Diversification and Simulation of Ground Motion Distribution Based on Data Mining and Machine Learning

Takahashi Yukihiro

A Dissertation
Presented to
Department of Engineering Science
Doctor Program in Graduate School of Engineering
Gifu University

March 25, 2023

Abstract

This study examined “Diversification of Ground Motion Distribution” which aims to cover various earthquakes, including the future earthquake, with the Scenario Earthquake Shaking Maps (SESMs). In addition, “Simulation of Ground Motion Distribution,” a part of the advanced use of SESMs (or already computed ground motion distribution data), is also the research objective. In particular, the former aims to achieve the above objective by diversifying the set values of source parameters in the “strong ground motion prediction method for earthquakes with specified source faults (“recipe”),” which is the standard setting criterion for scenario earthquakes in Japan, namely, by expanding the “recipe”. One of the features of this paper is that it combines the “decomposition, synthesis, regression, prediction, and synthesis” methods of “data mining” and “machine learning” to carry out these studies. Here, these studies consistently focus on earthquakes occurring on a strike-slip fault. For the numerical example, this study used 600 strike-slip fault models with varying the “recipe” settings, and their ground motion distributions.

First, in “Simulation of Ground Motion Distribution,” this study proposes methods for generating ground motion distributions using the distributions calculated by strong ground motion simulation. In particular, this study proposed two methods. One is “Chapter 2: Simulation of Strong Ground Motion Distributions of Two Horizontal Components,” which can randomly generate many ground motion distributions with the spatial variation structure of the original ground motion distributions as a constraint. The other is “Chapter 3: Simplified Prediction Model for Ground Motion Distribution,” which can predict ground motion distribution for a specific fault rupture scenario.

Chapter 2 extends the simulation method for single-component ground motion distributions proposed in the previous study to two horizontal component distributions. This method simultaneously generates the two-component distributions, preserving the mutual spatial variability structure of the two components. Specifically, singular value decomposition analysis was applied to the ground motion distributions of two horizontal components by period. Then, the mode shapes that reflect the mutual spatial variability structure of two components and the principal component scores (PCSs) representing the case-dependent mode weights were extracted for each mode. Next, the cross-covariance structure represented by the PCSs of the two components was reproduced by correlation simulation using the Cholesky decomposition; the ground motion distributions of the two components were generated. The simulated distributions adequately produced the two-component ground motion distribution and their interactive spatial variation structure.

Chapter 3 proposed a simplified prediction model of ground motion distribution based on mode decomposition and machine learning as a surrogate model for a detailed based on strong ground motion simulation. The modeling used “multiple linear regression,” “support vector regression,” “random forest,” and “gradient boosting decision trees,” respectively. The predicted distribution shows spatial characteristics of ground motion corresponding to the fault rupture scenario, which roughly reproduces the ground motion distribution calculated by the detailed method. The differences in prediction accuracy among the four models were not seen in period 0.1 s where the effect of distance attenuation was dominant. However, they were seen at long periods where the directivity effect and radiation pattern were more significant than at short periods. The models with the highest prediction accuracy among the four models were GBDT.

“Diversification of Ground Motion Distribution” is a study on set scenario earthquakes that can cover various ground motion distributions. This study is socially significant because the adequacy of the scenario earthquake setting is directly related to disaster prevention and risk assessment quality. “Chapter 4: Analysis of Period and Component Dependent Spatial Characteristics” clarifies the relationship between the source parameters and the spatial characteristics of ground motions. Next, “Chapter 5: Diversification of Source Parameters” examines the method of expanding the “recipe” for efficiently reproducing the diversity of spatial characteristics.

In Chapter4, the mode decomposition is applied to the ground motion distributions, then mode shapes that represent the spatial characteristics and the weight coefficients that characterize each case were obtained. Chapter 3 constructs the weight coefficient predictor with the source parameters as explanatory variables. Here, the input-output relationship of this predictor was analyzed for each mode by applying the machine learning model’s interpretation techniques, explainable AI, to the spatial characteristics and its dominant source parameters. As a result, there are three main modes that contribute significantly to the ground motion distributions. Firstly, the modes represent distance attenuation characteristics by period and component. Secondly, the modes increase or decrease the ground intensity of the entire area depending on the seismic moment and rupture velocity. Finally, the modes relate to the location of the hypocenter and the asperities.

First, Chapter 5 shows the diversity of ground motion distributions for 600 fault models which was set up with a variation in the set values for the “recipe”. Next, the relative position of the ground motion distributions for the fault models based on the “recipe” (the recipe cases) with respect to the 600 cases was presented. The results show that the ground motion distributions of the recipe cases are average and limited. Then, based on the results in Chapter 4, fault models with varying “recipe” settings (the extended cases) were set up to analyze the spread of spatial characteristics represented by the recipe case and the extended cases. Here, the parameters varied were the seismic moment, the inner fault parameters (e.g., arrangement of asperities, rupture velocity), and the extra fault parameters (e.g., position of hypocenter). As a result, it is found that the diversification of ground motion distributions can be reproduced by setting appropriate extended cases.

Finally, Chapter 6 clarifies the applicability and usefulness of data mining and machine learning methods used in this study to multi-hazards. This chapter applied the “decomposition” methods to tsunami inundation depth distribution and analyzed the spatial characteristics of tsunami inundation depth. As a result, it is found that the spatial characteristics of the tsunami are dominated by the large-slip area of the tsunami fault model.

Acknowledgments

This paper is the result of research conducted by the author under Professor Nobuoto NOJIMA, Faculty of Engineering, Gifu University, while enrolled in the Doctoral Program of the Graduate School of Engineering, Gifu University.

I would like to express my deepest gratitude to my supervisor, Professor NOJIMA, for his kind and careful guidance and much valuable advice since I was assigned to the Earthquake Engineering Laboratory. I have learned how to conduct research, write papers and research plans, make presentations, and keep in mind what it means to be a researcher. All the things he taught me became the foundation of my life and became important assets and weapons in my future research life. I would also like to express my sincere gratitude for his encouragement, advice, and support in selecting a laboratory, entering a doctoral program, job hunting, and many other situations. In addition, despite his busy schedule, he provided me with sincere guidance and advice not only on weekdays but also during the year-end and New Year holidays and late at night, which enabled me to be appointed as a JSPS Research Fellow DC2 and to write this paper. Furthermore, he has given me opportunities to visit disaster-affected areas, participate in domestic and international conferences, and study abroad, and I have learned to always be inquisitive. I would like to express my deepest gratitude for your guidance.

I would like to express my deepest gratitude to Professor Takao KAGAWA, Faculty of Engineering, Tottori University, for providing the fault models and ground motions data used in this paper and for his great expertise and valuable advice and recommendation on this research. Furthermore, despite his busy schedule, he answered my questions honestly, even though I had no knowledge of the subject. His advice and suggestions helped me learn about strong ground motion prediction and clarified the positioning of my research.

I am very grateful to Assoc. Prof. Masumitsu KUSE, the River Basin Research Center, Gifu University, for his guidance since I was an undergraduate student. He always answered my questions with great care. Furthermore, in addition to making the environment easy to consult, I have received various kinds of support regarding computers; this helped me to handle multiple problems related to my research.

I would like to acknowledge Professor Tomoyuki OHTANI, Faculty of Engineering, Gifu University, for his role in the review of this research. His comments and advice at the seminar enabled me to rethink this research from a new perspective and improve this paper's quality.

I would like to express my special thanks to Professor Masata SUGITO, Project Professor at Gifu University and Director of the Gifu Disaster Mitigation Center, for the many suggestions and advice he gave me in the seminar. In seminars and lectures, I was impressed not only by the content of his research, but also by his own work and valuable experiences.

I am very grateful to Assoc. Prof. Maki KOYAMA, the River Basin Research Center, Gifu University, for her many suggestions and advice in the seminar. In particular, she always asked questions from the perspective of the purposefulness and social significance of the research, which enriched the content of the research background in Chapter 1.

I am deeply grateful to Mr. Hiroki KATO, a researcher at the Faculty of Engineering, Gifu University, for his support in many situations, from research to daily life. In particular, his discussions were very stimulating and led to self-improvement. Furthermore, he always cared for me and gave me much advice when writing this paper.

I am grateful to Ms. Yunalia MUNTAFI, Universitas Islam Indonesia (at that time, Doctor Program at the Graduate School of Engineering, Gifu University), for her guidance and advice not only on my research but also on my English writing and presentations. Her curiosity and willingness to work on her research in a foreign country have been a great source of inspiration for me.

I would also like to thank my seniors and juniors in the earthquake engineering laboratory at Gifu University for their great help managing the laboratory and conducting research. Especially, Mr. LE GUANG DUC guided me, an undergraduate student, on everything from programming to how to proceed with research. I would also like to thank Mr. Kotaro YOSHIDA, Ms. Ayumi NISHIJIMA, and Mr. Taisyu KATAYAMA for great cooperation in discussing Chapter 5 of this paper.

I would also like to express my gratitude for the opportunities to participate in conferences and joint seminars of four universities and one technical college. Through these activities, I have received advice, guidance, and encouragement from many professors and researchers.

Finally, I am deeply grateful to my parents for providing me with a rich educational experience and supporting me daily.

March 25, 2023
Yukihiro TAKAHASHI

Contents

1 Introduction

1.1 Outline

It is estimated that there are thousands of active faults throughout Japan, and all cities have the potential to be affected by earthquakes occurring on some faults. The National Research Institute for Earth Science and Disaster Resilience (NIED) [1] and the Headquarters for Earthquake Research Promotion (HERP) [2] designate 114 active fault zones that have a total length of 20 km or more and a high activity level and great social influence, as "major active fault zones." They have compiled and published the "Scenario Earthquake Shaking Maps (SESMs)" showing earthquakes which can be potentially caused by these faults. In SESMs, several cases of fault rupture scenarios (characterized source models) and ground motion distributions are set based on a "strong ground motion prediction method for earthquakes with specified source faults ("recipe") [3]," which is a standard method for setting up models for earthquake scenarios in Japan. However, actual earthquakes do not occur according to the assumed scenarios, and ground motion distribution trends often differ between the actual and the assumed earthquakes. Therefore, it is desirable to assume a wider variety of scenario earthquakes as SESMs and to represent a wider variety of ground motion distribution.

This study attempts to give variations to the "recipe" setting conditions and to efficiently represent various ground motion distributions, in other words, to diversify SESMs. Furthermore, a simulation method of ground motion distributions is proposed based on the ground motion distributions calculated by strong ground motion (SGM) simulation as an example of effective use of existing calculation results, i.e., advanced use of SESMs.

1.2 summarizes the research background and clarifies the objectives of the study. ***1.3*** describes the recipe, and ***1.4*** describes the seismic hazard maps in Japan and abroad. Next, ***1.5*** summarizes the previous studies related to this study; ***1.6*** summarizes the authors' previous studies, which are the predecessors of this study. Finally, ***1.7*** outlines the methodology used in this study, and ***1.8*** describes the structure of this study.

1.2 Research Background

The Southern Hyogo Prefecture Earthquake that occurred on January 17, 1995, was the first earthquake in Japan to be measured at seismic intensity of 7 on the Japan Meteorological Agency (JMA) scale, and it caused a great deal of seismic damage. Since this earthquake, NIED [1] has developed and operated Hi-net (High Sensitive Seismograph Network Japan) [4], K-Net (Kyoshin Network) [5], KiK-net (Kiban Kyoshin Network) [5], and F-net (Full Range Seismograph Network of Japan) [6], which cover all land areas in Japan. A large amount of strong-motion data has been accumulated, and the detailed rupture process of faults and the generation mechanism of strong-motion have been clarified. In June 1995, HERP [2] was established, and various seismic surveys are ongoing.

In SESMs published by HERP [2] and the Japan Seismic Hazard Information Station (J-SHIS) [7] established by IED [1], several fault rupture scenarios are set up based on the “recipe” [3], and their ground motion distributions are calculated by the SGM simulation. The SGM simulation using fault rupture scenarios can consider the effects of the fault rupture process and the heterogeneity of the fault plane (asperity and background) and can represent strong ground motion areas that appear near the fault and in the direction of rupture.

When designing a city for disaster prevention or conducting seismic risk assessment, it is very important to clarify and correctly assess the earthquakes that can be potentially caused by active faults near the target area and the ground motion distribution, especially the spatial distribution and correlation of the strong ground motions. Therefore, the adequacy of scenario earthquake setting is directly related to disaster prevention and risk assessment quality. However, it is very difficult to set up a fault rupture scenario that perfectly matches the actual earthquake in advance or even to set up a fault rupture scenario with conditions close to the actual earthquake. For example, the crustal earthquakes which caused significant damage in recent years, such as the 2000 Western Tottori earthquake, the 2016 Kumamoto earthquake, the 2018 Northern Osaka earthquake, and the 2018 Hokkaido Eastern Iburi earthquake, were those that differed from (or were not based on) the assumed fault rupture scenarios. (The 2016 Kumamoto earthquake: The length of the Futagawa section of the Futagawa fault zone that caused the main shock was longer than that assumed by HERP [2] and extended into the Aso caldera. HERP set the scenario earthquakes at M_J 7.0, but the magnitude of the main shock was M_J 7.3.) In addition, the distribution trend of the ground motions for the actual earthquake was also different from the assumed trend (or could not be assumed).

As described above, the limited cases that follow the "recipe" cannot cover the diversity of ground motion distributions, which contain various uncertainties. Considering the application of SESMs to disaster prevention, it is desirable to be able to describe the ground motion distribution of actual or future earthquakes by the spatial distribution trends (spatial characteristics) of ground motions represented by several cases of assumed scenario earthquakes. For this objective, it is necessary to give variations to the set values of fault rupture scenarios in the current recipe [3] and to extend the variety of spatial characteristics represented by scenario earthquakes. Therefore, even at the beginning of the recipe, it is described as follows:

- “It is assumed that the “recipe” will be modified and revised by the considerations in the evaluation of strong ground motion.” [3]

- “When it is necessary to consider variations and uncertainties in phenomena, it is desirable to set the source fault with due consideration of these points.” [3]

which suggests the need to update the “recipe” (i.e., to diversify the setting conditions). Here, regarding understanding and using SESMs, a relatively small number of published cases of scenario earthquakes is preferable to a large number, and the recipe should be diversified accordingly.

From the above background, this study attempts to efficiently cover the spread of various ground motions, including those of possible future earthquakes, with a relatively small number of scenario earthquakes. Specifically, this study aims to extend the diversity of spatial characteristics represented by scenario earthquakes, i.e., “diversification of ground motion distributions (or diversification/expansion of the “recipe” setting conditions)”,” by giving appropriate variations to the set values of fault rupture scenarios in the “recipe” [3]. Furthermore, as an example of the advanced use of SESMs, this study proposes a method to simulate new ground motion distributions using the ground motion distribution calculated by SGM simulations.

1.3 Strong Ground Motion Prediction Method for Earthquakes with Specified Source Faults ("Recipe") [3]

HERP [2] defines the recipe as follows:

> "This summarizes the current methods of source characterization, subsurface structure modeling, calculation of strong ground motions, and validation of prediction results, which are the components of strong motion prediction methods, as well as the concept of setting parameters for source characterization." [3]

It also aims to establish a "standard methodology that will give the same answer no matter who does it [3]" in predicting scenario earthquakes and assumes that it will be revised through future surveys and studies.

The recipe is made up of four processes:

1. Setup of Characterized Source Model (Fig. 1-1)

Set characterized source parameters considering the following three fault parameters.

- Outer fault parameters : Geometry and location of a fault (e.g., fault model size, dip angle, and seismic moment)
- Inner fault parameters : Heterogeneity of source faults (e.g., number and arrangement of asperities and rupture velocity)
- Extra fault parameters : Propagation pattern of rupture (e.g., hypocenter)

2. Preparation of Subsurface Structure Model

Model the following three geotechnical structures.

- Surface soil layers : From the upper plane of engineering bedrock to the ground surface
- Deep sedimentary layers : From the upper plane of seismic bedrock to the upper plane of engineering bedrock
- Crustal structure : Deeper than the upper plane of seismic bedrock

3. Calculation of Strong Ground Motions

Calculate strong ground motions from the source fault to the engineering bedrock's upper plane and from the engineering bedrock's upper plane to the ground surface.

4. Verification of Predicted Results

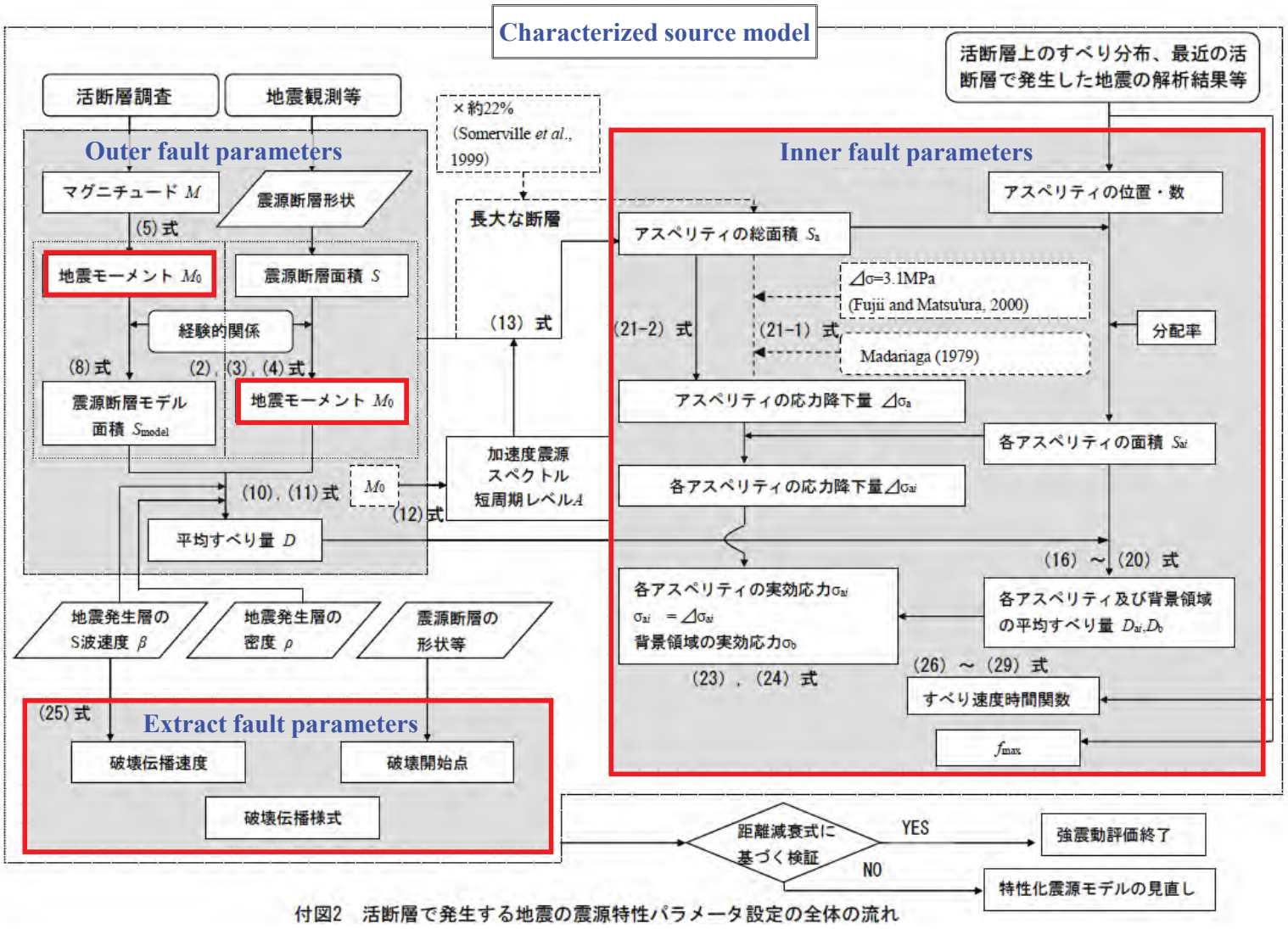
The validity of the predicted results is verified. In the case of an active fault, the verification is carried out by comparing the predicted results with the estimated value of the ground motion prediction equation and the observations from past earthquakes.

This study examines "1. Setup of Characterized Source Model" among the above four processes. Specifically, this study discusses how to assign appropriate variations to the setting conditions of "inner fault parameters" and "extra fault parameters" in addition to a seismic moment in "outer fault parameters (the other parameters are fixed)," which are shown in Fig. 1-1, circled in red. An example of setting source

parameters in a characterized source model is described below.

- Seismic moment : Calculation based on empirical formula between fault area and seismic moment
- Location of asperities : For estimating the average ground motions, the asperity is placed near the center of the fault when there is a single asperity, and the asperities are distributed in a balanced manner across the fault when there are several asperities.
- Average rupture velocity : Calculation based on empirical formula between S-wave velocity and average rupture velocity
- Hypocenter : Set at the left and right ends of the lower edge of asperity for strike-slip fault and the lower center of asperity for dip-slip fault.

Since most of these standards are empirical formulas based on observation data, the set values of source parameters in the recipe are “average” for various earthquakes. In other words, the set values include “variations,” and the set values within the range of variations are the parameters of an earthquake that have the potential to occur. Therefore, it is important to clarify the effect of this variation on the spatial distribution of ground motions in the “recipe diversification” of this study.



付図2　活断層で発生する地震の震源特性パラメータ設定の全体の流れ

(Red frame: Parameters examined in this study)

Figure 1-1. Overall flow of setting up a characterized source model in "recipe" (Note: Adapted from [3] and partially edited.).

1.4 Seismic Hazard Map around the World

In Japan, NEID [1] and HERP [2] have published the "National Seismic Hazard Maps for Japan [7, 8]". These hazard maps consist of two types of maps: the above "SESMs for Japan", and the "Probabilistic Seismic Hazard Maps for Japan", which show the relationship between the shaking intensity and the exceedance probability for all earthquakes in Japan. In SESMs, the detailed method of ground motion prediction based on the hybrid method is used to predict the ground motion distribution reflecting the rupture process of the source fault. Data on past earthquakes and the real-time ground motions are published by the Japan Meteorological Agency [9] and NIED [1], and other organizations.

In the United States, the United States Geological Survey (USGS) [10] publishes the "Seismic Hazard Maps [11]," which provide information on earthquake hazards and damage, and the "ShakeMap [12]," which provides information on the ground motion distributions. The ShakeMap consists of three primary types of shake maps: past large earthquakes (the ShakeMap Atlas), recent earthquakes (Real-time ShakeMaps) and a hypothetical earthquake that could occur on a specific fault (the ShakeMap Scenarios). In the ShakeMap Scenarios, ground motions are calculated using a conventional method based on a ground motion prediction equation (GMPE). Therefore, unlike the SESMs in Japan, the ground motion distributions of the ShakeMap Scenarios in the United States do not take into account the detailed rupture process of the fault, such as the rupture initiation point and rupture direction, but are average distributions that depends only on the location and dimensions of the fault, earthquake magnitude, distance, and site characteristics. The USGS states that the main reason for not setting up a detailed fault rupture model, such as a characterized source model, for the scenario earthquakes is as follows:

> "our approach is to generally show the average effect because it is difficult to justify a particular choice of hypocenter or to show the results for every possible hypocentral location." [13]

The USGS ShakeMap system is in operation in Switzerland [14], Italy [15], and other countries, and provides scenario earthquakes based on GMPEs and past observed earthquakes. In Italy [16] and New Zealand [17], probabilistic seismic hazard maps are provided. For global seismic hazard information, the Global Earthquake Model (GEM) Foundation, in collaboration with NIED, USGS, and others, has developed the "GEM Global Seismic Hazard Map," a global probabilistic seismic hazard map, and the "Global Seismic Risk Map," which shows the average annual loss to buildings from earthquakes [18].

As described above, various seismic hazard maps are also available in other countries. However, the Japanese SESMs are unique in the world in terms of scenario earthquake hazard maps (shake maps) that account for detailed rupture processes on specific source faults. In order to export the SESMs and the "recipe" to the rest of the world (for use by other countries), the key is to solve the problems identified by the USGS. The diversification of ground motion distributions (or diversification/expansion of the "recipe" setting conditions)" in this study is expected to be the basis for a solution to this problem.

1.5 Previous Studies on Fault Rupture Scenarios and Ground Motion Distribution

There have been many studies on setting up fault rupture scenarios (characterized source models) for scenario earthquakes. This section describes studies on the "recipe for predicting strong ground motions" and "effects of variations in source parameters on ground motions" among them.

