus one standard deviation of natural log unit) for N_{60} abscissa are shown. Median and median plus and minus one standard deviation of binned V_s are also shown.**

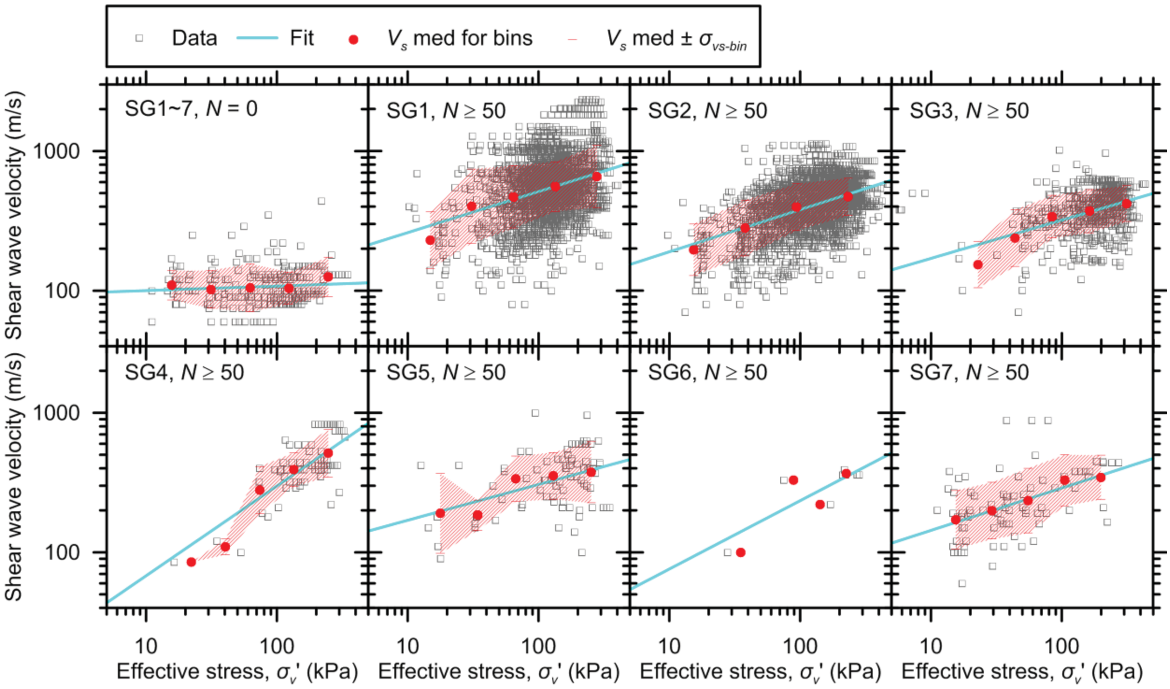

Figure 4.3 **V_s along with σ'_v for the case of N = 0 with all soil type and for the case of $N \geq 50$ for each soil group. Fits for σ_v' abscissa are shown. Median and median plus and minus one standard deviation for binned V_s are also shown.**

4.3.2 Influence of Geomorphic and Topographic Conditions

V_s is affected not only by soil type within a borehole but also by geologic setting. In this section I examine correlation between V_s and site feature by looking any trend of boring-to-boring residual (i.e., η in Eqn. 4.1) relative to geologic site conditions (i.e., geomorphic category, elevation, and ground slope). Model corrections are proposed to eliminate sources of bias relative to site conditions.

Figure 4.4 shows boring-to-boring residual (i.e., η) relative to JEGM group, ground slope shown as the tangent of slope angle, and elevation. The mean (or linear fit) and 90% confidence intervals are also shown. Stiffer geomorphology (i.e., JEGM group 1 to 4) is moderately biased in positive side, whereas softer geomorphology (i.e., Group 5 to 7) is in negative side. This means that sites with stiffer geomorphology tend to have higher V_s than predicted by the basic model, and softer sites tend to have lower V_s. The higher ground slope and elevation tend to have higher η, indicating higher V_s for high elevation and steeper slope. The slopes of fit lines for η in natural log unit (denoted as b_s and b_e, respectively in Figure 4.4) are approximately 0.03 for both ground slope and elevation. Standard deviations of η for mean of JEGM group and for fits of ground slope and elevation are 0.217, 0.225, and 0.223, respectively. This trend, which is the function of site conditions (referred to as correction factor), can reduce the uncertainty of boring-to-boring residual as follows:

$$\hat{\eta}_i = \eta_i - f(\text{site conditions}) \tag{4.2}$$

where $\hat{\eta}_i$ is corrected boring-to-boring residual by site conditions. The f must be determined to remove bias with respect to site conditions.

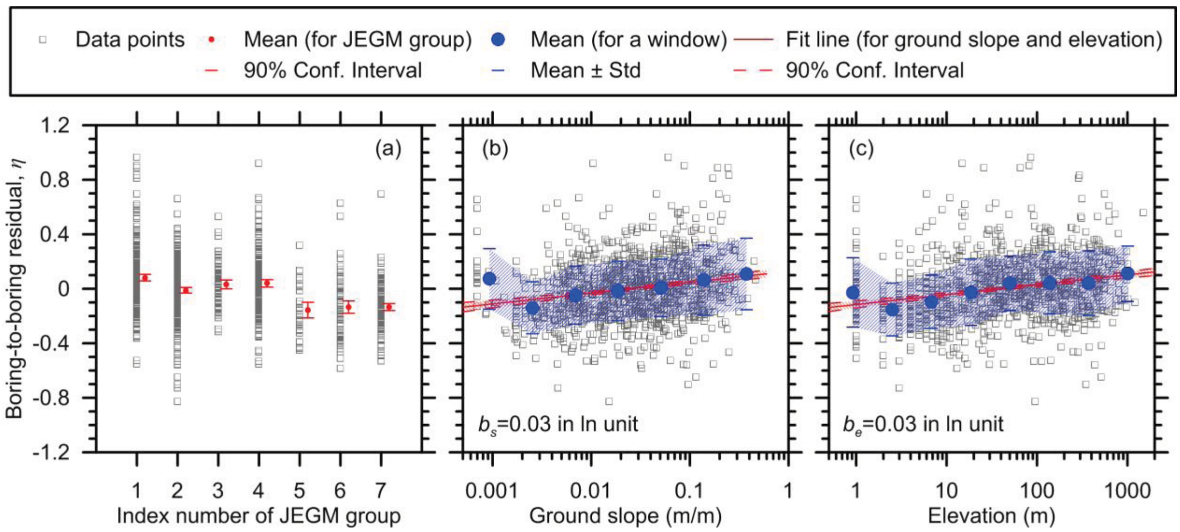

Figure 4.4 Boring-to-boring residuals (i.e., η) in terms of JEGM groups, ground slope, and elevation. (a) Mean and 90% confidence intervals for each JEGM group. (b and c) Linear fit lines with 90% confidence intervals and one positive and negative standard deviations for bins of ground slope and elevation, respectively.

In seeking the simplest possible form for f, I first explore correlation between JEGM, ground slope, and ground elevation. If the cross-correlation among these site conditions is strong, regression using all three of them could be redundant and unnecessarily complex. Figure 4.5 shows cross-correlation among JEGM group, ground slope, and elevation. As expected, stiffer JEGM groups have higher elevation and ground slope, and softer JEGM groups have lower elevation and ground slope. Ground slope and elevation have positive correlation where the correlation coefficient is 0.4.

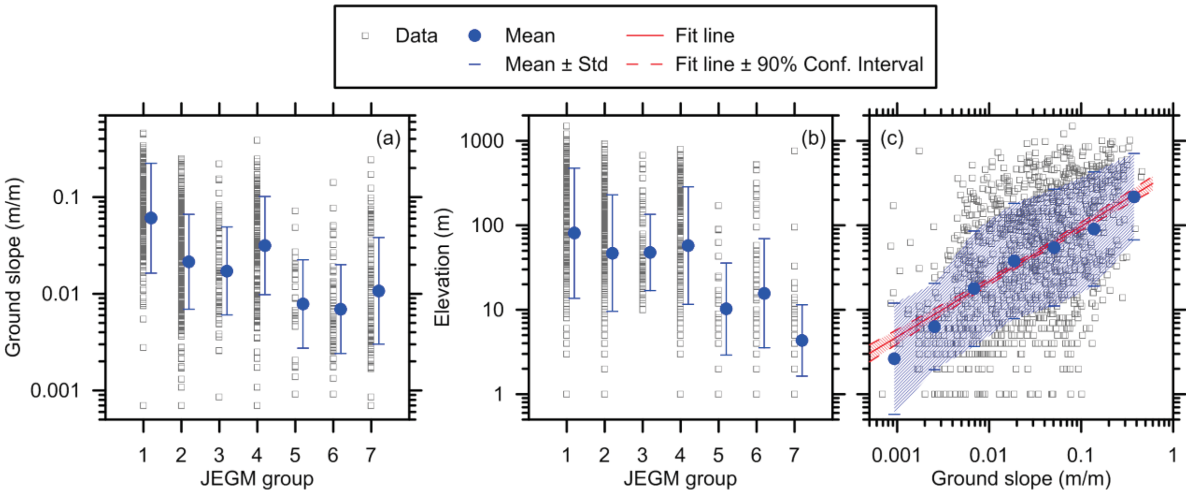

Figure 4.5 **Correlations among JEGM group, ground slope, and elevation. (a and b) Mean and one positive and negative standard deviations of ground slope and elevation against JEGM groups, and (c) fit line, 90% confidence interval, binned mean and one positive and negative standard deviations for between ground slope and elevation.**

I first postulate a functional form in which f is a constant that depends only on JEGM because JEGM is the simplest site condition to obtain for a particular site. I subsequently explore whether ground slope and elevation provide additional predictive power. The functional form is shown in Eqn. (4.3):

$$\hat{}_{i} = {}_{i} \qquad {}_{JEGM} \tag{4.3}$$

where $\Delta\eta_{JEGM}$ is a correction factor calculated as the mean of η for each JEGM group (Figure 4.4a). Figure 4.6 shows $\hat{\eta}_i$ along with JEGM group, ground slope, and elevation. The mean of $\hat{\eta}_i$ for JEGM groups are zero, as anticipated. Linear fit lines for ground slope and elevation become flatter (b_s and $b_e \approx 0.01$), which indicates that correcting for bias in JEGM also removes most of the bias in ground slope and elevation. Introducing a more complex form for f could further reduce this bias at the expense of making the model more complicated. For this reason, I explore whether this added complexity is justified using statistical tests. Based on the t-test on slopes of linear fits to evaluate the distinction of trends (Appendix B, H_{02} hypothesis), p-values are of 6% and 0.1% for ground slope and elevation, respectively. This indicates that the linear fits would be distinct at 10% significance level. However, although the trends are statistically significant, they are not particularly geotechnically significant. For example, suppose a site at an elevation of 1m has V_s = 200 m/s. The change of $\hat{\eta}_i$ of only 0.01 is associated with one natural-log cycle change in elevation, which means that V_s would be 212 m/s at 400 m elevation for the same JEGM category. Considering the uncertainty of V_s presented even in measured V_s (standard deviation ≈ 0.1; Seyhan et al., 2014), this difference is small enough to neglect, particularly considering that adding slope and elevation as predictive variables would significantly complicate the model.

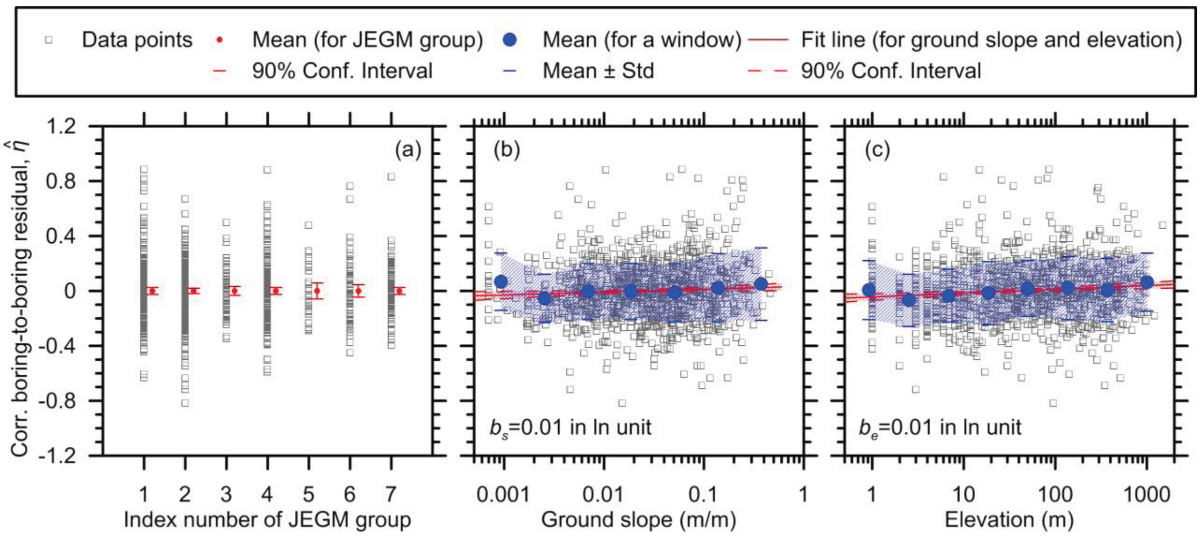

Figure 4.6 **JEGM corrected boring-to-boring residuals in terms of JEGM groups, ground slope, and elevation. (a) Mean and 90% confidence intervals for each JEGM group. (b and c) Linear fit lines with 90% confidence intervals and one positive and negative standard deviations for certain bins for ground slope and elevation, respectively.**

Therefore, I only select JEGM group for the correction factor of site terms. The α_{JEGM} for each JEGM group is listed in Table 4.5. The biggest change of α_{JEGM} is -0.157 for group 5, which would change a median prediction of V_s of 200 m/s to 171 m/s, and 400 m/s to 342 m/s.

Table 4.5 Correction factor (i.e., $\Delta\eta_{JEGM}$) for each JEGM group.

JEGM Group	1	2	3	4	5	6	7
$\Delta\eta_{JEGM}$	0.081	-0.011	0.033	0.040	-0.157	-0.135	-0.134

4.4 SPATIAL VARIATION OF BORING-TO-BORING RESIDUAL

Mixed effects regression is utilized to divide residuals, defined as data minus model, into a boring-to-boring residual (η_i) and a within-boring residual (ε_{ij}) as shown in Eqn. (4.1). The boring-to-boring residual is a measure of the average error for a particular boring in the shear wave velocity correlation equation, whereas the within-boring residuals define the variation in errors about the mean error for the particular depths within a single boring. These residuals exist because blow count and effective stress cannot be expected to be adequate predictors of shear wave velocity. If these residuals could be reduced, the predicted values would be improved to be closer to the data. Reducing the within-boring residual can only be achieved by adding more model parameters, which is not desired because it will unnecessarily complicate the model while perhaps not significantly enhancing predictive power. On the other hand, boring-to-boring residuals are an indication of spatial variability in the prediction errors. Spatial correlation of these residuals may therefore provide a means of reducing this source of error.

I evaluate spatially interpolated boring-to-boring residuals for Japan using the Kriging method (Appendix A). Figure 4.7 shows a semi-variogram of $\hat{\eta}_i$. I found that $\hat{\eta}_i$ are not strongly, but moderately correlated based on the separation distance between two borings. The nugget (i.e., the normalized semi-variance at a separation distance of 0) is set to 0.5 to facilitate a reasonable fit to the data. A non-zero nugget indicates that the residuals between two close sites are not the same, which is not ideal because the boring-to-boring residual should be the same for two measurements made at the same site (i.e., separation distance of zero). However, note that the resolution of the semi-variance data is not very good at very close separation distances, and the shortest spacing is 5 km. Considering the inherent spatial variability in alluvial deposits, I would anticipate that much shorter separation distances (perhaps on the order of meters or tens of meters) would be required to render zero semi-variance. This scale is too fine to explore using the available data. Despite this limitation, we are able to conclude that boring-to-boring residuals are correlated at separation distances less than about 27 km based on the semivariogram.

K-NET sites are spread out over Japan with approximately 20 km spacing. This difference is not dense enough to have well-established spatial correlation, but the 92 relocated sites provide relatively close separation distance from the original sites (the average spacing is 1 km). These pairs of original and relocated sites are utilized for semi-variance data less than 20 km.

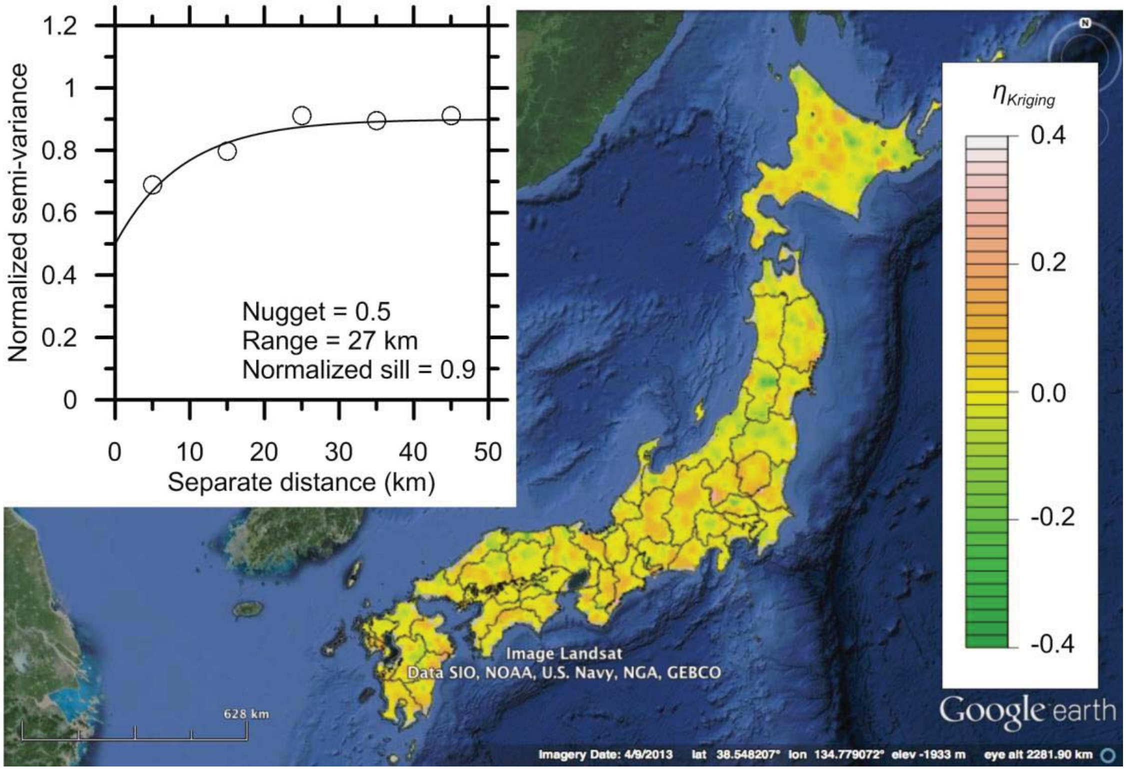

Figure 4.7 **Semi-variogram and contour map of boring-to-boring residuals corrected by JEGM group overlapping on satellite image of Google Earth.**

Boring-to-boring residuals were computed at each K-NET site in Japan, and a map of these residuals was obtained using the Kriging method in conjunction with the interpreted semivariogram. The resulting map of residuals, denoted as $\Delta\eta_{Kriging}$, is shown in Figure 4.7. Relatively low $\Delta\eta_{Kriging}$ is observed at north-middle Honshu, whereas high $\Delta\eta_{Kriging}$ is observed at middle Honshu near mountainous area.

Since it is possible to obtain $\Delta\eta_{Kriging}$ anywhere in Japan, subtracting $\Delta\eta_{Kriging}$ from $\hat{\eta}_i$ results in further reduction of overall boring-to-boring residual as follows:

$$\bar{}_i = \hat{}_i \quad {}_{Kriging} = {}_i \quad {}_{JEGM} \quad {}_{Kriging} \tag{4.6}$$

where $\bar{\eta}_i$ indicates second-level correction on boring-to-boring residual. $\bar{\eta}_i$ will be small for a site close to any K-NET sites, whereas will be equal to $\hat{\eta}_i$ if separate distance is large (> 27 km).

As a result of model development above, I classify V_s estimation as three levels:

1. Level 1: Basic model using geotechnical data only,
2. Level 2: Reduce boring-to-boring residual by accounting for geomorphology by applying the $\Delta\eta_{JEGM}$ term, and
3. Level 3: Further reduce boring-to-boring residual by accounting for spatial correlation by applying the $\Delta\eta_{Kriging}$ term.

4.5 COMPARISON OF V_{S30} VALUES COMPUTED USING VARIOUS METHODS

In this section I compare V_{s30} obtained from the correlation with geotechnical parameters developed in this study with those based on JEGM and topographic parameters (Matsuoka et al., 2006) and solely slope (Allan and Wald, 2009). In cases where the depth of the geotechnical data is less than 30 m, V_s is averaged up to boring depth z and extrapolated to 30 m depth (Boore, 2004; Boore et al., 2011) for V_{s30} calculation. I excluded sites with z less than 10 m depth.

Figure 4.8 shows V_{s30} residuals for K-NET sites. The correlations with geotechnical data provide estimates of V_{s30} with very small mean values, which indicates that these estimates are unbiased. The correlation with JEGM also provides an unbiased estimate, but the correlation with slope produces residuals with a significant negative mean value. Furthermore, the standard deviation of the residuals is significantly smaller for the correlations with geotechnical data compared with the JEGM- or slope-based predictions. This indicates that having geotechnical data can significantly improve estimates of V_{s30} compared with geology-based proxies. Of course, geophysical methods remain the best method for measuring V_{s30}, and the correlation with geotechnical data should only be performed when geophysical measurements are unavailable and cannot reasonably be obtained. Level 2 (JEGM correction) slightly improves from Level 1, but the Level 3 improvement is significant (σ ~ 0.16). This is because in this case V_{s30} is being computed at the same K-NET sites from which the boring-to-boring residuals were computed and interpolated. These are the very locations that will produce the smallest possible boring-to-boring residual terms, and the residuals will be higher as separation distance between a site of interest and the nearest K-NET site increases. .

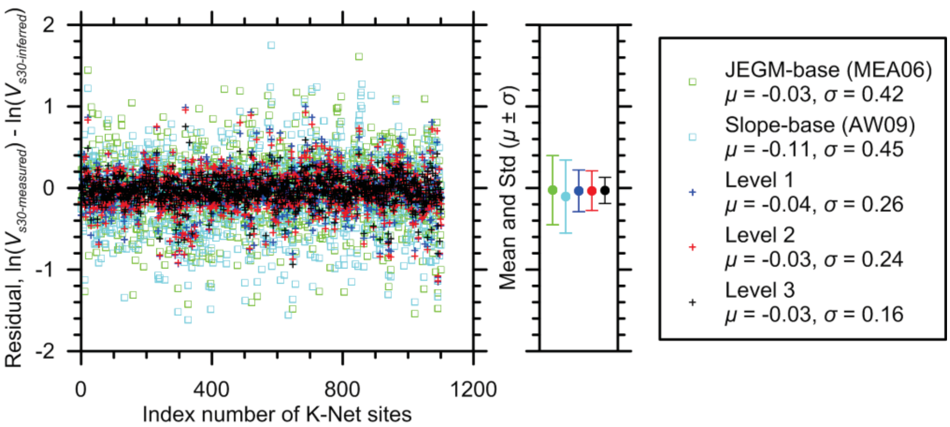

Figure 4.8 Residuals of inferred V_{s30}s to measured V_{s30}s for K-NET sites by various models: JEGM-base model (Matsuoka et al., 2006); slope-base model (Allan and Wald, 2009); proposed model with soil type only (Level 1); soil type and JEGM correction (Level 2); soil type, JEGM correction, and boring-to-boring residual correction (Level 3). Mean and one plus and minus standard deviations for each model are shown.

Figure 4.9 shows residuals of V_{s30} for PARI sites. This is an important validation exercise because the PARI sites were not included in model development. Again, the JEGM- and slope-based models produce estimates with higher standard deviations than the ones that include geotechnical data. In this case, the Level 1 prediction produces a biased estimate, but the Level 2 and Level 3 predictions are unbiased. This is evidence that corrections for JEGM and spatial correlation should be utilized. Note that 69% of the PARI sites are located in JEGM group 7 (Table 4.1) that has $\Delta\eta_{JEGM}$ = -0.134 (Table 4.5). Level 3 does not significantly improve the prediction from Level 2, but corrects the very slight bias (μ = 0 for Level 3 compared with -0.01

for Level 2). The average distance from PARI to the closest K-NET sites is 5 km for which the correlation is not strong.

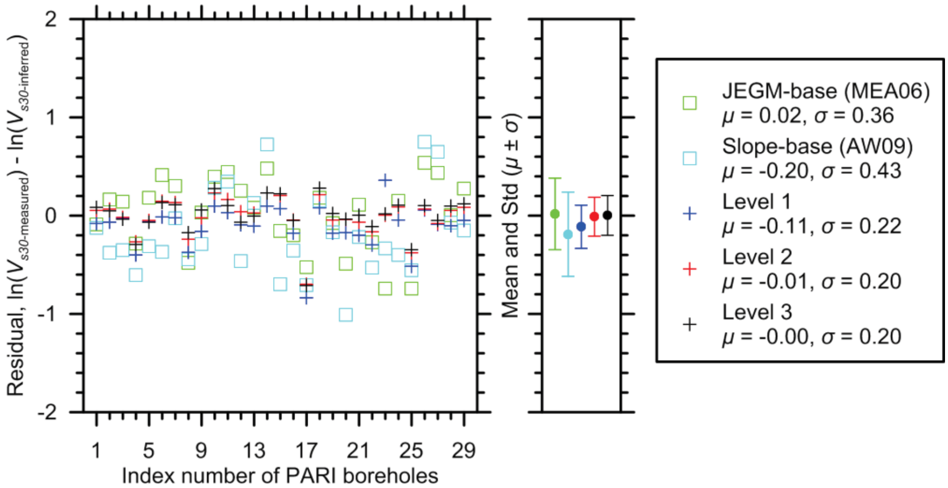

Figure 4.9 Residuals of inferred V_{s30}s to measured V_{s30}s at PARI sites by various models: JEGM-base model (Matsuoka et al., 2006); slope-base model (Allan and Wald, 2009); proposed model with soil type only; soil type and JEGM correction; soil type, JEGM correction, and boring-to-boring residual correction. Mean and one positive and negative standard deviations for each model are shown.

4.6 SITE CONDITIONS ON LEVEE FOUNDATIONS

The purpose of formulating correlations between V_s and geotechnical data is that I wish to improve estimates of shaking at levee sites where I have geotechnical data, but no geophysical measurements, based on ground motion measurements at recording stations site conditions that may differ from levee site conditions. As indicated in Section 2.5, penetration resistance data have been compiled along levees for the Shinano River system. Typically borings were performed at three locations: crest, slope, and toe (Figure 2.7). I desire a "free-field" ground motion record at the levee site, but acknowledge that levee fills alter vertical effective stress and stiffness of underlying soils. To predict V_{s30} of levee foundation using proposed model above, foundation conditions are converted to free-field conditions by following steps:

1) Remove over-burden pressure imposed by the embankment when computing σ_v' to input to the correlation with V_s, and

2) Adjust measured blow counts for removal of over-burden pressure to obtain the blow count that would be measured in the absence of levee fill.

This process is necessary because levee foundations have been compressed resulting in stiffer material beneath center of levees than free field conditions nearby, yet I desire to use a free-field ground motion intensity measure for consistency with the manner in which seismic hazard analysis is typically conducted. Figure 4.10 shows median of N_{60} and $(N_1)_{60}$ for three range of embankment heights (H) at the location of boreholes: H > 5 m (crest), H 3~5 m (slope), and H < 3 m (toe). Median of N_{60} is relatively high at crest for all soil types, whereas $(N_1)_{60}$ is similar among the three H ranges. This indicates that N_{60} beneath center of levee is positively biased due to overburden pressure. One exception is the case of H < 3m with sand.