1.5.1 Recipe for Predicting Strong Ground Motions

Strong ground motion prediction, which is the key to earthquake disaster countermeasures, consists of four processes as described in ***1.3***: "Setup of Characterized Source Model," "Preparation of Subsurface Structure Model," "Calculation of Strong Ground Motions," and "Verification of Predicted Results." In particular, "setup of characterized source model" is one of the most important tasks in strong ground motion prediction. In Japan, the characterized source models are often set up based on the recipe [3] of HERP, and the prediction quality is largely dependent on this recipe. This paragraph outlines the flow of research on the recipe in Japan.

In 1994, Irikura reviewed domestic and international papers on deterministic source modeling (setting up characterized source model) in strong ground motion prediction and mentioned the importance of examining the method of setting up source parameters [19]. In 1999, the paper [20] attempted to develop the "recipe for predicting strong ground motions" by compiling the procedures from source modeling to calculating strong ground motions for scenario earthquakes. In 2001, the paper [21] systematized the recipe with the aim of "everyone gets the same results when you assume a scenario earthquake." Furthermore, the setup of a source model and calculation of strong ground motions for the 1995 Southern Hyogo Prefecture Earthquake and the 1948 Fukui Earthquake were carried out, and it was shown that the prediction results were consistent with the observation records and the actual damage distribution, thereby clarifying the validity of the recipe [21]. In 2004, the papers [22, 23] proposed a revised recipe, which introduced new knowledge on setting up characterized source models, for example, the empirical relationship between seismic moment and acceleration source spectrum level [24]. In addition, these papers [20–24] pointed out the need for verification of settings such as "number of asperities" and "positions of asperities and hypocenter" included in "inner fault parameters" and "extra fault parameters" when modeling seismic sources. Therefore, in 2011, the paper [25] incorporated the setting of "rupture nucleation and termination" into the recipe. In addition, the paper [25] provided examples of setting "the number and arrangement of asperities" and "the location of the hypocenter."

In 2005, HERP published the "strong ground motion prediction method for earthquakes with specified source faults ("recipe") [26]" based on the recipe by Irikura et al. [23]. The recipe has also been continuously revised after its publication. For example, in 2008, the "setting of asperities and hypocenter" for evaluating the average earthquakes was added [27].

1.5.2 Effects of Variations in Source Parameters on Ground Motions

a) Spatial Distribution of Variation

Variations and uncertainties in source parameter settings can be classified into two categories [28]:

- Epistemic uncertainty: These uncertainties are based on a lack of scientific understanding and may be reduced in the future as knowledge accumulates.
- Aleatory variability (or Random uncertainty): These uncertainties cannot realistically be reduced even as knowledge accumulates.

The objective of this study is more narrowly defined as "to set up a scenario earthquake that is consistent with a future earthquake." Therefore, the uncertainties in this study are aleatory variabilities.

Evaluating the effect of variations in source parameters in conformity with the recipes (aleatory variabilities) on the spatial distribution of variations in ground motions is an important issue in probabilistic seismic hazard assessment and setting scenario earthquakes. Many studies carried out in the past have attempted to clarify this effect. Some of these studies are reviewed below.

Nozu et al. [29] pointed out that for a period 0.5 s and longer, the variability of ground motions between scenario earthquakes increases due to the directivity effects related to the arrangement of asperities and hypocenter (target: Fourier spectrum acceleration at period T = 0.1–1.0 s)).

Okumura et al. [30] pointed out that the variability varied among observation sites and was longer for longer periods (target: acceleration response spectrum S_A (T = 0.1, 2.0 s))

Ripperger et al. [31] found that the variability was dominated by the effect of the positional relationship with the hypocenter (target: peak ground velocity (PGV), S_A (T = 0.4, 0.5, 1.0, 4.0s)).

Yamada et al. [32–34] showed that the spatial distribution of variation in the short-period band is dominated by the location of asperity and hypocenter [32], that in the long-period band is dominated by the rupture velocity [32], and that the variation level is strongly influenced by the intensity of asperity (short period level) [33]. The papers [32–34] showed that the spatial distribution of variability is non-uniform regardless of fault type (target [32]: peak ground acceleration (PGA), PGV, S_A (T = 0.2, 0.5, 1.0, 2.0, 5.0s), target [33]: PGA, PGV, S_A (T = 0.2, 0.5, 1.0s), target [34]: PGV, S_A (T = 1.0, 2.0, 5.0s)).

Itoi et al. [35] found that the variability was larger for longer periods and that the distance dependence of the variability showed a negative dependence for shorter periods (the closer the distance between the fault and the site, the larger the variability at the site) and a positive dependence for longer periods (the farther away, the larger the variability) (target: S_A (T = 0.01 – 10 s)).

Kagawa [36, 37] pointed out that the variability is larger in the fault strike direction for the strike-slip fault and the dip direction from the bottom to the top of the fault for the reverse fault due to the directivity effect [36], and the effect is significant for a longer period [37] (target [36]: PGC, target [37]: S_A (T = 0.1, 0.5, 2.0s)).

Iwaki et al. [38] modeled the variability using the fault distance and the directivity parameter $\cos^2\theta$ (θ: the angle between the fault and the observation point). The results show that as the fault distance increases, the variability decreases for short periods and increases for long periods (especially significant at locations strongly affected by the directivity effect) (target: PGA, PGV, S_A (T = 0.1, 0.2, 1.0, 2.0, 5.0s)).

Hikita et al. [39, 40] pointed out the heterogeneous spatial distribution of variations. They found that variations in the fault strike direction are affected by parameters related to rupture pattern (e.g., asperity, hypocenter, and rupture velocity). In contrast, the variations in the fault orthogonal direction are strongly affected by major fault parameters (e.g., seismic moment, short period level) [39]. It was also noted that for earthquakes with M_{JMA} 5.2 or greater, the periodic band over in which the variability increases differs by seismic magnitude [40] (targets [39, 40]: S_A (T = 0.02–4.0s)).

These previous studies [29–40] revealed the heterogeneous spatial distribution in event-to-event (events or earthquakes) variability, such that the variability varies depending on the positional relationship between the fault (or hypocenter) and the observation site and shows different characteristics depending on the period.

Inter-event variability, a fundamental element in probabilistic seismic hazard assessment, has also been analyzed in various ways. For example, Iwaki et al. [38] found that the distance dependence of the inter-event is similar to the above events. Imtiaz et al. [41] revealed that the distance dependence of the inter-event depends on the rupture type (target: PGV). It was also noted that the intra-event variability increased with distance for the bilateral rupture, while the variability decreased for the unilateral rupture [41].

b) Spatial Correlation of Variation

The above studies [29–40] did not discuss the correlation of the ground motion variations among sites (site-to-site correlation). The distribution of site-to-site correlation (spatial correlation) of ground motion has a key role in considering the seismic risk assessment of portfolios of buildings in large areas. The spatial correlation of ground motions has also been studied extensively.

A typical evaluation method for inter-ground correlation is the macro spatial correlation model depending on the separation distance between two sites, which is a non-negative exponential function [42–44]. For example, Takada et al. [42] constructed a spatial correlation model based on the seismic records of the 1999 Chi-Chi earthquake in Taiwan, Hayashi et al. [43] constructed a model based on the seismic records in Kanto district (1996–2004, $M > 5.0$), and Sokolov et al. [44] constructed a model based on the seismic records in Taiwan (1993–2004, $M_L > 4.5$). However, Sokolov et al. [44] pointed out that the correlation structure depends on the regional ground structure's peculiarities and the rupture propagation path; a single generalized spatial correlation model may not adequately represent the correlation structure for the entire Taiwan region or large areas in other regions.

Hayashi et al. [43] performed the spatial correlation simulations based on the Cholesky decomposition of correlation matrices (variance-covariance matrices) according to the separation distances between sites and a Monte Carlo method; the simulation results are applied to the seismic risk analysis of building portfolios. Although this simulation method is simple and efficient, the correlation matrix becomes large for wide-area disasters, resulting in a high computational load. Abe et al. [45] proposed a method to reduce the computational load by making the correlation matrix sparse by setting the correlation coefficient to 0 above a certain separation distance. Apart from using a correlation matrix, Infantino et al. [46] showed that the 3D physics-based numerical simulations (3DPBNSs) can adequately reproduce the actual correlation structure from short to long periods. The paper also emphasizes the advantage of 3DPBNSs in reproducing spatial correlations even in regions where correlation models are difficult to define due to limited or no observational records [46].

1.6 Basic Research on "Diversification and Simulation of Ground Motion Distribution"

This section summarizes the basic research on "diversification of ground motions distribution (or diversification/expansion of the "recipe" setting conditions)" and" simulation of ground motion distribution" that the authors have carried out.

1.6.1 Analysis of Spatial Characteristics in Scenario Earthquake Shaking Maps

As mentioned above, in "diversification/expansion of the "recipe" setting conditions," it is necessary to clarify the effect of the variation range of source parameters on the ground motion distributions, in other words, the variation in the spatial distribution trend of ground motions. ***1.4.2*** reviewed the previous studies on the analysis of the effect of different multiple fault rupture scenarios (source parameter variations) on the spatial distribution of ground motion variability and the modeling and simulation of spatial correlation of variability. The studies [29–40] in ***1.4.2 a)*** analyze the "spatial distribution of variation" and do not focus on the "spatial correlation of variation." The spatial correlation model [42–44] in ***1.4.2 b)*** is a non-negative function based on the separation distance, assuming spatial uniformity (uniformity and isotropy) of the variation. Hence, it is not appropriate to apply this spatial correlation model to several ground motion distributions where the variation is spatially variable depending on the source parameter settings.

Nojima et al. [47] proposed a mode decomposition method using singular value decomposition for evaluating the "spatial distribution of variations and its spatial correlation (spatial characteristics)." This method is unique because it decomposes the ground motion distributions into orthogonally mode spaces and visualizes spatial characteristics without directly dealing with a large variance-covariance matrix. The authors have applied this method to SESMs for seven faults (reverse fault: 5, strike-slip fault: 2) and specified the spatial characteristics represented by each fault and their dominant source parameters [47–49]. As a result, it was revealed that the most dominant spatial characteristics that compose the ground motion distributions set based on the "recipe" are "distance attenuation characteristics and seismic amplification characteristics of the ground surface." Moreover, the spatial characteristics related to the location of the asperities and the hypocenter are dominant to the ground motion distributions, and the influence of the hypocenter-dependent directivity effect on the spatial characteristics is more apparent for strike-slip fault than for reverse fault. The spatial characteristics of faults with a similar arrangement of asperities are a similar trend.

1.6.2 Simulation of Ground Motion Distributions

Nojima et al. [47] proposed a method to generate a large number of ground motion distributions by combining mode decomposition and synthesis with an efficient Monte Carlo simulation method [50]. This simulation method uses the spatial characteristics of the original ground motion distributions as a constraint condition. It has the advantage that it can efficiently generate many cases of ground motion distributions that keep the spatial correlation structure with a small amount of computation. The authors have applied the simulation results of this method to the earthquake damage prediction [47] and the seismic risk

assessment of portfolios of buildings [51], demonstrating the method's usefulness. These studies [47, 51] were an advanced use of SESMs which are publicly available data and provides an example of the use of SESMs for disaster prevention and seismic risk assessment.

1.6.3 Classification of Characteristics of Multiple Scenario Earthquake Shaking Maps

The previous studies [47–49] in ***1.5.1*** mainly focused on the relationship between the spatial characteristics and the setting conditions of the source parameters. However, no similarity or characteristic evaluation was done based on the intercomparison of the predicted ground motion maps (SESMs). Therefore, Yoshida et al. [52] combined cluster analysis and principal component analysis to understand and analyze the effects of parameter settings on the characteristics of ground motion distributions in terms of similarity (or dissimilarity = features). The results show that the ground motion distributions caused by the strike-slip fault are similar in cases in which the hypocenter's location and rupture propagation direction are close to each other. For reverse fault, the similarity was observed in cases where the conditions for setting asperities (location, size, and number of asperities) were common. Using the same method, the authors also evaluated the relative relationship between the actual earthquake (the 2018 Hokkaido Eastern Iburi earthquake) and the scenario earthquake for a nearby fault (the Southern Part of the Eastern Boundary Fault Zone of the Ishikari Lowland (hereinafter referred to as Ishikari fault)) in terms of similarity.

1.6.4 Approximation Representation of Observed Ground Motion Distribution of Actual Earthquake using Scenario Earthquake Shaking Maps

Takahashi et al. [53] attempted to approximate the spatial distribution of ground motion of the actual earthquake (the 2018 Hokkaido Eastern Iburi earthquake) using the predicted ground motion distributions of SESMs (the Ishikari fault). As a result, the spatial distribution of the actual earthquake is interpreted based on the spatial characteristics of SESMs; this result provides an important insight into how the scenario earthquakes should be set up. This approximation method takes the opposite approach to the simulation method [47] described in ***1.4.2*** and is an advanced use of SESMs.

1.7 Diversification and Simulation of Ground Motion Distribution Based on Data Mining and Machine Learning

The studies [47–49, 51–53] in the previous sections are used as a basic study, and this study discusses the "diversification of ground motion distributions in scenario earthquakes." The study also proposes a simulation method of ground motion distribution using already computed ground motion distributions as part of the advanced use of SESMs, as described in the studies [47, 53]. This study set the research theme as "Diversification and Simulation of Ground Motion Distribution" and applies "Data Mining" and "Machine Learning" methods to the study. Specifically, as shown in Fig. 1-2, this study applies the "Decomposition, Synthesis, Classification, Regression (Modeling), and Prediction" methods to the ground motion distribution. The features of each method are described below.

- Decomposition (data mining) : Extract mode shapes representing spatial characteristics from ground motion distributions
- Synthesis (data mining) : Reproduce ground motion distribution by superimposing mode shapes
- Classification (data mining) : Understand characteristics of ground motion distributions
- Regression (machine learning) : Modeling relationship between source parameters and mode
- Prediction (machine learning) : Calculate weight of mode for any source parameters

This study combines these methods for diversification and simulation studies.

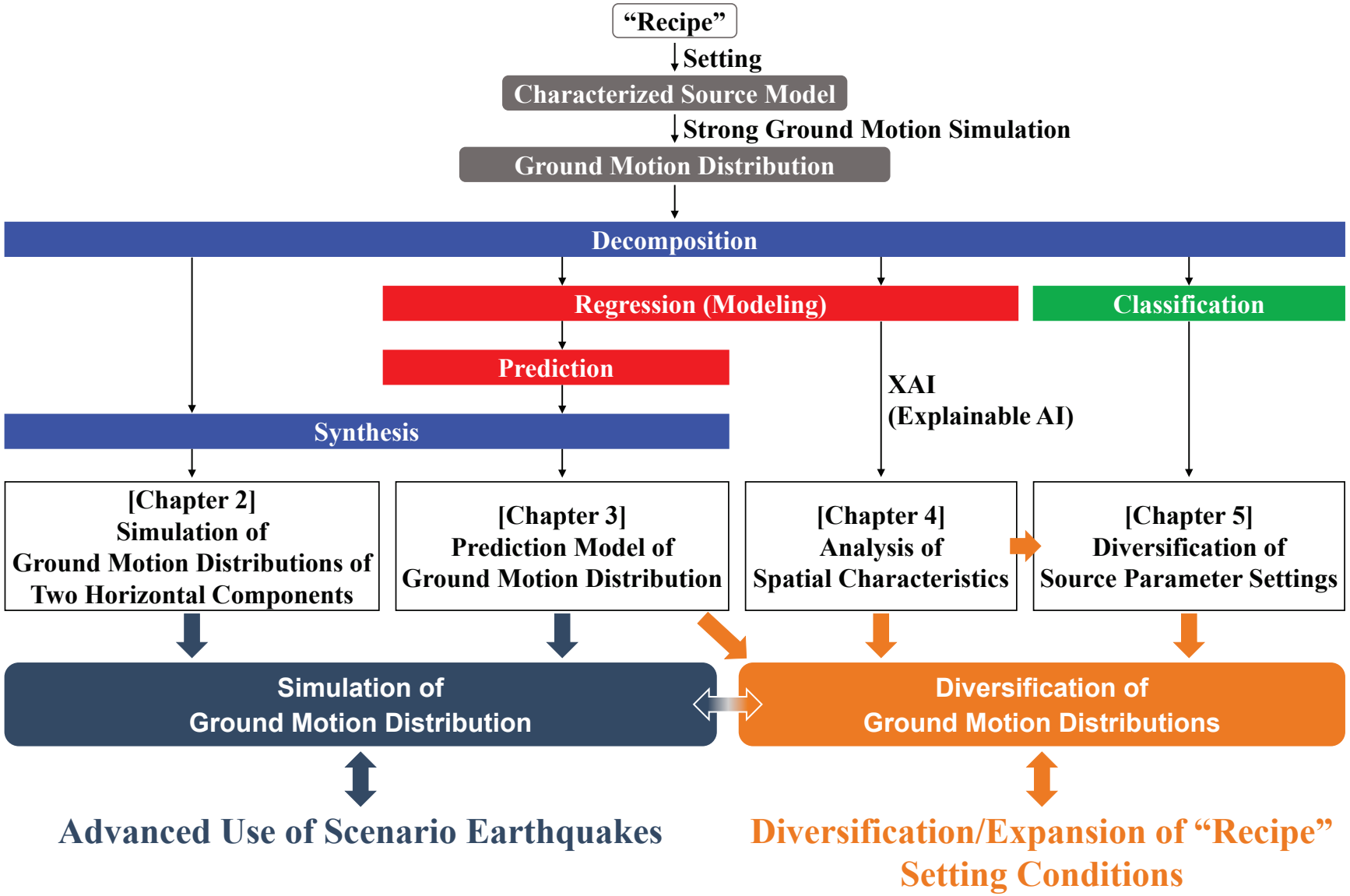


Figure 1-2. Research flow (Earthquake: Chapter 2–5): Diversification and simulation of ground motion distributions based on data mining and machine learning.

1.8 Structure of This Dissertation

The structure of this dissertation is as follows (Fig. 1-3).

Chapter 2 proposes a simulation method for ground motion distributions of two horizontal components (Fault-Normal: FN, Fault-Parallel: FP). The previous study [47] proposed a simulation method for single-component ground motion distribution. **Chapter 2** attempts to extend this single-component simulation method to two-components ground motion distributions. **Chapter 2** is a study on "Simulation" using "Decomposition and Synthesis." As a numerical example, this study uses a large number of cases of the ground motion distributions caused by the strike-slip fault, which was set up and calculated by Kagawa [36, 37] (the same applies to the following chapters).

Chapter 3 develops a simplified prediction model of ground motion distribution. The simulation methods in the previous study [47] and **Chapter 2** generate many ground motion distributions without relating them to fault models. In contrast, the prediction model of ground motion distribution in **Chapter 3** models the relation between source parameters and modes and predicts ground motion distribution by inputting any source model. **Chapter 3** is a study on "Simulation" using "Decomposition, Synthesis, Regression (Modeling), and Prediction."

Chapter 4 analyzes the period- and component-dependent spatial characteristics of ground motion distributions. This chapter reveals the combined effects of multiple source parameter settings on the ground motion distributions. Specifically, the input-output relationship of the predictor (input: source parameters, output; mode) constructed in **Chapter 3** is clarified by the Explainable AI, a machine learning model's interpretation method. Then, the relationship between the spatial characteristics and the source parameters is comprehensively analyzed. **Chapter 4** is a study on "Diversification" using "Decomposition and Regression."

Chapter 5 examines the diversification of ground motion distributions (or "recipe" setting conditions) based on **Chapter 4**. Specifically, it uses a classification method to visualize the diversity of spatial characteristics represented by multiple-case ground motion distributions. Next, the spatial characteristics of the fault models based on the current recipe (recipe cases) are compared with those of the multi-case. Based on the results of this comparison and the results of **Chapter 4**, some fault models (extended cases) are set up in which the set values of the source parameters in the recipe are varied, and an attempt is made to efficiently reproduce the diversity of spatial characteristics for multi-case by the recipe cases and the extended cases. The ground motion distributions for the recipe cases and extended cases are predicted using the prediction model for ground motion distribution in **Chapter 3**. **Chapter 4** is a study on "Diversification" using "Decomposition and Classification."

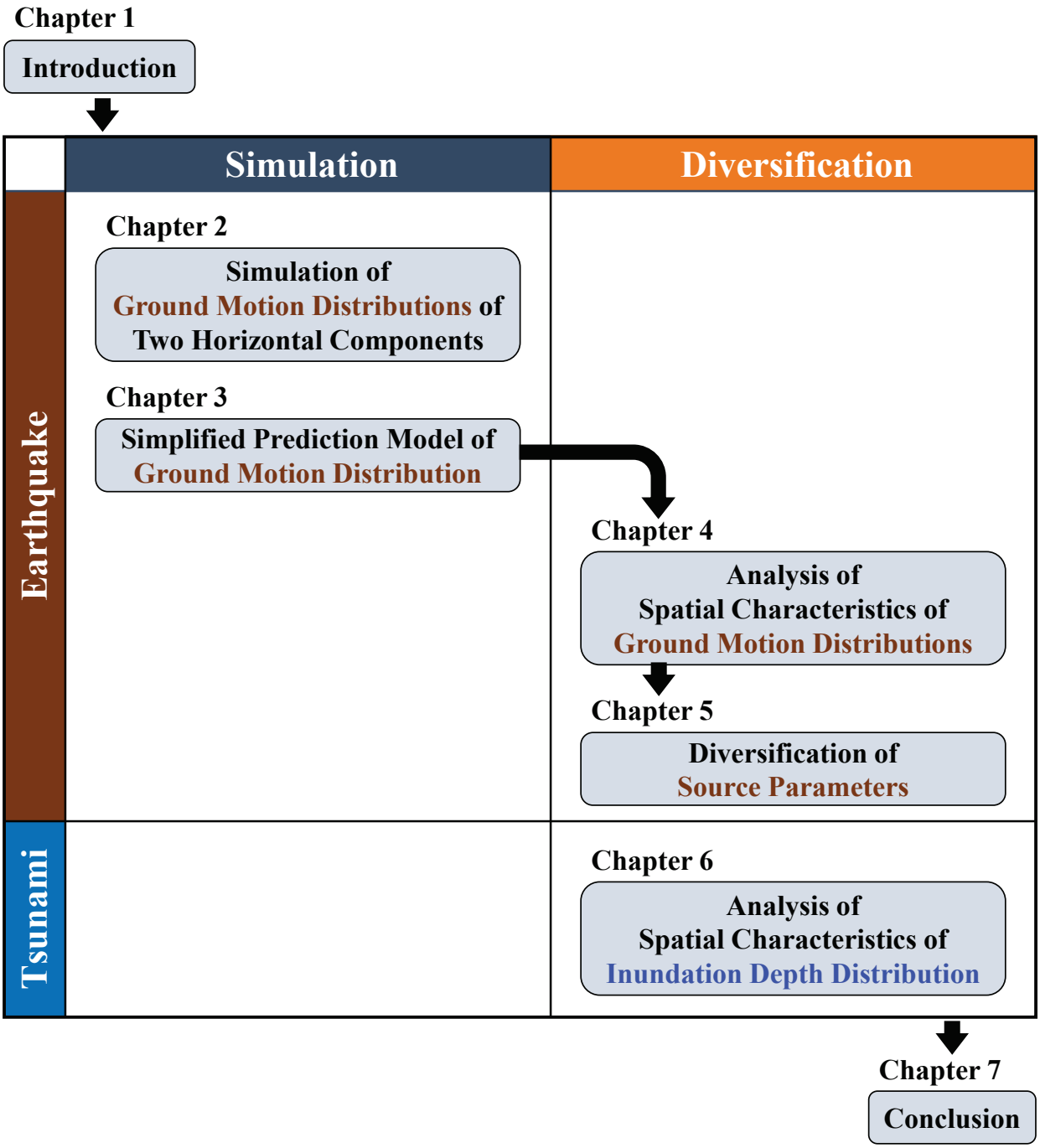
Chapter 6 analyzes the spatial characteristics of the tsunami as an example of the application of the proposed method. Specifically, the "decomposition" method is applied to the inundation depth distribution of tsunamis, and the spatial characteristics of the inundation depth are revealed with the parameters of the tsunami fault models. As a numerical example, **Chapter 6** uses tsunami inundation depth distributions for 11 cases caused by the Nankai megathrust earthquakes, which were set up and calculated by the Central

Disaster Management Council of the Cabinet Office, Government of Japan.

Chapter 7 summarizes the results of **Chapters 2–4** as the conclusion of this study. It also summarizes the issues of this study and describes future developments.

Each chapter in this dissertation has been edited and adapted from the following papers.