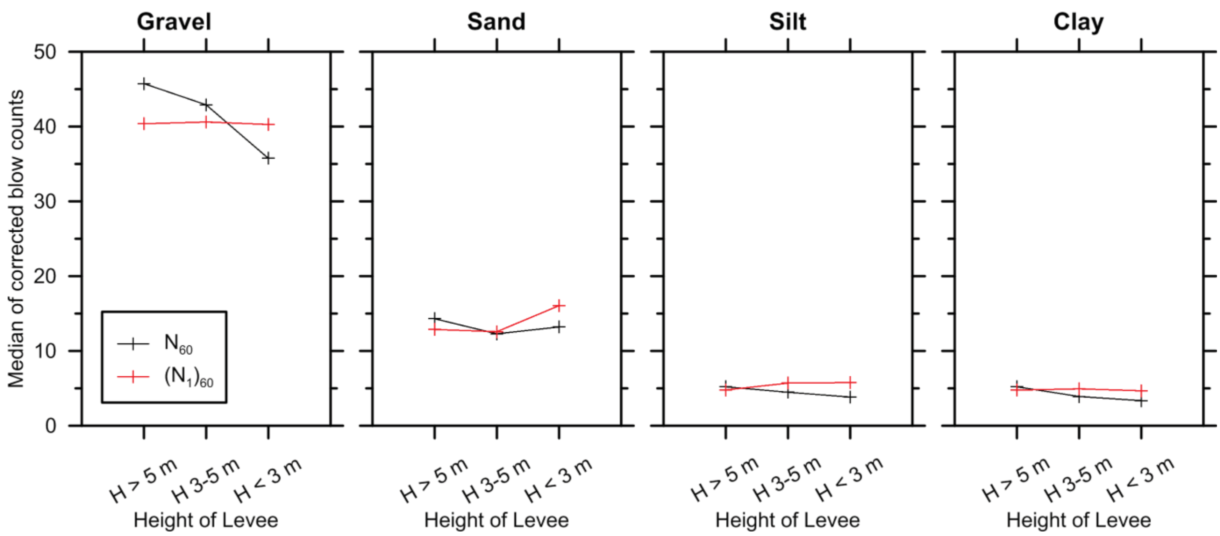

Figure 4.10 **Median of N_{60} and $(N_1)_{60}$ for embankment height on borehole locations (crest, H > 5 m; slope, H 3~5 m; toe, H < 3m).**

The scaled down N_{60} (denoted $N_{60\text{-}s}$) is calculated assuming that $(N_1)_{60}$ does not change by the overburden pressure change as follows:

$$N_{60-s} = N_{60} \frac{\left(p_a / \sigma_v'\right)^m}{\left(p_a / \sigma_v'_{-s}\right)^m} \tag{4.7}$$

where p_a is an atmospheric pressure (≈ 100 kPa), σ_v' is vertical effective stress, $\sigma_v'_{-s}$ is vertical effective stress after removal of levee embankment, and m is coefficient for $(N_1)_{60}$ calculation supposing m = 0.784 - 0.521×D_r for coarse-grain soil (Idriss and Boulanger, 2008), and m =1 for fine-grained soil. D_r is referred to as a relative density. Use of m = 1 for non-plastic silt may not be correct, but I stick with m = 1 due to absence of plasticity characteristics information. I use N_{60-s} instead of N_{60} for model input.

Figure 4.11 shows predicted V_{s30} on levees as well as V_{s30} on seismic stations in the vicinity of levees within 1 km spacing despite the fact that site conditions at these stations might be different from site conditions at the levees. NIED sites (i.e., K-NET and KiK-net) have measured V_{s30}, whereas V_{s30} of JMA sites are based on JEGM-base V_{s30} from J-SHIS database. For JMA sites close to NIED sites, V_{s30} of NIED sites are chosen. V_{s30} on levees and stations are generally similar except two stations at 17 km and 34 km. V_{s30} is smaller near the river mouth and higher in the upland areas, as predicted based on geologic conditions in these regions.

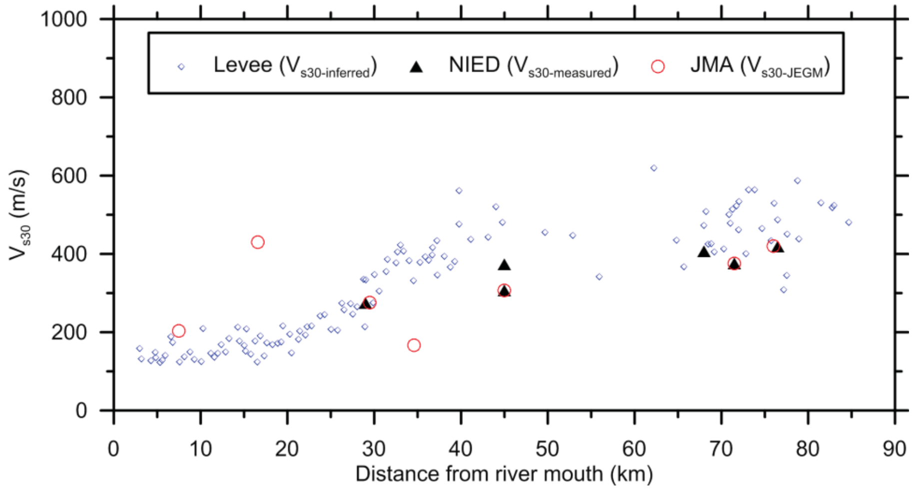

Figure 4.11 V_{s30}s for the Shinano River levees and adjacent NIED and JMA sites. Level 3 model is used for V_{s30} prediction on levees. Measured V_{s30} for NIED, and JEGM-base V_{s30} for JMA sites are used.

4.7 SUMMARY AND CONCLUSION

In this section I developed a predictive model for V_s based on soil type, penetration resistance, overburden pressure. Corrections to the model to account for geomorphic conditions were found to reduce modeling errors. Furthermore, spatial variation of boring-to-boring residuals was evaluated and the residual was interpolated spatially using the Kriging method. I suggest using the model that accounts for geomorphology and spatial correlation when this information is available.

I found that V_s of coarse-grained soil layer is more sensitive on overburden pressure, while V_s of fine-grained soil layer is more sensitive on penetration resistance. Geomorphology with stiffer site conditions such as mountain and hill indicate higher V_s than predicted by correlation with geotechnical parameters alone, whereas softer geomorphologic conditions (e.g., delta and abandoned river channel) result in lower V_s. Boring-to-boring residuals at sites near K-NET stations are reduced due to spatial correlation of the boring-to-boring residuals, and a map of these corrections was developed and presented. The proposed model provides significantly better predictive power on V_{s30} than other models regressed by site proxies alone, such as JEGM and ground slope. The proposed model produces estimates with prediction errors intermediate between geophysical measurements (the best method for obtaining V_{s30}) and predictions based on site proxies alone.

The proposed model was validated with a PARI dataset that was not used in development of the model, and worked quite well for this data set. The model was then utilized to predict V_{s30} corresponding to free-field conditions at levee sites along the Shinano River system. These estimated V_{s30}s are used for ground motion prediction on levees in Section 5.2.

5 GROUND MOTION AND GROUND WATER ELEVATION

5.1 INTRODUCTION

As described in the introduction, I seek to evaluate the dependence of levee damage on various predictive parameters related to seismic demands (ground motions), geotechnical / geological conditions, hydrological conditions, and levee geometrical parameters. Several of these parameters are straightforward to evaluate from the map resources and cross-sections presented in Chapter 2 (i.e., geotechnical / geological conditions; levee geometry). More challenging parameters are those related to ground motions and hydrological conditions, because we must rely on data that are either spatially or temporally separated from the specific levee segment locations and earthquake date. In this chapter, I present procedures developed to estimate ground motions at levee segments (originally presented in Kwak et al., 2012b) and to estimate ground water elevation relative to the levee base (originally presented in Kwak et al., 2012a).

5.2 GROUND MOTION ON LEVEES

Ground motion amplitude is expected to be a major predictor of damage, so I have dedicated substantial effort to obtaining reliable ground motion estimates from the network of recording stations around the levees. With few exceptions, seismographs are not located sufficiently near to levees to evaluate ground motion intensities at levee sections directly. Moreover, direct Kriging (i.e., simple interpolation) of intensity measures (e.g., *PGA*) is potentially problematic, because the site conditions at recording stations can be significantly different from those at levee sites (typically softer at levees). A detailed description of the problem and its solution is described in Section 5.2.1-5.2.4.

5.2.1 Description of the Problem

In Section 2.6, I reviewed the ground motion data available for the 2004 and 2007 earthquakes. The stations that produced usable recordings are marked in Figure 1.1 and listed in Table 2.2. In general, recording stations are not at the locations (along the rivers) where ground motion intensity measures are required for the present analysis, although stations Ookouzu, and Myoken are adjacent to levee segments (separate distance < 100 m). The most straightforward way to spatially interpolate ground motions is through direct Kriging analysis (Appendix A), which has been carried out for the two events as shown in Figure 2.4 for *PGV*.

Consider the portion of the strongly shaken region from the 2007 earthquake shown in Figure 5.1. A site shown in the figure (triangle) has an estimated *PGA* of 0.838g based on direct Kriging of *PGAs* at recording stations. As shown in Figure 5.1, the *PGA* at the site is strongly influenced by the closest recording (white circle), which is on stiff soil, and has *PGA* = 0.867g. The site at the triangle location has soft foundation soils with V_{s30} = 214 m/s, whereas the recording station with stiff soil has V_{s30} = 500 m/s. Based on empirical and semi-empirical site factors (e.g., Choi and Stewart, 2005; Walling et al., 2008; Boore and Atkinson, 2008), the ratio of soil/rock *PGA* for these velocities and the strength of the stiff soil motion ranges from around 0.55 to 0.60, suggesting that a better estimate of the motion on the relatively soft soil is about

$0.6 \times 0.87 = 0.52g$. Hence, the stiff soil recording is providing a biased estimate of the ground motions on the soft soil conditions.

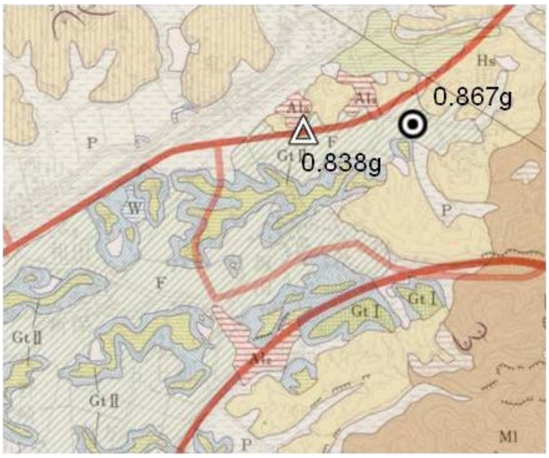

Figure 5.1 Geology map (NLSD, 2012) in highly shaken region (*PGA* > 0.8g) along with a seismic station provided by JMA (Station code 65059) marked as white donut and a borehole provided by NIED-Borehole Data Checker (NIED, 2012a) marked as white triangle. The V_{s30} for the seismic station is 500 m/s and 214 m/s for the borehole. Both sites are placed on alluvial fan (marked as F), but the seismic station is on stiff soil near hill (H_s) and mountain (Ml) and the borehole is on soft soil near valley plain (P) based on the geology (from Kwak et al., 2012b).

The condition illustrated in Figure 5.1 is not anomalous. The levees are preferentially located on soft materials along rivers, whereas ground motion stations tend to either be in urbanized regions (typically having soil conditions, but firmer ground than along rivers) or mountainous area (typically rock, stations are cited there deliberately to avoid large site effects). This is demonstrated in Figure 4.11 where V_{s30} on the station at 18 km from river mouth is higher than V_{s30} of levee. Hence, I postulate that simple interpolation will tend to produce systematically biased ground motion estimates that are too large in strongly shaken regions and too small in more weakly shaken regions (due to nonlinearity in site response).

5.2.2 Proposed Approach

The proposed methodology for estimating spatially distributed ground motion from recordings in a regional network is as follows:

1) Estimate V_{s30} for the foundation conditions beneath levees and recording sites using velocity measurements where available, and otherwise using correlations described in Chapter 4.

2) For earthquake i, compute within-event residuals as the difference between intensity measures from recording j and the median from a selected ground motion prediction equation (GMPE) computed for the magnitude, distance, and site conditions present at site j for event i. This residual is computed as follows:

$$R_{i,j} = \ln\left(IM_{i,j}^{rec}\right) - \left(\mu_{i,j} + \eta_i\right) \tag{5.1}$$

where $IM_{i,j}^{rec}$ denotes the intensity measure from recording j, $\mu_{i,j}$ denotes the GMPE mean in natural log units, and η_i denotes the event term (effectively the mean residual for event i for well-recorded events). I use the Boore et al. (2014) (BEA) GMPE.

3) Map the spatial variation of residuals R_i using the simple Kriging method (Appendix A).

4) Calculate ground motion IMs for sites of interest as:

$$\ln\left(IM_{i,k}^{K}\right) = R_{i,k}^{K} + \mu_{i,k} + \eta_i \tag{5.2}$$

where $R_{i,k}^{K}$ represents the mapped residual from (3), and index k refers to sites for which ground motions are to be estimated.

Relative to prior work (Yamazaki et al., 2000; Sawada et al., 2008), this procedure is different because it includes nonlinear site amplification factors, which is important due to the strength of the shaking and the softer site conditions beneath the levees compared with the recording stations.

Figure 5.2 shows within-event *PGV* residual contour maps produced in Step 3. The 2004 earthquake produces a patchwork of residuals, which are mostly positive in the near-fault region. For the 2007 earthquake, residuals are generally positive south of the hypocenter and negative to the north.

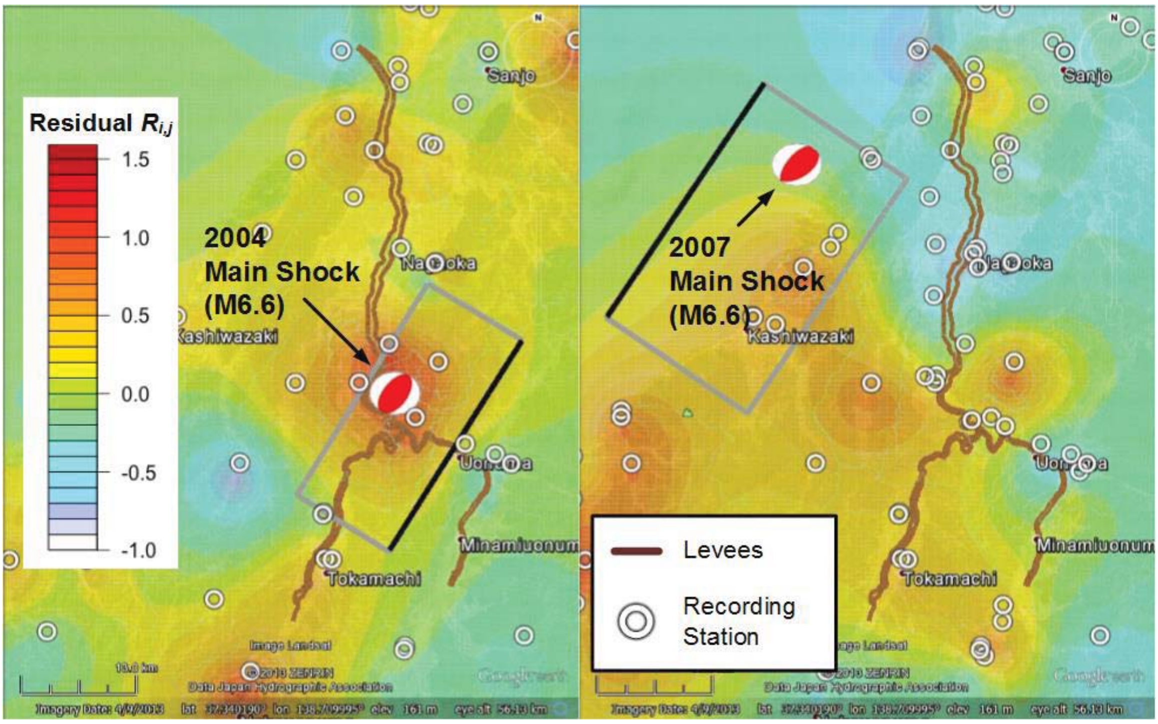

Figure 5.2 Contour maps of within-event *PGV* residuals from the Boore et al. (2014) GMPE for 2004 and 2007 earthquakes.

Figure 5.3 shows *PGA* and *PGV* profiles along the Shinano River levees produced by the proposed procedure and from relatively simple direct Kriging of ground motion data. The proposed procedure produces larger ground motion estimates (than those from direct Kriging) for levees near rock sites with moderate ground shaking (e.g., ~18 km from river mouth) and slightly smaller ground motions at most locations beyond 30 km from the river mouth. The larger ground motions near rock accelerograph sites result from relatively strong site responses at the levee sites, which amplify smaller estimated levee motions. In the regions beyond 30 km (from river mouth), typically the accelerographs and levee sites are both on soil, but small differences in the V_{S30} values (between accelerographs and levee sites) and the use of a nonlinear site term in the proposed procedure, produce the observed ground motion reductions.

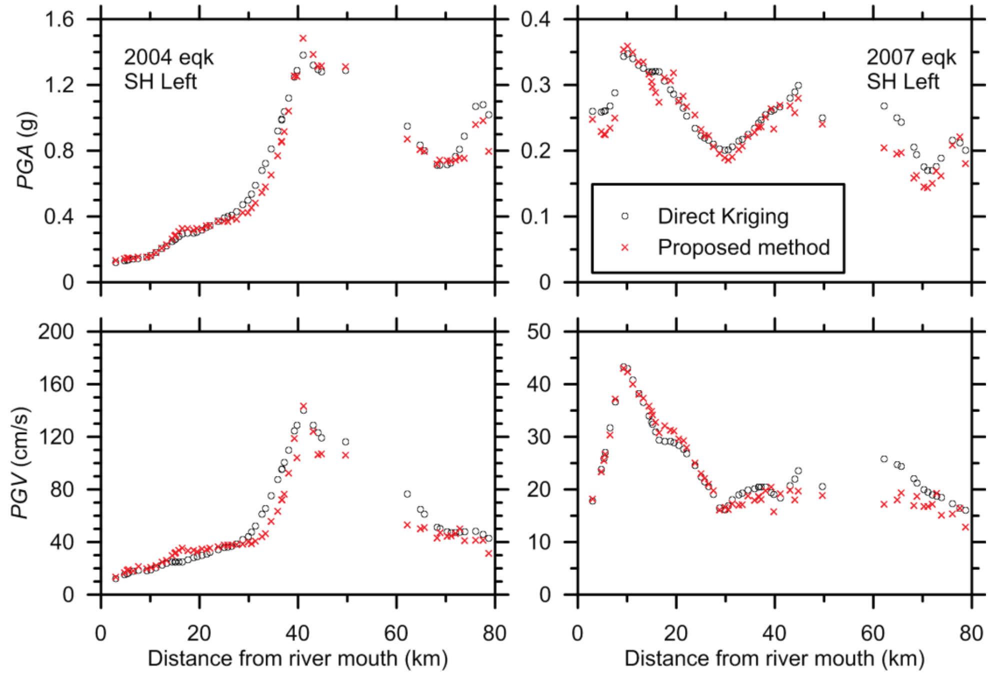

Figure 5.3 ***PGA* and *PGV* interpolated from seismic stations using direct Kriging and those estimated by proposed method using residuals analysis from the BEA GMPE.**

5.3 GROUND WATER ELEVATION ON EARTHQUAKE DATES

I describe procedures for estimating groundwater elevation on the earthquake dates. Groundwater levels were measured in geotechnical borings, but those water levels may not match those during earthquakes due to variations in the river water level over time and local agricultural practices. I estimate groundwater level on the earthquake date based on (*i*) measurements of levee groundwater elevation (*LGWE*) at the time of a geotechnical boring, (*ii*) measurements of river water elevation (*RWE*) from stream gauge stations on the borehole date, and (*iii*) *RWE* on the earthquake date.

My approach is to use available borehole data to evaluate the differential between *LGWE* and *RWE* at the time of subsurface exploration. This differential is then added to the *RWE* at the time of the earthquake to estimate *LGWE* on the earthquake date. A key assumption is that the *RWE* is directly related to *LGWE* since levees are adjacent to the river, but adjustments are made for levees with land-side irrigation. Measurement and selection of *LGWE* and *RWE* data are described in Section 2.6. I describe below how *RWE* was interpolated and analysis of the *LGWE*-*RWE* differential.

5.3.1 Interpolation of River Water Elevation

As described in Section 2.6.2, *RWE*s are measured from stream gauge stations hourly and daily; I sample the daily database on the earthquake date and the date of subsurface exploration. As shown in Figure 1.1, there are eleven stations along the study region, which is too sparse spatially (average distance between adjacent stations is 13 km) to provide accurate *RWE*s for each 50 m levee segment. For this reason, I also utilize *RWE* data from relatively detailed surveys performed after a flood (Oct 21 2004) and for maintenance purposes during a non-flood period having small *RWE* fluctuation (Oct 2009 ~ Feb 2010). These detailed surveys are used to improve the knowledge of the variation of *RWE* between stream gauges.

The relatively detailed surveys provide *RWE* profiles for portions of the levee system at a particular time; the data are not complete for the full 80 km of river length at any particular time, although data for the full river length are available for different times. The lengths of river for which the data at a given time apply are approximately 0.8 to 1.0 km (non-flood) and 10-30 km (flood). Given these complexities, the detailed survey data are best interpreted relative to coincident stream gauge measurements that are linearly interpolated between stream gauges. This approach is effective because the stream gauge data is available at regular time intervals and can be matched to the times of detailed *RWE* measurements. Residual elevations (*R*) at location *x* and time *t* are computed as follows:

$$R(x,t) = RWE_{data}(x,t) - RWE_{sg-li}(x_i, x_{i+1}, x, t) \tag{5.3}$$

where x_i and x_{i+1} indicate locations of the stream gauges immediately down- and upstream of *x*, $RWE_{data}(x,t)$ indicates a measured elevation from detailed surveys, and $RWE_{sg-li}(x_i,x_{i+1},x,t)$ indicates the linearly interpolated *RWE* at location *x* and time *t* from the nearest stream gauges.

Residuals are computed for both flood and non-flood conditions. Each set is smoothed using a running Hann window (Oppenheim and Schafer, 2010) of width 2.0 km. The smoothed residuals depend only on location and are denoted $\bar{R}(x,set)$, where '*set*' refers to the data set being evaluated (*fl* for flood or *nfl* for non-flood). Using these smoothed residuals, high-resolution *RWE* profiles can be evaluated through simple re-arrangement of Eqn. (5.3):

$$RWE(x,t) = RWE_{sg-li}(x_i, x_{i+1}, x, t) + \bar{R}(x, set) \tag{5.4}$$

Having established the above procedure to compute detailed *RWE* profiles, the next issue concerns applying these procedures to specific points in time; in particular dates of subsurface investigation along levees and the two earthquake dates. In general, a given date of interest corresponds to conditions intermediate between 'flood' and 'non-flood', so a weighted average value of $\bar{R}$ is computed for application in Eqn. (5.4):

$$\bar{R}(x,t) = w_{fl}(x,t)\bar{R}(x,fl) + w_{nfl}(x,t)\bar{R}(x,nfl) \tag{5.5}$$

where w_{fl} and w_{nfl} are location- and time-specific weights that reflect the probability of having *RWE* at location *x* and time *t* corresponding to *fl* and *nfl* conditions, respectively. Those weights are computed from stream gauge *RWE*s upstream and downstream of *x* at time *t*. Figure 5.4 illustrates the manner by which weights w_{fl} and w_{nfl} are computed. Essentially, weights are assigned to the flood or non-flood conditions on the basis of water elevation on the date of interest at streamgauges *i* and *i*+1. For example, high *RWE*s give more weight to the *fl* residual set. The weighting scheme also considers the location of the point of interest between the upstream and downstream streamgauges, giving more weight to data from the closer gauge. The weights are computed as:

$$\begin{aligned} w_{fl}(x,t) &= \frac{1}{L_i}\left(\begin{array}{l} (x_{i+1}-x)\dfrac{RWE(x_i,t)-RWE(x_i,nfl)}{RWE(x_i,fl)-RWE(x_i,nfl)} + \\ (x-x_i)\dfrac{RWE(x_{i+1},t)-RWE(x_{i+1},nfl)}{RWE(x_{i+1},fl)-RWE(x_{i+1},nfl)} \end{array} \right) \\ w_{nfl}(x,t) &= \frac{1}{L_i}\left(\begin{array}{l} (x_{i+1}-x)\dfrac{RWE(x_i,fl)-RWE(x_i,t)}{RWE(x_i,fl)-RWE(x_i,nfl)} + \\ (x-x_i)\dfrac{RWE(x_{i+1},fl)-RWE(x_{i+1},t)}{RWE(x_{i+1},fl)-RWE(x_{i+1},nfl)} \end{array} \right) \end{aligned} \tag{5.6}$$

where L_i is the distance between streamgauges *i* and *i*+1 and *RWE*s at indexed values of distance are data from streamgauges. Weights evaluated using this process for borehole exploration dates emphasize the *nfl* condition because borings were generally drilled in non-flood season ($w_{nfl} \approx 1.0$; $w_{fl} \approx 0.0$). Weights for the earthquake dates were approximately $w_{nfl} \approx 0.61$ and $w_{fl} \approx 0.39$ (2004 event) and $w_{nfl} \approx 0.77$ and $w_{fl} \approx 0.23$ (2007 event).

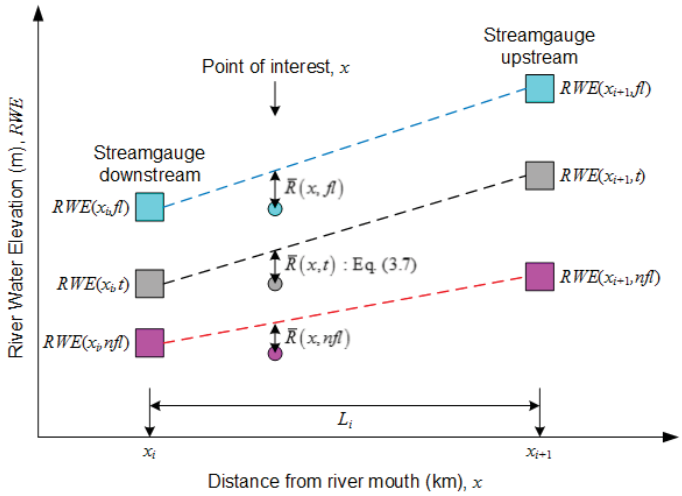

Figure 5.4 **Schematic drawing of geometric positions for variables used in weight calculation.**

Figure 5.5a shows *RWE*s during the flood event (Oct 21, 2004) and a representative date for the non-flood survey (Dec 1, 2009) along with linear interpolations between stream gauges. Those plots illustrate the poor fit of the linear interpolation function and the different shapes of the between-stream gauge *RWE* profiles for the *fl* and *nfl* conditions (particularly in the upstream region, > 60 km in SH2, and in the downstream region, < 10 km in SH1). These differences are what motivated the development of the interpolation scheme. A noteworthy feature of the *RWE* profiles occurs at x = 9.3 km, where the *nfl* *RWE*s abruptly drop 5 m but the *fl* *RWE* profiles are relatively flat. This difference occurs because of a weir at 9.3 km that retains a small reservoir under non-flood conditions and which overtops in floods.

Figure 5.5 b and c show *RWE* profiles for the 2004 and 2007 earthquake dates as given by the above procedure with linear interpolation shown for reference purposes. Special accommodations were needed for the Uono River (UO) because it was not surveyed during the Oct 21 2004 flood event, although non-flood surveys are available. Thus, I use w_{nfl} = 1.0 in Eqn. (5.5) and the same spatial interpolation scheme described above for the Shinano River.

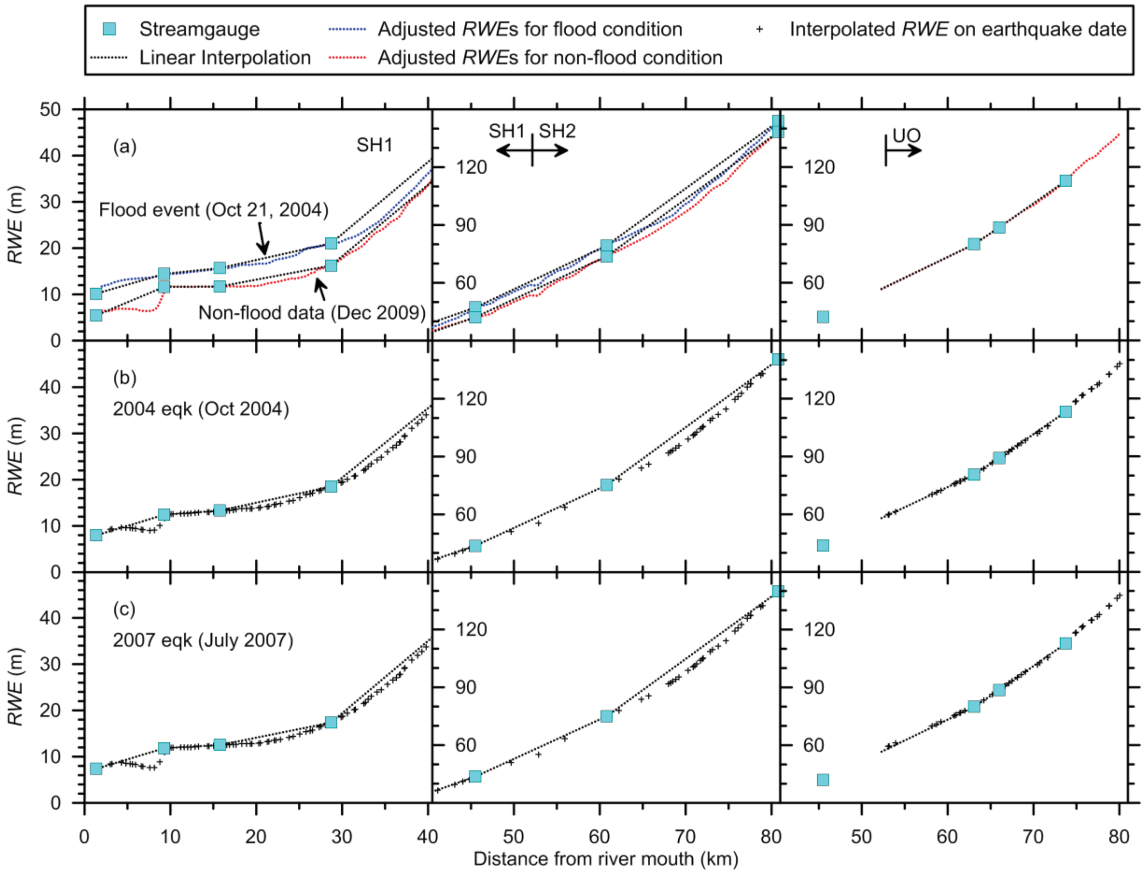

Figure 5.5 ***RWE*** **profiles for (a) dates of detailed surveys for flood and non-flood conditions, (b) the 2004 earthquake, and (c) the 2007 earthquake dates. Linear interpolations between stream gauges are also shown.**

5.3.2 Correlation between Groundwater Elevation and River Elevation

I hypothesize that the difference between the ground water elevation beneath the levee (*LGWE*) and the river water elevation (*RWE*) might vary seasonally due to local agricultural practices on the land-side of the levees, thereby requiring a time-dependent adjustment. In this section, I examine profiles of this differential elevation over the river length and test its stability relative to periods of time when agriculture-related irrigation is or is not occurring.

Figure 5.6 shows *LGWE*-*RWE* differentials from all observed *LGWE*s in boreholes and from *LGWE*s screened as described in Section 2.6.1 (i.e., data points meeting at least one of the three criteria). I plot the data separately for the growing and non-growing seasons. During the growing season (approximately June-September; FAO, 2004), there can be significant land-side irrigation for rice and other crops. As shown in Figure 5.6, in the mid- and downstream areas of the Shinano River (SH1) (river distance 15 ~ 25 km), the *LGWE-RWE* differential during the growing season is modestly greater than during the non-growing season for left-side levees, whereas both are similar for right-side levees. The differences between the two sides of the river can be explained based on the configuration of irrigation canals and other features described subsequently. In upstream areas (SH2 and UO; river distance > 50 km), borings were mostly performed during the non-growing season so differentials cannot be compared.

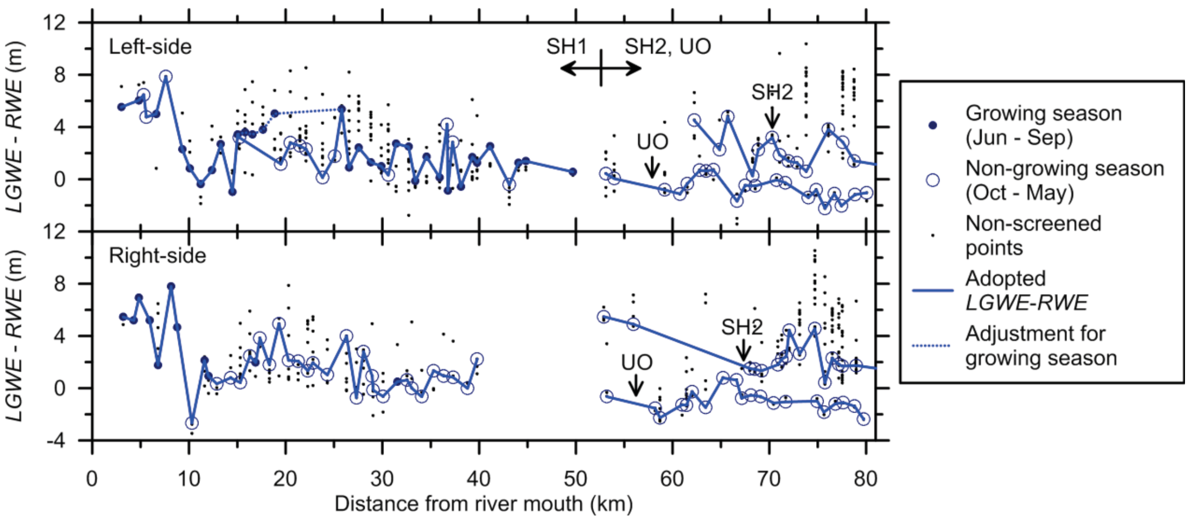

Figure 5.6 ***LGWE-RWE*** **differentials predicted in both growing season (June – September) and non-growing season (October – May) along SH and UO rivers. The data gap from 40-60 km corresponds to a lack of levees (natural channel).**

To provide insight into the possible cause of seasonal differences in *LGWE-RWE* on the left side of the river and the lack of such differences on the right side, I plot in Figure 5.7 an aerial view of the region showing land use. On the left-side, the river is relatively near a mountain, especially for river distances of 10 ~ 17 km, whereas the right side is a broad alluvial plain containing a stream flowing near the land-side toe of the levees for the river distance range of 10 ~ 20 km. I postulate that the lack of irrigation effect on the right side is caused in part by the adjacent stream isolating the levees from the irrigated farm areas (interestingly, there is also a lack of irrigation effect from 20-23 km where the stream is not adjacent to the river). Another stream flows near to left-side levees from 15 ~ 17 km, but there is farm land between the stream and land-side of the levees, so the isolation effect is not present. A residential area that is free from irrigation is present for river distances > 25 km on the left-side and > 23 km on the right-side.

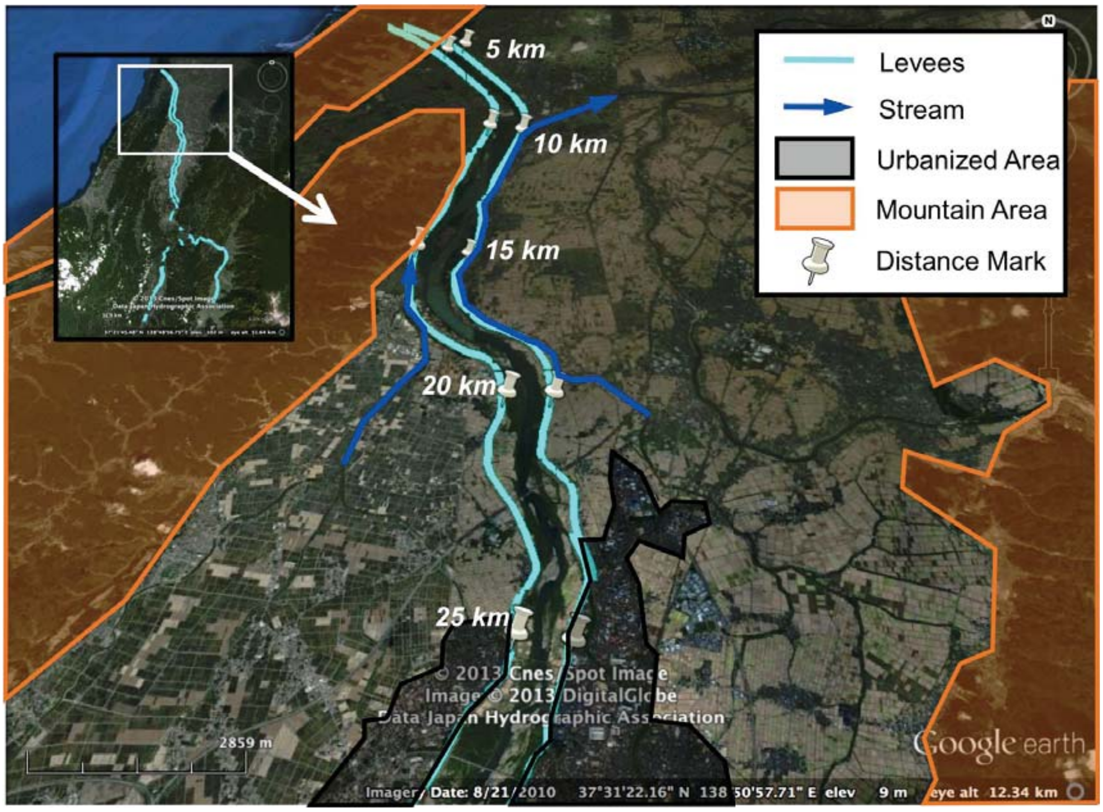

Figure 5.7 **Aerial view of the Shinano River downstream area through Google Earth with land uses in the vicinity of levees.**

Based on the above, I conclude that during periods of heavy irrigation, the *LGWE* is controlled by irrigation and less influenced by *RWE* over the 15 ~ 25 km interval on left side levees, but elsewhere there is no tangible irrigation effects. Figure 5.6 shows the *LGWE*-*RWE* profiles adopted for subsequent analysis.

5.3.3 Groundwater Elevation in Levees at the Time of Earthquake

*LGWE*s on the earthquake dates are computed as the sum of *RWE*s shown in Figure 5.5 and the differentials (*LGWE*-*RWE*) shown in Figure 5.6 [non-growing season differentials were used for the 2004 earthquake (October); growing season differentials were used for the 2007 earthquake (July)]. Levee base elevations (i.e., *LBE*s) are taken at the fill-native soil contact beneath the levee crest, as indicated from boreholes and cross sections (Figure 2.7). I then compute the differential $D_W = LGWE - LBE$, which is shown in Figure 5.8 for SH1, SH2 and UO. Note that D_W has a cap of 5 m, which is the average levee height in the study region (*LGWE* cannot be higher than the levee crest). This cap is applied near the river mouth.

*LBE*s are generally lower than *LGWE*s (positive D_W) at river mouth distances less than 30 km (indicating that levee fill in this region may be saturated over some depths), and are generally higher (negative D_W) at greater river distances. The 2004 earthquake occurred two days after a flood event, so D_W values were high, particularly in downstream areas. For the 2007 earthquake, left-side levees at river mouth distances of 15-25 km have elevated *LGWE*s due to land-side irrigation, which produces relatively high D_W values. Upstream areas have similar D_W values (generally negative) for both events.

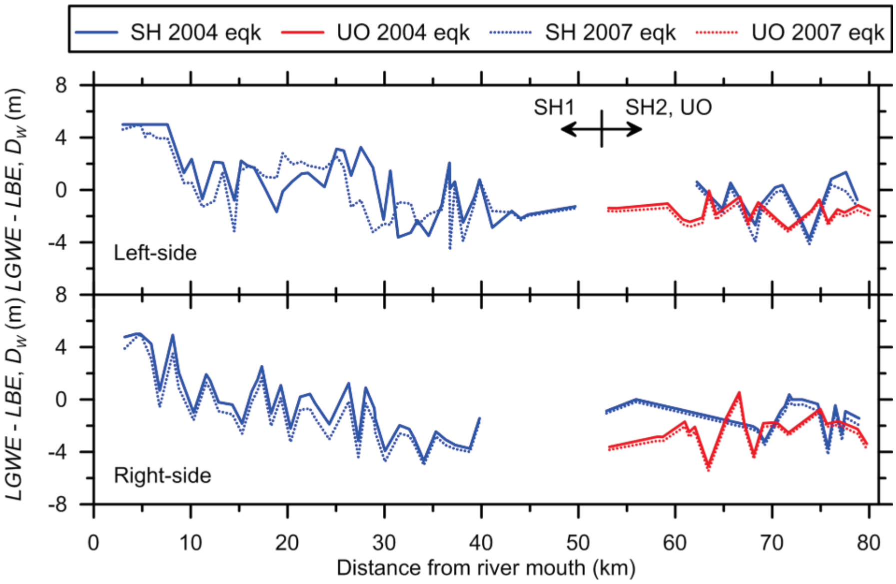

Figure 5.8 **Profiles of differential D_W between levee groundwater elevation and levee base elevation on dates of 2004 and 2007 earthquakes.**

6 DEVELOPMENT OF FRAGILITY FUNCTION

6.1 INTRODUCTION

In this chapter I construct damage models defining seismic fragility of levees as a function of ground motion intensity, ground water elevation relative to levee toe, levee shape, and mapped surface geology. These quantities are typically readily available without a site-specific geotechnical investigation.

I employ conventional probability concepts that are widely used in performance-based earthquake engineering (PBEE) for the damage model. The PBEE's methodology estimates the probability of exceeding decision variables (*DV*s, e.g., fatalities, financial losses and downtimes) by triple integration of damage measures (*DM*s, e.g., damage states of structural and nonstructural elements), engineering demand parameters (*EDP*s, e.g., interstory drift ratios, inelastic component deformations and floor acceleration spectra), and intensity measures (*IM*s). The general equation for calculating the exceedance rate of a *DV* in PBEE is as follows (Moehle and Deierlein, 2004; Porter, 2003):

$$\upsilon(DV) = \iiint G(DV \mid DM)\, dG(DM \mid EDP)\, dG(EDP \mid IM) \mid d\lambda(IM) \mid \tag{6.1}$$

where $\upsilon(DV)$ is the mean annual rate of exceeding a given level of $DV = dv$, and $G(DV|DM)$ is the probability of exceeding a given level of $DV = dv$ conditioned on $DM = dm$, i.e.,

$P(DV>dv|DM=dm)$, where the same format applies to other terms, $G(DM|EDP)$ and $G(EDP|IM)$. The remaining term $\lambda(IM)$ is the mean annual rate of exceedance for $IM = im$.

My application is relating IM (ground motion metrics) to DM (indicators of levee performance), which is commonly referred to as a *fragility function*. The forward application of such a function utilizes a simplification of Eqn. (6.1) as follows:

$$\upsilon(DV) = \int G(DM \mid IM)\, d\lambda(IM) \tag{6.2}$$

where $d\lambda(IM)$ represents the derivative of the hazard curve and $G(DM|IM)$ the fragility function that is developed in the remainder of this chapter.

Fragility functions are widely used in PBEE, typically for structural applications (e.g., Aslani and Miranda, 2005; Pagni and Lowes, 2006; Porter et al., 2007). As shown in Figure 6.1, a typical structural application of a fragility function is to relate a ground motion IM to a probability of a certain EDP (or relating EDP to DM, as in Figure 6.1). In structural applications, the fragility function is typically established from nonlinear analysis of the structural systems. In my application, I seek to evaluate this relationship empirically using the data presented in Chapters 2 and 5. I develop fragility functions that express the probability of $DL > dl$ conditioned on the IM (i.e., PGA and PGV) alone (Model 1) or IM in combination with other parameters descriptive of hydrological and geotechnical conditions (Model 2). Conditions investigated for Model 2 are IM, geomorphological and geological classification (G_N), levee groundwater elevation with respect to levee base (D_W), and shape factor (S_F), which is denoted as secondary parameters (SP).

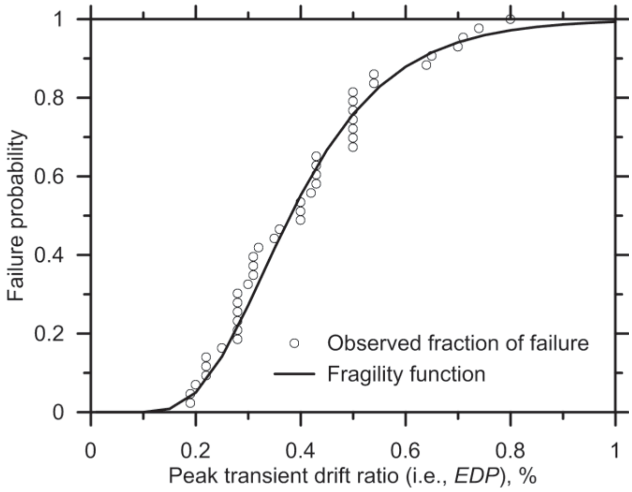

Figure 6.1 **An example of fragility function used in structural application (after Porter et al., 2007).**

Using probability notation, Models 1 and 2 can be expressed as:

$$\text{Model 1: } P\left(DL > dl \mid IM = im\right) \tag{6.3}$$

$$\text{Model 2: } P\left(DL > dl \mid IM = im, SP = sp\right) \tag{6.4}$$

Both Models 1 and 2 are constructed in two stages. In Stage 1, I compute the probability of $DL >$ 0 (i.e., the probability of some damage, regardless of its severity). In Stage 2, I compute the probability of exceeding particular damage levels conditional on some damage having occurred along with the additional metrics. Using the example of Model 1, these Stages can be expressed as:

$$\text{Stage 1: } P\left(DL > 0 \mid IM = im\right) \tag{6.5}$$

$$\text{Stage 2: } P\left(DL > dl \mid DL > 0, IM = im\right) \tag{6.6}$$

The desired Model 1 probability (Eqn. 6.3) is taken as the product of the Stage 1 and 2 probabilities using the total probability theorem (Ang and Tang, 2007):

$$P\left(DL > dl \mid IM\right) = P\left(DL > 0 \mid IM\right) \times P\left(DL > dl \mid DL > 0, IM\right) \tag{6.7}$$

In Section 6.2, I review the methodology of fragility function for my data set. In Section 6.3-6.4, I fit the data to various functional forms and develop the Stage 1 and 2 components of Models 1 and 2.

6.2 METHODOLOGY OF FRAGILITY FUNCTIONS

Methodologies for constructing fragility functions have been presented by Porter et al. (2007) and Baker (2014). An underlying assumption in those studies is that the functional form for a cumulative distribution function (CDF) (e.g., normal or log-normal) can be fit to data expressing probabilities of damage for various levels of seismic demand. The use of a CDF has the advantages of operating between the required probability range of zero to one, capturing commonly encountered data distributions, and being described by physically meaningful parameters (typically a mean and standard deviation). For example, if a log-normal CDF with mean (μ) and standard deviation (β) is fit to data on the probability of exceeding a damage level (dl) conditional on intensity measure IM, the fragility function can be defined as follows:

$$P\left(DL > dl \mid IM = im\right) = \Phi\left(\frac{\ln im - \mu}{\beta}\right) \tag{6.8}$$

where Φ represents the standard normal CDF with mean 0 and standard deviation 1.

6.2.1 Methodology for Defining Fragility Functions by Porter et al. (2007)

Porter et al. (2007) present methodologies for computing losses in a performance-based earthquake engineering (PBEE) framework given variable levels of data quality and availability (Methods A to E, and U). Method B describes a situation in which the peak engineering demand parameters (*EDP*s, e.g., interstory drift ratios) or intensity measures (*IM*s) to which specimens were subjected are known and there is knowledge about which specimens exceeded a damage state. Method B corresponds to the situation with the subject levees, since we know where damage occurred and the associated peak levels of ground shaking have been estimated for each segment. A common characteristic of Method B is that the data does not extend to sufficiently extreme demands that the *EDP* for high failure probabilities can be empirically defined. The Porter et al. (2007) methodology for Method B is as follows:

1) Data are grouped into bins defined by ranges of *EDP* (or *IM*). For each bin j, the number of damaged specimens (m_j) and total number of specimens (M_j) is identified along with the average *EDP* (or *IM*) for the bin (denoted x_j). Note that m_j/M_j is the fraction of damaged segments in bin j.

2) Transform the failure probability for each bin j to epsilon y using the inverse standard normal CDF operator (Φ^{-1})

$$y_j = \Phi^{-1}\left(\frac{m_j + 1}{M_j + 1}\right) \tag{6.9}$$

One is added to both the denominator and nominator to avoid negative infinity values of y.

3) Fit a line to the (x_j, y_j) pairs using the following equation:

$$\hat{y} = sx + c \tag{6.10}$$

where s and c are parameters established using least-squares regression. This is illustrated in Figure 6.2.

4) Calculate the mean (μ) and standard deviation (β) of the log-normal CDF (which is the fragility function) from regression parameters as follows:

$$\beta = 1/s$$
$$\mu = -c\beta \tag{6.11}$$

The resulting fragility function has the functional form and shape of a log-normal CDF with mean μ and standard deviation β. Note that x and μ used in Eqn. (6.10-11) are in natural log units for a log-normal CDF. This technique can be readily extended to normal CDFs by using x and μ in arithmetic units. Other CDFs can be used with this method, for example, beta CDF and gamma CDF with two variates. Distributions of those CDFs can be converted to a linear form as in Figure 6.2, so the regression results (i.e., s and c) can be converted to moments of the CDF (K. Porter, *personal communication*, 2013).

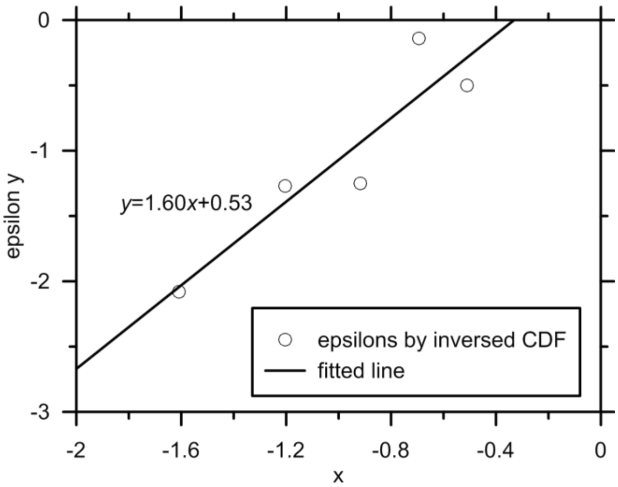

Figure 6.2 **An example of fitted line using least squares to epsilons transformed by inversed CDF (after Porter et al., 2007).**

6.2.2 Methodology for Defining Fragility Functions by Baker (2014)

Baker (2014) describes methods for defining fragility functions for data conditions analogous to those associated with Method B of Porter et al. (2007). Baker considered an *EDP*|*IM* relationship, with the *EDP* being collapse and *IM* being first-mode spectral acceleration. The "data" supporting the fragility functions were derived from structural simulations that were performed for scenario events (conditional spectra), and only certain fractions of the motions induced collapse even for large demands. Hence, the *IM*s required for high failure probabilities were often unknown (similar to the problem with Method B).

Baker's methodology for defining a log-normal fragility function begins, as with Porter et al. (2007), by grouping data into bins defined by ranges of *IM*. For each bin j, the number of collapse cases (m_j) and total number of simulations (M_j) is identified along with the average *IM* for the bin (denoted x_j). Figure 6.3 is effectively a plot of this data, where $P(collapse) = m_j/M_j$ and x_j is the spectral acceleration value on the abscissa.

I then recognize that the probability of observing m_j collapses out of M_j ground motions with $S_a = x_j$ is given by a binomial distribution:

$$\mathrm{P}\left(m_j \text{ collapses in } M_j \text{ motions}\right) = \binom{M_j}{m_j} P_j^{m_j} \left(1 - p_j\right)^{M_j - m_j} \tag{6.12}$$

where $\binom{n}{k}$ is known as the binomial coefficient as follows (Ang and Tang, 2007):

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{6.13}$$

Quantity p_j is the probability that a ground motion with $S_a = x_j$ will cause collapse, as specified by a certain fragility function, which is to be determined. The maximum likelihood estimation (MLE) provides a procedure to define a fragility function having the highest probability of predicting the observed data. As shown in Figure 6.3, I typically have data at multiple *IM* levels (let's suppose there are N such levels). The product of the binomial probabilities at each *IM* level is used to obtain the likelihood function

$$\text{Likelyhood} = \prod_{j=1}^{N} \binom{M_j}{m_j} P_j^{m_j} \left(1 - p_j\right)^{M_j - m_j} \tag{6.14}$$

where Π denotes a product over all N IM levels. Baker seeks a fragility function that maximizes this likelihood.

The maximization is performed by substituting an appropriate CDF into Eqn. (6.14), and then finding the moments of that CDF (i.e., $\hat{\mu}$ and $\hat{\beta}$ for a log-normal CDF) that maximize the likelihood. In the case of a log-normal CDF, the expression substituted for p_j is:

$$\hat{P}_j = \Phi\left(\frac{\ln x_j - \mu}{\beta}\right) \tag{6.15}$$

To be clear on nomenclature, the maximum likelihood operation finds the values of μ and β that maximize the likelihood function in Eqn. (6.14), and those values are denoted $\hat{\mu}$ and $\hat{\beta}$.

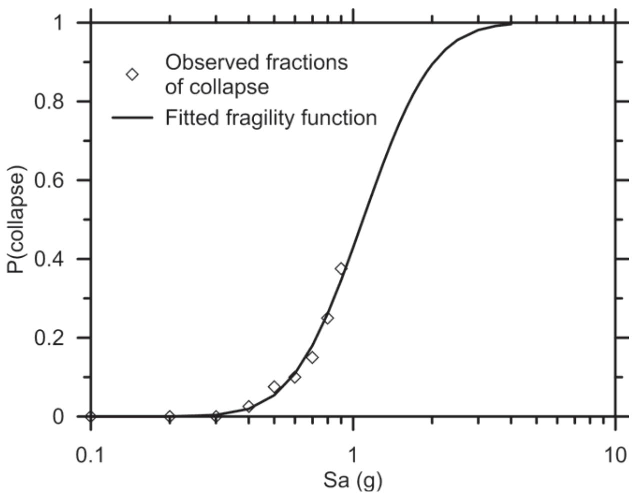

Figure 6.3 An example of fitted fragility function to observed fractions of collapse using MLE method (after Baker, 2014)

6.3 MODEL 1: PROBABILITY OF DAMAGE CONDITIONAL ON INTENSITY MEASURES ONLY

6.3.1 Selection of Fragility Function

In Section 6.2 I introduced two methodologies for constructing fragility functions, which are quantified in the form of moments of normal (or log-normal) CDFs. One method (Porter et al.,

2007) finds moments using a least squares regression on the data transformed via the standard normal variate (epsilon). The second method (Baker, 2014) finds moments using a maximum likelihood estimation (MLE) method. I fit data points from the data set (described in subsequent section) to a log-normal CDF using both fitting procedures, with the results shown in Figure 6.4. There are moderate differences between the two sets of moments, but the fitted fragility curves in Figure 6.4 are visually similar over the *PGA* range of the data. The standard deviations (σ) of residuals (i.e., observed data minus fitted curves) for *PGA* and *PGV* data points are 0.069 and 0.065 for MLE, and 0.071 and 0.069 for least-squares, respectively.

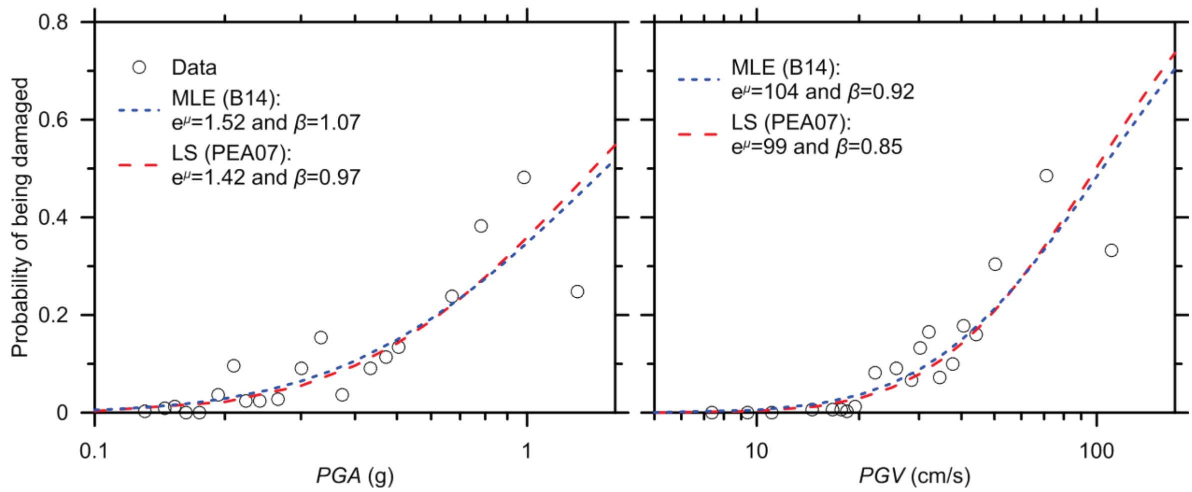

Figure 6.4 **Comparison of log-normal CDF forms fitted using Porter et al. (2007) (PEA07) and Baker (2014) (B14) methodologies.**

I select the MLE method to identify the parameters describing fragility functions since it is applicable to any functional form and produces lower σ of residuals for my data set. I next consider alternate fragility functions, in particular, normal CDF, log-normal CDF, and a linear function. The linear function is bracketed between zero and one as follows:

$$y = \begin{cases} 0 & \text{if } b_0 + b_1 IM + \varepsilon < 0 \\ b_0 + b_1 IM + \varepsilon & \\ 1 & \text{if } b_0 + b_1 IM + \varepsilon > 1 \end{cases} \tag{6.16}$$

where b_0 and b_1 are regression parameters and ε is error term with zero mean and a standard deviation of σ.

Figure 6.5 shows fits of normal CDF, log-normal CDF, and the linear function in Eqn. (6.16). Since each of these functions is fit to the data using the MLE method, the proper metric for goodness-of-fit is the maximized likelihood (Eqn. 6.14) instead of standard deviation of residuals. The log values of MLEs are -223.5, -153.1, and -163.8 for the normal, log-normal CDF, and linear functions for *PGA*, and -245.8, -132.5, and -151.2 for *PGV*, respectively. Recall that likelihood is maximized, so the larger values of likelihood indicate better fits.

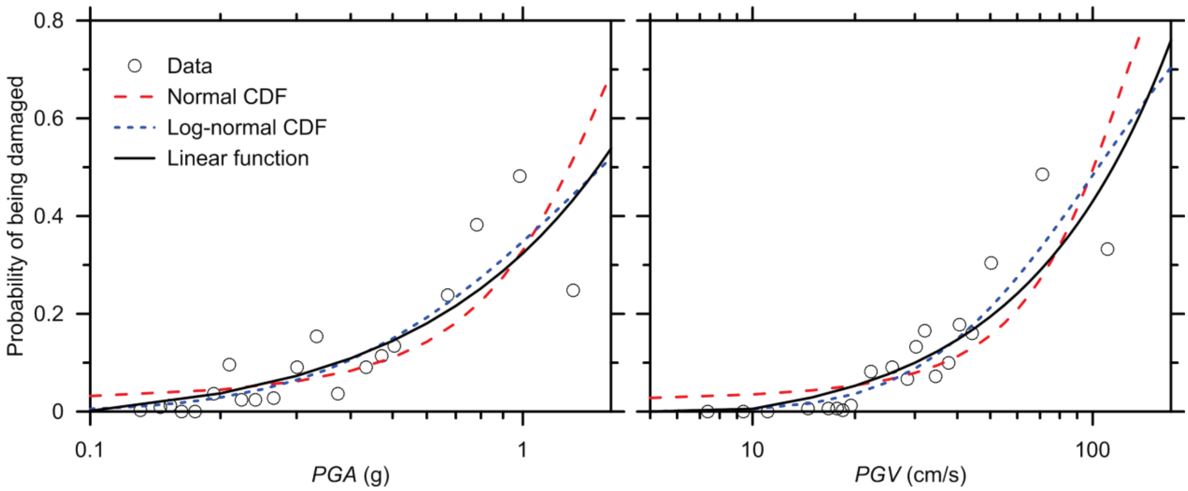

Figure 6.5 **Comparison of various functional forms (i.e., normal CDF, log-normal CDF, and linear function) for the best-fit fragility function using Baker (2011) methodology.**

The normal CDF has lower MLE than the other considered functions and a visually poorer fit to the data in Figure 6.5. The linear and log-normal CDF provide nearly equivalent MLE values and fragility function shapes over the *PGA* range 0.1-1.6g. I proceed using the log-normal CDF fragility function because its parameters are moments of the log-normal CDF (i.e., μ and β) having clear physical meaning.