- **Chapter 2**: Takahashi, Nojima and Kagawa (2022) [54]
- **Chapter 4**: Takahashi, Nojima and Kagawa (2022) [55]
- **Chapter 6**: Takahashi and Nojima (2020) [56], Takahashi and Nojima (2021) [57]



- Earthquake : Ground motion distribution caused by a strike-slip fault
- Tsunami : Inundation depth distribution cased by the Nankai megathrust earthquake

Figure 1-3. Outline in this paper.

References in Chapter 1

[1] The National Research Institute for Earth Science and Disaster Resilience. https://www.bosai.go.jp/e/

[2] The Headquarters for Earthquake Research Promotion. https://www.jishin.go.jp/main/index-e.html

[3] The Earthquake Research Committee, the Headquarters for Earthquake Research Promotion. (2020) Strong ground motion prediction method for earthquakes with specified source faults ("Recipe"). https://www.jishin.go.jp/main/chousa/20_yosokuchizu/recipe.pdf **(in Japanese)**

[4] The National Research Institute for Earth Science and Disaster Resilience. "High sensitive seismograph network Japan." https://www.hinet.bosai.go.jp/?LANG=en

[5] The National Research Institute for Earth Science and Disaster Resilience. "Strong-motion seismograph networks (K-NET, KiK-net)." https://www.kyoshin.bosai.go.jp/

[6] The National Research Institute for Earth Science and Disaster Resilience. "Broadband seismograph network." https://www.fnet.bosai.go.jp/top.php?LANG=en

[7] The National Research Institute for Earth Science and Disaster Resilience. "Japan seismic hazard information station (J-SHIS)." https://www.j-shis.bosai.go.jp/en/

[8] The Headquarters for Earthquake Research Promotion. "National Seismic Hazard Maps for Japan." https://www.jishin.go.jp/evaluation/seismic_hazard_map/shm_report/shm_report_2020/

[9] The Japan Meteorological Agency. https://www.jma.go.jp/jma/indexe.html

[10] The United States Geological Survey. https://www.usgs.gov/

[11] The United States Geological Survey. "Earthquake Hazards." https://www.usgs.gov/programs/earthquake-hazards

[12] The United States Geological Survey. "ShakeMap." https://earthquake.usgs.gov/data/shakemap/

[13] Worden, C. B., Thompson, E. M., Hearne, M. G., and Wald, D. J. (2020) ShakeMap Manual Online: technical manual, user's guide, and software guide. U. S. Geological Survey. https://doi.org/10.5066/F7D21VPQ.

[14] The Swiss Seismological Service. http://seismo.ethz.ch/en/home/

[15] The Italy ShakeMap. http://shakemap.ingv.it/shake/index.html

[16] The National Institute of Geophysics and Volcanology. https://esse1-gis.mi.ingv.it/mps04_eng.jsp

[17] The Aotearoa New Zealand National Seismic Hazard Model. https://nshm.gns.cri.nz/

[18] The Global Earthquake Model Foundation. https://www.globalquakemodel.org/

[19] Irikura, K. (1994) Earthquake source modeling for strong motion prediction. *Zishin*, **46**: pp. 495–512. **(in Japanese with English abstract)** https://doi.org/10.4294/zisin1948.46.4_495

[20] Irikura, K., Kagawa, T., Kamae, K. and Sekiguchi, H. (1999) Recipe of predicting strong ground motions for future large earthquakes. *Abstr., JEPS joint meeting*, Session ID Sg-013. **(in Japanese)**

[21] Irikura, K. and Miyake, H. (2001) Prediction of strong ground motions for scenario earthquakes. *J. Geogr*, **110** (6): pp. 849–875. https://doi.org/10.5026/jgeography.110.6_849 **(in Japanese with English abstract)**

[22] Irikura, K., Miyake, H., Iwata, T., Kamae, K., Kawabe, H. and Dalguer, L. A. (2004) Recipe for predicting ground motions from future large earthquakes. *Proc. of 13WCEE*: Paper No. 1371.

[23] Irikura, K. (2004) Recipe for predicting strong ground motion from future large earthquake. *Annuals of Disas. Prev. Inst., Kyoto Univ.*, **47 A**: 21 p. **(in Japanese with English abstract)**

[24] Dan, K., Watanabe, M., Sato, T. and Ishii, T. (2001) Short-period source spectra inferred from variable-slip rupture models and modeling of earthquake faults for strong motion prediction by semi-empirical method. *J. Struct. Constr. Eng. Trans. AIJ*, **66** (545): pp. 51–62. https://doi.org/10.3130/aijs.66.51_4 **(in Japanese with English abstract)**

[25] Irikura, K. and Miyake, H. (2010) Recipe for predicting strong ground motion from crustal earthquake scenarios.

Pure Appl. Geophys., **168** (2011): pp. 85–104. https://doi.org/10.1007/s00024-010-0150-9

[26] The Earthquake Research Committee, the Headquarters for Earthquake Research Promotion. (2005) Part 2: Explanation of scenario earthquake shaking maps. https://jishin.go.jp/main/chousa/05mar_yosokuchizu/bunsatsu2.pdf **(in Japanese)**

[27] The Earthquake Research Committee, the Headquarters for Earthquake Research Promotion. (2008) Appendix 3: the strong ground motion estimation method (Recipe) for earthquakes with specified source faults. https://jishin.go.jp/main/chousa/08_yosokuchizu/a3.pdf **(in Japanese)**

[28] Lawrence Livermore National Laboratory. (2002) Guidance for performing probabilistic seismic hazard analysis for a nuclear plant site: example application to the Southeastern United States. NUREG/CR-6607, UCRL-ID-133494.

[29] Nozu, A. and Morikawa, H. (2003) Some characteristics of seismic hazard curve under the predominant influence of active fault. *Journal of JCOSSAR 2003*: pp. 659–664. **(in Japanese with English abstract)**

[30] Okumura, T., Sato, T., Ishii, T., Somerville, P., Graves, R. and Collins, N. (2005) Study on variability of earthquake ground motion due to variation of source parameters. *JSCE Journal of Earthquake Engineering*, **28**: Paper No. P167. https://doi.org/10.11532/proee2005a.28.167 **(in Japanese with English abstract)**

[31] Ripperger, J., Mai, P. M. and Ampuero, J.-P. (2008) Variability of near-field ground motion from dynamic earthquake rupture simulations. *Bull. Seism. Soc. Am.*, **98** (3): pp. 1207–1228. https://doi.org/10.1785/0120070076

[32] Yamada, M., Senna, S. and Fujiwara, H. (2007) Statistical analysis of predicted ground motions on the basis of a recipe for strong-motion prediction. *Journal of JAEE*, 7 (1): pp. 43–60. https://doi.org/10.5610/jaee.7.43 **(in Japanese with English abstract)**

[33] Yamada, M., Senna, S. and Fujiwara, H. (2007) statistical analysis of predicted ground motions on the basis of a recipe for strong-motion prediction for -dip-slip fault and strike slip fault-. *JSCE Journal of Earthquake Engineering*, **29**: pp. 104–113. https://doi.org/10.11532/proee2005a.29.104 **(in Japanese with English abstract)**

[34] Yamada, M., Senna, S. and Fujiwara, H. (2011) Statistical analysis of ground motions estimated on the basis of a recipe for strong-motion prediction: approach to quantitative evaluation of average and standard deviation of ground motion distribution. *Pure Appl. Geophys.*, **168**: pp. 141–153. https://doi.org/10.1007/s00024-010-0159-0

[35] Itoi, T. Midorikawa, S. Kito, J. Miura, H. Uchiyama, Y. and Sakamoto, S. (2009) Variability in response spectra for crustal earthquake using Stochastic Green's Function approach. *Journal of JAEE*, **9** (1): pp. 128–142. https://doi.org/10.5610/jaee.9.1_128 **(in Japanese with English abstract)**

[36] Kagawa, T. (2015) A study on spatial and temporal variabilities of strong ground motions based on multiple fault rupture scenarios. *Journal of JSCE A1*, **71** (4): pp. I_191–I_197. https://doi.org/10.2208/jscejseee.71.I_191 **(in Japanese with English abstract)**

[37] Kagawa, T. (2015) Spatial variability of periodic characteristics among strong ground motions derived from multiple fault rupture scenarios. *Journal of JAEE*, **15** (7): pp. 90–99. https://doi.org/10.5610/jaee.15.7_90 **(in Japanese with English abstract)**

[38] Iwaki, A., Morikawa, N., Maeda, T. and Fujiwara, H. (2017) Spatial distribution of ground-motion variability in broad-band ground-motion simulations. *Bull. Seism. Soc. Am.*, **107** (6): pp. 2963–2979. https://doi.org/10.1785/0120170150

[39] Hikita, T., Ikemura, T., Tomozawa, I. and Ishida, H. (2015) Study on prediction method of ground motion response spectrum and variability in prediction results based on fault models. *J. Struct. Constr. Eng. Trans. AIJ*, **80** (715): pp. 1403–1413. https://doi.org/10.3130/aijs.80.1403 **(in Japanese with English abstract)**

[40] Hikita, T., Koketsu, K. and Miyake, H. (2020) Variability of ground motion simulation due to aleatory uncertainty of source parameters. *Journal of JAEE*, **20** (3): pp. 21–34. https://doi.org/10.5610/jaee.20.3_21 **(in Japanese with English abstract)**

[41] Imtiaz, A., Causse, M., Chaljub, E. and Cotton, F. (2015) Is ground-motion variability distance dependent? Insight from finite-source rupture simulations. *Bull. Seism. Soc. Am.*, **105** (2A): pp. 950–962. https://doi.org/10.1785/0120140107

[42] Takada, T. and Shimomura, T. (2003) Macro-spatial correlation of seismic ground motion on strong motion records of the 1999 CHI-CHI earthquake. *J. Struct. Constr. Eng. Trans. AIJ*, **68** (565): pp. 41–48. https://doi.org/10.3130/aijs.68.41_1 **(in Japanese with English abstract)**

[43] Hayashi, T., Fukushima, S. and Yashiro, H. (2006) Effects of the spatial correlation between ground motion intensities on the seismic risk of portfolio of buildings. *J. Struct. Constr. Eng. Trans. AIJ*, **71** (600): pp. 203–210. https://doi.org/10.3130/aijs.71.203_1 **(in Japanese with English abstract)**

[44] Sokolov, V., Wenzel, F., Jean, W. Y. and Wen, K-L. (2010) Uncertainty and spatial correlation of earthquake ground motion in Taiwan. *Terr. Atmos. Ocean. Sci.*, **21** (6): pp. 905–921. https://doi.org/10.3319/TAO.2010.05.03.01(T)

[45] Abe, Y., Yamamoto, H., Nakamura, M., Akiyama, S. and Inoue, T. (2017) Effect of considering spatial correlation of ground motion intensity on seismic damage estimation. *Journal of JAEE*, **17** (2): pp. 74–87. https://doi.org/10.5610/jaee.17.2_74 **(in Japanese with English abstract)**

[46] Infantino, M. Paolucci, R., Smerzini, C. and Stupazzini, M. (2018) Study of the spatial correlation of earthquake ground motion by means of physics-based numerical scenarios. *Proc. of 16ECEE*.

[47] Nojima, N. Kuse, M. and Duc, L. Q. (2018) Mode decomposition and simulation of strong ground motion distribution using singular value decomposition. *Journal of JAEE*, **18** (2): pp. 95–114. https://doi.org/10.5610/jaee.18.2_95 **(in Japanese with English abstract)**

[48] Nojima, N. Kuse, M. and Takahashi, Y. (2019) Analysis of spatial variation and correlation of predicted seismic intensity maps for scenario earthquake by use of mode decomposition. *Journal of JAEE*, **19** (1): pp. 121–135. https://doi.org/10.5610/jaee.19.1_121 **(in Japanese with English abstract)**

[49] Nojima, N. and Takahashi, Y. (2020) Effect of source parameter settings on spatial variation and correlation of predicted seismic intensity maps for scenario earthquake. *Report of Tono RIES*, **45**: pp. 53–66. **(in Japanese with English abstract)**

[50] Yamazaki, F. and Shinozuka, M. (1990) Simulation of stochastic fields by statistical preconditioning. *Journal of Engineering Mechanics*, **116**, Issue 2: pp. 268-287. https://doi.org/10.1061/(ASCE)0733-9399(1990)116:2(268)

[51] Takahashi, Y. and Nojima, N. (2019) Analysis of spatial correlation in seismic intensity maps predicted for a scenario earthquake. *Journal of JCOSSAR 2019*: pp. 117–124. **(in Japanese with English abstract)**

[52] Yoshida, K., Nojima, N. and Takahashi, Y. (2020) Classification of characteristics of multiple predicted seismic intensity distributions for a scenario earthquake. *Journal of JSCE A1*, **76** (4): pp. I_420–I_429. https://doi.org/10.2208/jscejseee.76.4_I_420 **(in Japanese with English abstract)**

[53] Takahashi, Y. and Nojima, N. (2022) Approximation representation of observed strong ground motion distribution using mode synthesis of predicted distributions. *Proc. of ISRERM 2022*, Paper ID GS-03-101: pp. 771–778. https://doi.org/10.3850/978-981-18-5184-1_GS-03-101-cd

[54] Takahashi, Y., Nojima, N. and Kagawa, T. (2022) Simulation of strong ground motion distributions of horizontal two components based on singular value decomposition analysis. *Journal of JSCE A1*, **78** (4): pp. I_445–I_458. https://doi.org/10.2208/jscejseee.78.4_I_445 **(in Japanese with English abstract)**

[55] Takahashi, Y., Nojima, N. and Kagawa, T. (2022) Analysis of period and component dependent spatial characteristics of strong ground motion distributions using mode decomposition and machine learning. *Journal*

of JSCE A1, **78** (4): pp. I_487–I_493. https://doi.org/10.2208/jscejseee.78.4_I_478 **(in Japanese with English abstract)**

[56] Takahashi, Y. and Nojima, N. (2020) Evaluating spatial correlation and simulation of tsunami inundation depth distribution caused by the Nankai megathrust earthquakes. *Journal of ISSS*, **36**: pp. 75–82. https://doi.org/10.11314/jisss.36.75 **(in Japanese with English abstract)**

[57] Takahashi, Y. and Nojima, N. (2021) Non-negative matrix factorization of spatial distribution of tsunami inundation depth caused by the Nankai megathrust earthquakes. *Journal of ISSS*, **38**: pp. 163–172. https://doi.org/10.11314/jisss.38.163 **(in Japanese with English abstract)**

2 Simulation of Strong Ground Motion Distributions of Two Horizontal Components Base on Singular Value Decomposition Analysis

2.1 Introduction

In ground motion prediction for an earthquake with specified source faults (setting of characterized source model and strong ground motion simulation based on a detailed method), the heterogeneity on the fault plane and the direction of rupture propagation is taken into account by setting source parameters such as magnitude and location of asperities and hypocenter. As a result, the ground motion distribution can greatly vary depending on the fault rupture scenario. In earthquake risk assessment, therefore, an integrated assessment based on various ground motion distributions for a large number of fault rupture scenarios is preferable to an assessment based on limited ground motion distributions. However, the high computational cost of the detailed method of ground motion prediction makes it impractical to set up a variety of fault rupture scenarios and to predict their ground motion distributions.

The previous study [1] proposed a method of simulating ground motion distributions by applying mode decomposition and synthesis to several ground motion distributions. The simulation method randomly generates a large number of ground motion distributions within the spatial structure of ground motion variability indicated by the original data. In the previous study [1], 192 and 960 ground motion distributions were newly generated from 12 cases; the results were applied to seismic damage prediction to clarify the usefulness of its simulation method.

The simulation method in the previous study [1] is designed for a single (or one-component) ground motion index (e.g., instrumental seismic intensity). In other words, when this simulation method is applied to ground motion distributions of two horizontal components (Fault-Normal: FN, Fault-Parallel: FP), the relationship between two components within the same case is ignored, simulated ground motion distributions are generated independently for each component. Therefore, when this method is applied to simulate the distribution of two components, it is necessary to consider the reciprocal spatial variability of both components.

As an example of dealing with spatial variability of different physical quantities, the spatial interrelationship between atmospheric circulation field and surface air (or sea surface) temperature has been analyzed in the meteorological field [2]; singular value decomposition analysis [2–5] has been adopted as an analysis method. Singular value decomposition analysis is a method to extract the spatial variation structure that is dominant in two physical quantities from the spatial distribution of both quantities.

Therefore, in this chapter, singular value decomposition analysis is applied to the ground motion distribution of two horizontal components, homogeneous as physical quantities, to extract the spatial structure of the mutual variation of the two components. Then, using the results of singular value decomposition analysis, the simulation method for the single-component ground motion distributions proposed in the previous study [1] is extended to the two-component distributions.

This chapter presents a numerical example using the distribution of the absolute acceleration response spectrum (period 0.1-2.0 s, two horizontal components (FN, FP)) in a strike-slip fault as set up and calculated in the references [6, 7].

2.2 outlines the application of singular value decomposition analysis to ground motion distributions. ***2.3*** describes the ground motion distribution used as a numerical example, and ***2.4*** shows the results of singular value decomposition analysis. ***2.5*** describes a simulation method for ground motion distributions of two horizontal components, and ***2.6*** presents a numerical simulation example. Finally, ***2.7*** summarizes the conclusions obtained in this chapter.

2.2 Singular Value Decomposition Analysis

2.2.1 Singular Value Decomposition and Singular Value Decomposition Analysis

The previous study [1] used singular value decomposition (SVD) [8] to analyze the spatial characteristics in single-component ground motion distributions. In contrast, this chapter uses singular value decomposition analysis (SVD analysis) [3–5] to analyze the interrelationships in the spatial variation of the two components. While SVD is a method to extract dominant spatial characteristics from the variance-covariance structure in a single component of the ground motion distributions, SVD analysis is a method to extract the spatial characteristics that are strongly correlated with each other from the cross-covariance structure in two horizontal components of the ground motion distributions. The following shows how SVD analysis is applied to ground motion distributions.

2.2.2 SVD Analysis of Ground Motion Distributions

Let x and y denote variables representing the common logarithmic values of absolute acceleration response spectrum (S_A) of FN and FP components, respectively; the spatial distribution of S_A (M: total number of computation points in the target area) is represented by M-dimensional feature vectors $\mathbf{x} = (x_1, \ldots, x_M)^T$, $\mathbf{y} = (y_1, \ldots, y_M)^T$, respectively. Thus, S_A distributions for N cases ($M \gg N$) are described by the $M \times N$ matrices $\mathbf{X}$ and $\mathbf{Y}$ (the row i – column j entry is the S_A for Case j at point i) as follows:

$$\mathbf{X} = \left(\mathbf{x}_1, \ldots, \mathbf{x}_N\right) = \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{M1} & \cdots & x_{MN} \end{pmatrix} \tag{2-1}$$

$$\mathbf{Y} = \left(\mathbf{y}_1, \ldots, \mathbf{y}_N\right) = \begin{pmatrix} y_{11} & \cdots & y_{1N} \\ \vdots & \ddots & \vdots \\ y_{M1} & \cdots & y_{MN} \end{pmatrix} \tag{2-2}$$

Here, the ranks of $\mathbf{X}$ and $\mathbf{Y}$ are both rank ($\mathbf{X}$) = rank ($\mathbf{Y}$) = N. $\mathbf{X}$ and $\mathbf{Y}$ are normalized by using the mean and standard deviation of all entries of each matrix, respectively. The normalized matrices are expressed as follows:

$$\mathbf{X_0} = \frac{\mathbf{X} - \mu_{\mathbf{X}}\mathbf{H}}{\sigma_{\mathbf{X}}} \tag{2-3}$$

$$\mathbf{Y_0} = \frac{\mathbf{Y} - \mu_{\mathbf{Y}}\mathbf{H}}{\sigma_{\mathbf{Y}}} \tag{2-4}$$

where $\mu_{\mathbf{X}}$, $\mu_{\mathbf{Y}}$ and $\sigma_{\mathbf{X}}$, $\sigma_{\mathbf{Y}}$ are the mean and standard deviation of all entries of $\mathbf{X}$, $\mathbf{Y}$, respectively. $\mathbf{H}$ is the $M \times N$ matrix in which all entries are 1.

Let $\mathbf{C_{XY}}$ denote the $M \times N$ cross-covariance matrix with $\mathbf{X_0}$ and $\mathbf{Y_0}$.

$$\mathbf{C_{XY}} = \mathbf{X_0}\mathbf{Y_0}^T \tag{2-5}$$

Here, the rank of $\mathbf{C_{XY}}$ is rank ($\mathbf{C_{XY}}$) = N. In the previous study targeting single-component ground motion [1], SVD was applied to the variance-covariance matrix $\mathbf{C_{XX}}$ (or $\mathbf{C_{YY}}$) with $\mathbf{X}$ (or $\mathbf{Y}$) to extract the spatial structure of a single component. In this chapter, to extract the reciprocal spatial variation structure of two horizontal components, SVD analysis is applied to the cross-covariance matrix $\mathbf{C_{XY}}$, and mode decomposition is performed on the product of the three matrices. SVD analysis is a process of applying SVD to the cross-covariance matrix.

$$\begin{aligned}\mathbf{C_{XY}} &= \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}} \\ &= \begin{pmatrix} u_{11} & \cdots & u_{1N} \\ \vdots & \ddots & \vdots \\ u_{M1} & \cdots & u_{MN} \end{pmatrix} \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_N \end{pmatrix} \begin{pmatrix} v_{11} & \cdots & v_{1N} \\ \vdots & \ddots & \vdots \\ v_{M1} & \cdots & v_{MN} \end{pmatrix}^{\mathrm{T}}\end{aligned} \tag{2-6}$$

$$\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I} \tag{2-7}$$

The matrices **U**, **D**, **V,** and **I** in Eq. (2-6) and Eq. (2-7) have the following meanings.

- The matrices **U**, **V**: the $M \times N$ matrices composed of orthonormal bases stretching an N-dimensional subspace out of an M-dimensional space. The column vectors of the k-th column ($1 \le k \le N$), $\mathbf{u}_k = (u_{1k}, \ldots, u_{Mk})^{\mathrm{T}}$ and $\mathbf{v}_k = (v_{1k}, \ldots, v_{Mk})^{\mathrm{T}}$, are the singular vectors in Mode k of $\mathbf{X_0}$ and $\mathbf{Y_0}$, respectively. The singular vector represents the mode form that defines the spatial distribution and spatial correlation of the features at the computation points; its map display allows visualization of the spatial correlation between the computation points.
- The matrix **D**: The $N \times N$ diagonal matrix. The diagonal entry d_k is the singular value, equivalent to the square root of the eigenvalue, and represents the weight of Mode k.
- The matrix **I**: The $N \times N$ identity matrix.

In Eq. (2-6), **U**, **D**, and **V** are originally $M \times M$ matrices, but since rank ($\mathbf{C_{XY}}$) = N, there are N nonzero singular values, and the part concerning zero singular values (N+1 – M columns for **U** and **V**, N+1 – M rows and columns for **D**) does not affect the results. For this reason, **U**, **D**, and **V** are expressed as $M \times N$ matrices.

The principal component scores (PCSs) of the matrices $\mathbf{X_0}$ and $\mathbf{Y_0}$ corresponding to the matrices **U** and **V** in Eq. (2-6) are defined as the matrices **P** and **Q**, respectively; Eq. (2-10) follows from the relationship between **P** and **Q**.

$$\mathbf{P} = \mathbf{U}^{\mathrm{T}}\mathbf{X_0} \tag{2-8}$$

$$\mathbf{Q} = \mathbf{V}^{\mathrm{T}}\mathbf{Y_0} \tag{2-9}$$

$$\mathbf{P}\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}\mathbf{P}^{\mathrm{T}} = \mathbf{D} \tag{2-10}$$

P and **Q** are $N \times N$ matrices in Eqs. (2-8)–(2-10). The row vectors of the k-th column, $\mathbf{p}_{(k)} = (p_{1k}, \ldots, p_{Nk})$, $\mathbf{q}_{(k)} = (q_{1k}, \ldots, q_{Nk})$, represent the PCSs for each case in Mode k for $\mathbf{X_0}$, $\mathbf{Y_0}$, respectively. Let l ($1 \le l \le N$) denote the modal number; then the cross-covariance structure of the PCSs with two horizontal components, Eq. (2-10), can be described as follows:

$$\mathbf{p}_{(k)}\mathbf{q}_{(l)}^{\mathrm{T}} = \mathbf{q}_{(k)}\mathbf{p}_{(l)}^{\mathrm{T}} = \begin{cases} d_k & (k = l) \\ 0 & (k \ne l) \end{cases} \tag{2-11}$$

For $k \neq l$, The PCSs between different components are orthogonal. In contrast, the PCSs between same components ($\mathbf{p}_{(k)}$-$\mathbf{p}_{(l)}$, or $\mathbf{q}_{(k)}$-$\mathbf{q}_{(l)}$) are not orthogonal to each other.