6.3.2 Stage 1: Probability of Damage Occurrence

The most basic fragility function describes the probability of experiencing damage at any level [i.e., P(DL > 0)] conditioned only on a ground motion IM (I have considered PGA and PGV). I denote this approach as Model 1-Stage 1. Probabilities of exceeding higher damage levels are addressed later.

Data on levee performance is segregated by IM level by organizing the data into discrete bins. The probability of levee damage for a bin j with median im_j can be calculated as follows:

$$P_j\left(DL > dl \mid IM = im_j\right) = \frac{ND_j}{N_j} \tag{6.8}$$

where ND_j is the number of damaged segments and N_j is the number of total segments in bin j. Figure 6.6 shows distributions of damage probabilities for the IMs of PGA and PGV. The plot shows both damage probabilities and ND_j for each bin, which are directly related because the bins are of equal size in terms of numbers of samples, which in turn requires unequal bin width. Bin size can be related to the square-root of the total number of data points (M) as follows:

$$N_{bin} = \frac{\sqrt{M}}{4} \tag{6.9}$$

The use of four in the denominator is a modification of the recommendations of Porter et al. (2007), which had one in the denominator. The modification is motivated by the small number of observations for high IM and the need to reduce the data count requirements for those bins. My dataset has M = 6636 levee segments, which results in N_{bin} = 20, and N_j = 6636/20 = 332 segments per bin. An advantage to this approach is that equal weight is given to each bin in the maximum likelihood estimation of the CDF moments. The fragility generally monotonically increases with PGA from 0.14g to 1.0g and with PGV from 14 cm/s to 80 cm/s. No damage occurs below approximately 0.14g or 14 cm/s and the damage probabilities reach as high as approximately 0.5 for large *IM*.

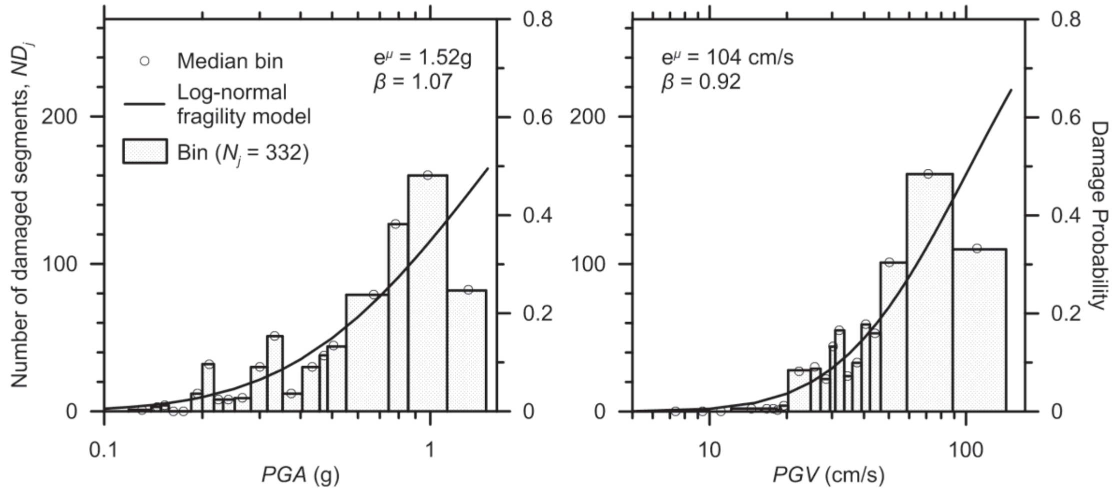

Figure 6.6 Probability of damage at any level conditional on intensity measures *PGA* and *PGV*. Results expressed using number of damaged segments in bins of unequal width (left) and probabilities (right). Log-normal CDF fit to data using MLE.

Figure 6.6 also shows the fit of the log-normal CDF to the data along with the identified μ and β values. Lower values of dispersion (β) indicate increased predictive power of the *IM*. In my case, *PGV* produces modestly lower dispersion (0.92) than does *PGA* (1.07); accordingly, I utilize *PGV* as the *IM* in Model 2 analysis. The dispersions shown in Figure 6.6 are high compared to those found in other earthquake engineering applications (e.g., values of 0.4-0.5 for many structural applications, e.g., Aslani and Miranda, 2005; Pagni and Lowes, 2006). I suspect that the relatively high β occurs because of uncertainties associated with analysis of empirical field performance data (prior studies are either analytical or use data from laboratory-scale testing); in particular, the estimation of *IM*s (not measured on-site) and the lack of detailed, section-specific, information on levee characteristics.

6.3.3 Stage 2: Probability of Exceeding Damage Level within Damaged Segments

In Stage 2, I evaluate the probability of exceeding particular damage levels among damaged segments. The probability of exceeding a damage level (dl) in bin j with median im_j conditioned on $DL > 0$ is calculated as follows:

$$P_j\left(DL > dl \mid IM = im_j, DL > 0\right) = \frac{ND_j^{dl}}{N_j^{dl}} \tag{6.10}$$

where ND^{dl}_j is the number of damaged segments with damage level $DL > dl$ in bin j, and N^{dl}_j is the number of total damaged segments in bin j. As shown in Figure 6.7, unlike Stage 1, the resulting conditional probabilities in Stage 2 do not exhibit a clear dependence on IMs (Figure 6.7a for PGA and Figure 6.7b for PGV). I performed t-test to evaluate the statistical significance of a linear trend line (Appendix B) relative to an IM-independent mean. The slope of the linear function is parameter b_1. If the slope is not statistically significant, an IM-independent mean probability will be used for Stage 2.

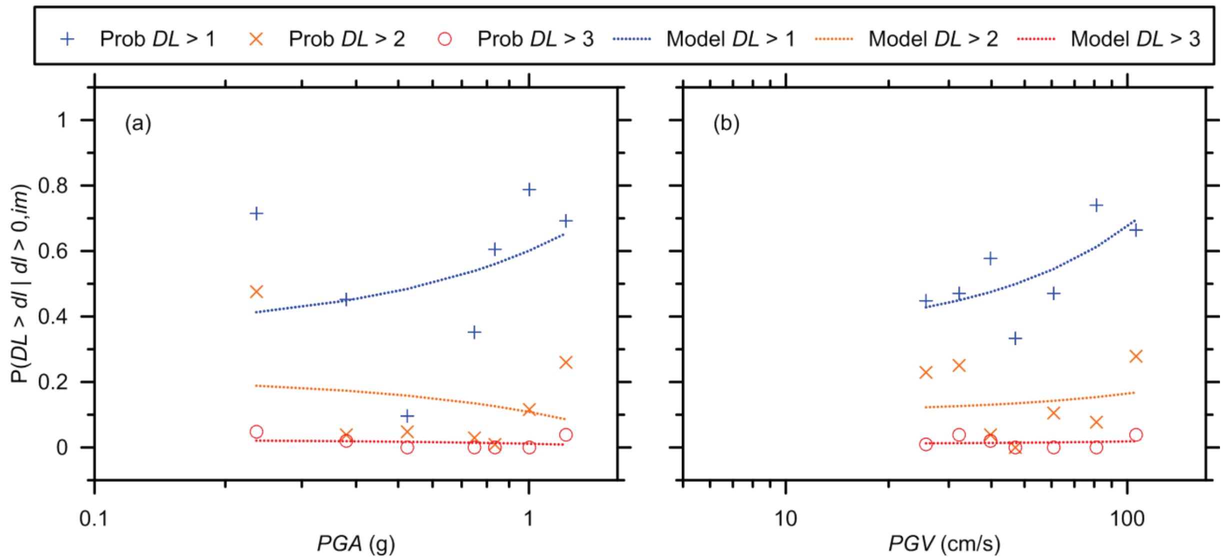

Figure 6.7 **Probability of exceeding damage levels above one (Stage 2) with linear fits regressed by least squares for *PGA* and *PGV*.**

A p-value, the probability of exceeding t in the t-distribution, of 10% or less is often used to identify a statistically significant non-zero b_1 value. The resulting p-values for DL > 1, 2, and 3 are 0.442, 0.647, and 0.651 for PGA, and 0.088, 0.756, and 0.779 for PGV, respectively, which indicates that none of cases are statistically significant. Accordingly, I utilize IM-independent mean probabilities for Stage 2 analysis, as shown in Table 6.1. Note that mean probabilities are independent from PGA and PGV. One exception is DL >1 with PGV, which has 8.8% p-value and is dependent on PGV. I again use the log-normal CDF for the functional form of this case, for which the moments are e^{μ} = 43 cm/s and β = 2.2.

Table 6.1 Mean probabilities for exceeding various damage levels conditioned on *DL* > 0 as Stage 2 analysis.

Damage Level	$DL > 0 \mid DL > 0$	$DL > 1 \mid DL > 0$	$DL > 2 \mid DL > 0$	$DL > 3 \mid DL > 0$
Mean probability	1	0.529	0.140	0.015

The lack of IM-dependence in these damage probabilities may be related to the strong correlation between damage and liquefaction, as shown in Figure 2.3. Many of the damaged levee sections had surface manifestation of liquefaction, the effects of which may have only modest IM sensitivity for the relatively soft soil conditions at the investigated sites. In other words, once the IM is high enough to trigger liquefaction, the subsequent effects may not be strongly IM-dependent for the conditions at the subject sites. More detailed analysis of levee sections using site-specific geotechnical data will shed more light into this issue. That work is ongoing.

Stages 1 and 2 can be combined to estimate the probability of exceeding a damage level dl given IM. Using the total probability theorem, this is the product of the probabilities of damage occurring (Stage 1) and exceeding damage level dl conditional on damage occurring

(Stage 2) (Eqn. 6.7). For example, considering *PGA* = 0.4g, the log-normal CDF with $e^{\mu} = 1.52$ and $\beta = 1.07$ in Stage 1 gives a probability of 0.11 and the mean probability of *DL* > 2 in Stage 2 is 0.14. The product results in a probability of 0.015 for *DL* > 2.

As an alternative to the use of the two-stage total probability approach, a fragility function could be regressed directly from probabilities of *DL* > *dl*. As shown in Table 6.2, this produces extremely high means of log-normal CDFs for high damage levels (e.g., *DL* > 2 and *DL* > 3). For example, the mean for *DL* > 3 for *PGA* in arithmetic units is 3998*g*, which clearly has no physical meaning. This is caused by high damage levels having very low rates of occurrence for the conditions present in the Niigata data set. I prefer the staged approach to avoid the introduction of physically unrealistic values to high-damage level fragility functions.

Table 6.2 Moments of log-normal CDF directly regressed from data for each damage level.

Damage Level	Moments of log-normal CDF for *PGA*		Moments of log-normal CDF for *PGV*	
	e^{μ} (g)	β	e^{μ} (cm/s)	β
DL > 0	1.52	1.07	104	0.92
DL > 1	3.07	1.27	181	1.05
DL > 2	104	2.59	1167	1.63
DL > 3	3998	3.14	8234	1.85

Figure 6.8 compares the fragility functions obtained from the combination of Stage 1 and 2 models with those from direct regressions as described above. The curves for *DL* > 0 are identical and are very similar for other damage levels for *PGA* < 0.7g and *PGV* < 70 cm/s. The differences increase slightly for *PGA* > 0.7g and *PGV* > 70 cm/s but remain small over the range of data.

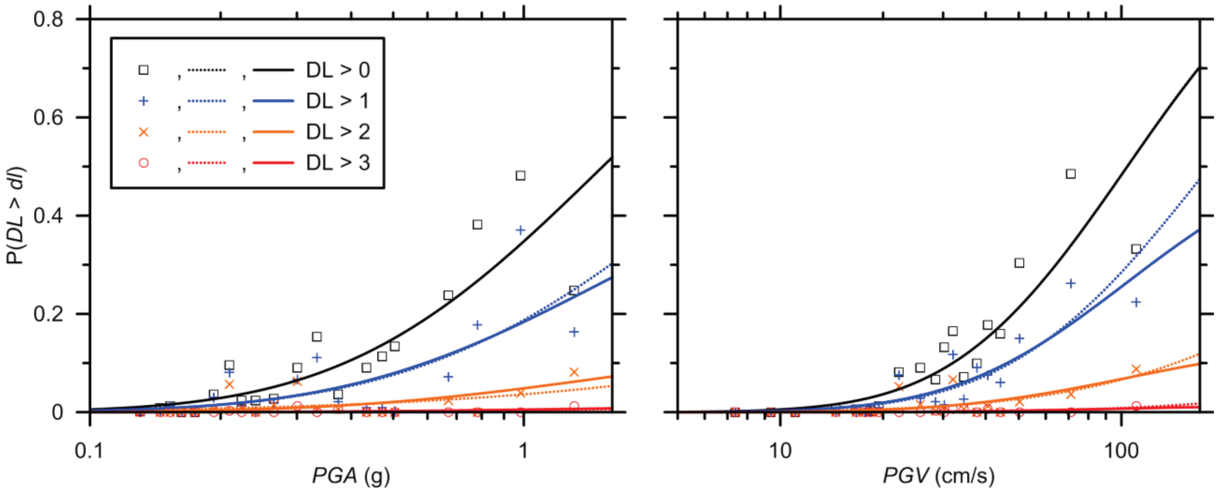

Figure 6.8 **Comparison between log-normal CDFs directly regressed (dotted line) and those scaled by mean probability (solid line) from *DL* > 0 curve.**

6.4 MODEL 2: PROBABILITY OF DAMAGE WITH SECONDARY CONDITIONS

In this section I evaluate *PGV*-dependent levee fragilities conditioned on secondary parameters: surface geology of foundation soils (G_N), ground water elevation relative to the levee base (D_W), or levee shape factor (S_F). I refer to these models collectively as Model 2. The fragility computed here is the probability of damage at any level [i.e., P(DL > 0)]; fragilities related to higher damage levels (Stage 2) are addressed in the next section. Note that only *PGV* is chosen as *IM* in Model 2 analysis due to better prediction power than *PGA* (Figure 6.6).

I considered developing a multi-parameter model using *PGV* with the secondary parameters, but instead chose to evaluate the parameters one at a time to see if they have predictive power for levee fragility. This is done by evaluating *PGV*-dependent levee fragility for selected ranges of the secondary parameters. When the data are conditioned according to these secondary parameters, there is a loss of resolution on two levels: (1) the number of *PGV*-bins is reduced per Eqn. (6.9), which can affect the regression of a fragility relation, particularly at high *PGV*s; (2) the number of data points per bin is reduced, decreasing the levels of confidence in the computed bin probabilities. While it is tempting to solve the second problem by using fewer bins (increasing the number of data points per bin), the first problem is then exacerbated. After some trial and error, I elected to maintain the binning criteria in Eqn. (6.9) for use in Model 2 regressions and not change the value '4' in the denominator.

6.4.1 Statistical Test for Selection of Distinct Secondary Conditions

I first seek to identify which of the secondary variables affect levee fragility by investigating conditions for which the differences between Model 2 and Model 1 are statistically significant. I define μ_{r1} and μ_{r2} as the means of residuals of Model 1 and Model 2 data points, respectively, relative to Model 1 predictions. If the residuals are plotted against *PGV*, the resulting slopes are b_{r1} and b_{r2} (using Model 1 and 2 data points, respectively). The distinction between Model 1 and Model 2 is then judged on the basis of *t*-tests with two null hypotheses:

1. H_{01}: hypothesis is that $\mu_{r1} - \mu_{r2} = 0$ (the two means are identical),
2. H_{02}: hypothesis is that $b_{r1} - b_{r2} = 0$ (the two slopes are identical).

Note that μ_{r1} and b_{r1} are approximately zero since Model 1 is regressed from Model 1 data points.

Rejection of the null hypothesis is expressed as a *p*-value indicating the level of significance. The *p*-value is the probability of exceeding the *t* variate in the *t*-distribution with *df* degrees of freedom; details of this calculation are provided in Appendix B. A *p*-value of 10% or less is often used to indicate that the null hypothesis can be rejected with confidence, although this 10% limit is arbitrary. The data for Models 1 and 2 can be considered as distinct if either one or both null hypotheses are rejected. The *p*-values in Table 6.3 show that among the 23 investigated conditions, one G_N condition (i.e., $G_N = 1$) and one D_W conditions (i.e., $D_W < -2.5$ m) result in $p < 0.10$. The D_W conditions indicating shallow ground water levels (i.e., $D_W > d_w$) are not strictly distinct from Model 1. Moreover, shape factor S_F was not found to be a significant secondary parameter. These results lead me to further examine G_N and D_W fragilities and to abandon S_F.

Table 6.3 Level of significance (*p*-value) by *t*-test between Model 1 and Model 2 for mean (H_{01}) and slope (H_{02}) with varying conditions and total number of data points for each condition.

Condition	Total number of data points	$DL > 0$		$DL > 1 \mid DL > 0$		$DL > 2 \mid DL > 0$	
		p-value (H_{01}), %	*p*-value (H_{02}), %	*p*-value (H_{01}), %	*p*-value (H_{02}), %	*p*-value (H_{01}), %	*p*-value (H_{02}), %
$G_N = 1$	528	25	3	65	29	88	13
$G_N = 2$	4970	80	75	55	87	69	73
$G_N = 3$	624	65	11	10	89	4	10
$D_W < -2.5$	1333	44	2	49	3	44	55
$D_W < -2$	2125	84	30	31	3	43	19
$D_W < -1.5$	3080	78	73	17	5	44	11
$D_W < -1$	4012	60	59	31	10	63	7
$D_W < -0.5$	4524	65	67	47	34	63	5
$D_W < 0$	5031	82	94	79	51	58	25
$D_W > -2.5$	5303	75	56	63	70	83	96
$D_W > -2.0$	4511	91	77	46	55	72	76
$D_W > -1.5$	3556	86	96	15	35	66	52
$D_W > -1$	2624	45	15	10	35	61	9
$D_W > -0.5$	2112	48	20	10	34	60	5
$D_W > 0$	1605	57	59	40	16	40	22
$S_F < 0.24$	1492	65	59	26	72	24	44
$S_F < 0.26$	2552	73	26	39	28	25	87
$S_F < 0.28$	3596	88	65	88	39	27	83
$S_F < 0.30$	4738	81	76	70	61	48	74
$S_F > 0.24$	5140	86	99	69	84	42	61
$S_F > 0.26$	4074	66	21	67	46	18	93
$S_F > 0.28$	3034	73	36	87	26	2	33
$S_F > 0.30$	1890	47	43	49	28	1	58

6.4.2 Stage 1: Fragility Function with Selected Conditions

Figure 6.9 shows fragility curves conditioned on G_N. The most common category comprising 81% of levee segments is G_N 2 (various alluvial sediments), which has fragilities nearly identical to Model 1. Fragilities for G_N 1 (mountain and terrace) are relatively low. The curve for G_N 3 (old river channel and back marsh) is relatively steep due to reduced fragility for $PGV \leq 30$ cm/s, but the data is sufficiently sparse that its distinction from Model 1 is not justified. Accordingly, I conclude that only G_N 1 can be said as distinct from Model 1, not G_N 2 and 3.

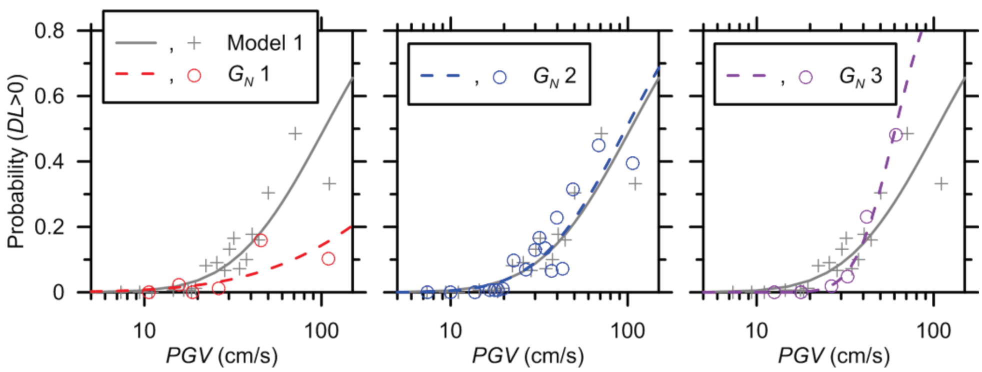

Figure 6.9 Model 2 fragility functions conditional on G_N groups.

Figure 6.10 shows fragility curves conditioned on D_W, with the upper set of plots corresponding to relative shallow groundwater ($D_W > d_w$) and the lower set corresponding to deep groundwater ($D_W < d_w$). I expect higher fragilities for the shallow groundwater case due to greater liquefaction susceptibility. The fragilities are nominally similar for $PGV < 30$ cm/s; these ground motion levels appear to be too low to induce significant liquefaction. For stronger shaking, shallow groundwater conditions ($D_W > -1$ m and $D_W > 0$) give rise to fragilities greater than Model 1, whereas results for deep groundwater conditions are more mixed. The greatest differences are for $D_W > -1$ m and $D_W < -1$ m, which fall consistently above and below Model 1 fragilities for $PGV > 30$ cm/s. While the fits for these cases are not statistically distinct relative to Model 1 (as shown in Table 6.3), the shallow and deep groundwater models are distinct from each other at 93% confidence (p-value = 0.07) using the slope-based t-test (i.e., H_{02} hypothesis; Appendix B). I adopt D_W as a Model 2 conditioning variable because: (1) it makes physical sense; (2) the fragility curves are indeed divergent and constrained by data for the important range of PGV > 30 cm/s; and (3) this conditioning has greater statistical significance in subsequent analysis involving higher damage level thresholds.

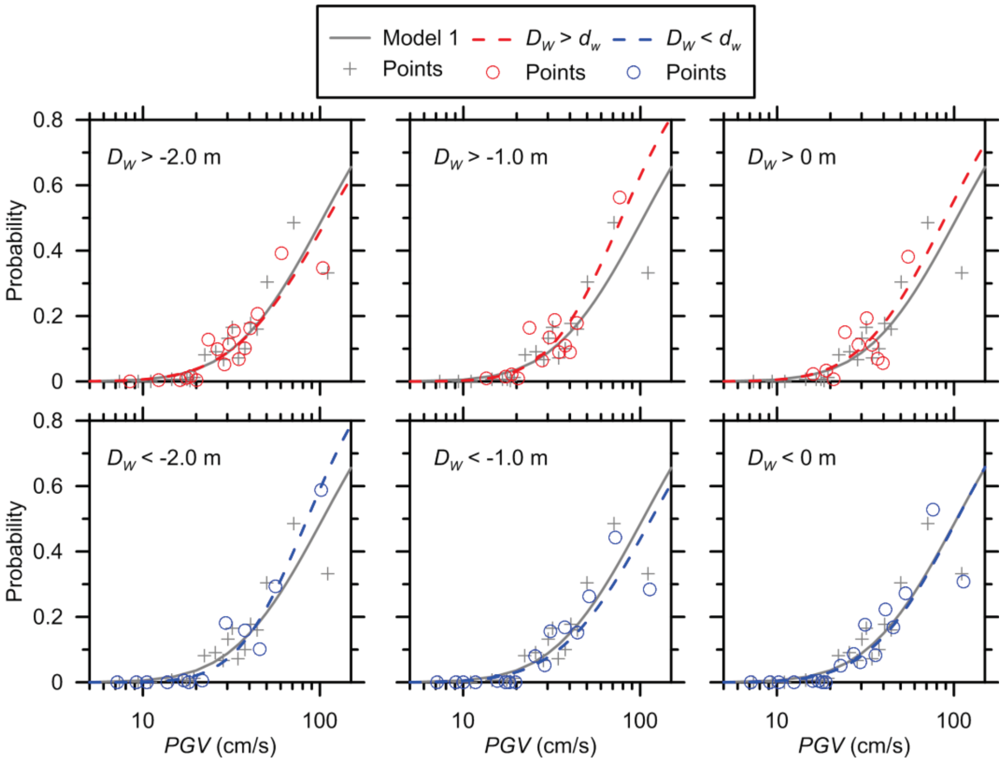

Figure 6.10 Model 2 fragility functions conditional on D_W groups.

Of the 2624 segments that have D_W > -1.0 m, 1605 (61%) have a fully saturated foundation (D_W > 0 m) and the median value of D_W is 0.4 m. Hence for practical purposes, the D_W > -1.0 m bin represents conditions with a reasonable probability of liquefaction susceptibility (provided that the soils are granular). For the deep ground water case of D_W < -1.0 m, the median D_W is -2.1 m for a large data population of 4012 segments.

Table 6.4 indicates moments of log-normal CDFs (i.e., mean μ and standard deviation β), standard deviation of residual (σ), and valid *PGV* range for versions of Model 2 based on G_N 1 and the recommended D_W limits. The corresponding values for Model 1 are also shown for reference purposes along with similarly derived results using the *IM* of *PGA*. The Model 2 β value for D_W > -1.0 m is smaller than that from Model 1, indicating improved resolution of the fragility function. There is practically no change in β for the deep groundwater case.

Table 6.4 Moments of log-normal CDFs (μ and β) for *PGV*- and *PGA*-based fragility curves standard deviation of residuals (σ), and valid IM ranges.

Model	*IM*	Condition	μ	e^{μ}	β	σ	Range
Model 1		*PGV*	4.64	104 cm/s	0.92	0.07	7 ~ 111 cm/s
		PGA	0.42	1.52g	1.07	0.07	0.13 ~ 1.31g
Model 2	*PGV*	G_N 1	6.42	611 cm/s	1.70	0.05	10 ~ 110 cm/s
		D_W < -1.0 m	4.75	116 cm/s	0.94	0.07	7 ~ 114 cm/s
		D_W > -1.0 m	4.36	78 cm/s	0.74	0.06	13 ~ 77 cm/s
	PGA	G_N 1	2.59	13.4g	2.02	0.03	0.14 ~ 1.29g
		D_W < -1.0 m	0.41	1.51g	0.92	0.08	0.13 ~ 1.33g
		D_W > -1.0 m	0.26	1.30g	1.12	0.07	0.14 ~ 1.07g

6.4.3 Stage 2: Probability of Severe Damage conditional on any Damage

For Stage 2 I examine the probabilities of exceeding various damage levels conditional on some damage having occurred [i.e., P(*DL* > *dl* | *DL* > 0)]. I examine the possible dependence of these failure probabilities on geomorphology (G_N) and groundwater level (D_W). I look for the possibility of *PGV*-dependent conditional damage probabilities, and when no such dependence is found, I provide *PGV*-independent mean probabilities (P_m). The *PGV*-dependent probabilities are described using a log-normal CDF (as above).

Figure 6.11 shows the conditional fragility data (damage thresholds of *DL* > 1, 2, and 3) for the full data set (no conditioning on secondary parameters) and binned according to the secondary parameters identified in the previous section. The full data set indicates *PGV*-independent fragility for all damage levels. For G_N 1, *DL* > 1 shows *PGV*-independent fragility, *DL* > 2 has finite probability only at high *PGV* (> 50 cm/s), and no instances of *DL* > 3 were reported. Cases with deep groundwater (D_W < -1 m; selected on basis of low *p*-values as indicated in Table 6.3) have *PGV*-dependent fragilities for *DL* > 1, 2 and 3. Deep groundwater conditions are less susceptible to liquefaction, so the principal damage mechanism for levees is expected to be slope deformation, which has been correlated to various intensity measures including *PGA* and *PGV* in past work (Saygili and Rathje, 2008). I find *PGV*-independent fragilities for shallow groundwater (D_W > -1 m). Because levee damage for these shallow groundwater cases is largely caused by liquefaction, the data indicate that the level of damage is not *PGV*-dependent once damage is triggered. Comparing the conditional fragilities for the deep and shallow ground water cases indicates increased probability of each higher damage threshold for shallow as compared to deep ground water. This shows that shallow ground water not only increases the probability of damage, but also the severity of damage.

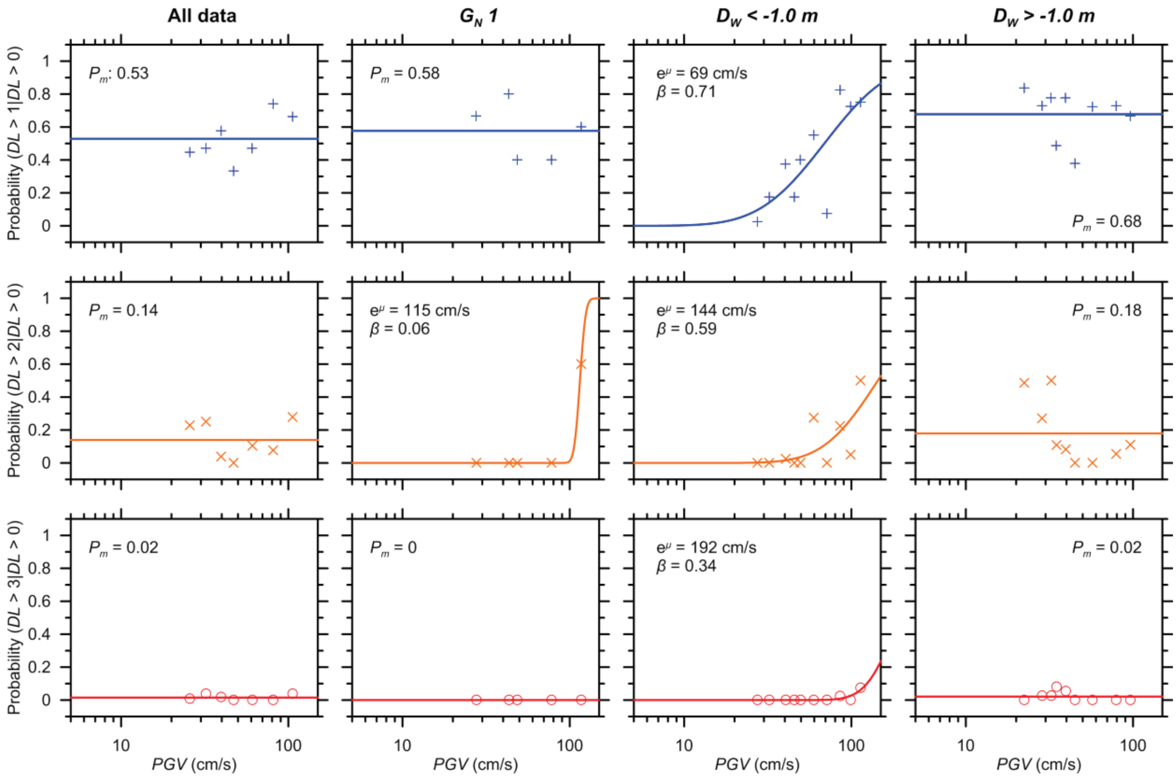

Figure 6.11 Probability of exceeding damage levels above one for the full data set and sets with geomorphic (G_N) and groundwater level (D_W) conditions. Moments of log-normal CDF (mean and standard deviation; μ and β) are indicated for *PGV*-dependent cases otherwise mean probabilities (P_m) are indicated.

These conditional damage fragilities can be combined with the fragilities from prior sections (for damage at any level) to develop a fragility function for any desired damage as follows:

$$P(DL > dl \mid IM, SP) = P(DL > 0 \mid IM, SP) \times P(DL > dl \mid DL > 0, IM, SP) \tag{6.11}$$

where *SP* represents secondary parameters used for Model 2 conditions. For Model 1, *SP* is disregarded. For an example of Model 2 probability, consider *PGV* = 40 cm/s and shallow ground water condition (i.e., D_W > -1.0 m). The probability of damage is 0.18 and the P_m for *DL* > 2 is 0.18. The product results in a probability of 0.03 for *DL* > 2.

7 LEVEE SYSTEM FRAGILITY CONSIDERING SPATIAL CORRELATION

7.1 INTRODUCTION

A levee system is comprised of earth embankments that protect a particular area such as a city, island, or agricultural area from flooding. One example is the levees that surround islands in the San Francisco Bay-Delta region of California. If we take a particular length of a levee as a segment (e.g., 50 m in length), then a levee system is a collection of levee segments in series. Levee systems are typically continuous and have lengths much greater (often measured in km) than their width or height (typically measured in m).

In Chapter 6, I developed fragility functions for 50 m long levee segments. I took the segment as damaged if any portion of the levee within the 50 m length suffered damage. Because these levee segments are connected in series, failure of one segment comprises failure of the system, because it would cause the protected region to flood. Hence, the levee system fragility problem essentially involves analysis of the probability of whether at least one levee segment in the series exceeds a target damage state. The solution of this problem depends strongly on the system length (i.e., number of segments) and correlation of damage between segments. The greater the number of segments, the higher is the opportunity for damage in at least one segment.

To illustrate the importance of damage correlation, consider two extreme cases: perfectly correlated and statistically independent. Perfect correlation of damage occurrence requires that all segments in a system are damaged, or not damaged, simultaneously. In this case, the segment having the highest failure probability (highest fragility) will control the system fragility. If the system failure is denoted $P(F_S)$ and the fragility of segment i as $P(f_i)$, we have:

$$P(F_S) = \max\left(P(f_i)\right) \tag{7.1}$$

On the other hand, statistical independence requires $P(F_S)$ to be computed as the complement of system survival, which in turn is the product of each individual segment surviving. Under these conditions, we have:

$$P(F_S) = 1 - \prod_{i=1}^{n}\left(1 - P(f_i)\right) \tag{7.2}$$

where n is the total number of segments. The statistically independent case will produce larger system fragilities. Perfect correlation and statistical independence comprise extreme cases known as simple bounds for a series system (called uni-modal bounds by Ang and Tang, 2007). Accordingly, the actual system fragility will be within these bounds, which can be represented as follows:

$$\max\left(P(f_i)\right) \le P(F_S) \le 1 - \prod_{i=1}^{n}\left(1 - P(f_i)\right) \tag{7.3}$$

Note that the term "failure" is used here instead of "damage," which was used in Chapter 6. The damage state that would be judged to comprise failure is application-dependent, with the principal consideration being the water level relative to the levee crest elevation.

The range of failure probabilities provided by Eqn. (7.3) is often very wide, so an ability to evaluate failure probabilities within the range is of considerable practical importance. This evaluation depends on the degree of damage correlation between segments. To my knowledge, prior research has not addressed the problem of analyzing system fragility given knowledge of spatially varying segment fragilities (per a model such as that in Chapter 6), damage correlations between segments (details below), and demand correlations. In this chapter, I present a methodology for computing this system fragility conditional on known segment fragilities $P(f_i)$ and between-segment correlation coefficients.

Since levee damage states are discrete variables (i.e., *DL* 0~4, or 0 and 1 for no-damage and damage), the direct use of damage states for analysis of correlation coefficients is problematic. Instead, I represent capacity using a damage capacity distribution, which can be obtained from the derivative of the fragility function, and is more amenable to correlation analysis. This correlation analysis is performed using the Shinano River levee data presented in prior chapters. Damage states are analyzed by comparing the capacity distribution to the demand distribution, which can be obtained from a ground motion prediction equation (GMPE) and which has its own correlation structure obtained from event-specific analysis (e.g., semi-variograms established for the 2004 and 2007 Niigata earthquakes in Section 5.2) or from the literature (Jayaram and Baker, 2009) for predictive analyses of future earthquakes. The proposed approach uses Monte-Carlo simulation to evaluate probabilities of exceeding damage states at the system level. I apply the approach to compute system fragility for components of the Shinano River levee system protecting a small town.

7.2 ANALYSIS OF SYSTEM FRAGILITY IN PRIOR WORK

Previous studies have solved the system fragility problem for relatively simplified conditions. For levee applications, previous studies have evaluated $P(F_S)$ by dividing the levee system into "reaches" within which the correlation is perfect, but correlations between reaches are taken as zero (i.e., statistically independent) (USACE, 2011b; Wolff, 2008). In these applications, a levee reach is a length of levee embankments judged to have similar geometry and foundation conditions. For each reach, $P(f_i)$ is evaluated using appropriate analytical methods and $P(F_S)$ is then computed using Eqn. (7.2). The length of levee reaches, referred to as the characteristic length, is recommended to be 100 to 300 m by USACE (2011b) and Wolff (2008). The principal limitations of this method are the non-physical nature of the underlying assumption (perfect correlation within segments, no correlation between) and the necessary arbitrariness of the characteristic length.

A mathematical solution for system fragility can in principal be developed using an n-dimensional joint standardized normal distribution function, Φ_n, in which the standard normal variate reflects the safety margin (Rackwitz and Krzykacz, 1978). The safety margin for system component i, M_i is defined as $M_i = C_i - D_i$, where C_i and D_i are random variables representing element capacity and demand, respectively. Both C_i and D_i are assumed to follow normal distributions with user-defined means and standard deviations for each element, so the differential M_i is also normally distributed. Since M_i is normally distributed by assumption, the probability of system survival can be obtained by finding the space in Φ_n where the n-dimensional standard normal variates are higher than the reliability index β_i, which is defined as the mean safety margin normalized by its standard deviation, throughout all components. The equation of β_i is as follows:

$$ {}_i = \frac{E\left(M_i\right)}{\sqrt{\mathrm{var}\left(M_i\right)}} \tag{7.4} $$

The system fragility $P(F_S)$ is the complement of probability of system survival. In practice, for systems with $n \geq 3$ components, the joint distribution Φ_n is difficult to solve for, so upper- and lower-bounded solutions based on limiting assumptions are most often used (Thoft-Christensen and Murotsu, 1986).

A solution for the joint distribution Φ_n was formulated by Dunnett and Sobel (1955) (for applications unrelated to system fragility) under the assumption that the correlation coefficient ρ between among all elements in the series is constant:

$$\underset{i=1\sim n}{\Phi_n}\left(x_i;\rho\right)=\int_{-\infty}^{\infty}\phi(t)\prod_{i=1}^{n}\phi\left(\frac{x_i-\sqrt{\rho}t}{\sqrt{1-\rho}}\right)dt \tag{7.5}$$

where x_i is a variate in the i^{th} normal distribution, i is an index for system components, ϕ is the PDF operator for the standard normal distribution, t is the standard normal variate and integrand, and n is the number of components. To apply the joint distribution solution in Eqn. (7.5) to fragility problems, as described above, it is useful to replace variate x_i with the reliability index $\beta_{i.}$ With the substation of β_i for x_i, Eqn. (7.5) represents the probability of survival for a series system with equally correlated elements. The system fragility $P(F_S)$ can then be computed as the complement of system survival:

$$P\left(F_S\right)=1-\underset{i=1\sim n}{\Phi_n}\left(\beta_i;\rho\right)=1-\int_{-\infty}^{\infty}\phi(t)\prod_{i=1}^{n}\phi\left(\frac{\beta_i-\sqrt{\rho}t}{\sqrt{1-\rho}}\right)dt \tag{7.6}$$

Grigoriu and Turkstra (1979) presented a simplified version of Eqn. (7.6) utilizing a second assumption that β_i is constant for all elements:

$$P\left(F_S\right)=1-\int_{-\infty}^{\infty}\phi(t)\left[\Phi\left(\frac{\beta_e+\sqrt{\rho}t}{\sqrt{1-\rho}}\right)\right]^n dt \tag{7.7}$$

where Φ is the CDF operator for a standard normal distribution and β_e is the constant reliability index for all elements. Eqn. (7.7) is a single integration problem and as such is amenable to numerical calculation. Thoft-Christensen and Sørensen (1982) extended this solution to include non-equally correlated elements by estimating an average correlation coefficient throughout the system and applying that value in Eqn. (7.7).

For applications to levees, there are significant limitations associated with the assumptions required to derive Eqn. (7.7). Constant reliability index will not apply to the levee segments within a spatially distributed system – some segments will have relatively low fragility (due to low demand or strong soils) while others will be higher. Likewise, I intuitively expect the correlation of safety margin to not be constant but to vary with separation distance (closer segments well correlated, distant segments uncorrelated).

The proposed approach, described and illustrated in subsequent sections of this chapter, was developed to overcome the limitations of previous methods for application to the levee system fragility problem.

7.3 DAMAGE STATE AND CORRELATION COEFFICIENT

Correlation of two random variables is often computed using residuals of the variables relative to a predictive model. For example, correlation coefficients between two ground intensity measures are computed as (Neter et al., 1996; Baker and Cornell, 2006):

$$\rho_{A,B} = \frac{\sum_{i=1}^{n}\left(A_i - \overline{A}\right)\left(B_i - \overline{B}\right)}{\sqrt{\sum_{i=1}^{n}\left(A_i - \overline{A}\right)^2 \sum_{i=1}^{n}\left(B_i - \overline{B}\right)^2}} \tag{7.8}$$

where A and B are residuals (data minus model) for random variables of interest (e.g., spectral accelerations at varying T, or at same T but varying locations), $\overline{A}$ and $\overline{B}$ are their sample means, subscript i denotes the i^{th} observation of variables A or B, and n is total number of observations. These correlations have been evaluated between spectral periods (Jayaram and Baker, 2008) and spatially (Jayaram and Baker, 2009). The form of correlation in Eqn. (7.8) provides a useful quantification of correlation for spatially variable ground motions, because such motions can be simulated by adding appropriately correlated residuals to model means (i.e., GMPEs).

The correlation of damage states is analyzed somewhat differently than indicated by Eqn. (7.8) because the damage is not expressed as continuous variables but as discrete variables (damage states). I describe this analysis in the following two sections, then describe its implementation for the simple case of a two-component system.

7.3.1 Correlation Coefficient of Damage States

The fragility functions developed in Section 6.3 utilize damage states expressed as Boolean variables (i.e., zero or one): zero for no-damage (i.e., $DL = 0$), and one for $DL > 0$. For the purpose of explanation, I use the terms "survival" for $DL = 0$, and "failure" for $DL > 0$ subsequently. Note that failure could easily be re-defined using a more severe damage state. Let define survival event s_i and failure event f_i for segment i as follows:

$$s_i = \begin{cases} 1 & \text{if survived} \\ 0 & \text{if failed} \end{cases} \tag{7.9}$$

$$f_i = \begin{cases} 0 & \text{if survived} \\ 1 & \text{if failed} \end{cases} \tag{7.10}$$

The probability of survival for segment i [i.e., $P(s_i)$] is the expected value of s_i [i.e., $E(s_i)$], and the probability of failure [i.e., $P(f_i)$] is $E(f_i)$. By definition, the correlation coefficients of damage states of s_i and s_j (i.e., $\rho_{DS,s}$) can be shown as follows (Neter et al., 1996):

$$\rho_{DS,s} = \frac{\operatorname{cov}\left(s_i s_j\right)}{\sqrt{\operatorname{var}\left(s_i\right)\operatorname{var}\left(s_j\right)}} \tag{7.11}$$

where $\text{cov}(s_i s_j)$ is the covariance of survival events for segments i and j, and $\text{var}(s_i)$ is the variance of s_i. Eqn. (7.11) is analogous to Eqn. (7.8), where the residual A_i or B_i is substituted by s_i or s_j. Based on the definition of covariance, the numerator in Eqn. (7.11) can be expanded as:

$$\begin{aligned}\text{cov}\left(s_i s_j\right) &= E\left[\left(s_i-\mu_{s,i}\right)\left(s_j-\mu_{s,j}\right)\right] \\ &= E\left(s_i s_j\right)-E\left(\mu_{s,i} s_j\right)-E\left(\mu_{s,j} s_i\right)+E\left(\mu_{s,i}\mu_{s,j}\right) \\ &= E\left(s_i s_j\right)-\mu_{s,i}\mu_{s,j}\end{aligned} \tag{7.12}$$

where $\mu_{s,i}$ represents the mean of s_i, which is analogous to $E(s_i)$ or $P(s_i)$. Note that the last three terms in the middle expression are all equal to $\mu_{s,i}\mu_{s,j}$. Since s_i is a Boolean variable, s_i^2 is equal to s_i resulting in $E(s_i^2) = E(s_i)$. The variance terms in Eqn. (7.11) [i.e., $\text{var}(s_i)$ and $\text{var}(s_j)$] can be similarly expanded as:

$$\text{var}\left(s_i\right)=E\left(s_i^2-\mu_{s,i}^2\right)=E\left(s_i^2\right)-E\left(\mu_{s,i}^2\right)=\mu_{s,i}-\mu_{s,i}^2 \tag{7.13}$$

Substituting $\text{cov}(s_i s_j)$, $\text{var}(s_i)$, and $\text{var}(s_j)$ from Eqns (7.12)-(7.13) into Eqn. (7.11), I obtain:

$$\rho_{DS,s}=\frac{E\left(s_i s_j\right)-\mu_{s,i}\mu_{s,j}}{\sqrt{\mu_{s,i}\left(1-\mu_{s,i}\right)\mu_{s,j}\left(1-\mu_{s,j}\right)}}=\frac{P\left(s_i\cap s_j\right)-P\left(s_i\right)P\left(s_j\right)}{\sqrt{P\left(s_i\right)\left(1-P\left(s_i\right)\right)P\left(s_j\right)\left(1-P\left(s_j\right)\right)}} \tag{7.14}$$

Eqn. (7.14) can be re-written in terms of failure states as follows:

$$\begin{aligned}\rho_{DS,s} &= \frac{E\left(s_i s_j\right)-\mu_{s,i}\mu_{s,j}}{\sqrt{\mu_{s,i}\left(1-\mu_{s,i}\right)\mu_{s,j}\left(1-\mu_{s,j}\right)}} \\ &= \frac{E\left[\left(1-f_i\right)\left(1-f_j\right)\right]-\left(1-\mu_{f,i}\right)\left(1-\mu_{f,j}\right)}{\sqrt{\left(1-\mu_{f,i}\right)\left(1-\left(1-\mu_{f,i}\right)\right)\left(1-\mu_{f,j}\right)\left(1-\left(1-\mu_{f,j}\right)\right)}} \\ &= \frac{1-E\left(f_i\right)-E\left(f_j\right)+E\left(f_i f_j\right)-1+\mu_{f,i}+\mu_{f,j}-\mu_{f,i}\mu_{f,j}}{\sqrt{\mu_{f,i}\left(1-\mu_{f,i}\right)\mu_{f,j}\left(1-\mu_{f,j}\right)}} \\ &= \frac{E\left(f_i f_j\right)-\mu_{f,i}\mu_{f,j}}{\sqrt{\mu_{f,i}\left(1-\mu_{f,i}\right)\mu_{f,j}\left(1-\mu_{f,j}\right)}}=\rho_{DS,f}\end{aligned} \tag{7.15}$$

where $\mu_{f,i}$ is the mean of f_i, which is equivalent to $E(f_i)$. Eqn. (7.15) shows that the correlation coefficients for survival and failure Boolean damage states are equivalent. Accordingly, I drop the '*s*' and '*f*' from the subscripts and refer to the correlation coefficient of damage states as ρ_{DS} hereafter.

Eqns (7.14) and (7.15) can be used to compute ρ_{DS} from damage state data without the use of an underlying model for levee fragility. As shown with Eqn. (7.8) and accompanying text, it is common in other applications (e.g., ground motions) for correlation coefficients to be computed from residuals of data relative to a model. To investigate this approach for calculation of correlation coefficient, we take the data for a given segment *i* as the state f_i and the model to be the fragility function from Chapter 6, which produces a probability of failure (fragility) denoted $P(f_i)$. A residual can then be calculated as:

$$R_i = f_i - P(f_i) \tag{7.16}$$

Adapting Eqn. (7.11) for this analysis, I compute the correlation coefficient of R_i and R_j (denoted ρ_{DSR}) as:

$$\rho_{DSR} = \frac{\mathrm{cov}(R_i R_j)}{\sqrt{\mathrm{var}(R_i)\,\mathrm{var}(R_j)}} \tag{7.17}$$

The numerator in Eqn. (7.17) can be written as:

$$\begin{aligned}\mathrm{cov}(R_i R_j) &= E\left[(R_i - \mu_{R,i})(R_j - \mu_{R,j})\right] \\ &= E(R_i R_j) \\ &= E\left[(f_i - P(f_i))(f_j - P(f_j))\right] = E\left[(f_i - \mu_{f,i})(f_j - \mu_{f,j})\right] \\ &= \mathrm{cov}(f_i f_j)\end{aligned} \tag{7.18}$$

The second line in Eqn. (7.18) takes the mean of residuals for a segment i (i.e., $\mu_{R,i}$) as zero, which is true if the model is unbiased for segment i. This would not be expected to strictly hold for the present fragility model (in Chapter 6), but could be envisioned as being more nearly true for future fragility models that take into account additional site-specific factors that are not presently considered (making the model non-ergodic). As shown in Eqn. (7.18), with that assumption, the covariance of residuals is equivalent to the covariance of damage states.

The variance of residuals R_i can also be shown to match the variance of f_i as follows:

$$\begin{aligned} \mathrm{var}(R_i) &= E\left(R_i^2 - \mu_{R,i}{}^2\right) = E\left(R_i^2\right) \\ &= E\left[\left(f_i - \mu_{f,i}\right)^2\right] = E\left(f_i^2\right) - 2\mu_{f,i}\mu_{f,i} + \mu_{f,i}{}^2 = \mu_{f,i} - \mu_{f,i}{}^2 \\ &= \mathrm{var}(f_i) \end{aligned} \tag{7.19}$$

Substituting Eqns. (7.18) and (7.19) into Eqn. (7.17), ρ_{DSR} becomes:

$$\rho_{DSR} = \frac{\mathrm{cov}\left(r_i r_j\right)}{\sqrt{\mathrm{var}\left(r_i\right)\mathrm{var}\left(r_j\right)}} = \frac{\mathrm{cov}\left(f_i f_j\right)}{\sqrt{\mathrm{var}\left(f_i\right)\mathrm{var}\left(f_j\right)}} = \rho_{DS} \tag{7.20}$$

Therefore, the correlation coefficient of damage states' residuals is equivalent to the correlation coefficients of damage states, provided the model used for residuals analysis is unbiased. In the case where the model has segment bias (ergodic model), these correlation coefficients would not be identical.

The computation of correlation coefficients from damage states (ρ_{DS}) between two segments using Eqn. (7.14) requires observations from many events. The probability of survival for a segment i [i.e., $P(s_i)$] is the mean of s_i from many samples, whose reliability is highly depending on the number of samples, which must be from events that produce shaking that is strong enough to cause damage. The joint distribution $P\left(s_i \cap s_j\right)$, which is the probability of survival of both segments i and j, similarly requires a large number of samples for a reliable estimate. In practice, data will seldom be available with which to compute ρ_{DS} from observed damage states. In the following section, I present an auto-correlation coefficient approach that relies on a large volume of data for a few events. This approach is investigated as an means by which to approximate ρ_{DS}.

7.3.2 Autocorrelation Coefficient as Approximation of Correlation Coefficient

The correlation coefficient between damage states in distinct system components, described in the previous section, can only be evaluated directly if there are tens of damage observations. Since that volume of data for these computations is seldom available, in this section I investigate the use of auto-correlation coefficients of damage states (ρ_{acc}) as an alternate.

Auto-correlation represents the correlation coefficient between a set of observations from an event and the same data set lagged in time or space. In this case the lag is in space; for example, a lag of one means that the damage states are assigned in exactly the same order but shifted over one levee segment. The correlation is computed between the original and shifted data sets, and the resulting value is assigned to the level of shift. This process is repeated for varying amounts of shift to produce an autocorrelation function.

Figure 7.1**Error! Reference source not found.** shows examples of ρ_{acc} for levee damage states along the three rivers within the Shinano River system (SH1, SH2, and UO; shorthand defined in Section 1.2). Levee damage occurred along SH1 in the 2004 and 2007 earthquakes and along SH2 and UO only in the 2004 earthquake. The variation of correlation coefficient ρ_{acc} with distance is regressed as follows:

$$\hat{\rho}_{acc} = \exp\left(-3 \times \frac{x}{\alpha_{DS}}\right) \tag{7.21}$$

where $\hat{\rho}_{acc}$ is the mean autocorrelation for lagged distance x and α_{DS} is a regression coefficient that is indicated in Figure 7.1. The use of multiplier 3 in Eqn. (7.21) causes α_{DS} to match the 'range' in a semi-variogram, which is the lag where $\hat{\rho}_{acc}$ becomes practically zero. As described in Section 2.4, levees are discontinuous along the rivers in some cases. In such cases, autocorrelations are computed using data for each continuous section greater than 0.5 km in length within a single calculation. Accordingly, there is only one autocorrelation result for a given river and earthquake. For a given level of lag, a continuous section of levee only contributes data to the autocorrelation calculation if its length is greater than the lag. Accordingly, for large lag distances, only a subset of the data having long continuous stretches of levee are used. The longest stretches are 798 for SH1, 221 for SH2, and 362 segments for UO.

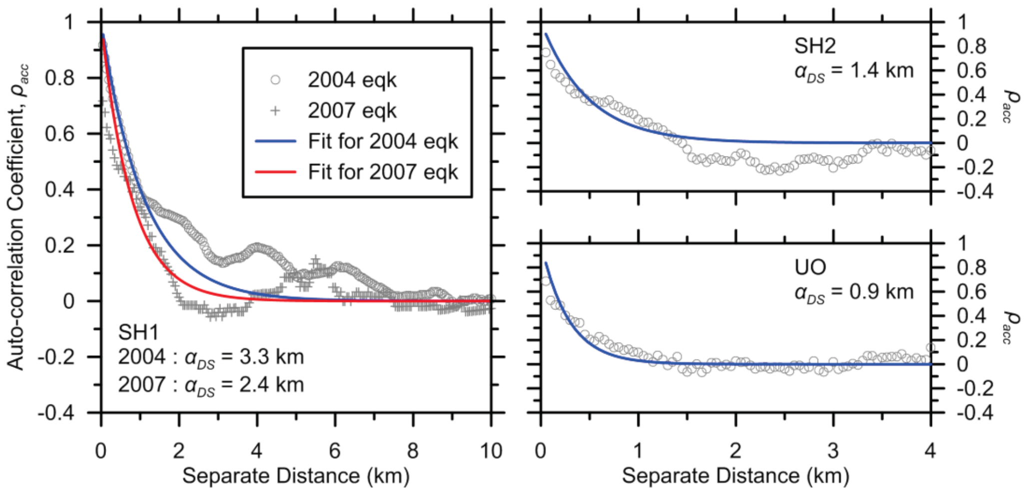

Figure 7.1 Auto-correlation coefficients of damage states for levee systems of SH1, SH2, and UO Rivers. Exponential function fits are shown.

The use of ρ_{acc} as an alternative to ρ_{DS} applies if damage states are stationary (in space) between segments; in other words, while the mean and standard deviation of damage states may vary in space, their correlation must depend only on separation distance and otherwise be independent of location. To check the use of ρ_{acc} for ρ_{DS} with the above assumption, I randomly generated Boolean variables (0 and 1) following a pre-defined function for describing ρ_{DS} as a function of separation distance, and check whether the ρ_{acc} calculated from the randomly generated dataset matches the values of ρ_{DS} used to generate the data. Figure 7.2 compares pre-defined ρ_{DS} and ρ_{acc} evaluated for 200, 1000, and 5000 segments using five random sets of observations. Each frame in the figure shows the pre-defined ρ_{DS}, five sets of results for ρ_{acc}, and the mean of ρ_{acc}. The agreement between the mean of ρ_{acc} and ρ_{DS} strengthens with the number of segments, with ρ_{acc} underestimating ρ_{DS} for the cases with 200 and 1000 segments. Note that this simulation does not validate the assumption of stationarity (in space) of the damage correlation; it merely shows that if this assumption is correct that ρ_{acc} can approximate a stationary ρ_{DS} given a sufficiently large data set.

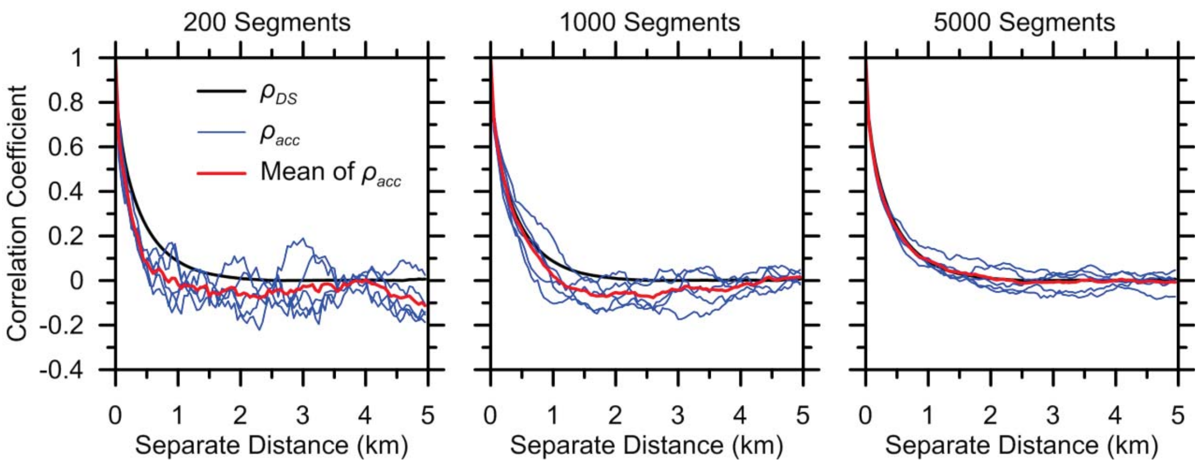

Figure 7.2 Comparison of pre-defined ρ_{DS} with five randomly generated realizations of auto-correlation ρ_{acc} with their mean. Results shown for cases having 200, 1000, and 5000 segments.

7.3.3 Evaluation of Two-segment System Fragility using Correlation Coefficient

For a two-segment system, the probability of survival [i.e., $P(S_S)$] is the probability of the intersection of s_i and s_j, written as $P\left(s_i \cap s_j\right)$. Because this is a term in Eqn. (7.14), a simple re-arrangement of that expression allows its computation when the correlation coefficient is known:

$$P(S_S) = P(s_i)P\left(s_j\right) + \rho_{DS}\sqrt{P(s_i)\left(1-P(s_i)\right)P\left(s_j\right)\left(1-P\left(s_j\right)\right)} \qquad (7.22)$$

Since the failure event is the complement of the survival event at both the system and component levels, the probability of failure for the system [i.e., $P(F_S)$] is readily evaluated from Eqn. (7.22) as follows:

$$P(F_S) = 1-\left(1-P(f_i)\right)\left(1-P\left(f_j\right)\right) - \rho_{DS}\sqrt{P(f_i)\left(1-P(f_i)\right)P\left(f_j\right)\left(1-P\left(f_j\right)\right)} \qquad (7.23)$$

Eqn. (7.23) indicates that the system fragility with two segments can be estimated once the probability of failure for each segment and the correlation coefficient of damage states ρ_{DS} between those segments are known.