From Eqs. (2-6), (2-8) and (2-9), $\mathbf{X_0}$ and $\mathbf{Y_0}$ can be described using **U**-**P**, and **V**-**Q**, respectively.

$$\mathbf{X_0} = \mathbf{UP} \tag{2-12}$$

$$\mathbf{Y_0} = \mathbf{VQ} \tag{2-13}$$

In the above calculations, SVD is applied to the cross-covariance matrix $\mathbf{C_{XY}}$ in Eq. (2-6); however, since $\mathbf{C_{XY}}$ is a huge matrix with M-rows and M-columns, it is impractical to compute it directly. Therefore, in actual calculations, the matrices **U**, **D**, **V**, **P**, and **Q** are obtained using a method called dual formalism [4]; SVD analysis based on dual formalism is described in ***Appendix I*** at the end of **Chapter 2**.

2.2.3 Singular Values and Mode CFoefficients for Each Component

By applying SVD to the $M \times N$ matrix $\mathbf{X_0}$ (or $\mathbf{Y_0}$), the mode shape ($M \times N$ matrix), singular values ($N \times N$ matrix), and mode coefficients ($N \times N$ matrix) are obtained [1]. Each column of the mode coefficients represents the case-dependent weight for each mode. The PCSs are obtained by multiplying them by the corresponding singular values. Similarly, the PCSs **P** and **Q** of SVD analysis are decomposed into singular values $\mathbf{D_P}$ and $\mathbf{D_Q}$ corresponding to $\mathbf{X_0}$ and $\mathbf{Y_0}$, and mode coefficients $\mathbf{P_0}$ and $\mathbf{Q_0}$, respectively.

$$\mathbf{P} = \mathbf{D_P P_0} = \begin{pmatrix} d_{p1} & & 0 \\ & \ddots & \\ 0 & & d_{pN} \end{pmatrix} \begin{pmatrix} p_{011} & \cdots & p_{01N} \\ \vdots & \ddots & \vdots \\ p_{0N1} & \cdots & p_{0NN} \end{pmatrix} \tag{2-14}$$

$$\mathbf{Q} = \mathbf{D_Q Q_0} = \begin{pmatrix} d_{q1} & & 0 \\ & \ddots & \\ 0 & & d_{qN} \end{pmatrix} \begin{pmatrix} q_{011} & \cdots & q_{01N} \\ \vdots & \ddots & \vdots \\ q_{0N1} & \cdots & q_{0NN} \end{pmatrix} \tag{2-15}$$

The mode coefficients of SVD are constructed on an orthonormal basis. In contrast, since PCSs between the same components are not orthogonal in SVD analysis, $\mathbf{P_0}$ and $\mathbf{Q_0}$ in Eqs. (2-14) and (2-15) are also composed of unit vectors (row vectors) that do not satisfy the orthogonality condition. The mean of the row vectors in $\mathbf{P_0}$ and $\mathbf{Q_0}$ is $N^{-1/2}$ for Mode 1 and 0 for Mode 2 and higher. The standard deviation for Mode 1 is data set dependent and is $N^{-1/2}$ for Mode 2 and higher.

2.3 Characterized Source Model and Ground Motion Distribution

This study targets and discusses the strike-slip fault and the ground motion distribution caused by its fault, which was set up and calculated by Kagawa [6, 7]. The following is a description of the using dataset.

2.3.1 Characterized Source Model

Figure. 2-1 shows the target strike-slip and the computational area of strong ground motion simulation. The geometry and location of the strike-slip fault are as follows.

- Length : 26 km
- Width : 16 km
- Depth of the fault top : 3 km
- Strike : 0 °
- Dip : 90 °
- Location : center in the computational area

The moment magnitude M_W of this fault, evaluated from the fault area based on the recipe for predicting strong ground motions [9, 10], is slightly lower than 6.6.

In the papers [6, 7], 300 cases of characterized source models with different source parameters are set up for this fault. The conditions for setting the source parameters are described below.

- Fixed parameter
  - Fault geometry (fault area) : 26 km × 16 km
  - Number of asperities : 2 (The large and small asperities with different areas)

- Variated Parameters:

  Set by taking the standard value for the recipe for predicting strong ground motions [9, 10] as the mean, and randomly giving a variation assuming a normal distribution [11, 12]. The following values are means and coefficients of variation.
  - Seismic moment M_0 : 9.63×10^{25} dyne・cm, 5%
  - Rupture velocity : 2.52 km/s, 5%
  - Asperities area ratio* : 0.22, 10%

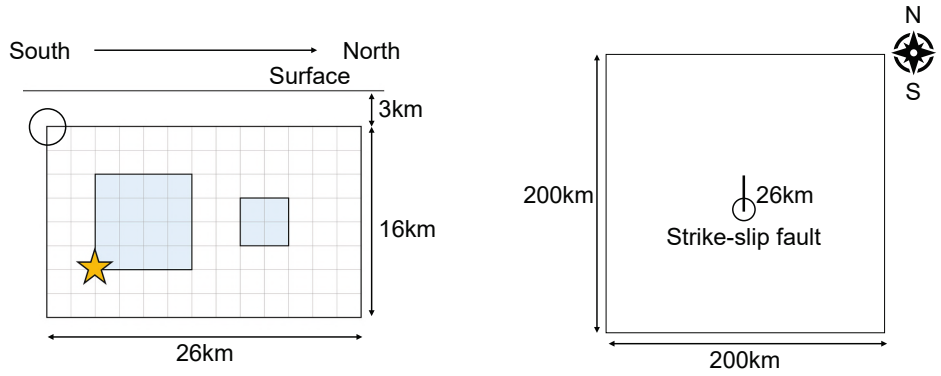
  (*Asperities are ratio: the ratio between the combined asperity area and total rupture area)

- Other variated parameters for the location of asperity and hypocenter
  - Location of asperities : set up by uniform distribution in both strike and depth on the fault plane with no overlap of each asperity
  - Location of hypocenter : set at either edge of the bottom of the large asperity

This study analyzes a total of 600 cases, including 300 cases in which the fault model is inverted around the perpendicular bisector of the fault model in consideration of the symmetry of the data. The inclusion of the inversion case has the merit of stabilizing the mode decomposition results in the next section since the

distribution of ground motion variation becomes line-symmetric across the perpendicular bisector of the fault, as shown in Fig. 2-4 and Fig. 2-5 below.

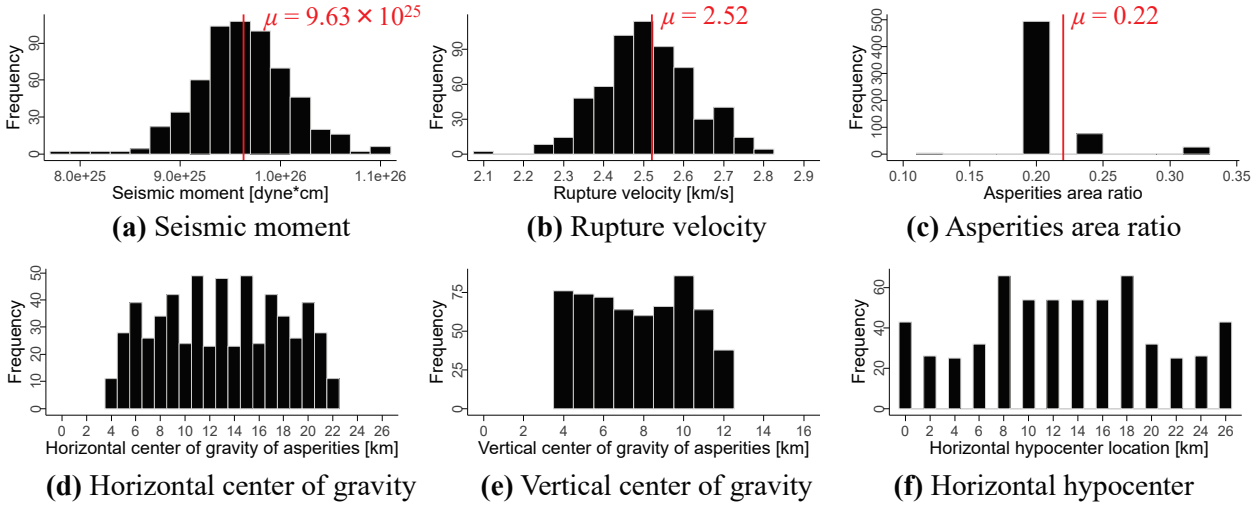
Figure. 2-2 shows the frequency distribution of the source parameters for the 600 cases. The asperity area ratios in Fig. 2-2 (c) were rounded to fit the 2 km square mesh of the fault model after giving the variation. Figs. 2-2 (d), (e) are weighted averages of the seismic moment ($\propto$ area to the power of 1.5) for the horizontal and vertical positions (distance from the southern corner of the fault top) of large and small asperities.



(a) Image of fault model **(b)** Computational area

(Blue rectangle: asperity, yellow star: hypocenter, circle: reference point)

Figure 2-1. Target strike-slip fault.



(a) Seismic moment **(b)** Rupture velocity **(c)** Asperities area ratio

(d) Horizontal center of gravity of asperities **(e)** Vertical center of gravity of asperities **(f)** Horizontal hypocenter location

Figure 2-2. Frequency distribution of source parameters (μ represents mean value).

2.3.2 Strong Ground Motion Simulation and Calculation Condition

The computational area (Fig. 2-1 (b)) is 200 km × 200 km, and the computational points are set every 2 km × 2 km for a total of 10,201 points (= 101 points × 101 points). The ground structure is assumed to be a typical sedimentary plain or basin. As shown in Table. 2-1 and Fig. 2-3, the ground model is a horizontally stratified ground model with the engineering bedrock (V_S = 350 m/s) as the ground surface, and the site amplification characteristics are the same at all computational points.

Under the above conditions, three-component waveforms (Fault-Normal: FN, Fault-Parallel: FP, Vertical: UD) were calculated at each computation point by strong ground motion (SGM) simulation based on the stochastic Green's function method [13]. The element waveforms at each point used in SGM simulation are different for each of the asperities and background areas, but they are all the same within the same area; the amplitude spectra are given their respective distance attenuation and azimuthal radiation characteristics. This treatment follows the empirical Green's function method and does not give different random element waveforms for each computation point and element fault. The waveforms in the short-period band at each point have no azimuthal characteristics [7], and their response spectra have been confirmed to be consistent with the attenuation relationship of the response spectrum [14]. In the long-period band, azimuthal characteristics reflecting the directivity effect and the radiation pattern appear [7, 14] and are consistent with the theoretical waveforms obtained by the discrete wavenumber method [13].

The stochastic Green's function method [13] applies to waveforms with periods 0.05–20 s. However, using a theoretical wave propagation calculation method for longer periods is preferable. Therefore, this study targets ground motions with periods up to 2.0 s.

Table 2-1. Horizontally stratified ground structure
(ρ: density, $Q(f) = Q_0 \times f^{Qf}$: attenuation) (Data source: [6])

Parameters					
Vp (km/s)	Vs (km/s)	ρ (g/cm^3)	Q_0	Q_f	Depth (km)
1.60	0.35	1.70	35	0.5	0.000
1.80	0.55	1.80	55	0.5	0.092
2.50	1.00	2.10	100	0.5	0.599
5.40	3.20	2.70	63.8	1.0	1.101
6.00	3.50	2.80	63.8	1.0	2.000

Depth of Layer Top

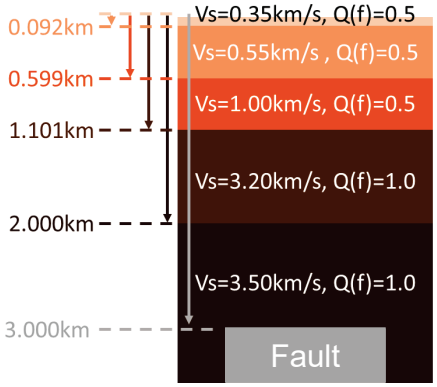


Figure 2-3. Ground structures model and fault.

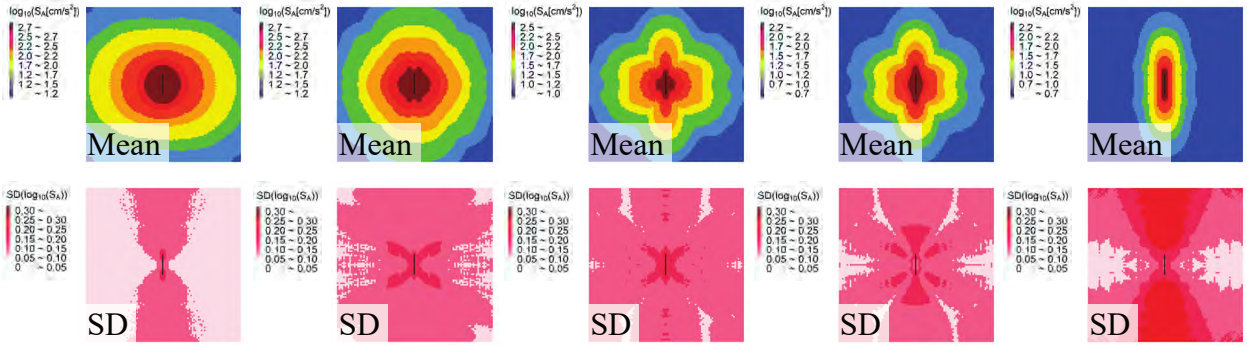
2.3.3 Distribution and Correlation of Ground Motions of Horizontal Two Components

The ground motion intensities used in this study are absolute acceleration response spectrum (S_A) for 5 % damping of two horizontal components (FN, FP) with periods 0.1–2.0 s (0.1, 0.5, 0.75, 1.0, and 2.0 s) calculated for the strike-slip fault.

Figure. 2-4 and Fig. 2-5 show the distribution of the common logarithmic mean and common logarithmic standard deviation of the S_A for the 600 cases. The mean at period 0.1 s of both components shows a distance-decaying distribution trend centered on the fault [7]. At period 2.0 s, the mean of FN component has large values in the fault strike (north and south) direction due to the forward directivity effect and the radiation pattern, while the mean of FP component has large values in the orthogonal (east and west) direction. The distributions of period 0.5–1.0 s tend to be intermediate between these two periods; the distribution trend changes from a distance-decaying distribution to an azimuthal distribution as short-period to long-period. A similar tendency is observed in the standard deviation, and the variation tends to be larger for longer periods.

Figure. 2-6 shows the spatial distribution of the correlation coefficient between S_A of FN and FP components (top of Fig. 2-4 and Fig. 2-5). In common with all periods, the S_A of FN and FP components are positively correlated in almost the entire map. The correlation tends to be weaker in the fault strike direction and its orthogonal direction than in the surrounding area. This trend may be because the S_A of FN and FP components are orthogonal.

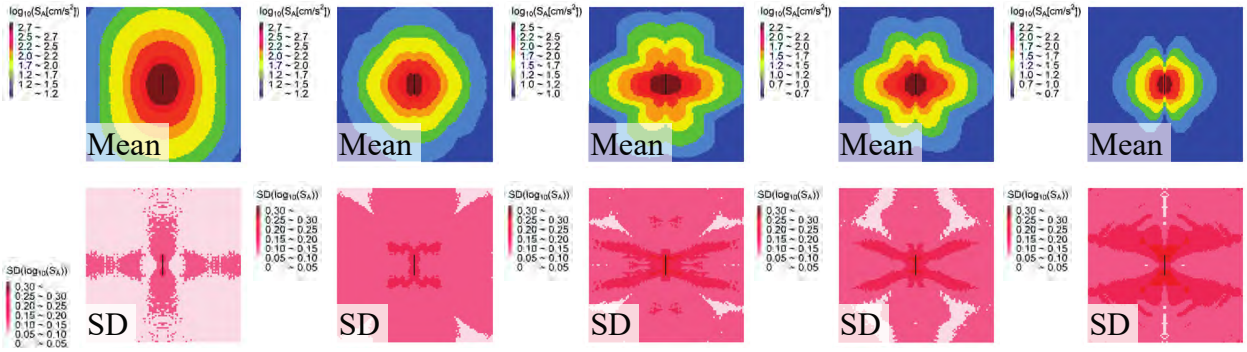
The correlation coefficients between components in each case (600 in total) are 0.77 – 0.94 for period 0.1 s, 0.94–0.98 for period 0.5 s, 0.85–0.95 for period 0.75 s, 0.86–0.95 for period 1.0 S, and 0.66–0.86 for period 2.0 s, with the lowest for period 2.0 s. These results are because the impact of the radiation pattern and forward directivity effect becomes stronger for longer periods, suggesting that the same case can clearly show differences among components.



(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

(Top: Spatial distribution of the common logarithmic mean of six hundred cases.
Bottom: Spatial distribution of the common logarithmic standard deviation of six hundred cases.)

Figure 2-4. Absolute acceleration response spectrum (S_A) for 5 % damping of FN component.



(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

(Top: Spatial distribution of the common logarithmic mean of six hundred cases.
Bottom: Spatial distribution of the common logarithmic standard deviation of six hundred cases.)

Figure 2-5. Absolute acceleration response spectrum (S_A) for 5 % damping of FP component.

(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

Figure 2-6. Spatial distribution of correlation coefficient between S_A of FN and FP components.

2.4 SVD Analysis of Distributions of S_A

2.4.1 Squared Covariance Fraction of Mode

SVD analysis is applied to the distributions of S_A (FN and FP components) for 600 cases by period. Fig. 2-7 shows the cumulative squared covariance fraction (cumulative SCF) of eigenvalues (square of the singular value, which is the diagonal entries of the matrix **D**), and Fig. 2-8 shows the SCF from the 2nd-order mode to the 10th-order mode. In common for all periods, the SCFs of Mode 1 are significantly high, showing more than 99.9 %. Modes 2–3 show constant SCFs, while Mode 4 and higher show almost 0 %. The SCF of Mode 1 at period 2.0 s is lower than that at other periods (0.1–1.0 s), while Modes 2–3 are higher than in other periods. The reason for this is discussed in the next paragraph.

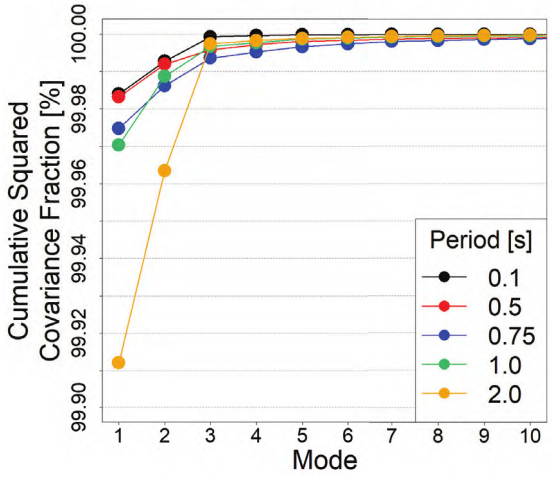


Figure 2-7. Cumulative squared covariance fraction.

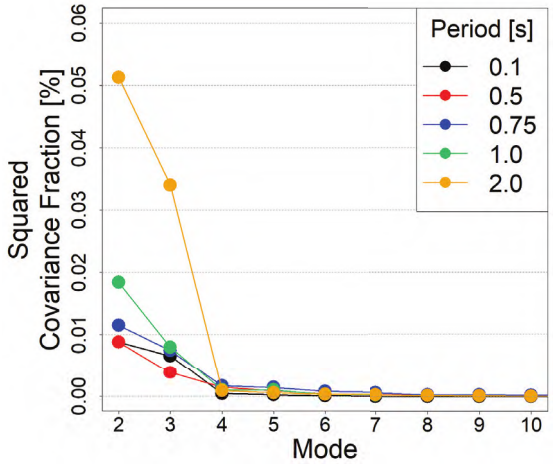


Figure 2-8. Squared covariance fraction.

2.4.2 Mode Shape

Figures. 2-9 (a), (b)–Figs. 2-11 (a), (b) show the mode shape in FN and FP components (only Modes 1–3 at period 0.5, 1.0, 2.0 s are shown due to the limitation of space). The mode shape represents the spatial characteristics of ground motions with strong interrelationships between the two components. A red point is a positive value, and a blue point is a negative value; a correlation between points with the same (or different) sign is a positive (or negative) correlation. The mode shapes of the two components at same-period and same mode show similar distribution trends; this indicates that the mutual spatial variation structure of two horizontal components can be extracted from the ground motion distribution of two components.

For Mode 1, the distributions are simple, with positive values near the fault and negative values farther away, corresponding to the mean distribution (top of Fig. 2-4 and Fig. 2-5). Thus, as for the spatial

interrelationship between the two horizontal components, the most dominant spatial variability structure (mode form) is the spatial distribution of the mean ground motion (attenuation effect). The cross-covariance matrix with two components is 99.9 % represented by this spatial structure.

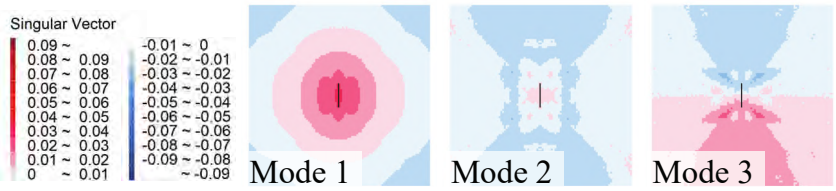
For Mode 2 and Mode 3, the spatial distribution can be classified into two categories: “spatial variation with positive correlation in an almost entire area (e.g., Fig. 2-9: Mode 2)” and “spatial variation negatively correlated north and south of the map (e.g., Fig. 2-9: Mode 3).” The former represents a uniform variation of ground motion across the entire area. In addition, The FN component has a larger absolute value in the fault strike (north-south) direction, and the FP component has a larger absolute value in the orthogonal (east-west) direction, which is considered to be related to the bilateral rupture propagation of the fault. The latter emphasizes the difference in seismic intensity between the north and south of the map, i.e., the contrast in spatial seismic intensity, which may be related to the unilateral rupture propagation of the fault.

In ***2.4.2***, it was pointed out that the SCFs of Modes 2–3 at period 2.0 s are higher than those of other periods. This result may be because the directivity effect represented by Modes 2–3 is stronger for longer periods.

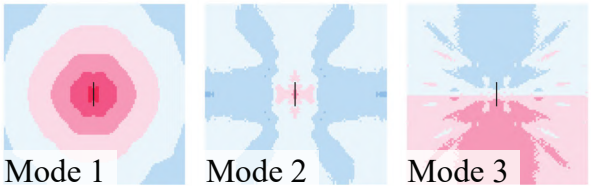
2.4.3 Principal Component Score of Mode

Figures. 2-9 (c)– Fig. 2-11 (c) show the scatter plot of the PCSs between FN and FP components. The mean PCSs of Mode1 are 96.7–98.6, and the mean mode coefficients are 0.0408 ($= N^{-1/2} = 600^{-1/2}$). The standard deviations of the mode coefficient are 0.00075–0.0026, and the coefficient of variation are 1.8 – 6.4%, with the coefficient being larger for longer periods. The mean of the PCSs and mode coefficients for Mode 2 and higher are 0; the standard deviations of the mode coefficients are 0.0408 ($= N^{-1/2}$).

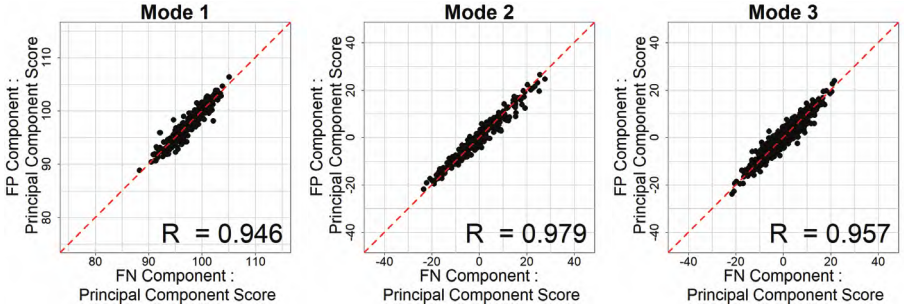
Table. 2-2 shows the two horizontal components’ correlation coefficients of the PCSs (mode coefficients). The correlation coefficients for Modes 1–3 are all greater than 0.8, indicating that the two components are strongly correlated.



(a) Mode shape in FN component.