The calculation of $P(F_S)$ for systems having many segments is relatively complex, especially when the damage capacity and demand vary across segments. This issue is resolved using correlations of damage capacities and demands and is the subject of the following section.

7.4 SYSTEM FRAGILITY UTILIZING DAMAGE CAPACITIES AND DAMAGE DEMANDS

We begin this section by defining the concepts of damage capacities and damage demands that are expressed as random variables from which segment damage probabilities can be computed. The analysis of system level fragility requires correlation coefficients for those distributions. Section 7.4.2 extends the capacity-related correlation coefficients from Section 7.3 to correlations of PDF-based damage capacities. Demand-related correlations are described in Section 7.4.3. In Section 7.4.4, I introduce a numerical procedure for evaluating system fragility.

7.4.1 Damage Capacities and Damage Demands

Damage capacity represents some measure of the “strength” or “resistance” of a segment against being damaged by an earthquake. In the present context, capacity is quantified as the ground motion intensity measure, which if exceeded, causes the levee segment to experience damage exceeding a damage state. Based on the fragility work in Chapter 6, the selected intensity measure for representing capacity is *PGV*, which can be taken as a random variable having a PDF that is the derivative of the fragility curve (Baker, 2008).

For a given segment, the uncertain damage capacity is compared to a *damage demand*, represented by the same intensity measure type (in this case *PGV*), to determine whether damage occurs. For a past earthquake with a ground motion recorded at the levee, demand is known. Otherwise, for future events, demand too is a random variable having a PDF defined from a GMPE. Figure 7.3 shows an example of capacity and demand described by log-normal PDFs for a particular segment. The PDFs have log means of μ_D for demand and μ_C for capacity, and log standard deviations of σ_D for demand and σ_C for capacity.

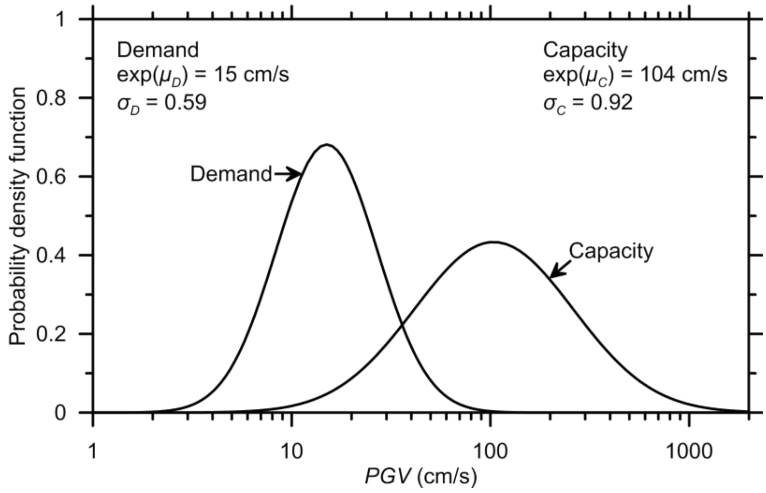

Figure 7.3 Probability density function of capacity and demand expressed as *PGV*.

Failure occurs when demand exceeds capacity. The probability of failure, P_f, for a segment is computed from the PDFs as (adapted from Melchers, 1999):

$$\begin{aligned} P_f &= P(C - D \leq 0) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{c \geq d} f_C(c) f_D(d)\, dc\, dd \\ &= \int_{-\infty}^{\infty} F_C(t) f_D(t)\, dt \end{aligned} \tag{7.24}$$

where [f_C f_D] and [F_C F_D] represent probability density functions (PDF) and cumulative distribution functions (CDF) of capacity and demand, respectively. Integrand t is the variate for the product of F_C and f_D.

Figure 7.4 shows an example fragility function (CDF) and its derivative representing F_C and f_C, respectively. The value of P_f = 0.15 marked on the figure applies for a deterministic demand of *PGV* = 40 cm/s. Figure 7.5 shows an example in which the same capacity distribution is combined with a distributed demand having a median of 40 cm/s and a standard deviation of σ_D = 0.65 (a typical value from GMPEs). The figure illustrates the two terms in the integrand (F_C and f_D) and their product, which is a failure density function having an underlying area equivalent to P_f. In the present case, P_f rises from 0.15 (case of deterministic demand) to 0.20.

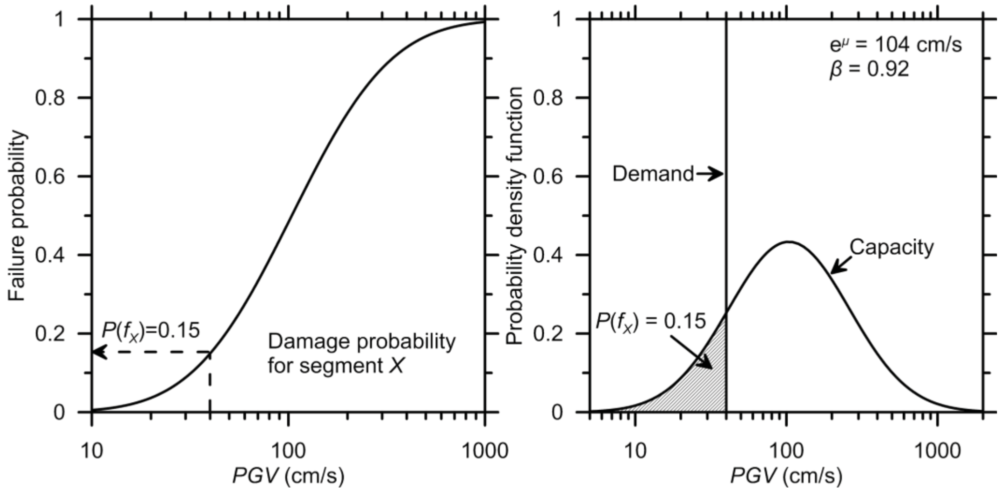

Figure 7.4 Transformation of fragility function from CDF to PDF that is damage capacity distribution (after Baker, 2008).

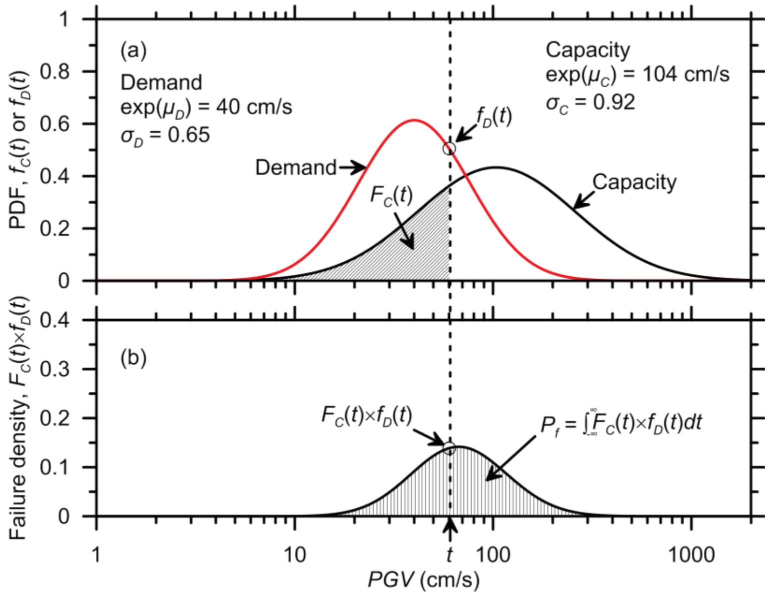

Figure 7.5 (a) Example PDFs of capacity (f_C) and demand (f_D). (b) Failure density computed as product of capacity CDF (F_C) and f_D. The area below the failure density function is the probability of failure, P_f (after Melchers, 1999).

7.4.2 Correlation of Damage Capacity

In Section 7.3.3, I showed that the probability of failure for a system [i.e., $P(F_S)$] consisting of two segments can be computed by knowing P_f for each segment and the correlation coefficient of damage states (ρ_{DS}). A similar computation is possible if we have the correlation coefficient of damage capacities (ρ_{DC}) instead of ρ_{DS}. In this section, I undertake two tasks associated with the use of ρ_{DC}. First, I describe how system failure probabilities can be computed given PDFs of capacity for two segments and ρ_{DC}. Second, I describe how ρ_{DC}, which is unknown, can be estimated from ρ_{DS}, which was estimated in Section 7.3.2.

Suppose we have two levee segments (*i* and *j*) with given damage capacities based on the fragility curves in Chapter 6. Segment *i* has exp(μ_C) of 116 m/s and σ_C of 0.94, whereas segment *j* is somewhat weaker and with less dispersion, having exp(μ_C) of 78 m/s and σ_C of 0.74. Figure 7.6 shows randomly generated sets of capacities for both segments plotted together. In one case (Figure 7.6a) the capacities are uncorrelated (ρ_{DC} = 0), whereas in the other (Figure 7.6b) they are strongly correlated with ρ_{DC} = 0.8. If these segments are subjected to the levels of deterministic demand marked in the figure (*PGV* = 35 cm/s for levee *i* and 42 cm/s for levee *j*), the capacity realizations (dots) that fall within the shaded areas indicate the conditions where the damage capacity is less than the demand for at least one of the segments, which constitutes system failure. The failure probability for the two-segment-system is equal to the ratio of the number of points in the shaded areas to the total number of points. As shown in Figure 7.6, the number of points in the shaded areas depends on the correlation coefficient, being higher for the uncorrelated case than the partially correlated case.

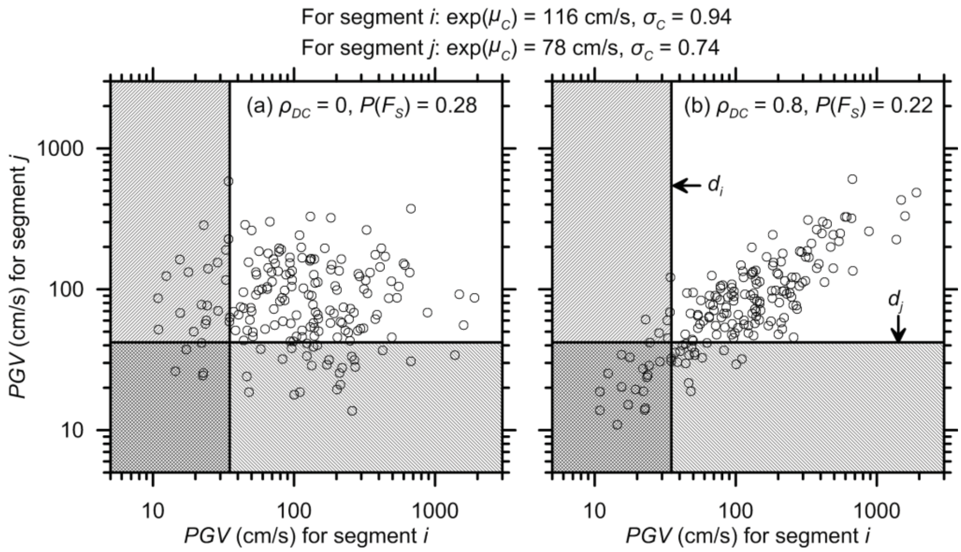

Figure 7.6 Randomly sampled damage capacities for two segments having different statistical moments. The capacities are (a) uncorrelated (ρ = 0) and (b) correlated (ρ = 0.8). Modified from Baker (2008).

Because the capacities for both segments are log-normally distributed, by changing the capacities to normal variates, $P(F_S)$ can be solved for using the bivariate normal distribution (BND) theorem. Normal variates for capacities, x_i, are computed as:

$$x_i = \frac{\ln(c_i) - \mu_C}{\sigma_C} \tag{7.25}$$

where c_i is a sampled capacity, and μ_C and σ_C are the log mean and standard deviation of capacity. Figure 7.7 shows probability density contour maps for standard BNDs having correlation coefficients of ρ = 0 and 0.8.

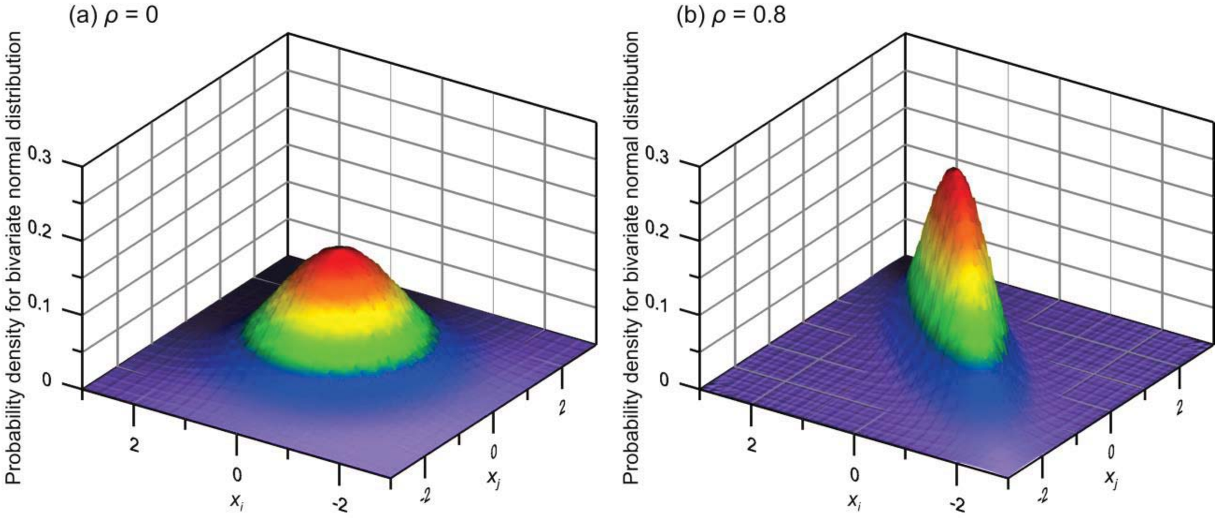

Figure 7.7 Probability density surfaces for standard bivariate normal distribution for case of (a) uncorrelated (ρ = 0) and (b) correlated (ρ = 0.8) normalized capacities.

The deterministic demands d_i and d_j in Figure 7.6 can be transformed to standard normal variates as in Eqn. (7.25) by simply substituting d_i for c_i. The normalized demands are denoted y_i and y_j for segment i and j, respectively. By integrating the volume beneath the probability density surface defined by the BND for conditions where capacity exceeds demand for both segments (indicating survival), the probability of survival for a two-segment-system [i.e., $P(S_S)$] can be calculated. This can be represented as:

$$P(S_S) = P\left(x_i \geq y_i, x_j \geq y_j; \rho\right) \tag{7.26}$$

This is analogous to the percentage of dots in the non-shaded area in Figure 7.6. Denoting the probability density of the BND as $p(x_i, x_j; \rho)$, $P(S_S)$ can be computed as:

$$\begin{aligned} P(S_S) &= P\left(x_i \geq y_i, x_j \geq y_j; \rho\right) \\ &= P\left(x_i \leq -y_i, x_j \leq -y_j; \rho\right) \\ &= \int_{-\infty}^{-y_i} \int_{-\infty}^{-y_j} p\left(x_i, x_j; \rho\right) dx_i dx_j \end{aligned} \tag{7.27}$$

The system fragility $P(F_S)$ is the complement of $P(S_S)$:

$$P(F_S) = 1 - P(S_S) = 1 - \int_{-\infty}^{-y_i} \int_{-\infty}^{-y_j} p(x_i, x_j; \rho)\, dx_i dx_j \tag{7.28}$$

The second issue addressed in this section is how ρ_{DC}, which is unknown, can be estimated from ρ_{DS}, which was estimated in Section 7.3.2. The correlation coefficient on damage capacity, ρ_{DC}, is unknown because the capacities themselves are unknown; they are only available as random variables described by a distribution. For this reason, ρ_{DC} cannot be evaluated from damage observations directly.

Instead, I infer ρ_{DC} by equating the system fragility from Eqn. (7.28) with that computed using the correlation coefficient of damage states (ρ_{DS}) in Eqn. (7.23). Combining those expressions for $P(F_S)$, we have:

$$\begin{aligned} P(F_S) &= 1 - \int_{-\infty}^{-y_i} \int_{-\infty}^{-y_j} p(x_i, x_j; \rho_{DC})\, dx_i dx_j \\ &= 1 - (1 - P(f_i))(1 - P(f_j)) - \rho_{DS} \sqrt{P(f_i)(1 - P(f_i)) P(f_j)(1 - P(f_j))} \end{aligned} \tag{7.29}$$

Eqn. (7.29) shows that ρ_{DC} is a function of ρ_{DS} and failure probabilities P_f for segments i and j. It is difficult to develop a closed-form solution for ρ_{DC}, but it can be inferred from numerical inversion. Figure 7.8 shows examples of ρ_{DC} back-calculated from given ρ_{DS}, $P(f_i)$, and $P(f_j)$. The inversion analyses seek the value of ρ_{DC} that minimizes the difference between $P(F_S)$ values computed from Eqns. (7.23) and (7.28). Figure 7.8 shows that ρ_{DC} depends strongly on ρ_{DS}, with second-order dependencies on P_f. Nevertheless, the important point to make at this stage is that ρ_{DC} can be estimated from these three variables. This approach is used in Section 7.4.4 for system fragility estimation within Monte-Carlo simulations.

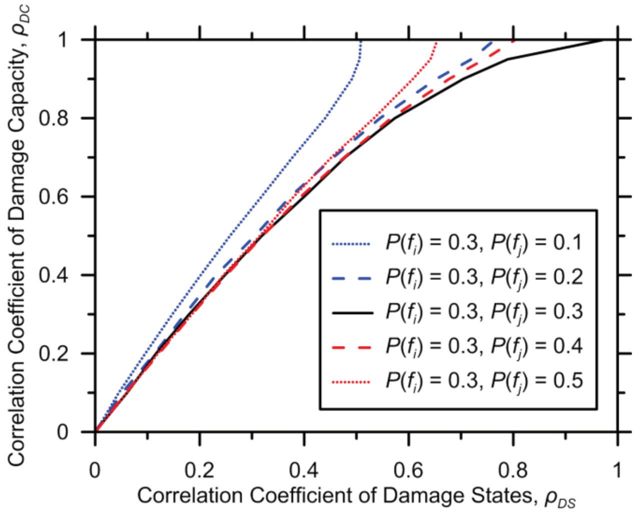

Figure 7.8 **Correlation coefficient of damage capacity ρ_{DC} as a function of ρ_{DS}, $P(f_i)$, and $P(f_j)$.**

7.4.3 Correlation of Damage Demands

The ground shaking experienced by a distributed levee system will exhibit spatially variable intensity measures (IMs). Some of those spatial variations will follow well-understood trends with respect to site-source distance (IMs decay with distance) or site condition (IMs will often, but not always, decrease as V_{S30} drops). However, some of those variations are, for practical purposes, random. The relatively 'deterministic' features can usually be described by a GMPE, whereas the random features can be represented by spatial variations of residuals. To facilitate this analysis, recall the notation from Section 5.2.2 in which the event term for earthquake i is denoted η_i and the residual between observation j and an event term-adjusted GMPE (also known as a within-event residual) is denoted R_{ij}. The issue considered in this section is the spatial correlation of residuals R_{ij}.

As described in Section 5.2.2 and Appendix A, the spatial variation of within-event residuals is described using a semi-variogram, an example of which is shown in Figure 5.2. As will be shown in this section, the correlation coefficient of damage demand (ρ_{DD}), which describes the correlation coefficients of within-event residuals, can be reproduced from the semi-variogram.

By definition, the semi-variance $\gamma_i(\mathbf{h})$ (Eqn. A.1) for event i can be computed as (Isaaks and Srivastava, 1990):

$$\begin{aligned}\gamma_i(\mathbf{h}) &= \frac{1}{2}\left[E\left(R_{ij}-R_{ij+\mathbf{h}}\right)^2\right] \\ &= \frac{1}{2}E\left(R_{ij}{}^2\right)-E\left(R_{ij}R_{ij+\mathbf{h}}\right)+\frac{1}{2}E\left(R_{ij+\mathbf{h}}{}^2\right) \\ &= E\left(R_{ij}{}^2\right)-E\left(R_{ij}R_{ij+\mathbf{h}}\right)\end{aligned} \tag{7.30}$$

where j+$\mathbf{h}$ is a segment separated by distance $\mathbf{h}$ from segment j. Note that $\mathbf{h}$ is the direct separation distance between levee segments, which may be shorter than the separation distance measured along the alignment of the levees. The covariance [i.e., $C_i(\mathbf{h})$] can be similarly computed as (Isaaks and Srivastava, 1990):

$$\begin{aligned}C_i(\mathbf{h}) &= E\left[\left(R_{ij+\mathbf{h}}-\mu_R\right)\left(R_{ij}-\mu_R\right)\right] \\ &= E\left(R_{ij+\mathbf{h}}R_{ij}\right)-\mu_R{}^2 \\ &= E\left(R_{ij}{}^2\right)-\gamma_i(\mathbf{h})-\mu_R{}^2 = C_i(0)-\gamma_i(\mathbf{h})\end{aligned} \tag{7.31}$$

where μ_R is the mean of R_{ij} across all segments, and $C_i(0)$ is the variance of R_{ij}. The distance-dependent correlation coefficient of residuals for event i, denoted $\rho_{DD,i}$, is the ratio of the co-variance to the variance of R_{ij}:

$$\rho_{DD,i}(\mathbf{h}) = \frac{C_i(\mathbf{h})}{C_i(0)} = 1-\frac{\gamma_i(\mathbf{h})}{C_i(0)} \tag{7.32}$$

Considering two segments separated by a large enough distance to be regarded as statistically independent [i.e., separation distance > range (α)], the semi-variance will be as follows:

$$\gamma_i(\mathbf{h}') = E\left(R_{ij}^{\ 2}\right) - E\left(R_{ij+\mathbf{h}'}R_{ij}\right)$$
$$= E\left(R_{ij}^{\ 2}\right) - E\left(R_{ij+\mathbf{h}'}\right)E\left(R_{ij}\right) = E\left(R_{ij}^{\ 2}\right) - \mu_R^{\ 2}$$
$$= C_i(0) = S_i \tag{7.33}$$

where **h'** is a separation distance larger than the range, and S_i is the sill, which is a maximum value of semi-variance corresponding to statistically independent conditions.

As shown in Appendix A, a general expression for fitting semi-variance data is:

$$\hat{\gamma}(\mathbf{h}) = c_0 + c_1\left(1 - \exp\left(-\frac{3\mathbf{h}}{\alpha}\right)\right) \qquad \text{(A.2, 7.34)}$$

where α is the range, and c_0+c_1 is equivalent to the sill, S_i. Substituting $\hat{}(\mathbf{h})$ and $\hat{\rho}$ into Eqn. (7.32) and replacing $C_i(0)$ with S_i, I find the following if the nugget (c_0) is zero:

$$\hat{\rho}_{DD,i}(\mathbf{h}) = 1 - \frac{\hat{\gamma}_i(\mathbf{h})}{S_i} = \exp\left(-3 \times \frac{\mathbf{h}}{\alpha_i}\right) \tag{7.35}$$

where α is the range, as before, which can be established by regression. The subscript i is retained because α could potentially vary by event.

Regression analyses based on semi-variance data from the 2004 and 2007 earthquakes provide α estimates of 21 and 27 km. Using those values, the variation of $\hat{\rho}_{DD}$ with distance is as shown in Figure 7.9. **Error! Reference source not found.** Jayaram and Baker (2009) have analyzed semi-variance data from several California earthquakes and the Chi Chi Taiwan earthquake, from which a general model similar to Eqn. (7.34) was derived. While they did not consider the IM of PGV, for the related IM of 1.0 sec pseudo spectral acceleration, they find a range of 25.7 km. The correlation based on this model is also shown in Figure 7.9, and is very close to that evaluated for the Japan earthquakes. It is important to point out that the range for these ground motion models is approximately an order of magnitude greater than the range for the damage state correlation (shown in Figure 7.1). Hence, the capacity correlation is in general much weaker than the demand correlation.

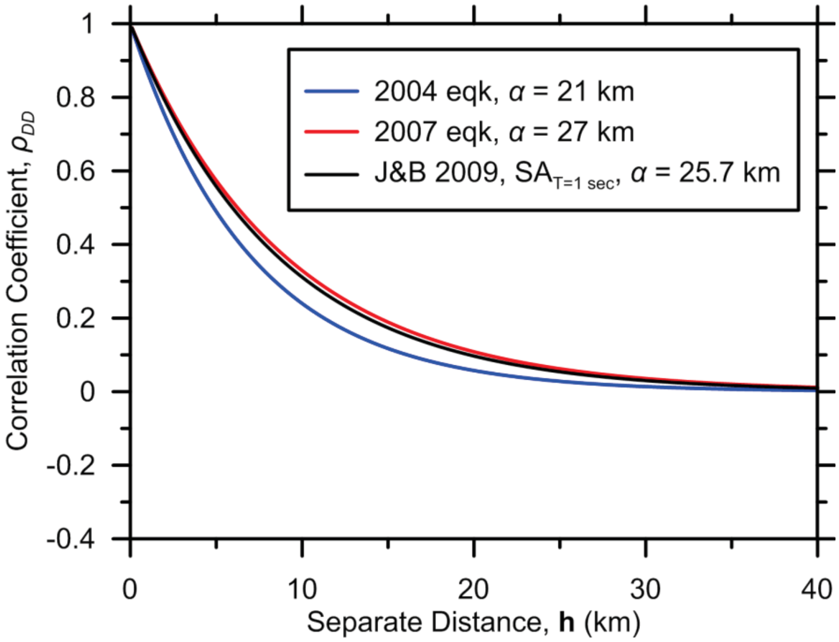

Figure 7.9 Correlation coefficients of damage demands (i.e., within-event residuals for *PGV*) by 2004 and 2007 Niigata earthquakes.

7.4.4 Monte-Carlo Simulation for Evaluating System Fragility

With the capacity and demand correlation coefficients (ρ_{DC} and ρ_{DD}) now having been defined, we can now propose a Monte-Carlo simulation approach for analysis of system fragility. The steps in this procedure are as follow:

1. Generate two matrices with normally distributed random variables with zero mean and standard deviation of unity. One matrix ($\mathbf{Z}_{DC}$) contains the random field that will be used to compute damage capacities (elements in the matrix are denoted z_{ki}). The other matrix ($\mathbf{Z}_{DD}$) contains entries r_{ki} that are taken as modified within-event residuals for event k and segment i (modification is described subsequently). The two matrices are written as:

$$\mathbf{Z}_{DC} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N1} & z_{N2} & \cdots & z_{Nn} \end{bmatrix} \tag{7.36}$$

$$\mathbf{Z}_{DD} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1} & r_{N2} & \cdots & r_{Nn} \end{bmatrix} \tag{7.37}$$

Within the matrices, each row represents an event, and each column represents a segment. There are n segments and N events that are simulated in $\mathbf{Z}_{DC}$ and $\mathbf{Z}_{DD}$.

2. Construct matrices of correlation coefficient for damage capacity ($\mathbf{K}_{DC}$) and demand ($\mathbf{K}_{DD}$), which are as follows:

$$\mathbf{K}_{DC} = \begin{bmatrix} 1 & (\rho_{DC})_{12} & \cdots & (\rho_{DC})_{1n} \\ (\rho_{DC})_{21} & 1 & \cdots & (\rho_{DC})_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (\rho_{DC})_{n1} & (\rho_{DC})_{n2} & \cdots & 1 \end{bmatrix} \tag{7.38}$$

$$\mathbf{K}_{DD} = \begin{bmatrix} 1 & (\rho_{DD})_{12} & \cdots & (\rho_{DD})_{1n} \\ (\rho_{DD})_{21} & 1 & \cdots & (\rho_{DD})_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (\rho_{DD})_{n1} & (\rho_{DD})_{n2} & \cdots & 1 \end{bmatrix} \tag{7.39}$$

where $(\rho_{DC})_{ij}$ and $(\rho_{DD})_{ij}$ represents correlation coefficients of damage capacity and demand between segment i and j, respectively.

3. Using the Cholesky decomposition method (Baecher and Christian, 2003), estimate matrices $\mathbf{Y}_{DC}$ and $\mathbf{Y}_{DD}$ containing correlated random variables. The entries in these matrices are denoted z'_{ki} (used for correlated capacity) and r'_{ki} (used for correlated demand). The decomposition is performed as follows:

$$\mathbf{Y}_{DC} = \mathbf{S}_{DC} \times \mathbf{Z}_{DC} \tag{7.40}$$

$$\mathbf{Y}_{DD} = \mathbf{S}_{DD} \times \mathbf{Z}_{DD} \tag{7.41}$$

where $\mathbf{S}_{DC}$ and $\mathbf{S}_{DD}$ are lower triangular matrices from Cholesky decomposition, whose multiplication with their transpose [i.e., $(\mathbf{S}_{DC})^{\mathrm{T}}$ and $(\mathbf{S}_{\mathrm{DD}})^{\mathrm{T}}$] results in the correlation matrices, $\mathbf{K}_{DC}$ and $\mathbf{K}_{DD}$.

4. Transform z'_{ki} and r'_{ki} to damage capacities (c_{ki}) and demands (d_{ki}) in the units of log velocity:

$$c_{ki} = \exp\left(z'_{ki} \times \sigma_{C,i} + \mu_{C,i}\right) \tag{7.42}$$

$$d_{ki} = \exp\left(r'_{ki} \times \sigma_{GMPE,i} + \mu_{GMPE,i}\right) \tag{7.43}$$

where $\mu_{C,i}$ and $\sigma_{C,i}$ are the log mean and standard deviation of damage capacity for segment i evaluated from a fragility function, whereas $\mu_{GMPE,i}$ and $\sigma_{GMPE,i}$ are the log mean and standard deviation of IMs from a GMPE.

5. Find the damage state for segment i and event k (f_{ki}) by comparing c_{ki} and d_{ki} as follows:

$$f_{ki} = \begin{cases} 0 & \text{if } c_{ki} > d_{ki} \\ 1 & \text{if } c_{ki} \le d_{ki} \end{cases} \tag{7.44}$$

6. Find the damage state for the system, considering the damage states of all n components for event k (F_k):

$$F_k = 1 - \prod_{i=1}^{n} \left(1 - f_{ki}\right) \tag{7.45}$$

7. Estimate probability of system survival $P(F_S)$ as the complement of the mean survival across the N events:

$$P(F_S) = 1 - \frac{1}{N}\sum_{k=1}^{N}(1 - F_k) \tag{7.46}$$

Following the above procedure, $P(F_S)$ can be calculated for a system given the statistical moments of damage capacity and demand distributions, as well as the capacity and demanc correlation coefficients. In Section 7.5, I estimate $P(F_S)$ for a portion of the Shinano River levee system using this procedure.

7.5 EXAMPLE APPLICATION FOR SHINANO RIVER SYSTEM

7.5.1 Correlation Coefficients for Shinano River Levee System

For system fragility it is needed to quantify the correlation coefficients of damage capacities ρ_{DC} and damage demands ρ_{DD} as well as moments of capacity and demand. I calculate ρ_{DC} and ρ_{DD} for the Shinano River down- to mid-stream (SH1), upstream (SH2), and Uono River (UO) levees from the observed damage by 2004 and 2007 earthquakes following the procedures described in Section 7.4.2 and 7.4.3.

As shown in Section 7.3.2, correlation coefficient of damage states ρ_{DS} can be substituted by the auto-correlation coefficient ρ_{acc} if damage occurrence is stationary in space and number of segments are enough. With assumption of stationarity, I suppose that ρ_{acc} can be used for ρ_{DS} because the number of segments for the longest stretch is more than 200 for each region and averaging multiple stretches reduces the scatter of ρ_{acc} as shown in Figure 7.2.

As described in Section 7.4.2, ρ_{DC} is a function of ρ_{DS} and P_f at two segments [i.e., $P(f_i)$ and $P(f_j)$]. Utilizing fragility function developed in Chapter 6 (Table 6.4) with ground motions (i.e., PGV) and hydrologic conditions (i.e., D_W), P_f can be evaluated for each segment. ρ_{DS} for the Shinano River is estimated from ρ_{acc} as shown in Figure 7.1. Then, ρ_{DC} can be calculated by inversion analysis with ρ_{DS} and P_fs as described in Section 7.4.2. Figure 7.10 shows ρ_{DC} between segments with certain separation distance x. Although $\hat{\rho}_{acc}$, which is the representative of ρ_{DS}, is constant for separation distance x, ρ_{DC} varies because P_f for each segment changes depending on the metrics for fragility function (i.e., PGV and D_W). Nevertheless, ρ_{DC}s are placed within moderately narrow ranges. I fit a function ($\hat{\rho}_{DC}$) to mean values of ρ_{DC} for each separation distance x, where the fit form is as follows:

$$\hat{\rho}_{DC} = \exp\left(-3\times\left(\frac{x}{\alpha_{DC}}\right)^{1.5}\right) \tag{7.46}$$

where α_{DC} is the regression parameter at which change of $\hat{\rho}_{DC}$ relative to zero is negligible beyond. Figure 7.10 shows that α_{DC} is longer than α_{DS} (Eqn. 7.21).

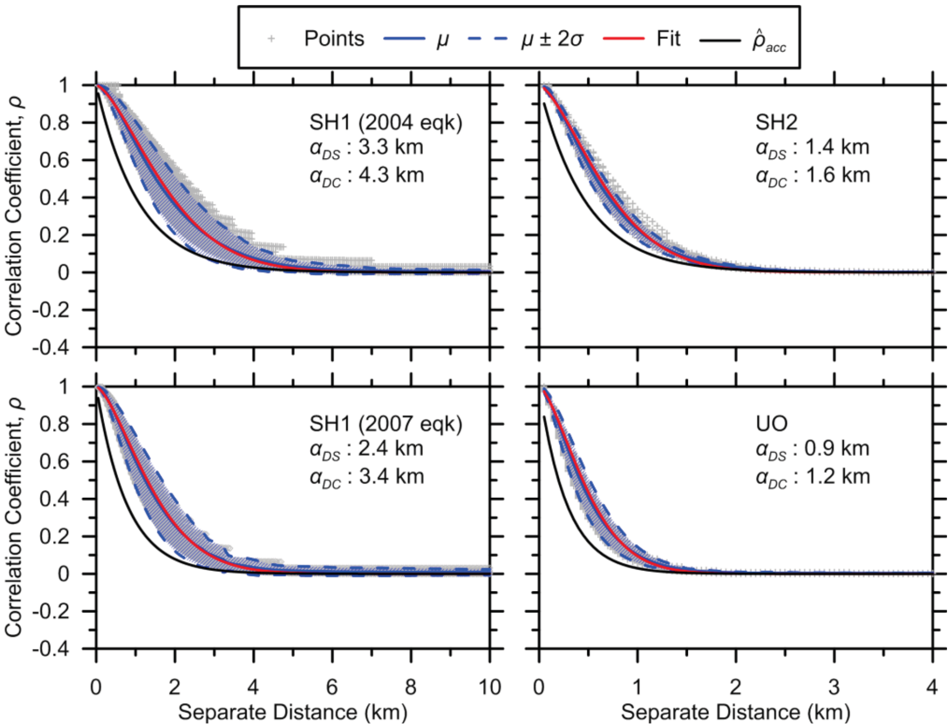

Figure 7.10 **Correlation coefficients of damage capacities for levee systems within SH1, SH2, and UO Rivers. Mean (μ) and two plus and minus standard deviations ($\mu \pm 2\sigma$) are shown.**

Correlation coefficient of within-event *PGV* residuals, which is utilized for damage demand by summation with GMPE, is derived from semi-variograms as described in Section 7.4.3. The regression parameter α_{DD} in Eqn. (7.34) are 21 and 27 km for 2004 and 2007 earthquake, respectively, which are much longer than α_{DC}. This indicates that the correlation for damage demand is high at wider range than one of the damage capacity.

7.5.2 System Fragility for Shinano River System

Following the procedure described in Section 7.3.4, system fragility is calculated for a sub-set of Shinano River levees, SH1 left-side perspective to river mouth from 2.4 to 3.4 km river distance. I use $\hat{}_{DC}$ in Eqn. (7.46) for variables of $\mathbf{K}_{DC}$ in Eqn. (7.37) and $\hat{}_{DD}$ in Eqn. (7.34) for $\mathbf{K}_{DD}$ in Eqn. (7.38). I also utilize moments of log-normal CDF from fragility function (Table 6.4) in Eqn. (7.41) and mean and standard deviation of *PGV* from BEA GMPE in Eqn. (7.42).

Figure 7.11 shows the system fragility for selected levees having 1 km length using the **M** 6.6 2004 earthquake fault information (i.e., source-to-sited distance and fault mechanism). The maximum and minimum fragility bounds from Eqn. (7.3) are also shown, where P_f is calculated from Eqn. (7.24), not from fragility function. By varying α_{DC} and α_{DD}, I found that both α_{DC} and α_{DD} affect system fragility. The higher level of correlation, which corresponds to the longer α_{DC} and α_{DD}, causes the lower fragility. This result is not surprising because we know that perfect correlation predicts the minimum $P(F_S)$ and no correlation predicts the maximum $P(F_S)$ as shown in Eqn. (7.3). If both damage capacity and damage demand are perfectly correlated by themselves (i.e., infinite α_{DC} and α_{DD}), $P(F_S)$ will be the minimum, whereas both are statistically independent (i.e., zero α_{DC} and α_{DD}), the $P(F_S)$ will be the maximum. Note that α_{DC} and α_{DD} for the pair of SH1 and 2004 earthquake are 4.3 and 21 km, respectively. The fragility is more sensitive on α_{DD} than α_{DC}, indicating the significance of damage demand correlation more on system fragility than capacity.

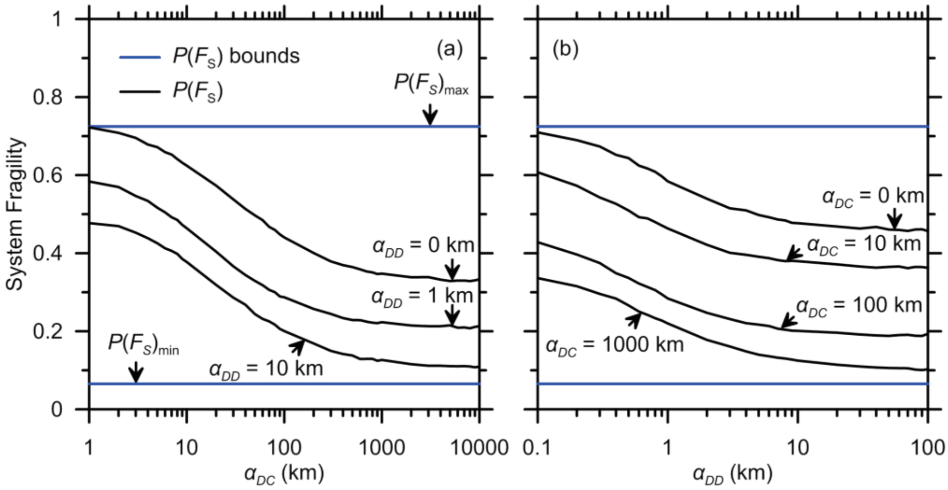

Figure 7.11 Fragility for levees at SH1 Left 2.4 ~ 3.4 km by varying (a) α_{DC} and (b) α_{DD} with the 2004 earthquake fault information (i.e., source-to-site distance and fault mechanism).

Figure 7.12 shows $P(F_S)$ of selected levee system with different magnitude in the range of **M** = 4 ~ 7 assuming the same source-to-side distance and the fault mechanism with the **M** 6.6 2007 earthquake. The original α_{DD} for the 2007 earthquake (27 km) is used for $P(F_S)$ calculation, whereas α_{DC} is changing from 0 to 1000 km. Note that α_{DC} is 3.4 km for the pair of SH1 and 2007 earthquake. Maximum and minimum bounds [i.e., $P(F_S)_{max}$ and $P(F_S)_{min}$] are also shown. $P(F_S)$ increases after **M** 5 event with high rate, whereas the rate is getting slow after **M** 6 event. Again, we can see that the lower correlation level predicts the higher $P(F_S)$. For example, **M** 6 event with α_{DC} = 100 km causes any damage within the selected 1 km levees with 20% chance, whereas the same event with α_{DC} = 10 km has 40% chance to occur any damage. This indicates that correlation level is significant for estimating system fragility.

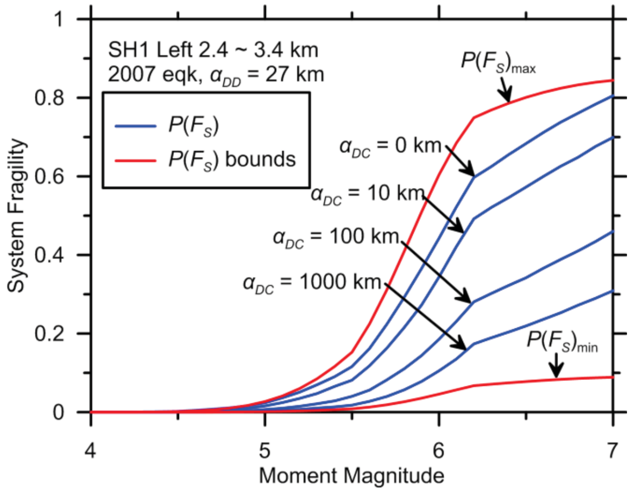

Figure 7.12 Fragility for levees at SH1 Left 2.4 ~ 3.4 km with varying α_{DC} and varying moment magnitudes. Fault information (i.e., source-to-site distance and fault mechanism) is assumed as the same with the M 6.6 2007 earthquake.

7.6 SUMMARY AND CONCLUSION

I derived a methodology evaluating system fragility considering spatial correlation of damage capacity and demand. The damage state, which is discrete variable, was represented as damage capacity and damage demand, of which both are continuous variables. The auto-correlation coefficient (ρ_{acc}) was used for the correlation coefficient of damage states (ρ_{DS}) due to lack of events. After evaluating ρ_{acc} from observed damage, correlation coefficient of damage capacity (ρ_{DC}) was derived by performing inversion analysis utilizing ρ_{acc} and segment damage probabilities obtained from the fragility function. Correlation coefficient of damage demand (i.e., within-event residual of PGV) (ρ_{DD}) was calculated from semi-variogram established in Section 5.2.2. A Monte-Carlo simulation procedure was proposed evaluating system fragility with correlated damage capacity and correlated damage demand.

I applied the proposed methodology to the Shinano River levee system damaged by 2004 and 2007 earthquakes. Correlations of both damage capacity and damage demand change the severity of the system fragility. The lower correlation coefficient caused the higher probability of damage for a system. For example, for 1 km length of levees near river mouth, correlation coefficient of damage capacity with α_{DC} =100 km caused 20% chance of damage, whereas α_{DC} =10 km caused 40% chance of damage by the **M** 6.6 2007 earthquake. I expect that the proposed procedure is useful to predict the system fragility for levee systems for future earthquakes.

8 CONCLUSION AND RECOMMENDATION

8.1 SCOPE OF RESEARCH AND DATA PREPARATION

My research goal is to learn from experience in regions of the world where levee systems have been strongly shaken in a manner similar to what would be expected during an earthquake for urban levees in California's Central Valley. I seek to evaluate under what conditions (i.e., ground shaking levels, soil conditions) levee systems similar to those in California experience permanent deformations and how severe those deformations are.

I have collected information from various branches under the Ministry of Land, Infrastructure, Transportation and Tourism (MLIT) in Japan regarding levee performance during the 2004 **M** 6.6 Niigata-ken Chuetsu earthquake and the 2007 **M** 6.6 Niigata-ken Chuetsu-oki earthquake. The MLIT branches I referred to for data collection are the Shinano River Work Office (SWO) for damage statements, geotechnical survey data, and levee geometry; Geological Survey of Japan (GSJ) for geology map; Geospatial Information Authority of Japan (GSI) for geomorphic classifications; MLIT Water Information System (WIS) for river water level; and several seismic networks including National Research Institute for Earth Science and Disaster Prevention (NIED) networks (i.e., K-net, KiK-net), Japan Meteorological Agency (JMA), National Institute for Land and Infrastructure Manage (NILIM), and local agencies (i.e., East Japan Railway Company, NEXCO East Japan, and Kashiwazaki City) for ground motion intensity. This information is described for the levees of the Shinano River system, which passes through regions strongly shaken during the 2004 and 2007 Niigata earthquakes.

Empirical analysis of levee fragility requires information on the spatial extent of levee damage and its severity, along with various metrics that may correlate to damage. Levee damage was quantified by measuring crack depth and width, the amount of slip (offsets across cracks), and the amount of relative settlement between damaged and undamaged levee sections by SWO. There are various types of damage, but crack depth, crack width, and crest subsidence are the most commonly available parameters to quantify damage severity. The metrics possibly correlating to damage are ground shaking intensity, geotechnical/geological conditions in the levee and its foundation, hydrological conditions (water level), and geometric parameters related to the levee cross section.

Geomorphological and geological conditions were obtained from the Geospatial Information Authority of Japan (GSI) that provides geologic and geomorphologic units with 1:25,000 resolution covering most of the study region. Eight units from GSI are present in the study region, which can be grouped further into three bins described in Figure 2.6: (1) mountain, terrace; (2) alluvial fan, natural levee, old river highland, alluvial plain; and (3) old river channel, back marsh. A levee geometric parameter (called shape factor) was defined as average levee height over average levee width. The shape factor explains the approximate aspect ratio of the levee, which might contribute to seismic deformations.

I developed a procedure to evaluate ground motions along the levee system that is sensitive to the site conditions related to varying shear wave velocity (V_s) of levee foundation materials. Geophysical measurements were not available at levee sites, whereas standard penetration test (SPT) blow counts were available at many levee sites. Furthermore V_s and SPT measurements were both made at a number of strong motion recording stations in Niigata and throughout Japan. Therefore, I developed a procedure to evaluate shear wave velocity as a function of SPT blow count, vertical effective stress, and soil type for the Niigata region. I collected a large V_s data set and performed regressions according to the functional form of Brandenberg et al. (2010) (Eqn. 4.1) to evaluate the effects of blow count (N_{60}) and overburden pressure (σ_v') on V_s. In addition, I interpolated boring-to-boring residuals using the Kriging method and related these spatially distributed values to geomorphic categories (JEGM), and developed a map of Japan that permits a reduction in prediction error. The time-averaged V_s to 30 m depth, V_{s30}, was estimated from the available blow count, effective stress, and JEGM data, and utilized to model site conditions on levees for ground motion prediction.

Shaking demands along levees are represented by *PGA* and *PGV*, which were estimated based on a spatial interpolation procedure that accounts for differences in site effects between recording stations and levee sites. A ground motion prediction equation, GMPE, was used to predict motions at strong recording stations, an inter-event residual (event term) was computed, and the remaining ground motion residuals were mapped using the Kriging method. These residuals were spatially interpolated at levee locations, and ground motions were estimated at the levee locations by adding the event-term and residual term to the median predicted ground motion intensity. The site term used in the GMPE prediction was consistent with levee site conditions, which were generally softer than conditions at strong motion recording stations. Direct Kriging interpolation of the ground motions themselves might result in bias due to differences in site conditions, whereas the procedure I developed accounts for differences in site conditions. Details of this procedure were described in Section 5.2.

The groundwater level at the time of the earthquake (*LGWE*) was estimated at each levee site based on river water elevations (*RWE*) obtained from eleven streamgauge stations and densely sampled surveys of river elevation at certain times, and borehole groundwater elevation at the time of the geotechnical site investigation. The difference *LGWE-RWE* was evaluated on the date of borehole explorations performed along the levee alignments. Corrections were made during rice growing season in irrigated regions because irrigation would be anticipated to influence *LGWE-RWE* for levees adjacent to farm land. Details of this procedure are described in Section 5.3.

I compiled statistics on penetration resistance conditional on soil type for the study area in Japan and a subset of data made available from the files of the California Department of Water Resources (CDWR) for urban levees in the Central Valley region of California. The data were compared to evaluate the degree of similarity of the soil conditions for these two levee systems, which is needed to judge whether the fragility models developed for Japan can be applied in the Central Valley region of California.

Fragility functions for levees were developed utilizing the Niigata-region empirical dataset, which is comprised of 3318 levee segments 50 m in length for the Shinano River system. Data sets are available for both the 2004 and 2007 earthquakes. Each 50 m segment has post-earthquake observations of performance following each event from which a damage level assignment was made, an estimated peak ground acceleration, and a series of secondary parameters including ground water elevation relative to levee base (D_W), levee shape (S_F), and mapped geomorphology (G_N). Using this dataset, I developed fragility functions that express the probability of exceeding a damage level (*DL*) conditioned on intensity measure (*IM*) alone (i.e., Model 1) and *IM* in combination with other parameters descriptive of geomorphic, hydrological and geometric conditions (i.e., Model 2). Model 2 conditions that significantly affect fragility were identified through visual inspection and formal statistical testing (*t*-test).

Both Models 1 and 2 were constructed in two stages. In Stage 1, I computed the probability of $DL > 0$ (i.e., the probability of some damage, regardless of its severity). In Stage 2, I computed the probability of exceeding particular damage levels conditional on some damage having occurred along with the additional metrics. Damage levels were assigned based on measured crest subsidence, crack width, and crack depth. Per the total probability theorem, the product of Stage 1 and 2 probabilities provides the desired probability of exceeding a given damage level. Liquefaction was responsible for about 80% of the most severe damage observed during the earthquakes. High groundwater elevation exhibited a significant influence on the fragility curves, and levees with high groundwater elevation were damaged more frequently than levees with deep groundwater.

The fragility functions represent the probability of damage exceedance for a specific 50m long segment of levee. However, system damage state depends on the length of levee within a particular system and the spatial correlation of damage occurrence. I developed a procedure to evaluate system fragility considering the damage prediction for each segment obtained from the fragility functions, and spatial correlation of damage among segments in the system. This procedure is essential for properly modeling system performance, and spatial correlation is very important in this context because levees are a series system (i.e. controlled by the weakest link).

8.2 RESEARCH FINDINGS AND RECOMMENDATIONS

The data collection and synthesis phase of work led to the following conclusions:

- At levee sites founded on Holocene floodplain deposits, the Central Valley north, central, and south regions levees have, on average, slightly lower liquefaction susceptibilities than comparable materials in the Shinano River region. For each of the major material types investigated other than gravel, penetration resistances for the Niigata-region materials are lower than those for the California regions. Moreover, the proportion of coarse-grained soils encountered in levees and levee foundations is less in the Central Valley regions than in the Niigata region. California levees founded on older basin deposits rather than Holocene floodplain deposits exhibited higher penetration resistance compared with Japanese levee sites.
- The models for prediction of V_s from N_{60} and σ_v' considering geomorphology and spatial variation of boring-to-boring residuals predict sufficiently more accurate Vs_{30} values than other V_{s30} models using geomorphologic and topographic proxies. Including geotechnical site investigation data results in less biased and less scattered residuals compared with other proxies, which indicates more prediction power. Predicted V_{s30}s along levees were similar with measured V_{s30}s at adjacent sites. Nevertheless, the most accurate V_{s30} can be achieved by geophysical measurements rather than by correlation with penetration resistance.
- The ground motion interpolated scheme proposed in this work provides substantial benefit to the estimation of ground motion intensity measures under conditions where nonlinear site response is likely to be present. Because levees are often located on soft soils, and the levels of shaking considered in this study are high, this is a critical detail in the development of the empirical models.

The analysis of fragility functions for levees described in Chapter 6 led to the following findings and preliminary conclusions:

- I recommend the maximized likelihood estimates (MLE) method for regressing the parameters describing a fragility function. After testing various alternate functional forms, I selected the log-normal CDF because it provides relatively high MLE, and because moments of log-normal CDF (i.e., mean and standard deviation) have an easily understood physical meaning.
- The Model 1 (*IM*s only), Stage 1 (probability of damage, regardless of damage level) analysis showed that *PGV* correlates more strongly to damage occurrence than *PGA*. The fragility generally monotonically increases with *PGV* between 14 cm/s and 80 cm/s. No damage occurred below approximately 14 cm/s, and the damage probabilities saturated at approximately 0.4 for $PGV > 80$ cm/s. Model 1, Stage 2 (conditional probability of a particular damage state for a damaged segment) results indicated that the damage severity, conditional on damage having occurred, is effectively independent of *PGV*. The damage probabilities of Stage 2 are approximately 0.53 for $DL > 1$, 0.14 for $DL > 2$, and 0.02 for $DL > 3$.
- The development of Model 2 (*IM*s plus geologic conditions, groundwater level, and levee aspect ratio) began with consideration of which among several secondary parameters produce statistically significant departures in *PGV*-dependent damage probabilities from those given in Model 1. The most important secondary parameter in this regard is the elevation of groundwater in levees relative to the levee base (D_W). Shallow ground water conditions ($D_W > -1.0$ m) provide higher damage probabilities, by approximately 40 percent, than those for

deep ground water conditions (D_W < -1.0 m). This indicates that groundwater elevation plays an important role in seismic levee fragility. Geomorphic category and geometric conditions are generally not significant at the level of Stage 1 fragility, although the relatively competent materials in G_N 1 produce markedly lower fragility. Stage 2 results indicate that the damage severity is generally *PGV*-dependent for non-liquefaction sites (typically D_W < -1.0 m) and *PGV*-independent for liquefaction sites (i.e., D_W > -1.0 m).

The analyses of spatial correlation of damage and the levee system probability described in Chapter 7 led to following findings and preliminary conclusions:

- The levee system probability depends on the spatial correlation of damage capacity and damage demand. Higher correlation leads to the lower system damage probability, converging to the levee segment fragility in the extreme case of perfect correlation among the segments. Spatial correlation of damage demand more significantly affected system fragility than spatial correlation of damage capacity.
- Fragility of a 1 km long levee system was evaluated for various scenario moment magnitudes based on the simplifying assumption that the source-to-site distances and the fault mechanisms for those events are fixed as the same with the 2007 Niigata earthquake. For 1 km length levee system, there was no damage predicted under **M** 5. The system fragility increases for **M** > 5, whereas the increasing rate decreases after M **6**.

The levee fragility models developed in this work are strictly applicable for the conditions along the Shinano River in Japan for *PGV* ≤ 140 cm/s and **M** 6.6 earthquakes. Their applicability to other regions, such as those present along flood control levees in California's Central Valley, are dependent on the similarity of the seismological, geotechnical, and hydrological conditions as well as the spatial correlation of damage occurrence. Where this compatibility can be demonstrated, the proposed fragility relations are useful for preliminary levee risk assessments in which detailed geotechnical data are unavailable. The fragility functions presented in this work are not applicable to levees that constantly impound water, such as those in the Sacramento – San Joaquin Delta, which we anticipate to be more susceptible to earthquake-induced damage due to the predominance of shallow saturated sediments.

8.3 FUTURE WORK

Completion of this dissertation marks an important, but intermediate, milestone in the continuing research effort, and as such, several significant tasks are incomplete and not discussed completely in this document. A summary of major objectives that have not yet been completed is given below:

i. I am working on a damage model for cases in which site-specific geotechnical data is available, which will be denoted Model 3. I anticipate relating observed damage to lateral displacement index (LDI) and/or and slope displacement (SD) from Newmark analysis.

ii. Data from additional river systems will be evaluated, in particular the Sabaishi River and U River located southwest of the Shinano River. I also anticipate evaluating data from the 1993 Kushiro-oki earthquake and 2011 Tohoku earthquake. This will extend the data set and fill in some data gaps.

APPENDIX A: METHODOLOGY OF KRIGING

In this Appendix I summarize the methodology of the Kriging utilized for spatial variations of boring-to-boring residuals in Section 4.4 and with-event residuals in Section 5.2.2. Semi-variogram which is used for Kriging analysis is also presented.

Semi-variogram is a function of semi-variances (i.e., half of variance) of data spatially separated by a distance **h**. The equation of semi-variogram [i.e., $\gamma(\mathbf{h})$] is as follows:

$$\gamma(\mathbf{h}) = \frac{1}{2}\left[E\left\{ Z(u_i) - Z(u_{i+\mathbf{h}}) \right\}^2 \right] \quad \text{(A.1)}$$

where a vector **Z** [i.e., $Z(u_1)$, $Z(u_2)$, ..., $Z(u_n)$] is random values spatially distributed, and $Z(u_{i+\mathbf{h}})$ denotes a variable separated by **h** from $Z(u_i)$. This semi-variogram is typically zero when the separate distance is zero, and is increased by **h** because usually data are scattered more by distance. The semi-variogram has three parameters: nugget, sill (or partial sill), and range. The nugget and sill are the semi-variograms at zero distance and infinite distance, respectively, and range is the distance where the difference of the semi-variogram from the sill becomes negligible. Figure A.1 shows semi-variograms for within-event PGV residuals for the 2004 and 2007 Niigata earthquakes.

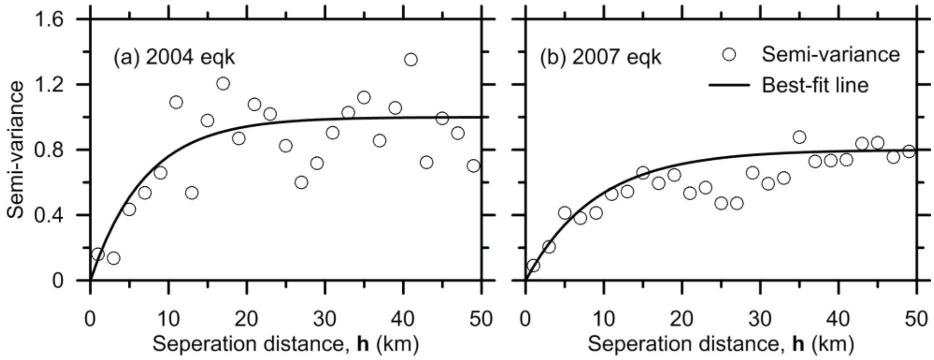

Figure A.1 **Semi-variograms for within-event *PGV* residuals by (a) 2004 and (b) 2007 Niigata earthquakes. The ranges are 21 and 27 km for 2004 and 2004 events, respectively.**

There are several functions used for fit models to semi-variograms, while a from often used is the exponential function as follows:

$$\hat{\gamma}(\mathbf{h}) = c_0 + c_1\left(1 - \exp\left(-\frac{3\mathbf{h}}{\alpha}\right)\right) \tag{A.2}$$

where c_0 is the nugget, $c_0 + c_1$ is the sill, and α is the range. Figure A.2 shows an example of semi-variogram of $\mathbf{Z}$ (i.e., with-in event residuals for the 2004 earthquake *PGV* as shown in Figure A.1a) and a fitted model using Eqn. (A.2). Note that nugget, sill, and range are [0, 1, 21 km] for 2004 event, and [0, 0.8, 27 km] for 2007 event, respectively. Sill was found by normalizing within-event residuals by one standard deviation.