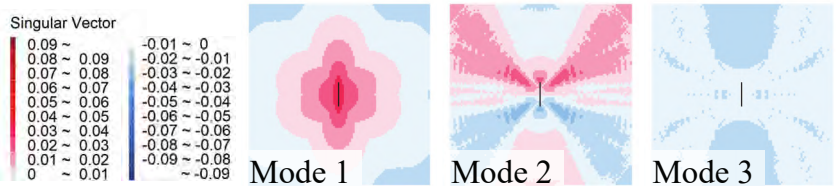


(b) Mode shape in FP component.

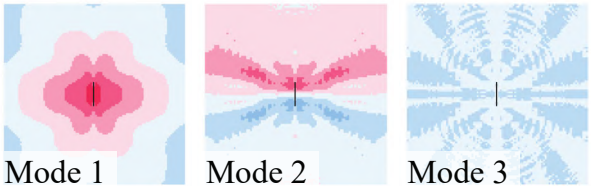


(c) Scatter plot between PCS of FN (horizontal axis) and FP (vertical axis). (R: correlation coefficient)

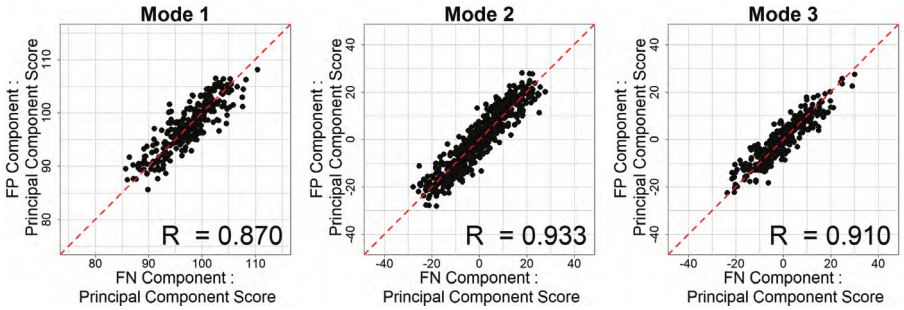
Figure 2-9. Result of SVD analysis in period 0.5 s.



(a) Mode shape in FN component.

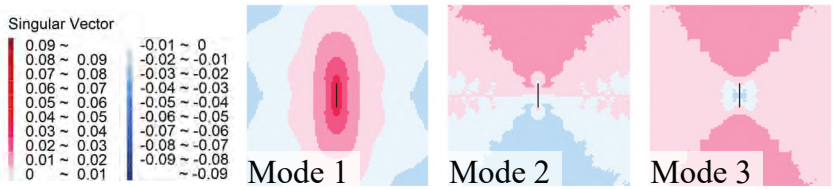


(b) Mode shape in FP component.

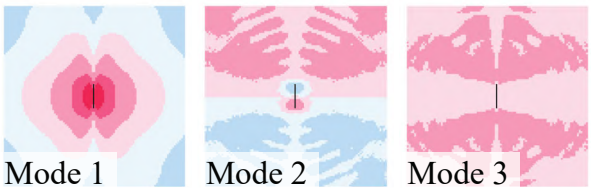


(c) Scatter plot between PCS of FN (horizontal axis) and FP (vertical axis). (R: correlation coefficient)

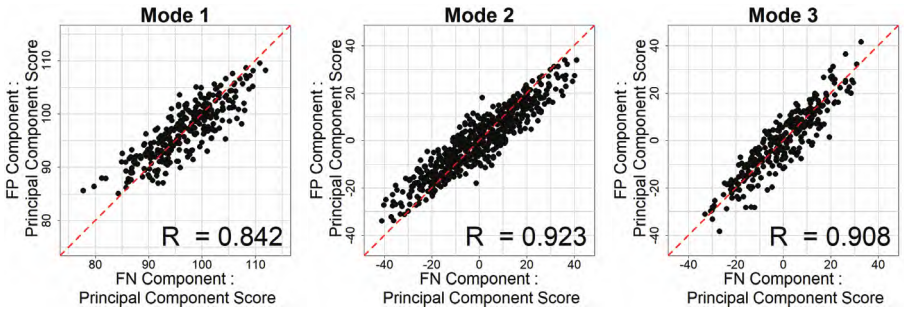
Figure 2-10. Result of SVD analysis in period 1.0 s.



(a) Mode shape in FN component.



(b) Mode shape in FP component.



(c) Scatter plot between PCS of FN (horizontal axis) and FP (vertical axis). (R: correlation coefficient)
Figure 2-11. Result of SVD analysis in period 2.0 s.

Table 2-2. The correlation coefficient of PCSs of two horizontal components.
(Red: correlation coefficient is 0.9 – 1.0, blue: correlation coefficient is 0.8 – 0.9)

Period [s]	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7	Mode 8	Mode 9	Mode 10
0.1	**0.806**	**0.950**	**0.965**	**0.823**	**0.937**	0.796	**0.838**	**0.864**	**0.955**	**0.946**
0.5	**0.946**	**0.979**	**0.957**	**0.960**	**0.955**	**0.908**	**0.900**	**0.916**	**0.943**	**0.900**
0.75	**0.906**	**0.966**	**0.943**	**0.941**	**0.904**	**0.832**	**0.945**	**0.832**	**0.910**	**0.918**
1.0	**0.870**	**0.933**	**0.910**	0.728	**0.926**	0.755	0.765	0.704	**0.848**	**0.840**
2.0	**0.842**	**0.923**	**0.908**	0.775	**0.857**	0.687	0.756	0.775	0.724	0.679

2.5 Simulation Method of Ground Motion Distributions Considering Interrelationship between Horizontal Two Components

2.5.1 Simulation Conditions for Ground Motion Distributions of Horizontal Two Components

The previous study [1] proposed a method of simulating ground motion distributions for a single component based on mode decomposition and synthesis, which combines SVD and an efficient Monte Carlo method [15]. This section extends this simulation method to simultaneously generate ground motion distributions of the two components while maintaining the spatial correlation structure of ground motion between the two components. The following conditions must be satisfied by the ground motion distribution simulation of the two horizontal components based on the SVD analysis proposed in this section.

- Condition I: For reproducing the relationship between cases within the same component

  The mean of mode coefficients and the coefficient of variation or standard deviation of the mode coefficient must be consistent with the following values.

  - Mode 1 : mean $N^{-1/2}$, coefficient of variation δ
  - Modes 2–N (= 600) : mean 0, standard deviation $N^{-1/2}$

- Condition II: For reproducing the cross-covariance structure and correlation coefficients of the two horizontal components

  The following equation must hold.

  - Eq. (2-10) : $\mathbf{PQ}^{\mathrm{T}} = \mathbf{QP}^{\mathrm{T}} = \mathbf{D}$
  - Eq. (2-31) in the later : $r_{k\,(\geq 2)} = \mathbf{p}_{\mathbf{0}\,(k)}\,\mathbf{q}_{\mathbf{0}\,(k)}^{\mathrm{T}}$

2.5.2 Simulation of Ground Motion Distributions of Horizontal Two Components Base on Singular Value Decomposition Analysis and Mode Synthesis

Let S be the number of simulation trials, and replace the $N \times N$ matrix $\mathbf{P}$ representing the PCS of the FN component in Eq. (2-12) and Eq. (2-14) with the $N \times S$ matrix $\mathbf{T}$; define the $M \times S$ matrix $\mathbf{Z}_{\mathbf{X}\,\mathbf{0}}$ as follows:

$$\mathbf{Z}_{\mathbf{X0}} = \mathbf{UT} = \mathbf{UD}_{\mathbf{P}}\mathbf{T}_{\mathbf{0}} \tag{2-16}$$

$$\mathbf{T} = \begin{pmatrix} \mathbf{t}_{(1)} \\ \vdots \\ \mathbf{t}_{(N)} \end{pmatrix} = \begin{pmatrix} t_{11} & \cdots & t_{1S} \\ \vdots & \ddots & \vdots \\ t_{N1} & \cdots & t_{NS} \end{pmatrix} \tag{2-17}$$

Where $\mathbf{T}_{\mathbf{0}}$ is the $N \times S$ matrix representing the simulation mode coefficients of FN component, and $\mathbf{t}_{(k)} = (t_{1k}, \ldots, t_{Sk})$ is the S-dimensional row vector. For FP component in Eq. (2-13) and Eq. (2-15), define the $M \times S$ matrix $\mathbf{Z}_{\mathbf{Y}\,\mathbf{0}}$ in the same way.

$$\mathbf{Z}_{\mathbf{Y0}} = \mathbf{V}\,\mathbf{W} = \mathbf{VD}_{\mathbf{Q}}\mathbf{W}_{\mathbf{0}} \tag{2-18}$$

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}_{(1)} \\ \vdots \\ \mathbf{w}_{(N)} \end{pmatrix} = \begin{pmatrix} w_{11} & \cdots & w_{1S} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{NS} \end{pmatrix} \tag{2-19}$$

where $\mathbf{W}$ and $\mathbf{W_0}$ are the $N \times S$ matrices representing the simulation PCSs and mode coefficients of FP components, respectively. $\mathbf{w}_{(k)} = (w_{1k}, \ldots, w_{Sk})$ is the S-dimensional row vector. Set the condition that the cross-covariance matrices between the modes in $\mathbf{T}$ and $\mathbf{W}$ equal the cross-covariance matrices between the modes in $\mathbf{P}$ and $\mathbf{Q}$.

$$\frac{\mathbf{PQ}^{\mathrm{T}}}{N} = \frac{\mathbf{QP}^{\mathrm{T}}}{N} \tag{2-20}$$

$$\frac{\mathbf{TW}^{\mathrm{T}}}{S} = \frac{\mathbf{WT}^{\mathrm{T}}}{S} \tag{2-21}$$

Assuming that Eq. (2-20) and Eq. (2-21) are equal, Eq. (2-10) in condition II can be described as follows:

$$\mathbf{TW}^{\mathrm{T}} = \mathbf{WT}^{\mathrm{T}} = \mathbf{D} \times \frac{S}{N} \tag{2-22}$$

The diagonal entry of $\mathbf{TW}^{\mathrm{T}}$ in Eq. (2-22) is described as follows:

$$\sum_{i=1}^{S} t_{ki} w_{ki} = d_k \times \frac{S}{N} \tag{2-23}$$

The simulation methods for the matrices $\mathbf{T}$ and $\mathbf{W}$ satisfying conditions 1 and 2 are described in ***2.4.3*** – ***2.4.4***.

Using the matrices $\mathbf{Z_{X\,0}}$ in Eq. (2-16) and $\mathbf{Z_{Y\,0}}$ in Eq. (2-18), the simulation ground motion distributions of the two horizontal components are described as follows:

$$\mathbf{Z_X} = \mu_{\mathbf{X}} \mathbf{H_Z} + \sigma_{\mathbf{X}} \mathbf{Z_{X0}} \tag{2-24}$$

$$\mathbf{Z_Y} = \mu_{\mathbf{Y}} \mathbf{H_Z} + \sigma_{\mathbf{Y}} \mathbf{Z_{Y0}} \tag{2-25}$$

where $\mathbf{H_Z}$ is the $M \times S$ matrix in which all entries are 1. Accordingly, the simulation method of ground motion distribution for a single component in the previous study [1] can be extended to distributions of two components.

2.5.3 Simulation Method of Matrices T and W

The abovementioned simulated matrices $\mathbf{T}$ and $\mathbf{W}$ were used to generate the ground motion distributions of two horizontal components. From Eqs. (2-16) and (2-18), $\mathbf{T}$ and $\mathbf{W}$ are composed of the diagonal matrices $\mathbf{D_P}$ and $\mathbf{D_Q}$, whose diagonal entries are singular values (constants) for each component, and the matrices $\mathbf{T_0}$ and $\mathbf{W_0}$, whose entries are mode coefficients. In practice, $\mathbf{T}$ and $\mathbf{W}$ are generated by simulating the mode coefficients $\mathbf{T_0}$ and $\mathbf{W_0}$ and synthesizing them with the singular values $\mathbf{D_P}$ and $\mathbf{D_Q}$, respectively. As simulation methods for $\mathbf{T_0}$ and $\mathbf{W_0}$, four methods are proposed and compared with reference to the previous study [1]: a method that assumes no correlation between two components, a method that makes them correlated, a method that assumes perfect correlation, and a method that considers correlation by giving variation.

a) Method A: Method to give two components independently as uncorrelated

The relationship between two components is assumed to be uncorrelated, and $N \times S$ standard normal random numbers are generated independently for each component, which are the elements of the $N \times S$

matrices $\mathbf{T_{0\,R}}$ and $\mathbf{W_{0\,R}}$. For $\mathbf{T_{0\,R}}$ and $\mathbf{W_{0\,R}}$, the following equations hold [1].

$$\begin{aligned} \mathrm{E}\left[\frac{\mathbf{T_{0R}}\mathbf{T_{0R}}^{\mathrm{T}}}{S}\right] &= \mathrm{E}\left[\frac{\mathbf{W_{0R}}\mathbf{W_{0R}}^{\mathrm{T}}}{S}\right] = \mathbf{I} \\ \mathrm{E}\left[\frac{\mathbf{T_{0R}}\mathbf{W_{0R}}^{\mathrm{T}}}{S}\right] &= \mathrm{E}\left[\frac{\mathbf{W_{0R}}\mathbf{T_{0R}}^{\mathrm{T}}}{S}\right] = \mathbf{0} \end{aligned} \tag{2-26}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix, and 0 is the $N \times N$ zero matrix. The first equation in Eq. (2-26) represents that the norm of the row vectors $\mathbf{t_{0\,R}}(k)$, $\mathbf{w_{0\,R}}(k)$ of the matrices $\mathbf{T_{0\,R}}$, $\mathbf{W_{0\,R}}$ is on average $S^{1/2}$. The first and second equations indicate that the row vectors are, on average orthogonal across modes and components.

Since the row vectors $\mathbf{p_0}$ and $\mathbf{q_0}$ are N-dimensional unit vectors, $\mathbf{T_0}$ and $\mathbf{W_0}$ are obtained by adjusting the norm of $\mathbf{t_{0\,R}}$ and $\mathbf{w_{0\,R}}$ as follows:

$$\mathbf{T_0} = \frac{\mathbf{T_{0\,R}}}{\sqrt{N}} \qquad \mathbf{W_0} = \frac{\mathbf{W_{0R}}}{\sqrt{N}} \tag{2-27}$$

b) Method B: Method to consider the correlation of two components by correlation simulation

In Method B, Cholesky decomposition is applied to the correlation matrix $\mathbf{R}_k$ between the mode coefficients $\mathbf{p_0}(k)$ and $\mathbf{q_0}(k)$, and correlation simulations are performed on the randomly sampled row vectors $\mathbf{t_{0\,R}}(k)$ and $\mathbf{w_{0\,R}}(k)$.

$$\begin{aligned} \mathbf{R}_k &= \mathbf{L}_k\mathbf{L}_k^{\mathrm{T}} \\ \begin{pmatrix} 1 & r_k \\ r_k & 1 \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ r_k & \sqrt{1-r_k^2} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ r_k & \sqrt{1-r_k^2} \end{pmatrix}^{\mathrm{T}} \end{aligned} \tag{2-28}$$

$$\begin{pmatrix} \mathbf{w_{0\,Chol}}(k) \\ \mathbf{t_{0\,Chol}}(k) \end{pmatrix} = \mathbf{L}_k\begin{pmatrix} \mathbf{w_{0\,R}}(k) \\ \mathbf{t_{0\,R}}(k) \end{pmatrix} = \begin{pmatrix} \mathbf{w_{0\,R}}(k) \\ r_k\mathbf{w_{0\,R}}(k) + \sqrt{1-r_k^2}\,\mathbf{t_{0\,R}}(k) \end{pmatrix} \tag{2-29}$$

Where r_k is a correlation coefficient between $\mathbf{p_0}(k)$ and $\mathbf{q_0}(k)$, $\mathbf{L}_k$ is the 2×2 lower triangular matrix of Mode k. $\mathbf{t_{0\,Chol}}(k)$ and $\mathbf{w_{0\,Chol}}(k)$ are the S-dimensional vectors in the k-th row of the $N \times S$ matrices $\mathbf{T_{0\,Chol}}$ and $\mathbf{W_{0\,Chol}}$, respectively. In this section, $\mathbf{t_{0\,Chol}}(k)$ and $\mathbf{w_{0\,Chol}}(k)$ are obtained by giving a correlation to $\mathbf{t_{0\,R}}(k)$ of FN component based on $\mathbf{w_{0\,R}}(k)$ of FP component. As in Eq. (2-27), $\mathbf{T_0}$ and $\mathbf{W_0}$ are finally obtained as follows:

$$\mathbf{T_0} = \frac{\mathbf{T_{0\,Chol}}}{\sqrt{N}} \qquad \mathbf{W_0} = \frac{\mathbf{W_{0\,Chol}}}{\sqrt{N}} \tag{2-30}$$

$\mathbf{p_0}(k)$ and $\mathbf{q_0}(k)$ are the unit vectors, and the mean of the vector entries is zero for Mode 2 and higher. Thus, the correlation coefficient $r_{k\ (\geq 2)}$ is described as follows:

$$r_{k(\geq 2)} = \sum_{i=1}^{N} p_{0ki}q_{0ki} = \mathbf{p_{0(}}_{k)}\mathbf{q_{0(}}_{k)}^{\mathrm{T}} \tag{2-31}$$

In Mode 2 and higher, the row vectors $\mathbf{t_0}$ and $\mathbf{w_0}$ of the matrices $\mathbf{T_0}$ and $\mathbf{W_0}$ in Eq. (2-30) hold as follows:

$$\mathrm{E}\left[\mathbf{t}_{\mathbf{0}(k)}\mathbf{w}_{\mathbf{0}(k)}{}^{\mathrm{T}}\,\middle|\,k\geq 2\right]=\sum_{i=1}^{S}\frac{t_{0\mathrm{Chol}\,ki}}{\sqrt{N}}\frac{w_{0\mathrm{Chol}\,ki}}{\sqrt{N}}$$
$$=\sum_{i=1}^{S}\frac{r_k w_{0\mathrm{R}ki}{}^2+\sqrt{1-r_k{}^2}\,t_{0\mathrm{R}ki}w_{0\mathrm{R}ki}}{N} \tag{2-32}$$
$$=r_k\frac{\mathbf{w}_{\mathbf{0}\mathrm{R}(k)}\mathbf{w}_{\mathbf{0}\mathrm{R}(k)}{}^{\mathrm{T}}}{N}=\frac{S}{N}\mathbf{p}_{\mathbf{0}(k)}\mathbf{q}_{\mathbf{0}(k)}{}^{\mathrm{T}}$$

Eq. (2-32) indicates that the diagonal entries of the matrix $\mathbf{T_0}\mathbf{W_0}^{\mathrm{T}}$ in Method B satisfy Eq. (2-23) on average.

c) Method C: Method to make the two components perfectly correlated

Based on the randomly sampled $N \times S$ matrix $\mathbf{W_{0\,R}}$, $\mathbf{T_0}$, and $\mathbf{W_0}$ are obtained with the two components as perfect correlations.

$$\frac{\mathbf{W_{0R}}}{\sqrt{N}}=\mathbf{W_0}=\mathbf{T_0} \tag{2-33}$$

d) Method D: Method to randomly give the difference between two components

In method D, the sampled $\mathbf{W_{0\,R}}$ is given a normal random number based on the difference between **P** and **Q** as follows:

$$\mathbf{w}_{\mathbf{0}\mathrm{R}(k)}=\mathbf{w}_{\mathbf{0}\mathrm{P}(k)}\qquad \mathbf{t}_{\mathbf{0}\mathrm{R}(k)}=\mathbf{w}_{\mathbf{0}\mathrm{R}(k)}+N(\mu_k,\sigma_k{}^2) \tag{2-34}$$

$$\mu_k=\frac{1}{N}\sum\nolimits_{i=1}^{N}\left(q_{ki}-p_{ki}\right)$$
$$\sigma_k=\sqrt{\frac{1}{N}\sum\nolimits_{i=1}^{N}\left\{\left(q_{ki}-p_{ki}\right)-\mu_k\right\}^2} \tag{2-35}$$

where $\mathbf{t_{0\,R\,(}}_{k)}$ and $\mathbf{w_{0\,R\,(}}_{k)}$ are the S-dimensional vectors in the k-th row of the $N \times S$ matrices $\mathbf{T_{0\,R}}$ and $\mathbf{W_{0\,R}}$, respectively. $N(\mu_k, \sigma_k{}^2)$ is a normal random number following mean μ_k and variance $\sigma_k{}^2$. In this section, $\mathbf{t_{0\,R}}$ of FN component is obtained by giving the variation to $\mathbf{w_{0\,R}}$ of FP component. Finally, as in Methods A – C, $\mathbf{T_0}$ and $\mathbf{W_0}$ are obtained as follows:

$$\mathbf{T_0}=\frac{\mathbf{T_{0R}}}{\sqrt{N}}\qquad \mathbf{W_0}=\frac{\mathbf{W_{0R}}}{\sqrt{N}} \tag{2-36}$$

2.5.4 Adjustment of Vector Entries for Mode 1

In order to satisfy condition I, the row vectors $\mathbf{t_{0\,(}}_{k)}$ and $\mathbf{w_{0\,(}}_{k)}$ are adjusted. The row vectors $\mathbf{p_0}$ and $\mathbf{q_0}$ for Mode 2 and higher ($k \geq 2$) have mean $\mu_k = 0$ and standard deviation $\sigma_k = 0.0408$ ($= N^{-1/2}$); $\mathbf{t_{0\,(}}_{k\geq 2)}$ and $\mathbf{w_{0\,(}}_{k\geq 2)}$ in Eqs. (2-27), (2-30), (2-33), and (2-36) satisfy condition I. In contrast, $\mathbf{t_{0\,(1)}}$ and $\mathbf{w_{0\,(1)}}$ for Mode 1 have mean $\mu_1 = 0.0408$ ($= N^{-1/2}$), standard deviation $\sigma_1 = 0.00075 – 0.0026$, and coefficient of variation $\delta_1 = 1.8 – 6.4$ %. This coefficient of variation δ_1 depends on the distributions of ground motion for the 600 cases

for each period and component. Therefore, using this as a parameter, the first row $\mathbf{t_0}_{(1)}$ and $\mathbf{w_0}_{(1)}$ of matrices $\mathbf{T_0}$ and $\mathbf{W_0}$ is adjusted as follows [1].

$$\begin{aligned} t_{00\,1i} &= \mu_{\mathrm{P}\,1} + \sigma_{\mathrm{P}\,1} \cdot \sqrt{N} t_{01i} \\ &= \mu_{\mathrm{P}\,1} + \mu_{\mathrm{P}\,1}\delta_{\mathrm{P}\,1} \cdot \sqrt{N} t_{01i} = \frac{1}{\sqrt{N}} + \delta_{\mathrm{P}\,1} t_{01i} \\ w_{00\,1i} &= \mu_{\mathrm{Q}\,1} + \sigma_{\mathrm{Q}\,1} \cdot \sqrt{N} w_{01i} = \frac{1}{\sqrt{N}} + \delta_{\mathrm{Q}\,1} w_{01i} \end{aligned} \tag{2-37}$$

2.6 Numerical Example of Simulating Ground Motion Distributions of Two Horizontal Components

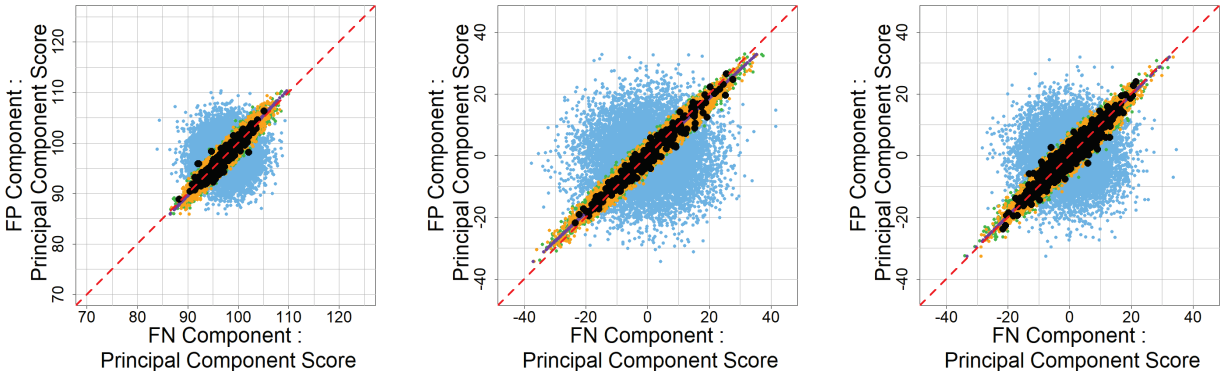
2.6.1 Simulation of Matrices T and W

Based on the results of the SVD analysis applied in ***2.4***, simulate the distribution of S_A of two horizontal components. The matrices **T** and **W** are simulated by each of the four methods in ***2.5.3***, and the results are compared. The number of simulation trials is set to $S = 9600$. The common logarithmic mean at point i for 600 cases of two components S_A (matrices **X** and **Y**) are $\mu_{\mathbf{X}i}$ and $\mu_{\mathbf{Y}i}$, and the standard deviations are $\sigma_{\mathbf{X}i}$ and $\sigma_{\mathbf{Y}i}$, respectively. The range of the simulation results at point i is set from $\mu_{\mathbf{X}i} - 3\sigma_{\mathbf{X}i}$ to $\mu_{\mathbf{X}i} + 3\sigma_{\mathbf{X}i}$ (from $\mu_{\mathbf{Y}i} + 3\sigma_{\mathbf{Y}i}$ to $\mu_{\mathbf{Y}i} - 3\sigma_{\mathbf{Y}i}$). Note that the upper and lower limits are provisional values set in an exploratory manner by comparing the distributions of the original data and the simulation results.