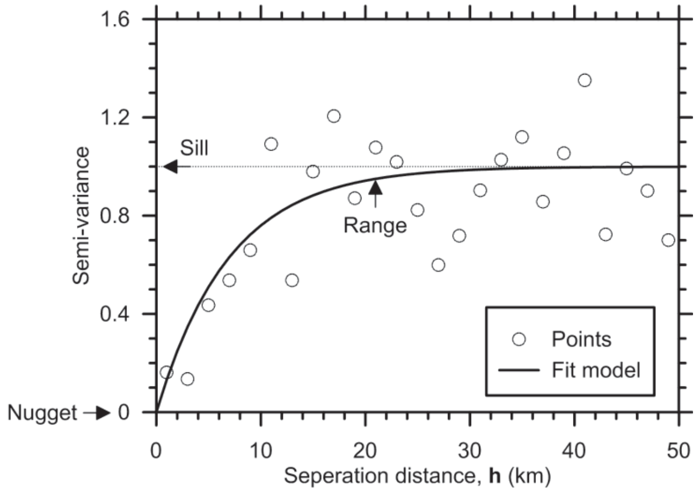

Figure A.2 Example of semi-variogram with the exponential fit to data points.

Once the semi-variogram of a spatial data is given, we can approximate the value at unknown points using Kriging method that widely used for spatial variation (Oliver and Webster, 1990). Kriging is a linear interpolation method between known data points to estimate values at locations without data. The basic equation is

$$\hat{Z}(u_0) = \sum_{i=1}^{n} w_i Z(u_i) \tag{A.3}$$

where $\hat{Z}(u_0)$ is a value at a target unknown point u_0, $Z(u_i)$ is a value at a known data point i with separate distance **h** between u_0 and u_i, and w_i is a weight, which depends on **h** and a semi-variogram. Known data points closer to the target point have larger weights (lower variance), and vice versa. The sum of weights is unity.

The weight w is calculated from a semi-variogram model as follows:

$$\begin{Bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{Bmatrix} = \begin{bmatrix} C(u_1-u_1) & C(u_1-u_2) & \dots & C(u_1-u_n) \\ C(u_2-u_1) & C(u_2-u_2) & \dots & C(u_2-u_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(u_n-u_1) & C(u_n-u_2) & \dots & C(u_n-u_n) \end{bmatrix}^{-1} \begin{Bmatrix} C(u_0-u_1) \\ C(u_0-u_2) \\ \vdots \\ C(u_0-u_n) \end{Bmatrix} \tag{A.4}$$

where u_i-u_j within the covariance function C indicates separation distance between the points u_i and u_j. $C(u_i$-$u_j)$ is the covariance between u_i and u_j points, which is calculated as $C(\mathbf{h}) = \text{Sill} - \gamma(\mathbf{h})$ utilizing the semi-variogram model (Eqn. A.2). Eqn (A.4) is derived by minimizing the mean squared error of prediction, $E\left[Z(u_0) - \hat{Z}(u_0)\right]^2$.

Using the above procedure, we can predict a Z variable at any unknown point using spatially correlated $\mathbf{Z}$ for which a semi-variogram model is fitted. This technique was applied in the main text in Section 4.4 for boring-to-boring residuals and in Section 5.2.2 for with-in event residuals.

APPENDIX B: STATISTICAL TEST FOR MODEL DISTINCTION

Herein I illustrate statistical tests used to evaluate the distinction of a model from another model. For illustrate purpose, I utilize Model 2 and Model 1 fragilities described in the main text in Sections 6.3 and 6.4.

The dependencies of Model 2 on secondary parameters can be illustrated visually by segregating the data according ranges of a secondary parameter. I select $G_N = 1$ and $G_N = 3$, which are two geomorphic categories. Figure B.1a shows the G_N-independent fragility (Model 1) along with the two G_N subsets. Figure B.1b shows Model 1 fragility residuals for each subset, a linear fit to the residuals, and the 90% confidence intervals on the fit. Figure B.1c shows the mean residuals of all data points within a subset. These results are useful to gain a qualitative sense for the data-to-model misfit. In the case of $G_N = 1$, Figure B.1a shows that data fall below the Model 1 fragility, with the degree of misfit growing with PGV. The residuals in Figure B.1b have a strongly negative trend that causes the confidence intervals to not include zero for $PGV >$ 30 cm/s. The mean of overall residual (μ_{r2}) is negative but its confidence intervals include zero. Taken collectively, these results suggest that the $G_N = 1$ data are distinct relative to Model 1 principally on the basis of the strong slope of the residuals. On the other hand, the $G_N = 3$ data produce Model 1 residuals having 90% confidence intervals that encompass zero over the full range of PGV. For this reason, the $G_N = 3$ data appears to be less statistically distinct from Model

1. This issue of data distinction relative to a model is more formally investigated through statistical testing below.

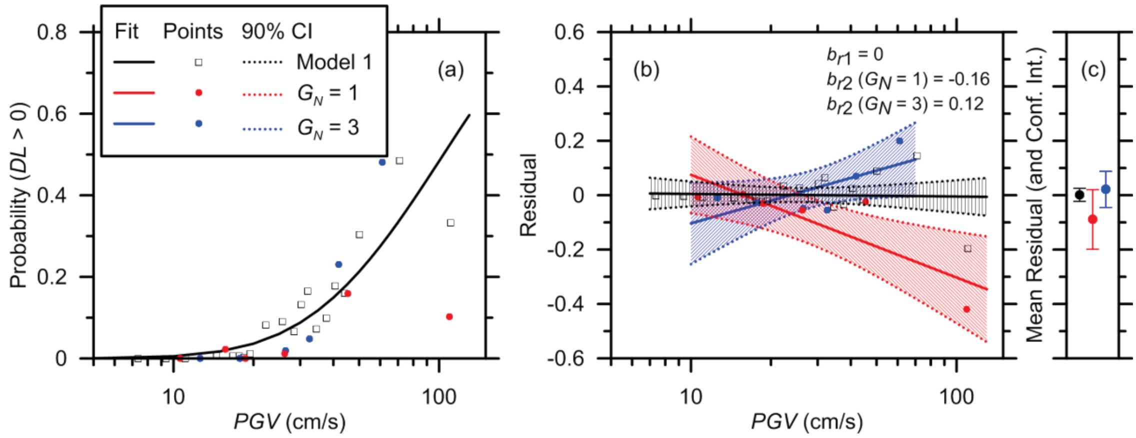

Figure B.1 **Examples of (a) data points of Model 2 conditioned by G_N = 1 and G_N = 3 along with Model 1 fragility and (b) residuals relative to Model 1 predictions. (b) Linear fit lines and (c) means of residuals with 90% confidence intervals, respectively.**

I use two types of t-tests to investigate the distinction between Model 1 and Model 2. Recall that μ_{r1} and μ_{r2} are defined as the means of Model 1 residuals relative to Model 1 and 2 data points, respectively as in Figure B.1c. If the Model 1 residuals are plotted against PGV, the resulting slopes are b_{r1} and b_{r2} (using Model 1 and 2 data points, respectively) as in Figure B.1b. The t-tests used to evaluate the following null hypotheses are as follows:

1. H_{01}: hypothesis is that μ_{r1} - μ_{r2} = 0 (the two means are identical),

2. H_{02}: hypothesis is that b_{r1} - b_{r2} = 0 (the two slopes are identical).

Rejection of the null hypothesis indicates that Model 2 is distinct from Model 1.

The t statistic for H_{01} is calculated as follows (Cook and Weisberg, 1999):

$$t = \frac{{}_{r1} \qquad {}_{r2}}{\sqrt{\frac{s_{r1}^2}{n_{r1}} + \frac{s_{r2}^2}{n_{r2}}}} \tag{B.1}$$

where μ_{r1} and μ_{r2} are as defined above, s_{r1} and s_{r2} are the corresponding standard deviations, and n_{r1} and n_{r2} are the number of residuals from Model 1 and 2 data, respectively. Note that μ_{r1} is approximately zero since Model 1 is regressed from Model 1 data points. Null hypothesis H_{01} is evaluated with the two-sample *t*-test, which takes as input the *t* value from Eqn. (B.1) and the approximate degree of freedom of the *t* distribution (*df*) as follows:

$$df = \frac{\left(\frac{s_{r1}^2}{n_{r1}} + \frac{s_{r2}^2}{n_{r2}}\right)^2}{\frac{\left(s_{r1}^2/n_{r1}\right)^2}{n_{r1} \quad 1} + \frac{\left(s_{r2}^2/n_{r2}\right)^2}{n_{r2} \quad 1}} \tag{B.2}$$

This expression is known as Welch–Satterthwaite equation which calculates combined degree of freedom of two unequal sample variances (Satterthwaite, 1946; Welch, 1947). The two-sample *t*-test is used to consider variances of both sets of data points (i.e., Model 1 and Model 2). The case of $G_N = 1$ results in [$t = 1.32$, $df = 5$], and the case of $G_N = 3$ results in [$t = -0.48$, $df = 6$], respectively.

Null hypothesis H_{02} is computed using the two-sample *t*-test as with H_{01}. The *t* statistic for H_{02} is calculated as follows:

$$t = \frac{b_{r1} \quad b_{r2}}{\sqrt{se_{b1}^2 + se_{b2}^2}} \tag{B.3}$$

where se_{b1} and se_{b2} are the standard errors of b_{r1} and b_{r2}, respectively, calculated as follows:

$$se_b = \sqrt{\frac{\frac{1}{n_r \quad 2} \; {}_{i=1}^{n} \; \hat{}_i^{2}}{{}_{i=1}^{n} \left(x_i \quad \bar{x}\right)^2}} \tag{B.4}$$

where x is PGV in natural log units, $\hat{\ }$ is the residual [i.e., difference between data and linear fit (Figure B.1b)], and $\overline{x}$ is the mean of x. I use the two-sample t-test to consider variances of b_{r1} and b_{r2}. The degree of freedom of t is then as follows:

$$df = \frac{\left(se_{b1}^{2} + se_{b2}^{2}\right)^{2}}{\dfrac{\left(se_{b1}^{2}\right)^{2}}{n_{r1} \quad 2} + \dfrac{\left(se_{b2}^{2}\right)^{2}}{n_{r2} \quad 2}} \qquad \text{(B.5)}$$

The case of $G_N = 1$ results in [t = 2.80, df = 6], and the case of $G_N = 3$ results in [t = -1.92, df = 5], respectively.

The p-value is the probability of exceeding the t variate in the t-distribution with df degrees of freedom. Considering t and df for $G_N = 1$, p-values are 0.25 for H_{01} and 0.03 for H_{02}. For $G_N = 3$, p-values are 0.65 for H_{01} and 0.11 for H_{02}, respectively. These values support the findings, given in the main text, that the $G_N = 1$ data are distinct from Model 1 on the basis of the slope criterion and that the $G_N = 3$ data are not distinct from Model 1. These same findings were evident from the qualitative results given in Figure B.1 and discussed previously.

In addition to the application described above for model distinctions on fragility functions used in Section 6.4.1, the t-test for H_{02} hypothesis is applied to evaluate the distinction of ground slope and elevation with respect to boring-to-boring residuals in the main text of Section 4.3.2.

REFERENCES

Allen, T. I., and Wald, D. J., 2009. On the use of high-resolution topographic data as a proxy for seismic site conditions (V_{S30}), *Bull. Seism. Soc. Am.* **99**(2A), 935-943.

Andrus, R. D., Stokoe, K. H., and Juang, C. H., 2004. Guide for Shear-Wave-Based Liquefaction Potential Evaluation, *Earthquake Spectra* **20**(2), 285-308.

Ang, A. H-S., and Tang, W. H., 2007. *Probability concepts in engineering: emphasis on applications to Civil and Environmental Engineering*, *2nd Edition*, Wiley, New York, 420p.

Apel, H., Thieken, A. H., Merz, B., and Blöschl, G., 2004. Flood risk assessment and associated uncertainty, *Nat. Hazards Earth Syst. Sci.* **4**(2), 295–308.

Asano, K., and Iwata, T., 2009. Source rupture process of the 2004 Chuetsu, Mid-Niigata Prefecture, Japan, Earthquake inferred from waveform inversion with dense strong-motion data, *Bull. Seism. Soc. Am.* **99**(1), 123–140.

Aslani, H., and Miranda, E., 2005. Fragility assessment of slab-column connections in existing non-ductile reinforced concrete buildings, *J. Eqk. Eng.* **9**(6), 777–804.

Atwater, B. F., 1982. Geologic maps of the Sacramento-San Joaquin Delta, California, *U.S. Geological Survey Miscellaneous Field Studies Map MF-1401*.

Baecher, G. B., and Christian, J. T., 2003. *Reliability and Statistics in Geotechnical Engineering*, John Wiley & Sons Ltd, England.

Baker, J. W., 2008. Introducing correlation among fragility functions for multiple components, *Proc. 14th World Conf. Eqk. Eng.*, Oct 12-17, 2008, Beijing, China.

Baker, J. W., 2014. Efficient analytical fragility function fitting using dynamic structural analysis, *Earthquake Spectra*, in-press.

Baker, J. W., and Cornell, C. A., 2006. Correlation of response spectral values for multicomponent ground motions, *Bull. Seism. Soc. Am.* **96**(1), 215–227.

Boore, D. M., 2004. Estimating $V_s(30)$ (or NEHRP site classes) from shallow velocity models (depths < 30 m), *Bull. Seism. Soc. Am.* **94**(2), 591–597.

Boore, D. M., and Atkinson, G. M., 2008. Ground-Motion Prediction Equations for the Average Horizontal Component of PGA, PGV, and 5%-Damped PSA at Spectral Periods between 0.01s and 10.0s, *Earthquake Spectra* **24**(1), 99–138.

Boore, D. M., Stewart, J. P., Seyhan, E., and Atkinson, G. M., 2014. NGA-West 2 equations for predicting PGA, PGV, and 5%-damped PSA for shallow crustal earthquakes, *Earthquake Spectra*, in-press.

Boore, D. M., Thompson, E. M., and Cade, H., 2011. Regional correlations of VS30 and velocities averaged over depths less than and greater than 30 meters, *Bull. Seism. Soc. Am.* **101**(6), 3046–3059.

Brandenberg, S. J., Bellana, N., and Shantz, T., 2010. Shear wave velocity as function of standard penetration test resistance and vertical effective stress at California bridge sites, *Soil Dyn. Eqk. Eng.* **30**, 1026-1035.

Central Development (CD), 2004. *Detailed investigation of the Shinano River levees: the second*, Report, Central Development Co., Ltd. (in Japanese).

Choi, Y., and Stewart, J. P., 2005. Nonlinear site amplification as function of 30 m shear wave velocity, *Earthquake Spectra* **21**(1), 1–30.

Cook, R. D., and Weisberg, S., 1999. *Applied regression including computing and graphics*, Wiley & Sons, New York.

Daiwa Exploration Technology (DET), 2007. *Exploration of the geology of the Shinano River levees: Kamiya district*, Report, Daiwa Exploration Technology Co., Ltd. (in Japanese).

Danielson, J. J., and Gesch, D. B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010), *U.S. Geological Survey Open-File Report 2011–1073*, 26p.

Dikmen, U., 2009. Statistical correlations of shear wave velocity and penetration resistance for soils, *J. Geop. Eng.* **6**(1), 61–72.

Dunnett, C. W., and Sobel, M., 1955. Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution, *Biometrika*, 258-260.

Food and Agriculture Organization of the United Nations (FAO), 2004. *International year of rice 2004*. (last accessed from http://www.fao.org/rice2004/index_en.htm at February 2014)

Foster, J. L., et al., 2009. *Performance evaluation of the New Orleans and Southeast Louisiana Hurricane Protection System*, *vol. VIII–Engineering and operational risk and reliability analysis*, *Final Report of the Interagency Performance Evaluation Task Force*, U.S. Army Corps of Engineers, Washington, DC.

Foundation ground Consultants (FC), 2009. *Exploration of the geology of the Uono River levees: upstream district*, Report, Foundation ground Consultants Co., Ltd. (in Japanese)

Geological Survey of Japan (GSJ), AIST (ed.) 2014. Geological map of Japan 1:50,000. (last accessed from https://www.gsj.jp/en/index.html at April 2014)

Geospatial Information Authority of Japan (GSI), 1977. *Geomorphological map for flood control use* (in Japanese). (last accessed from http://www1.gsi.go.jp/geowww/lcmfc/lcmfc.html at February 2014)

Grigoriu, M., and Turkstra, C., 1979. Safety of Structural Systems with Correlated Resistances, *Appl. Math. Modelling* **3**, 130–136.

Hasancebi, N., and Ulusay, R., 2006. Empirical correlations between shear wave velocity and penetration resistance for ground shaking assessments, *Bull. Eng. Geol. Env.* **66**(2), 203–213.

Helley, E. J., 1979. Preliminary geologic map of Cenozoic deposits of the Davis, Knights Landing, Lincoln, and Fair Oaks quadrangles, California, *U.S. Geological Survey Open-File Report OF-79-583*.

Helley, E. J., and Harwood, D. S., 1985. Geologic map of the Late Cenozoic deposits of the Sacramento Valley and northern Sierran Foothills, California, *U.S. Geological Survey Miscellaneous Field Studies Map MF-1790.*

Hijmans, R. J., 2013. Raster: Geographic data analysis and modeling, *R package version 2.1-49*.

Horn, B. K. P., 1981. Hill shading and the reflectance map. *Proc. IEEE* **69**, 14-47.

Idriss, I. M., and Boulanger, R. W., 2008. *Soil liquefaction during earthquakes*, Earthquake Engineering Research Institute.

INA, 2008. *Detailed investigation of the Shinano River levees: Koide district*, Report, INA Co., Ltd. (in Japanese)

Isaaks, E. H., and Srivastava, R. M., 1990. *An introduction to applied geostatistics*, Oxford University Press, New York, 592p.

Japan Highway Public Corporation (JHPC), 2005. *Guidelines for Electronic Delivery of Surveys: Appendix of Geological Survey Part* (in Japanese).

Japan Meteorological Agency (JMA), 2012. *Database of Observed Earthquakes* (in Japanese). (last accessed from http://www.seisvol.kishou.go.jp/eq/kyoshin/jishin/index.html at December 2012)

Japanese Society of Civil Engineers (JSCE), 2011. *Earthquake Strong Motion Data Download Site*. (last accessed from http://download.jsce.or.jp/ at February 2011)

Jayaram, N., and Baker, J. W., 2008. Statistical tests of the joint distribution of spectral acceleration values, Bulletin of the Seismological Society of America, *Bull. Seism. Soc. Am.* **98**(5), 2231–2243

Jayaram, N., and Baker, J. W., 2009. Correlation model for spatially distributed ground-motion intensities, *Eqk. Engng Struct. Dyn.* **38**, 1687–1708.

Joyner, W. B., and Boore, D. M., 1981. Peak horizontal acceleration and velocity from strong motion records including records from the 1979 Imperial Valley, California, earthquake, *Bull. Seism. Soc. Am.* **71**, 2011-2038.

Kitakku, 2004. *Detailed investigation of the Shinano River levees: the third*, Report, Kitakku Co., Ltd. (in Japanese).

Kitakku, 2007a. *Detailed investigation of the Shinano River levees: Yoita, Atsushikesone, Hasugata district*, Report, Kitakku Co., Ltd. (in Japanese).

Kitakku, 2007b. *Exploration of the geology of the Shinano River levees: Koshijitakanashi district*, Report, Kitakku Co., Ltd. (in Japanese).

Kitakku, 2007c. *Exploration of the geology of the Shinano River levees: Manodainakajo district*, Report, Kitakku Co., Ltd. (in Japanese).

Kwak, D. Y., Brandenberg, S. J., Stewart, J. P., and Mikami, A., 2012a. Groundwater Level Evaluation for River Flood Control Levees and its Effect on Seismic Performance, *Proc. 15th World Conf. Eqk. Eng.*, Sep 24-28, Lisbon, Portugal.

Kwak, D. Y., Mikami, A., Brandenberg, S. J., and Stewart, J. P., 2012b. Ground motion estimation for evaluation of levee performance in past earthquakes, *9th Int. Conf. Urban Eqk. Eng./4th Asia Conf. Eqk. Eng.*, Tokyo Institute of Technology, Tokyo, Japan.

Matsuoka, M., Wakamatsu, K., Fujimoto, K., and Midorikawa, S., 2006. Average shear-wave velocity mapping using Japan engineering geomorphologic classification map, JSCE, *Str Eng./Eqk. Eng.* **23**(1), 57-68.

Miller, E. A., and Roycroft, G. A., 2004. Seismic performance and deformation of levees: four case studies, *J. Geotech. & Geoenvir. Eng.* **130**(4), 344–354.

Ministry of Land, Infrastructure, Transport and Tourism (MLIT), 2012a. *Inspection manual of seismic levee performance regarding Level 2 earthquake* (in Japanese), MLIT.

Ministry of Land, Infrastructure, Transport and Tourism (MLIT) 2012b. *Water Information System* (in Japanese). (last accessed from http://www1.river.go.jp at December 2012)

Miyake, H., Koketsu, K., Hikima, K., Shinohara, M., and Kanazawa, T., 2010. Source fault of the 2007 Chuetsu-oki, Japan, Earthquake, *Bull. Seism. Soc. Am.* **100**(1), 384–391.

Moehle, J., and Deierlein, G. G. 2004. A framework methodology for performance-based earthquake engineering, *International Workshop on Performance-Based Design*, Bled, Slovenia.

Moss, R. E. S., 2008. Quantifying measurement uncertainty of thirty-meter shear-wave velocity, *Bull. Seism. Soc. Am.* **98**(3), 1399-1411.

National Institute for Land and Infrastructure Management (NILIM), 2013. *Seismic Network for River and Road Facilities managed by MLIT* (in Japanese). (last accessed from http://www.nilim.go.jp/japanese/database/nwdb/html/newearthquake.htm at August 2013)

National Research Institute for Earth Science and Disaster Prevention (NIED), 2012a. *Borehole Data Checker*.

(last accessed from http://www.geo-stn.bosai.go.jp/software/boring/index.html at December 2012)

National Research Institute for Earth Science and Disaster Prevention (NIED), 2012b. *Strong-motion Seismograph Networks (K-NET, KiK-net)*.
(last accessed from http://www.kyoshin.bosai.go.jp/ at December 2012)

National Research Institute for Earth Science and Disaster Prevention (NIED), 2013. *Japan seismic hazard information station* (J-SHIS).

National Research Institute for Earth Science and Disaster Prevention (NIED), 2014. *Geo-Station: Integrated Geophysical and Geological Information Database*. (last accessed from http://www.geo-stn.bosai.go.jp/jps_e/ at April 2014)

Naval Facilities Engineering Command (NAVFAC), 1986. *Design Manual 7.01: Soil Mechanics*.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W., 1996. *Applied Linear Statistical Models*, McGraw-Hill, Boston, Massachusetts, 1408p.

New Cooperation Geology (NCG), 2007. *Exploration of the geology of the Shinano River levees: Wakikawa district*, Report, New Cooperation Geology, Co., Ltd. (in Japanese)

Nippon Koei (NK), 2009. *Exploration of the geology of the Shinano River levees: Tokamachi downstream district*, Report, Nippon Koei Co., Ltd. (in Japanese)

Oliver, M. A., and Webster, R., 1990. Kriging: a method of interpolation for geographical information systems, *Int. J. Geog. Inf. Sci.* **4**(3), 313–332.

Oppenheim, A. V., and Schafer, R. W., 2010. *Discrete-time signal processing, 3rd Edition*, Prentice Hall, 1120p.

OYO, 2008. *Investigation of liquefaction in the Shinano River*, OYO, Co., Ltd. (in Japanese)

OYO, 2009. *Exploration of the geology of the Shinano River levees: Tokamachi upstream district*, Report, OYO, Co., Ltd. (in Japanese)

Pagni, C. A., and Lowes, L. N., 2006. Fragility functions for older reinforced concrete beam-column joints, *Earthquake Spectra* **22**(1), 215–238.

Port and Airport Research Institute (PARI), 2014. *Strong-Motion Earthquake Records in Japanese Ports*. (in Japanese) (last accessed from http://www.eq.pari.go.jp/kyosin/ at April 2014)

Porter, K., 2003. An overview of PEER's Performance-based earthquake engineering methodology, *Ninth International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP9)*, San Francisco, California, U.S.A.

Porter, K., Kennedy, R., and Bachman, R., 2007. Creating fragility functions for Performance-Based Earthquake Engineering, *Earthquake Spectra* **23**(2), 471–489.

R Core Team (RCT), 2013. R: A language and environment for statistical computing, *R Foundation for Statistical Computing*, Vienna, Austria.

Rackwitz, R., and Krzykacz, B., 1978. Structural reliability of reactor systems, *Probabilistic analysis of nuclear reactor safety*, American Nuclear Society, Los Angeles, California.

Rosidi, D., 2007. Seismic risk assessment of levees, *Civil Engineering Dimension* **9**(2), 57–63.

Salah-Mars, S., Rajendram, A., Kulkarni, R., McCann, M. W. Jr., Logeswaran, S., Thangalingam, K., Svetich, R., and Bagheban, S., 2008. Seismic vulnerability of the Sacramento-San Joaquin Delta levees, *Geotech. Eng. & Soil Dyn. IV*, May 18-22, 2008, Sacramento, CA, 1–10.

Sanko Consultant (SC), 2007. *Exploration of the geology of the Shinano River levees: Ojiya district*, Report, Sanko Consultant Co., Ltd. (in Japanese)

Sasaki, Y., 2009. River dike failures during the 1993 Kushiro-oki Earthquake and the 2003 Tokachi-oki Earthquake, *Earthquake Geotechnical Case Histories for Performance-Based Design*, T. Kokusho (editor), 131–157.

Satterthwaite, F. E., 1946. An Approximate Distribution of Estimates of Variance Components, *Biometrics Bulletin* **2**, 110–114.

Sawada, S., Suetomi, I., Fukushima, Y., and Goto, H., 2008. Characteristics and distribution of strong ground motion during the 2004 Niigata-ken Chuetsu and 2007 Niigata-ken Chuetsu-oki Earthquake in Japan, *Proc. 14th World Conf. Eqk. Eng.*, Oct 12-17, 2008, Beijing, China.

Saygili, G., and Rathje, E. M., 2008. Empirical predictive models for earthquake-induced sliding displacements of slopes, *J. Geotech. & Geoenvir. Eng.* **134**(6), 790–803.

Seed, H. B., Tokimatsu, K., Harder, L. F., and Chung, R. M., 1985. The influence of SPT procedures in soil liquefaction resistance evaluations, *J. Geotech. Engrg.* **111**(12), 1425–1445.

Seyhan, E., Stewart, J. P., Ancheta, T. D., Darragh, R. B., and Graves, R. W., 2014. NGA-West 2 Site Database, *Earthquake Spectra*, in-press.

Shinano River Work Office (SWO), 2007. *Report of damage survey and estimation of the Shinano River by 2007 Niigata-ken Chuetsu-oki Earthquake.* (in Japanese)

Shinano River Work Office (SWO), 2008. *Log of the 2004 Niigata-ken Chuetsu Earthquake by Shinano River Work Office*. (in Japanese)

Sugita, H., and Tamura, K., 2008. Development of seismic design criteria for river facilities against large earthquakes, *Proc. 14th World Conf. Eqk. Eng.*, Oct 12-17, 2008, Beijing, China.

Thoft-Christensen, P., and Murotsu, Y., 1986. *Application of structural systems reliability theory*, Springer-Verlag Berlin Heidelberg.

Thoft-Christensen, P., and Sørensen, J. D., 1982. Reliability of structural systems with correlated elements, *Appl. Math. Modelling* **6**, 171–178.

U.S. Army Corps of Engineers (USACE), 2011a. *Guidelines for seismic stability evaluation of levees*, USACE Sacramento District.

U.S. Army Corps of Engineers (USACE), 2011b. *Design of I-walls, Appendix I: Length effects in levee system reliability*, EC 1110-2-6066, Washington, DC.

Vorogushyn, S., Merz, B., and Apel, H., 2009. Development of dike fragility curves for piping and micro-instability breach mechanisms, *Nat. Hazards Earth Syst. Sci.* **9**(4), 1383–1401.

Wakamatsu, K., and Matsuoka, M., 2011. Developing a 7.5-sec site-condition map for Japan based on geomorphologic classification, *Eqk. Res. Eng. Str. VIII*, 101–112.

Walling, M., Silva, W. J., and Abrahamson, N. A., 2008. Nonlinear Site Amplification Factors for Constraining the NGA Models, *Earthquake Spectra* **24**(1), 243–255.

Welch, B. L., 1947. The generalization of "student's" problem when several different population variances are involved, *Biometrika* **34**, 28–35.

Wolff, T. F., 2008. Chapter 12: Reliability of levee systems, *Reliability-based design in geotechnical engineering*, K. Phoon (editor), 448–496.

Yachiyo Engineering (YE), 2004. *Detailed investigation of the Shinano River levees: the first*, Report, Yachiyo Engineering Co., Ltd. (in Japanese)

Yachiyo Engineering (YE), 2009. *Exploration of the geology of the Uono River levees: downstream district*, Report, Yachiyo Engineering Co., Ltd. (in Japanese)

Yamazaki, F., Motomura, H., and Hamada, T., 2000a. Damage assessment of expressway networks in Japan based on seismic monitoring, *Proc. 12th World Conf. Eqk. Eng.,* Auckland, New Zealand.

Yong, A., Hough, S. E., Iwahashi, J., and Braverman, A., 2012. Terrain-based site conditions map of California with implications for the contiguous United States. *Bull. Seism. Soc. Am.* **102**, 114-128.