First, examine how well the PCSs for the 9600 cases of Methods A-D reproduce the 600 cases. Fig. 2-12–Fig. 2-14 show the PCSs of the 600 cases (the matrices **P** and **Q**: Fig. 2-9 (c)–Fig. 2-11 (c)) in black dots, and the simulated PCSs of the 9600 cases by Methods A-D (A, B, C, and D) in light blue, green, purple, and orange, respectively. The simulated PCSs of FP component are the same for all methods, while the simulated PCSs of FN component differ for each method. Method A (light blue) and Method C (purple) do not reproduce the correlation between two components in the original 600 cases because they treat them as uncorrelated and perfectly correlated, respectively. In contrast, Method B (green) and Method D (orange) adequately reproduce the correlation between the two components.

Table. 2-3 shows the correlation coefficients of simulated PCSs for two horizontal components by period. The correlation coefficients of Methods B and D are almost the same as those between **P** and **D** for the 600 cases, suggesting that Methods B and D are almost equivalent. On the other hand, the correlation coefficients of Methods A and C are uncorrelated and perfectly correlated, respectively, and the correlation relationship between the two components cannot be reproduced.

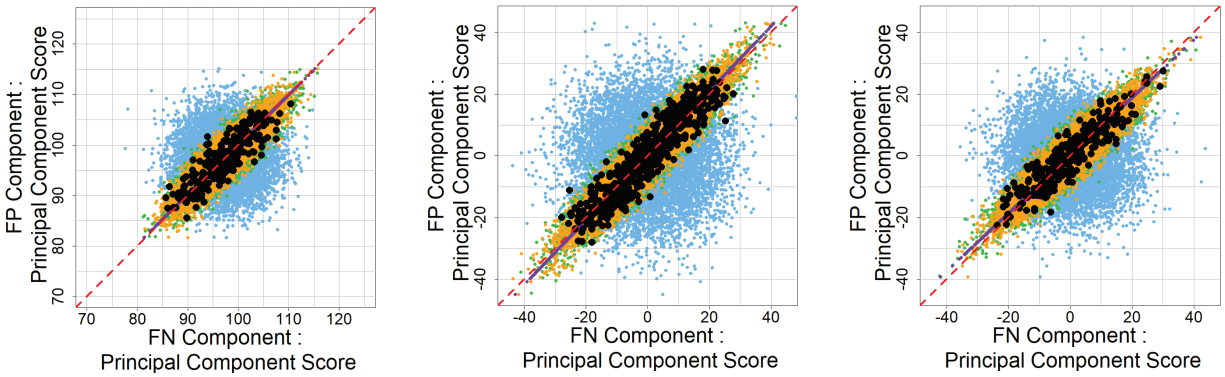
Next, focus on condition II "$\mathbf{PQ}^{\mathrm{T}} = \mathbf{QP}^{\mathrm{T}} = \mathbf{D}$." Table. 2-4 shows the ratio of the simulated values of Methods A – D to the target value of the diagonal elements (singular values) of the matrix $\mathbf{TW}^{\mathrm{T}}$ in Eq. (2-23). Red indicates the method with the ratio closest to 1 in each mode; blue indicates the method closest to 2 in each mode. For Mode 1, the mean mode coefficient is $\mu_1 = 0.0408$, and the coefficient of variation is $\delta_1 = 1.8 – 6.4\%$, with the difference in results between Methods A – D due to the small variability. Among Methods A – D, it is Method B that satisfies Eq. (2-23) sufficiently well and reproduces the relationship between the two horizontal components. This result is because, as shown in Eq. (2-32), Method B satisfies Eq. (2-23) on average for Mode 2 and higher and converges to the right-hand side of Eq. (2-23) for a sufficient number of samples S. The next best reproducing the relationship between two components is Method D. On the other hand, Methods A and C have low reproducibility. Therefore, they are not appropriate as a simulation method for the two components' PCSs (or mode coefficients).



(a) Mode 1 **(b)** Mode 2 **(c)** Mode 3

(Horizontal axis: FN component, vertical axis: FP component, black:600 cases of matrix **P**, **Q** (Fig. 2-9 (c)), blue: Method A, green: Method B, purple: Method C, and orange: Method D)

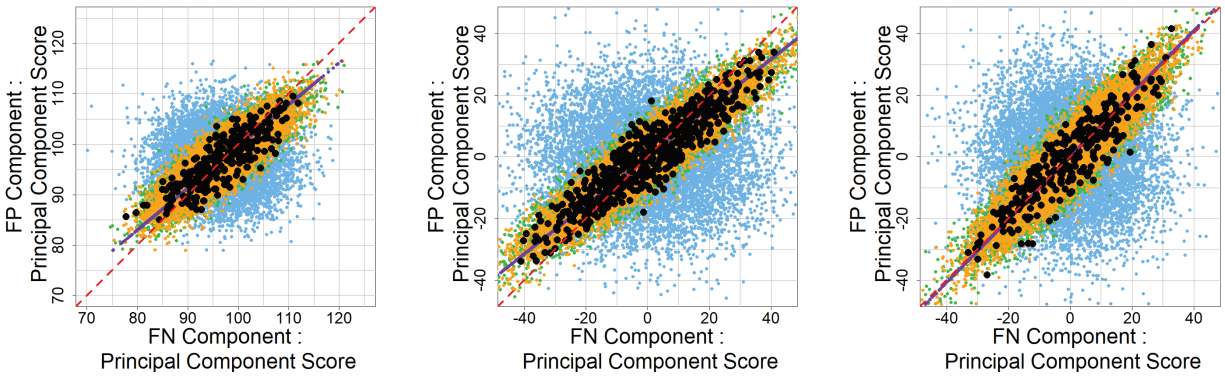
Figure 2-12. Scatter plot of simulated PCSs for two horizontal components in period 0.5 s.



(a) Mode 1 **(b)** Mode 2 **(c)** Mode 3

(Same as Fig. 2-12, 600cases of matrix **P**, **Q**: Fig. 2-10 (c))

Figure 2-13. Scatter plot of simulated PCSs for two horizontal components in period 1.0 s.



(a) Mode 1 **(b)** Mode 2 **(c)** Mode 3

(Same as Fig. 2-13, 600cases of matrix **P**, **Q**: Fig. 2-11 (c))

Figure 2-14. Scatter plot of simulated PCSs for horizontal two components in period 2.0 s.

Table 2-3. Correlation coefficients of simulated PCSs for horizontal two components.

(a) Period 0.5 s

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7	Mode 8	Mode 9	Mode 10
Matrix P,Q	0.946	0.979	0.957	0.960	0.955	0.908	0.900	0.916	0.943	0.900
Method A	-0.004	0.019	-0.014	-0.015	0.002	0.020	0.003	0.007	-0.010	0.003
Method B	0.947	0.979	0.958	0.959	0.955	0.910	0.899	0.918	0.944	0.902
Method C	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Method D	0.953	0.980	0.960	0.963	0.960	0.921	0.913	0.928	0.950	0.915

(b) Period 1.0 s

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7	Mode 8	Mode 9	Mode 10
Matrix P,Q	0.870	0.933	0.910	0.728	0.926	0.755	0.765	0.704	0.848	0.840
Method A	-0.004	0.019	-0.014	-0.015	0.002	0.020	0.003	0.007	-0.010	0.003
Method B	0.872	0.933	0.912	0.721	0.925	0.762	0.763	0.709	0.849	0.843
Method C	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Method D	0.893	0.941	0.923	0.807	0.936	0.821	0.825	0.800	0.881	0.873

(c) Period 2.0 s

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7	Mode 8	Mode 9	Mode 10
Matrix P,Q	0.842	0.923	0.908	0.775	0.857	0.687	0.756	0.775	0.724	0.679
Method A	-0.004	0.019	-0.014	-0.015	0.002	0.020	0.003	0.007	-0.010	0.003
Method B	0.845	0.923	0.910	0.769	0.857	0.696	0.755	0.778	0.725	0.684
Method C	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Method D	0.844	0.933	0.921	0.832	0.888	0.786	0.820	0.836	0.811	0.784

Table 2-4. The ratio of simulated value to the target value of the diagonal element (singular value) of matrix $\mathbf{TW}^{T}$. (For each mode, a red number indicates the method with a ratio close to 1, and a blue number indicates the second-closest method.)

(a) Period 0.5 s

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7	Mode 8	Mode 9	Mode 10
Method A	**1.00**	0.02	-0.01	-0.01	0.00	0.02	0.00	0.01	-0.01	0.00
Method B	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Method C	**1.00**	1.02	1.05	1.04	1.05	1.10	1.11	1.09	1.06	1.11
Method D	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.01**	**1.02**	**1.01**	**1.01**	**1.02**

(b) Period 1.0 s

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7	Mode 8	Mode 9	Mode 10
Method A	**1.00**	0.02	-0.02	-0.02	0.00	0.03	0.00	0.01	-0.01	0.00
Method B	**1.00**	**1.00**	**1.00**	**0.99**	**1.00**	**1.01**	**1.00**	**1.01**	**1.00**	**1.00**
Method C	**1.00**	1.07	1.10	1.37	1.08	1.32	1.31	1.42	1.18	1.19
Method D	**1.00**	**1.01**	**1.01**	**1.11**	**1.01**	**1.09**	**1.09**	**1.13**	**1.03**	**1.04**

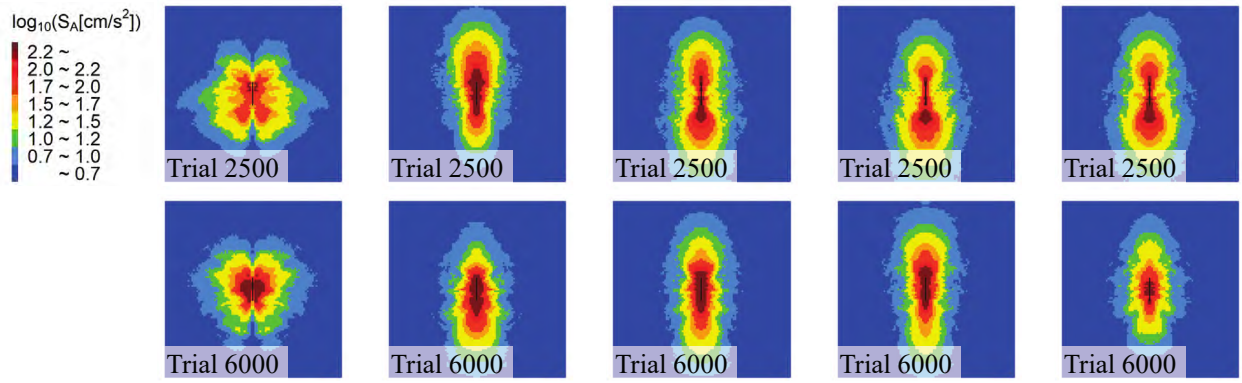
(c) Period 2.0 s

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7	Mode 8	Mode 9	Mode 10
Method A	**1.00**	0.02	-0.02	-0.02	0.00	0.03	0.00	0.01	-0.01	0.00
Method B	**1.00**	**1.00**	**1.00**	**0.99**	**1.00**	**1.01**	**1.00**	**1.00**	**1.00**	**1.01**
Method C	**1.00**	1.08	1.10	1.29	1.17	1.45	1.32	1.29	1.38	1.47
Method D	**1.00**	**1.01**	**1.01**	**1.08**	**1.03**	**1.15**	**1.09**	**1.07**	**1.11**	**1.15**

2.6.2 Simulated Ground Motion Distributions of Horizontal Two Components

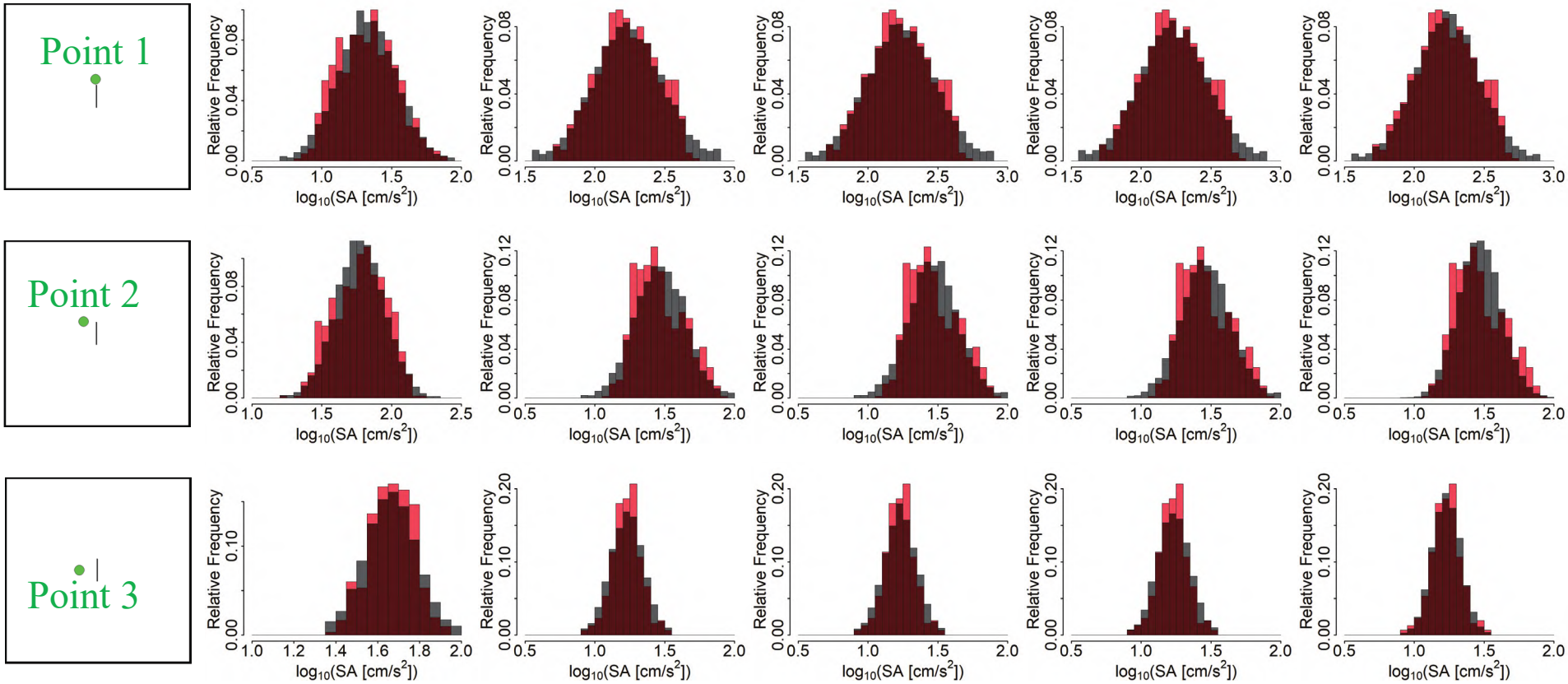
Figure. 2-15 shows examples of the distributions of S_A of two horizontal components for the 9600 cases t in Methods A – D (due to space limitation, only two cases in period 2.0 s are shown). Note that the matrix **W** for FP component is the same for all four methods. Because two components are treated independently in Method A, the cross-covariance structure between the two components is not reflected in the ground motion distributions of FN component. In contrast, the ground motion distributions of FN component in Method B reflect the cross-covariance structure of the two components. The ground motion distributions of FN component in Methods C and D also show similar trends to those of Method B. However, the difference appears between the case where two components are perfectly correlated and the case where correlation is considered.

Next, the simulated ground motions of the 9600 cases at the selected points are compared with the original 600 cases. For symmetry, the computational area (Fig. 2-1 (b)) is divided into four quadrants, and seven points are selected in the second quadrant. Fig. 2-16 shows the relative frequency distribution of S_A for the 600 and 9600 cases at three points (Fig. 2-16 (a), Points 1 – 3). The simulation's relative frequency distribution adequately reproduces the distribution's shape in the original data. In addition, as shown in the upper panels of Figs. 2-16 (d), (f), the hem of the relative frequency distribution is wider than that of the original data, indicating that the simulation method generates ground motions with large amplitude levels that are not included in the original data. These trends were also found at four other sites not shown in the figure and other periods. The results indicate that the ground motion distributions of two horizontal components generated by the simulation method in this section reproduce the trend of the distribution calculated by the stochastic Green's function method [13].



(a) FP comp. (Method A–D) **(b)** FN comp. (Method A) **(c)** FN comp. (Method B) **(d)** FN comp. (Method C) **(e)** FN comp. (Method D)

Figure 2-15. Simulated distributions of S_A (T = 2.0s).



(a) Selected point. (Green circle)

(b) FP comp. (Method A – D)

(c) FN comp. (Method A)

(d) FN comp. (Method B)

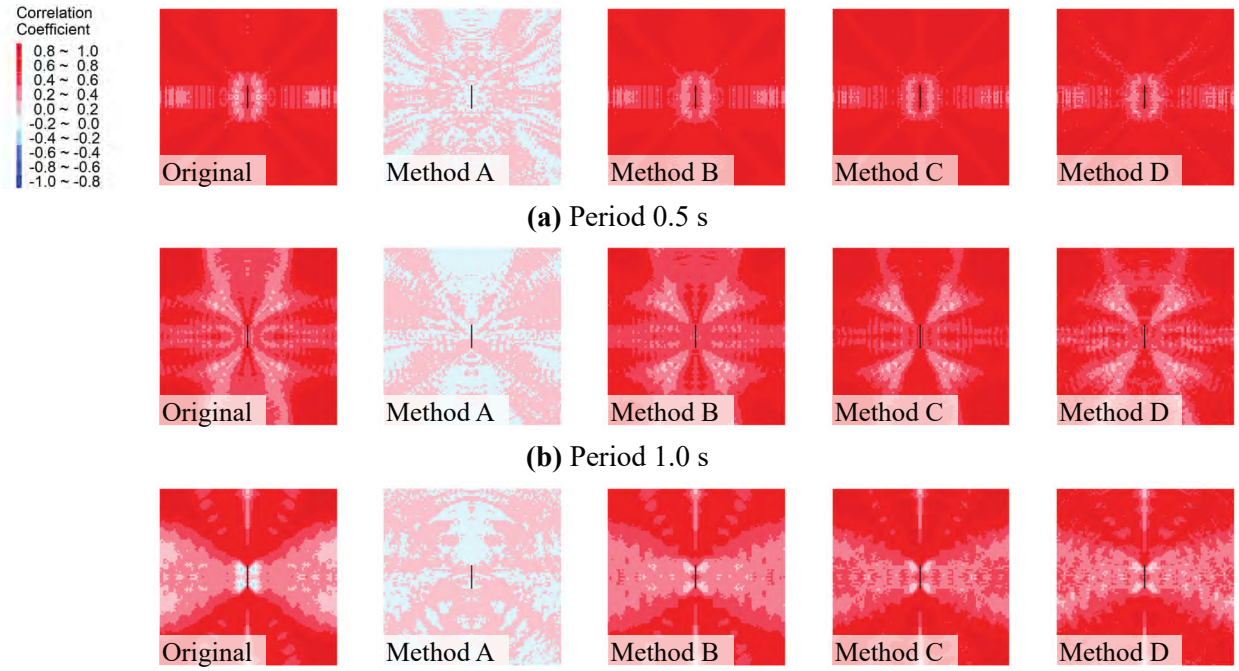
(e) FN comp. (Method C)

(f) FN comp. (Method D)

(Red: original 600 cases, black: simulated 9600 cases.)

Figure 2-16. Relative frequency of S_A (T = 2.0s) at selected point (green point in **(a)**).

Evaluate the spatial correlation of the simulated horizontal two components. Fig. 2-17 shows the distribution of the correlation coefficients of two components for each point in the original 600 cases (reiteration of Fig. 2-6) and the simulated 9600 cases of Methods A – D. Method A, which treats the relationship between two components as uncorrelated, is almost uncorrelated over the entire area and does not reproduce the correlation distribution of two components in the 600 cases. The correlation coefficients of Method C, which treats the two components as perfectly correlated, show generally stronger positive correlations than those of the 600 cases. Methods B and D, which consider the correlation of two components, reproduce the correlation distribution of the 600 cases well enough, and there is no significant difference.



(a) Period 0.5 s

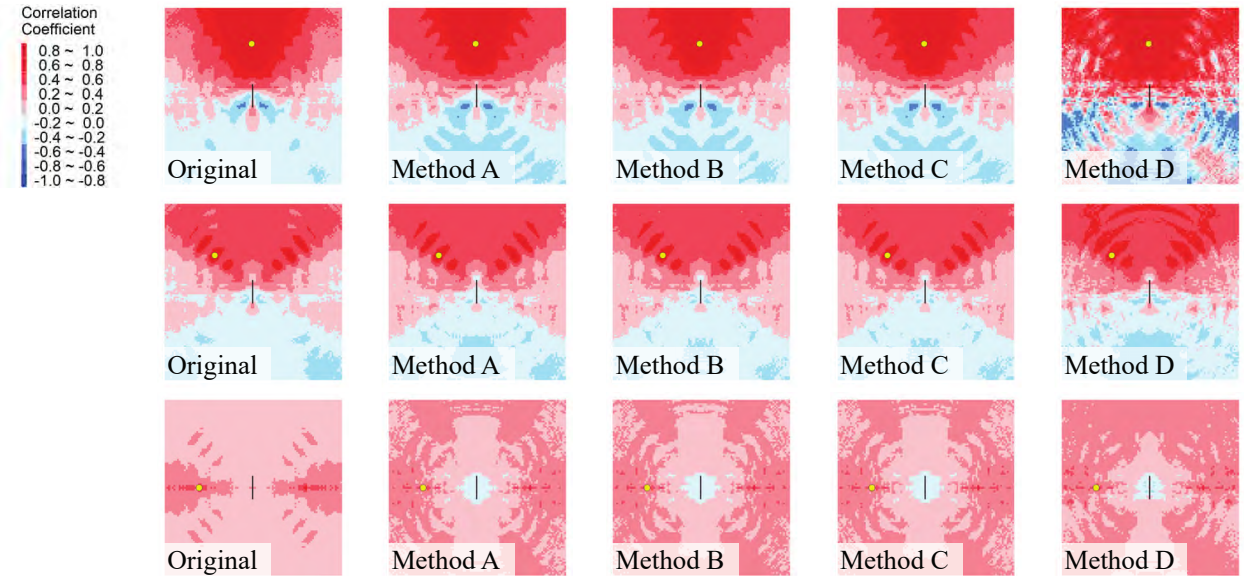
(b) Period 1.0 s

(c) Period 2.0 s

("Original" on the left are restatements of Fig. 2-3.)

Figure 2-17. Spatial distribution of correlation coefficient between FN and FP components S_A for simulated 9600 cases.

Finally, evaluate the spatial structure of the site-to-site correlations for each component. This condition regarding site-to-site correlation is not included in conditions I and II; however, this is a condition in single-component ground motion distribution simulation [1], and is an important requirement in the portfolio risk assessment for facility siting, a typical use case of simulation results. Fig. 2-18 shows the spatial distribution of the correlation between the selected point and each point surrounding the selected point for FN component in the original 600 cases and the simulated 9600 cases. The ground motions of FP component are the same for Methods A – D and are essentially the same as those of Method A in Fig. 2-18 (b) and are therefore not shown in the figure. Methods A – C can adequately reproduce the spatial structure of the site-to-site correlations in the original 600 cases. In contrast, Method D, which randomly gives the differences between two components, has strong correlations and fails to reproduce the 600 cases, as shown in the upper panel of Fig. 2-18 (e). The above trend was also found in other periods.



(a) Original 600 cases **(b)** Method A **(c)** Method B **(d)** Method C **(e)** Method D

Figure 2-18. Spatial correlations of S_A (T = 2.0s, FN component) between the selected point (yellow circle) and each computational point surrounding the selected point.

2.6.3 Comparison of Simulation Methods

The relationship between two horizontal components is not reproduced by Methods A and C. In contrast, Methods B and D reproduce the correlation well, satisfying conditions I and II. Focusing on the site-to-site correlation of ground motions within the same component, Methods A, B, and C can reproduce the spatial structure of the site-to-site correlation, but Method D cannot reproduce it. Therefore, it is preferable to use Method B as a simulation method of ground motion distributions that preserves the spatial interrelationship between two horizontal components.

The number of trials for the simulations in this chapter was large ($S = 9600$); Method B satisfied Eq. (2-23) and may have sufficiently reproduced the relationship between the two horizontal components. Thus, the same study was conducted with $S = 1000$ and 2400 trials. As a result, it was found that the reproducibility of the spatial correlation structure of two components (and the spatial structure of the site-to-site correlation within the same component) in the simulation results of Method B was sufficient, although it was lower than $S = 9600$.

2.6.4 Comparison with Simulation Method for Single Component

The method for simulating ground motion distributions using mode decomposition and synthesis based on SVD proposed in the previous study [1] is for a single component. Therefore, when this method is applied to the distribution of S_A of two horizontal components, correlation cannot be considered, and it can only be assumed to be uncorrelated or perfectly correlated. This case corresponds to Method A or Method C (light blue or purple in Fig. 2-12–Fig. 2-14). In contrast, the simulation method proposed in this study, based on SVD analysis that can simultaneously decompose two horizontal components, can adequately reproduce the correlation between the components.

2.7 Conclusions

This chapter proposes a simulation method for ground motion distribution that considers the spatial interrelationship of two horizontal components. As an example of numerical simulation, the distributions of the absolute acceleration response spectrum of horizontal two components (FN, FP) and period 0.1 – 2.0 s for 600 cases of earthquakes caused by a strike-slip fault were used. The major conclusions are listed below.

1) Singular value decomposition analysis was applied to the ground motion distributions of two horizontal components by period, and the spatial variations related to the two components were extracted for each mode as mode form. The Mode shape of Mode 1 corresponds to the mean ground motion distribution and shows spatial variations attributed mainly to the distance attenuation effect. The mode shape of Mode 2 and Mode 3 could be classified as the "spatial variation with positive correlation in almost entire area" and the "spatial variation negatively correlated north and south of the map."

2) By replacing the matrices of principal component scores obtained from SVD analysis with matrices that reproduce the cross-covariance structure of the two horizontal component ground motion distributions, a method is proposed to simulate the ground motion distributions of two components while preserving the spatial interrelationship of the original two-component distribution. Four simulation methods were used: "Method A: treating the two components independently as uncorrelated," "Method B: considering correlation by correlation simulation," "Method C: making the two components perfectly correlated," and "Method D: randomly giving the difference between the two components."

3) The simulation method using SVD analysis generated the distributions of S_A of two horizontal components. The simulation results of Methods A – D were compared by the PCSs, the spatial distribution of the correlation between two horizontal components, and the site-to-site correlation distribution within the same component. The results showed that Method B adequately reproduced the ground motion distribution of two-component and their interactive spatial variation structure.

The simulation method proposed in this chapter can simultaneously simulate the spatial distribution of two physical indicators. For future developments, this method should be applied to the simultaneous simulation of ground motion distributions in different periodic bands caused by the same earthquake is the subject of future work.

Appendix I: SVD Analysis Based on Dual Formalism

The following is the procedure for calculating the SVD analysis based on dual formalism [4].

From Eqs. (2-7) and (2-10), the variance-covariance matrix of $\mathbf{X_0}$ is as follows:

$$\mathbf{X_0}^{\mathrm{T}}\mathbf{X_0} = \mathbf{P}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{P} = \mathbf{P}^{\mathrm{T}}\mathbf{P} \tag{I.1}$$

Right-multiplying Eq. (I.1) by the matrix $\mathbf{Q}^{\mathrm{T}}$.

$$(\mathbf{X_0}^{\mathrm{T}}\mathbf{X_0})\mathbf{Q}^{\mathrm{T}} = (\mathbf{P}^{\mathrm{T}}\mathbf{P})\mathbf{Q}^{\mathrm{T}} = \mathbf{P}^{\mathrm{T}}\mathbf{D} \tag{I.2}$$

Right-multiplying Eq. (I.2) by the matrix $\mathbf{D}^{-1}$.

$$(\mathbf{X_0}^{\mathrm{T}}\mathbf{X_0})\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1} = \mathbf{P}^{\mathrm{T}}\mathbf{D}\mathbf{D}^{-1} = \mathbf{P}^{\mathrm{T}} \tag{I.3}$$

The same is done for $\mathbf{Y_0}$.

$$(\mathbf{Y_0}^{\mathrm{T}}\mathbf{Y_0})\mathbf{P}^{\mathrm{T}} = (\mathbf{Q}^{\mathrm{T}}\mathbf{Q})\mathbf{P}^{\mathrm{T}} = \mathbf{Q}^{\mathrm{T}}\mathbf{D} \tag{I.4}$$

$$(\mathbf{Y_0}^{\mathrm{T}}\mathbf{Y_0})\mathbf{P}\mathbf{D}^{-1} = \mathbf{Q}^{\mathrm{T}} \tag{I.5}$$

Substitute Eq. (I.5) for Eqs. (I.2) and (I.4) for Eq. (I.3) as follows:

$$(\mathbf{X_0}^{\mathrm{T}}\mathbf{X_0})(\mathbf{Y_0}^{\mathrm{T}}\mathbf{Y_0})\mathbf{P}^{\mathrm{T}} = \mathbf{P}^{\mathrm{T}}\mathbf{D}^2 \tag{I.6}$$

$$(\mathbf{Y_0}^{\mathrm{T}}\mathbf{Y_0})(\mathbf{X_0}^{\mathrm{T}}\mathbf{X_0})\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}^{\mathrm{T}}\mathbf{D}^2 \tag{I.7}$$

From Eqs. (I.6) and (I.7), the matrices **P**, **Q** and **D** are obtained as solutions to an $N \times N$ (= $\{(N \times M) \times (M \times N)\} \times \{(N \times M) \times (M \times N)\}$) matrix eigenvalue problem.

Next, the matrices **U** and **V** are computed. Since the PCSs between the same components are not orthogonal in SVD analysis (Eq. (2-11)), it is necessary to compute **U** from the PCSs **Q** of different components. Eqs. (2-10) and (2-12) are transformed as follows:

$$\mathbf{P}^{-1} = \mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1} \tag{I.8}$$

$$\mathbf{U} = \mathbf{X_0}\mathbf{P}^{-1} \tag{I.9}$$

U is obtained by substituting Eq. (I.8) into Eq. (I.9).

$$\mathbf{U} = \mathbf{X_0}\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1} \tag{I.10}$$

The same is done for **V**.

$$\mathbf{V} = \mathbf{Y_0}\mathbf{P}^{\mathrm{T}}\mathbf{D}^{-1} \tag{I.11}$$

The above computational procedure (dual formalism) provides the matrices **U**, **D**, **V**, **P**, and **Q** without explicitly dealing with the *M*-by-*M* cross-covariance matrix $\mathbf{C_{XY}}$.

References in Chapter 2

[1] Nojima, N. Kuse, M. and Duc, L. Q. (2018) Mode decomposition and simulation of strong ground motion distribution using singular value decomposition. *Journal of JAEE*, **18** (2): pp. 95–114. https://doi.org/10.5610/jaee.18.2_95 **(in Japanese with English abstract)**

[2] Koide, H. and Kodera, K. (1999): A SVD analysis between the winter NH 500-hPa height and surface temperature fields. *JMSJ*, **77** (1): pp. 47–61. https://doi.org/10.2151/jmsj1965.77.1_47

[3] Tanimoto, Y. (1996) SVD (singular value decomposition) analysis. *Tenki*, **43** (4): pp. 243–245. **(in Japanese)**

[4] Minobe, S. (2001) Atmosphere and ocean statistics data analysis -chapter 6 multivariate time domain analysis part 1: SVD analysis-. *Hokkaido Univ.* **(in Japanese)** https://www.sci.hokudai.ac.jp/~minobe/data_anal/chap6.pdf

[5] Matsuyama, Y. and Tanimoto, Y. (2008) Using UNIX/Windows/Macintosh how to analyze climatological data. Tokyo, *Kokon shoin Co.*: 118 p. **(in Japanese)**

[6] Kagawa, T. (2015) Spatial variability of periodic characteristics among strong ground motions derived from multiple fault rupture scenarios. *Journal of JAEE*, **15** (7): pp. 90–99. https://doi.org/10.5610/jaee.15.7_90 **(in Japanese with English abstract)**

[7] Kagawa, T. (2015) A study on spatial and temporal variabilities of strong ground motions based on multiple fault rupture scenarios. *Journal of JSCE A1*, **71** (4): pp. I_191–I_197. https://doi.org/10.2208/jscejseee.71.I_191 **(in Japanese with English abstract)**

[8] Kanatani, K. (2003) Applied mathematics classroom -from least squares to wavelets-. Tokyo, *Kyoritsu Shuppan Co., LTD*: 270 p. **(in Japanese)**

[9] Irikura, K. and Miyake, H. (2001) Prediction of strong ground motions for scenario earthquakes. *J. Geogr*, **110** (6): pp. 849–875. https://doi.org/10.5026/jgeography.110.6_849 **(in Japanese with English abstract)**

[10] Irikura, K., Miyake, H., Iwata, T., Kamae, K., Kawabe, H. and Dalguer, D. L. (2004) Recipe for predicting ground motions from future large earthquakes. *Proc. of 13WCEE*: Paper No. 1371.

[11] Somerville, P. G., Irikura, K., Graves, R., Sawada, S., Wald, D., Abrahamson, N., Iwasaki, Y., Kagawa, T., Smith, N. and Kowada, A. (1999) Characterizing crustal earthquake slip models for the prediction of strong ground motion. *Seismological Research Letters*, **70** (1): pp. 59–80. https://doi.org/10.1785/gssrl.70.1.59

[12] Kagawa, T., Irikura, K. and Somerville, P. G. (2004) Differences in ground motion and fault rupture process between the surface and buried rupture earthquakes. *EPS*, **56**: pp. 3-14. https://doi.org/10.1186/BF03352486

[13] Kagawa, T. (2004) Developing a Stochastic Green's function method having more accuracy in long period range to be used in the Hybrid Method. *Journal of JAEE*, **4** (2): pp. 21–32. https://doi.org/10.5610/jaee.4.2_21 **(in Japanese with English abstract)**

[14] Kagawa, T. (2010) Probabilistic earthquake hazard evaluation adopting the fault rupture model under characteristic stress field. *Proc. of 13th JEE Symposium*: pp. 4165–4170. **(in Japanese with English abstract)**

[15] Yamazaki, F. and Shinozuka, M. (1990) Simulation of stochastic fields by statistical preconditioning. *Journal of Engineering Mechanics*, **116**, Issue 2: pp. 268-287. https://doi.org/10.1061/(ASCE)0733-9399(1990)116:2(268)

3 Development of Simplified Prediction Model for Ground Motion Distribution

3.1 Introduction

Ground motion prediction is one of the most important issues in seismic risk assessment and disaster prevention/mitigation. There are two typical methods for ground motion prediction: a conventional method based on ground motion prediction equation (GMPE) and a detailed method based on strong ground motion (SGM) simulation. The former conventional method (upper left in Fig. 3-1) is an empirical/statistical method that models ground motion intensity with a relatively small number of parameters based on observation records; the source effects are described by magnitude, propagation path effects by attenuation models, and site effects by ground conditions [1–4]. This method can produce average prediction results. However, the trade-off for this feature is that it cannot account for complex fault rupture processes and subsurface structures that greatly impact ground motions. Therefore, some studies have introduced hanging-wall effects, rupture directivity effects, and regional terms into GMPEs to reproduce more realistic ground motions (e.g., Japan: [5–7], the U.S.: the NGA-West2 (Enhancement of Next Generation of Attenuation Relationships for Western US) project [8–11]).

On the other hand, the latter detailed method (upper right in Fig. 3-1) considers the three-dimensional geometry and rupture process of the source fault by setting a characterized source model and propagation path effects and site effects by setting a subsurface structure model [12–15]. This method can reproduce the fault rupture processes and the influence of the subsurface structure on seismic motion, and provide the time history waveforms of ground motions. However, the detailed method requires more parameters, computer resources, and time than the conventional method.

In recent years, various attempts to use artificial intelligence (AI) have been actively made in the fields of seismology and earthquake engineering [16, 17]. In the aspect of ground motion prediction, prediction models for ground motion have been developed using AI (training data for the AI models: recorded data [18–26], simulation data [27]) (bottom right of Fig. 3-1). A hybrid predictor that combines AI with GMPE based on a physical model has also been proposed [28]. These AI models for ground motion prediction are an advanced version of GMPE; they can efficiently predict ground motion for various earthquakes with relatively high accuracy. However, these models target earthquakes caused by point-source; they cannot make predictions that directly reflect the detailed rupture process of the fault as the detailed method does. Therefore, it would be useful in various applications if the ground motion distribution for an arbitrary fault rupture scenario, as simulated by the detailed method, could be predicted easily and efficiently like the conventional method and AI models.

The study in this chapter applies “mode decomposition and synthesis” and “modeling by machine learning (ML)” to the ground motion distributions calculated by the detailed method and constructs a simplified prediction model for ground motion distribution (bottom left of Fig. 3-1). Ref. [29] and **Chapter 2** propose methods to simulate many ground motion distributions using mode decomposition and synthesis. These methods generate ground motion distributions by random simulation using the spatial characteristics of the original distributions as a constraint. For future developments, the simulation of ground motion distribution in relation to the fault rupture scenario is a key issue. Therefore, this chapter models the relationship between fault rupture scenarios and modes, and applies it to a predictive ground motion distribution model. This model is a surrogate model of the detailed method (simplification of detailed

method), and its advantage is that it can simply predict the ground motion distribution from any characterized source model. In this chapter, four types of modeling ("multiple regression," "support vector regression," "random forest," and "gradient boosting decision tree") are performed, and their prediction accuracy is compared.

3.2 outlines how to build a prediction model of ground motion distribution based on mode decomposition/synthesis and machine learning modeling. Then, ***3.3*** describes the mode decomposition of the ground motion distributions, and ***3.4*** describes the machine learning modeling. Next, ***3.5*** presents the results of applying the mode decomposition. ***3.6*** constructs the prediction model for ground motion distribution. ***3.7*** evaluates the validity of the proposed model by comparing the prediction results from the proposed prediction model with those from the detailed method. Finally, ***3.8*** summarizes the conclusions obtained in this chapter.

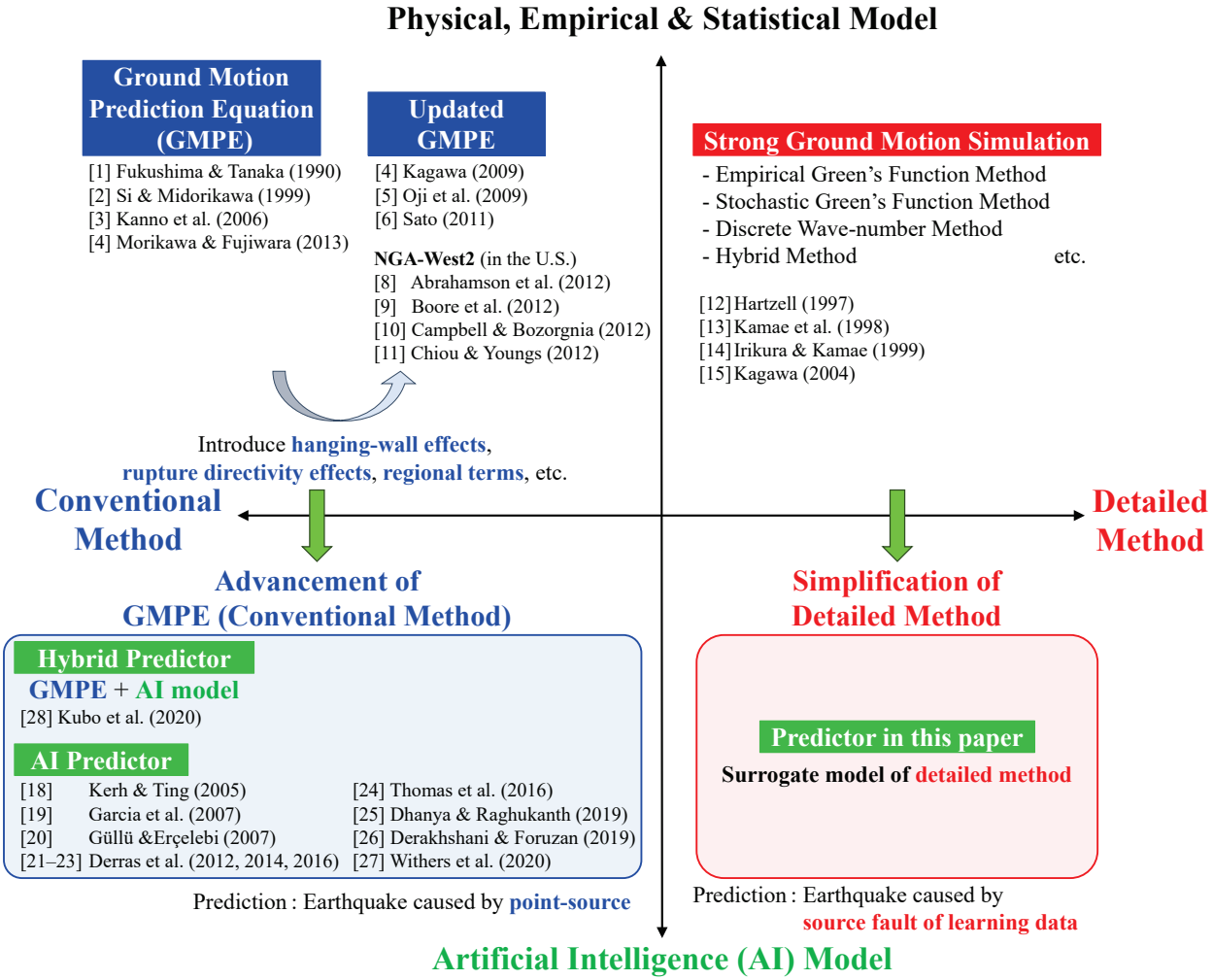


Figure 3-1. Overview of ground motion prediction methods.

3.2 Simplified Prediction Model for Ground Motion Distribution

Figure. 3-2 shows an image of the simplified prediction model for ground motion distribution proposed in this study; the construction procedure is described below. First, the mode decomposition is applied to the ground motion distributions calculated by SGM simulation on the specific source fault to extract three elements: "spatial distribution of modes (mode shape)" which represent the spatial characteristics of ground motion, "singular values" which are weights for each mode, and "weight coefficients" for each mode corresponding to each case. Next, the "weight coefficients" are modeled for each mode by the "parameters of the characterized source model (source parameters)" based on machine learning regression analysis, and a predictor of the weight coefficients is constructed. Finally, by inputting arbitrary source parameters (arbitrary case) into the weight coefficient predictor for each mode, its weight coefficients are predicted; the ground motion distribution is predicted by mode synthesis with predicted weight coefficients, the mode shapes, and singular values.

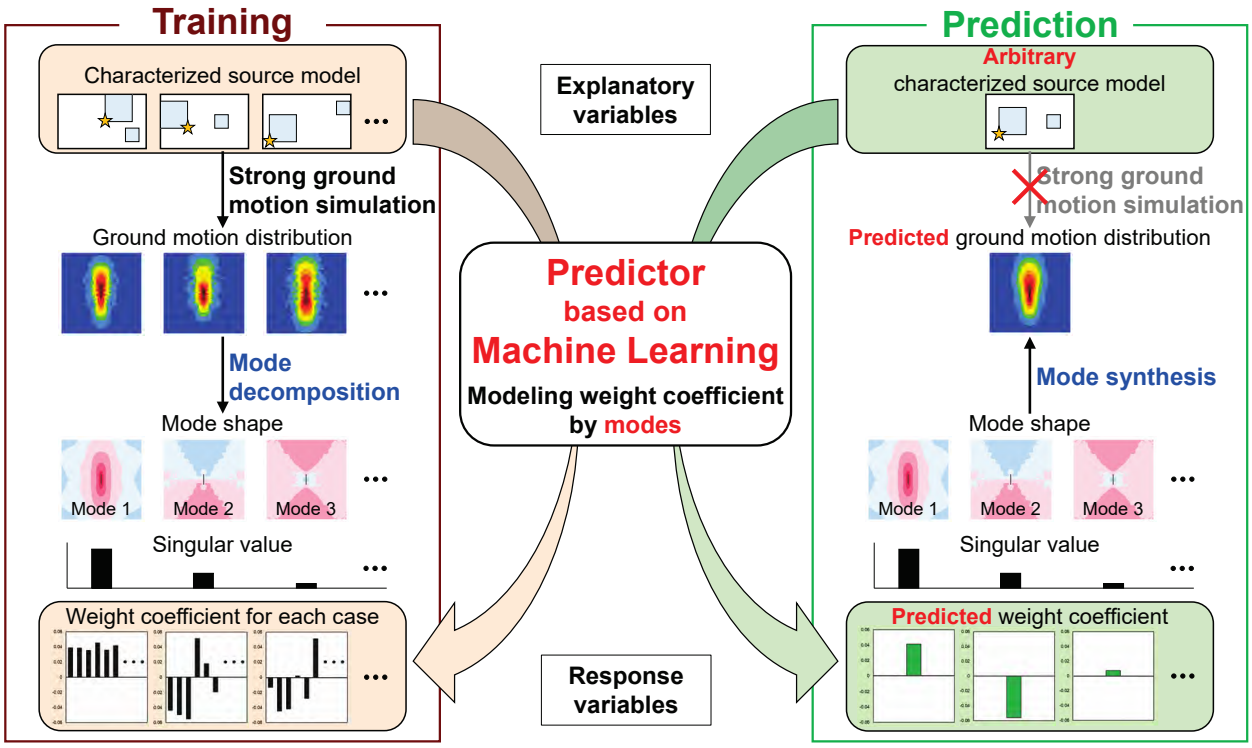


Figure 3-2. Prediction model of strong ground motion distribution using "mode decomposition and synthesis" and "modeling by machine learning."

3.3 Mode Decomposition and Synthesis of Ground Motion Distributions

Let x denote a variable representing ground motion, and a ground motion map consisting of M sites is represented by a column vector $\mathbf{x} = (x_1, \ldots, x_M)^{\mathrm{T}}$. Thus, ground motion distribution maps for N cases of different parameter settings are described by the $M \times N$ ($M > N$) matrix $\mathbf{X}$ as follows:

$$\mathbf{X} = \left(\mathbf{x}_1, \ldots, \mathbf{x}_N\right) = \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{M1} & \cdots & x_{MN} \end{pmatrix} \tag{3-1}$$

where x_{ij} ($1 \le i \le M$, $1 \le j \le N$) is ground motion at site i in the ground motion distribution of Case j. The rank of $\mathbf{X}$ is rank($\mathbf{X}$) = N.

Then the matrix $\mathbf{X}$ is normalized using the mean μ_G and the standard deviation σ_G of all N cases of ground motion maps. The normalized matrix $\mathbf{X_0}$ is expressed as Eq. (3-2).

$$\mathbf{X_0} = \frac{\mathbf{X} - \mu_G \mathbf{H}}{\sigma_G} \tag{3-2}$$

$$\mu_G = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} x_{ij}}{MN} \qquad \sigma_G = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(x_{ij} - \mu_G\right)^2}{MN}} \tag{3-3}$$

where $\mathbf{H}$ is the $M \times N$ matrix in which all entries are 1. Next, singular value decomposition is applied to the matrix $\mathbf{X_0}$ [29].

$$\mathbf{X_0} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}} = \begin{pmatrix} u_{11} & \cdots & u_{1N} \\ \vdots & \ddots & \vdots \\ u_{M1} & \cdots & u_{MN} \end{pmatrix} \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_N \end{pmatrix} \begin{pmatrix} v_{11} & \cdots & v_{1N} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{NN} \end{pmatrix}^{\mathrm{T}} \tag{3-4}$$

$$\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I} \qquad \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{V}\mathbf{V}^{\mathrm{T}} = \mathbf{I} \tag{3-5}$$

The matrices $\mathbf{U}$, $\mathbf{D}$, $\mathbf{V}$ and $\mathbf{I}$ in Eq. (3-4) and Eq. (3-5) have the following meanings.

- The matrix $\mathbf{U}$: The column vector of the k-th column ($1 \le k \le N$), $\mathbf{u}_k = (u_{1k}, \ldots, u_{Mk})^{\mathrm{T}}$, is the left singular vector that consists of orthogonal basis, where u_{ik} represents ground motion at the site i in Mode k. The left singular vector represents the mode shape that defines the spatial correlation in a set of ground motion distributions. A map of the left singular vector visualizes the spatial correlation between different sites.
- The matrix $\mathbf{D}$: The diagonal matrix whose diagonal entry d_k is the singular value equivalent to the square root of eigenvalue in Mode k. The singular value represents the case-independent weight for each mode.
- The matrix $\mathbf{V}$: The column vector of the k-th column, $\mathbf{v}_k = (v_{1k}, \ldots, v_{Nk})^{\mathrm{T}}$, is the right singular vector that consists of orthogonal basis, where v_{kj} is the weight of Mode k in case j. The right singular vector represents the case-dependent weight for each mode; thus, these entries are the weight coefficients that characterize each case.
- The matrix $\mathbf{I}$: $N \times N$ identity matrix.

The matrix **UD** represents the fundamental structure dominating all cases of ground motion distributions and their spatial correlation. Based on the fundamental structure, the ground motion distribution of Case *j* is characterized by the *j*-th row vector of the matrix **V**.

The weight coefficients of Mode 1–*N* for arbitrary characterized source model (arbitrary case) are described by an *N*-dimensional row vector $\mathbf{v}_\mathbf{y}$ as follows:

$$\mathbf{v}_\mathbf{y} = \left(v_{\mathrm{y}\,1}, \ldots, v_{\mathrm{y}\,N}\right) \tag{3-6}$$

The vector of weight coefficients is modally synthesized with **UD** in Eq. (3-4) to obtain an *M*-dimensional column vector $\mathbf{y_0}$.

$$\mathbf{y_0} = \mathbf{U}\mathbf{D}\mathbf{v}_\mathbf{y}^{\mathrm{T}} = \sum_{k=1}^{N} \mathbf{u}_k d_k v_{y\,k} \tag{3-7}$$

The ground motion distribution for arbitrary case is described as $\mathbf{y} = (y_1, \ldots, y_M)^{\mathrm{T}}$ using $\mathbf{y_0}$.

$$\mathbf{y} = \mu_G \mathbf{h} + \sigma_G \mathbf{y_0} \tag{3-8}$$

where **h** is the *M*-dimensional column vector in which all entries are 1. As shown on the right side of Fig. 3-2, if we can predict the weight coefficients for arbitrary characterized source model, we can avoid an SGM simulation and predict the ground motion distribution. The next section describes the weight coefficient prediction method.

3.4 Modeling Weight Coefficient

3.4.1 Weight Coefficient Predictor and Prediction Model of Ground Motion Distribution

The weight coefficients for each mode (element v of matrix $\mathbf{V}$ in Eq. (3-4)) are modeled by the parameters of the characterized source models (source parameters) based on the regression analysis of machine learning. The source parameters (number of parameters: p) are defined as the p-dimensional row vector $\mathbf{s} = (s_1, \ldots, s_p)$. The issue covered in this chapter can be described as constructing a predictor with the source parameters $\mathbf{s}$ as the input-vector and the weight coefficient $\hat{v}_k$ for Mode k as the output-value, as follows:

$$\hat{v}_k = f_k(\mathbf{s}) \tag{3-9}$$

From Eqs. (3-7) and (3-9), when the source parameters for arbitrary case is $\mathbf{s_y}$, its standardized ground motion distribution $\mathbf{y_0}$ is described as follows:

$$\hat{\mathbf{y}}_\mathbf{0} = \sum_{k=1}^{N} \mathbf{u}_k d_k f_k(\mathbf{s_y}) \tag{3-10}$$

The prediction model for ground motion distribution in Eq. (3-10) is a linear combination model of the mode shape.

A dataset including n^* case ($1 \le n^* < N$) of source parameters and weight coefficients is used as the training data to construct the predictor. The source parameters for arbitrary case in the training data are described by the p-dimensional row vector $\mathbf{s_L} = (s_{L\,1}, \ldots, s_{L\,p})$, and the training dataset of the source parameters is described by the $n^* \times p$ matrix $\mathbf{S_L}$. The weight coefficients for arbitrary case in the training data are by the n^*-dimensional column vector $\mathbf{v_L} = (v_{L\,1}, \ldots, v_{L\,n^*})^{\mathrm{T}}$, and the training dataset of the weight coefficients is described by the $n^* \times N$ matrix $\mathbf{V_L}$.

$$\mathbf{S_L} = \left(\mathbf{s}_{\mathbf{L}\,(1)}, \ldots, \mathbf{s}_{\mathbf{L}\,(n^*)}\right)^{\mathrm{T}} = \begin{pmatrix} s_{L\,11} & \cdots & s_{L\,1p} \\ \vdots & \ddots & \vdots \\ s_{L\,n^*1} & \cdots & s_{L\,n^*p} \end{pmatrix} \tag{3-11}$$

$$\mathbf{V_L} = \left(\mathbf{v}_{\mathbf{L}\,1}, \ldots, \mathbf{v}_{\mathbf{L}\,N}\right) = \begin{pmatrix} v_{L\,11} & \cdots & v_{L\,1N} \\ \vdots & \ddots & \vdots \\ v_{L\,n^*1} & \cdots & v_{L\,n^*N} \end{pmatrix} \tag{3-12}$$

This study uses four methods for modeling weight coefficient: “multiple regression,” which is the simplest method, “support vector regression,” which can handle linear and nonlinear data through kernel methods, and “random forest,” and “gradient boosting decision tree,” which often show high prediction accuracy. The following paragraphs outline each of these methods.

3.4.2 Multiple Regression [30, 31]

The first is a multiple regression (MR). MR-predictor is a linear combination model in which the objective variable is a linear function of the explanatory variables, and the relationship between the training data weight coefficients $\mathbf{v_L}$ and the source parameter $\mathbf{S_L}$ are described as follows:

$$\begin{aligned} \mathbf{v_L} &= b_0 + \mathbf{S_L}\mathbf{b} + \mathbf{e} \\ \begin{pmatrix} v_{L1} \\ \vdots \\ v_{L n^*} \end{pmatrix} &= b_0 + \begin{pmatrix} s_{L11} & \cdots & s_{L1p} \\ \vdots & \ddots & \vdots \\ s_{L n^* 1} & \cdots & s_{L n^* p} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_{n^*} \end{pmatrix} \end{aligned} \tag{3-13}$$

where b_0 is a scalar quantity called bias, and $\mathbf{b}$ is a p-dimensional column vector whose elements are the regression coefficients b_l ($1 \leq l \leq p$). $\mathbf{e}$ is a residual vector with n^* entries. The MR is formulated as an optimization problem (the least squares method) to minimize the sum of squares of the residual vectors $\|\mathbf{e}\|^2=\mathbf{e}^{\mathrm{T}}\mathbf{e}$, as follows:

$$\min\left(\|\mathbf{e}\|^2\right) = \min_{b_0,\mathbf{b}}\left[\sum_{j^*=1}^{n^*}\left\{v_{j^*} - \left(b_0 + \mathbf{s}_{\mathbf{L}(j^*)}\mathbf{b}\right)\right\}^2\right] \tag{3-14}$$

Thus, the MR-predictor f_{MR} ($\mathbf{s}$) is described as follows:

$$f_{MLR}(\mathbf{s}) = b_0 + \mathbf{s}\mathbf{b} = b_0 + b_1 s_1 + \cdots + b_p s_p \tag{3-15}$$

In this study, the explanatory variables are selected using a stepwise method, and the MR that minimizes Akaike's information criterion (AIC) is used as the final MR-predictor for the weight coefficients.

The main advantage of the MR-predictor is that the linearity of the regression coefficients makes the behavior of the predictor very easy to interpret (i.e., a white-box model). In contrast, it also has the fatal drawback of poor expressiveness due to the linearity concerning the explanatory variables, which is a trade-off for the interpretability of the regression model. Furthermore, there is also the risk of multicollinearity when correlations are observed among the explanatory variables, among other drawbacks.

3.4.3 Support Vector Regression [30, 32, 33]

Support vector regression (SVR) is an extension of support vector machine (SVM), a popular method for 2-class classification problems, to regression problems. The SVR-predictor f_{SVR} (**s**) is described as follows:

$$f_{SVR}(\mathbf{s}) = \phi(\mathbf{s})\mathbf{w} + b_{SVR} \tag{3-16}$$

where **w** is the weight vector and b_{SVR} is the bias. Φ is the mapping function, and a Φ (**s**) is obtained by a nonlinear mapping of the input vector **s** to a higher-dimensional feature space. SVR is formulated as an optimization problem using an ε-insensitive loss function based on the insensitivity parameter $\varepsilon \geq 0$, as shown in Fig. 3-3, instead of the squared error loss function used in MR.

$$\min_{\mathbf{w},b_{SVR}} \left[C\sum_{j^*=1}^{n^*} \max\left\{0, \left|v_{j^*} - f_{SVR}(\mathbf{s}_{\mathbf{L}(j^*)})\right| - \varepsilon\right\} + \frac{1}{2}\|\mathbf{w}\|^2 \right] \tag{3-17}$$

where C is the regularization parameter that determines the penalty for data exceeding ε and controls the trade-off between training data and model complexity. The tolerance error ε function (the first term in Eq. (3-17)) is consistent with the minimum absolute error method when = 0. Note that C and ε are hyperparameters that must be set a priori.

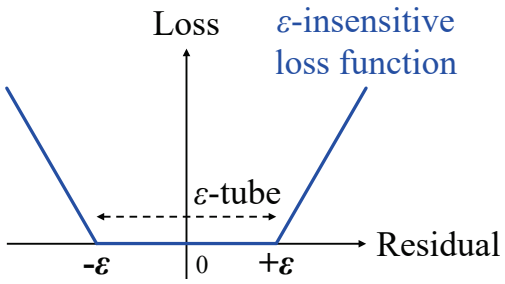


Figure 3-3. ε-insensitive loss function.

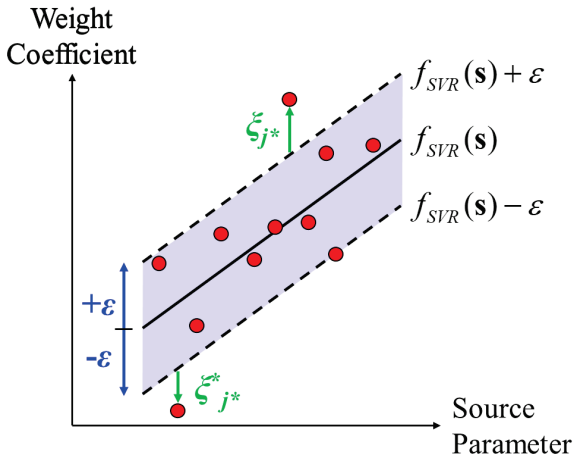


Figure 3-4. ε-tube and slack variables for SVR.

The optimization problem in Eq. (3-17) can be represented by introducing the slack variables $\xi_{j^*} \geq 0$ and $\xi^*_{j^*} \geq 0$ ($j^* = 1, \ldots, n^*$) as follows:

$$\begin{aligned} &\min_{\mathbf{w}, b_{SVR}} \left[C \sum_{j^*=1}^{n^*} \left(\xi_{j^*} + \xi^*_{j^*} \right) + \frac{1}{2} \|\mathbf{w}\|^2 \right] \\ &\text{subject to} \quad v_{j^*} \leq f_{SVR}(\mathbf{s}_{\mathbf{L}(j^*)}) + \varepsilon + \xi_{j^*} \\ &\qquad\qquad v_{j^*} \geq f_{SVR}(\mathbf{s}_{\mathbf{L}(j^*)}) - \varepsilon - \xi^*_{j^*} \end{aligned} \tag{3-18}$$

The constraint allows the points of the training data to lie outside the ε-tube (the area indicated by both arrows in Fig. 3-3), which is the error dead band. Fig. 3-4 shows an image of the relationship between the regression model, the ε-tube (blue area), and the slack coefficients. $\xi_{j^*} > 0$ and $\xi^*_{j^*} = 0$ hold for points located above the ε-tube, and $\xi_{j^*} = 0$ and $\xi^*_{j^*} > 0$ hold for points located below. For points inside the ε-tube, $\xi_{j^*} = \xi^*_{j^*} = 0$ holds. Eq. (3-18) can be optimized by solving the following Lagrange function L_{SVR} using the Lagrange multipliers $\alpha_{j^*} \geq 0$, $\alpha^*_{j^*} \geq 0$, $\mu_{j^*} \geq 0$, $\mu^*_{j^*} \geq 0$.

$$\begin{aligned} L_{SVR} = {} & C \sum_{j^*=1}^{n^*} \left(\xi_{j^*} + \xi^*_{j^*} \right) + \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{j^*=1}^{n^*} \left(\mu_{j^*} \xi_{j^*} + \mu^*_{j^*} \xi^*_{j^*} \right) \\ & - \sum_{j^*=1}^{n^*} \alpha_{j^*} \left(\varepsilon + \xi_{j^*} + f_{SVR}(\mathbf{s}_{\mathbf{L}(j^*)}) - v_{j^*} \right) - \sum_{j^*=1}^{n^*} \alpha^*_{j^*} \left(\varepsilon + \xi^*_{j^*} - f_{SVR}(\mathbf{s}_{\mathbf{L}(j^*)}) + v_{j^*} \right) \end{aligned} \tag{3-19}$$

Partial derivative Eq. (3-19) with $\mathbf{w}$:

$$\frac{\partial L_{SVR}}{\partial \mathbf{w}} = \mathbf{0} \Leftrightarrow \mathbf{w} = \sum_{j^*=1}^{n^*} \left(\alpha_{j^*} - \alpha^*_{j^*} \right) \phi(\mathbf{s}_{\mathbf{L}(j^*)})^{\mathrm{T}} \tag{3-20}$$

By performing partial differentiation for b_{SVR}, ξ_{j^*} and $\xi^*_{j^*}$ in the same way and transforming the Lagrange functions using them, the SVR optimization problem (Eqs. (3-17) and (3-18)) can be formulated as a quadratic programming problem that maximizes the following equation:

$$\begin{aligned} \tilde{L}_{SVR}(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = {} & -\frac{1}{2} \sum_{i^*=1}^{n^*} \sum_{j^*=1}^{n^*} \left(\alpha_{i^*} - \alpha^*_{i^*} \right) \left(\alpha_{j^*} - \alpha^*_{j^*} \right) K\left(\mathbf{s}_{\mathbf{L}(i^*)}, \mathbf{s}_{\mathbf{L}(j^*)} \right) \\ & - \varepsilon \sum_{j^*=1}^{n^*} \left(\alpha_{j^*} + \alpha^*_{j^*} \right) + \sum_{j^*=1}^{n^*} \left(\alpha_{j^*} - \alpha^*_{j^*} \right) v_{j^*} \\ \text{subject to} \quad & \sum_{j^*=1}^{n^*} \left(\alpha_{j^*} - \alpha^*_{j^*} \right) v_{j^*}, \ 0 \leq \alpha_{j^*} \leq C, \ 0 \leq \alpha^*_{j^*} \leq C \end{aligned} \tag{3-21}$$

where a kernel function $K()$ is the inner product of $\Phi(\mathbf{s})$ as follows:

$$K\left(\mathbf{s}_{\mathbf{L}(i^*)}, \mathbf{s}_{\mathbf{L}(j^*)} \right) = \phi\left(\mathbf{s}_{\mathbf{L}(i^*)} \right) \phi\left(\mathbf{s}_{\mathbf{L}(j^*)} \right)^{\mathrm{T}} \tag{3-22}$$

Assigning Eq. (3-16) to Eq. (3-20), the SVR-predictor can be described as follows:

$$f_{SVR}(\mathbf{s}) = \sum_{j^*=1}^{n^*} \left(\alpha_{j^*} - \alpha^*_{j^*} \right) K\left(\mathbf{s}, \mathbf{s}_{\mathbf{L}} \right) + b_{SVR} \tag{3-23}$$

This equation shows that SVR does not require direct computation of $\Phi(\mathbf{s})$ but only its inner product. This mapping of a higher-dimensional space using kernel functions is called a "kernel trick" or "kernel substitution." This study uses the radial basis function (RBF) kernel, which can realize nonlinear models and perform well for various data types.

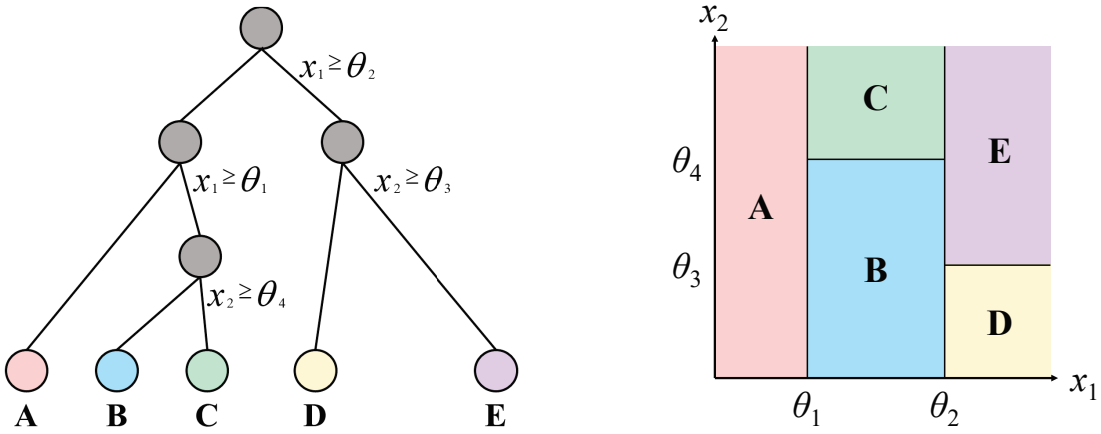
$$K\left(\mathbf{s}, \mathbf{s}_{\mathbf{L}} \right) = \exp\left(-\gamma \left\| \mathbf{s} - \mathbf{s}_{\mathbf{L}} \right\|^2 \right) \tag{3-24}$$

where γ is a hyperparameter representing the range of influence (precisely, the reciprocal of the radius of influence) for each case in the training data. In other words, the smaller γ is, the greater the influence of the individual data, resulting in a simple decision boundary, and the larger γ is, the smaller the range of influence of the data, resulting in a more complex decision boundary.

SVR has the advantage of handling a wide range of data, from linear to nonlinear, and is less susceptible to noise in the training data, making it difficult to over-learn. However, it has the disadvantage that the combination of hyperparameters and kernels is complex, making it difficult to construct an optimal model. In addition, the trade-off for high prediction accuracy is low model interpretability, as with the random forest and gradient boosting decision trees.

3.4.4 Decision Tree [30–32, 34]

Next, the decision tree, which is the basis of "random forest" and "gradient boosting decision tree" algorithms, is described in the outline. As shown in Fig. 3-5, a decision tree is a tree structure-based learner that performs classification (in this case, the decision tree is called a classification tree) or regression (regression tree) by dividing the input space into multi-dimensional cuboid regions with two branching nodes. Although decision trees do not have high predictive accuracy, they have a clear structure and can be interpreted intuitively. In contrast, decision trees can only consider one explanatory variable per partition and, therefore, can only draw partition lines orthogonal to the axis of the explanatory variable being considered, which may not reflect the relationship between the explanatory variable and the objective variable well. In the following, the decision tree-predictor for an input vector **s** is denoted by $f_{\mathrm{Tree}}(\mathbf{s})$.



(a) Tree structure **(b)** Divide data based on tree structure

Figure 3-5. Decision tree model.

3.4.5 Random Forest [30–32, 34]

Random forest (RF) is a model based on bootstrap aggregating (bagging), an ensemble method that combines multiple training units (in this case, decision trees) to achieve high prediction accuracy. Bagging is a method in which multiple models of the same type (decision trees in RF) are created in parallel for a training subset (the same number of samples as training data) created by the bootstrap method (overlapping sampling), and the average of their predictions is used as the prediction result. In addition to bootstrap sampling, RF constructs the training subset by sampling only some of the explanatory variables. As shown in Fig. 3-6, the RF-predictor f_{RF} (**S**) is described as follows:

$$f_{RF}(\mathbf{s}) = \frac{1}{K}\sum_{k=1}^{K} f_{Tree}{}^{(k)}(\mathbf{s}) \tag{3-25}$$

where K is the number of decision trees constituting the RF-predictor, and $f_{Tree}{}^{(k)}$ is the k-th decision tree. In RF, sampling training data and explanatory variables allow the construction of decision trees with diversity (i.e., decision trees that are low in correlation with each other), thus reducing the variance of the prediction results. Typical hyperparameters of RF are the number of decision trees K (Eq. (3-25)) and the number of explanatory variables used in the decision tree.

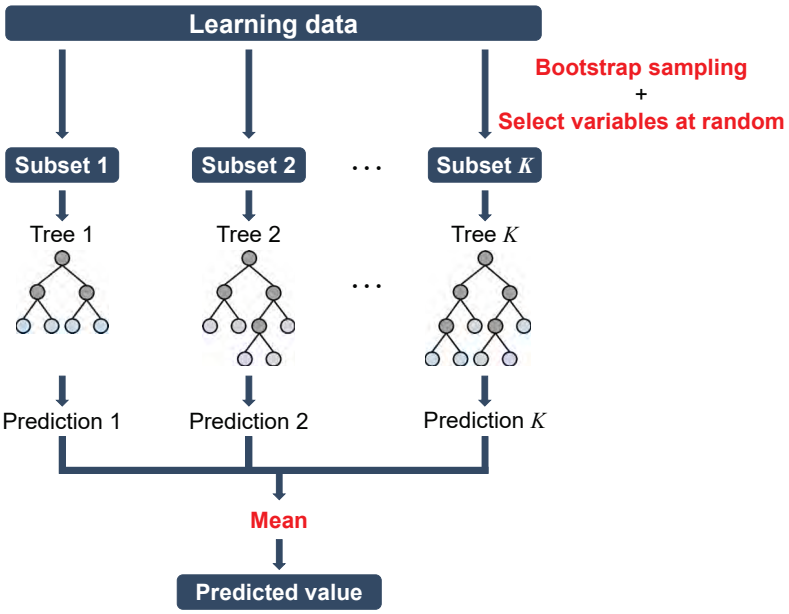


Figure 3-6. Random forest.

The decision tree's node splitting condition for RF in the regression problem is expressed as the minimization of the mean squared error between the training data and the predictions as follows:

$$\min\left\{\frac{1}{n_R}\sum_{j_R\in\mathbf{D_R}}\left(v_{j_R}-\hat{v}_{j_R}\right)^2+\frac{1}{n_L}\sum_{j_L\in\mathbf{D_L}}\left(v_{j_L}-\hat{v}_{j_L}\right)^2\right\} \tag{3-26}$$

where n_R and n_L are the total number of samples divided into two parts at the target node, respectively, and $\mathbf{D_R}$ and $\mathbf{D_L}$ are the sets of training subsets for each of the two parts. $\hat{v}$ is the predicted value of the weight coefficient at the target node.

The advantages of RF are its high performance for most problems, its ease of achieving high accuracy without adjusting hyperparameters, and its ability to account for interactions among explanatory variables. In addition, while the prediction accuracy of machine learning models is greatly affected by the standardization of explanatory variables, RF based on bifurcated decision trees focuses only on large and small values of explanatory variables, thus eliminating the need for standardization. One disadvantage is the relatively large number of hyperparameters, but a certain degree of prediction accuracy can be achieved without tuning.

3.4.6 Gradient Boosting Decision Tree [30–32, 34]

Gradient boosting decision tree (GBDT) is a model that achieves high performance by combining multiple decision trees based on boosting in the ensemble method. Specifically, as shown in Fig. 3-7, the first decision tree is constructed from training data; the second and subsequent decision trees are sequentially trained on the difference between the "objective variable" and the "sum of the predictions of the previously constructed decision trees." The more decision trees are created, the more the bias is reduced, and the weights (influence) of the decision trees gradually decrease. The GBDT-predictor f_{GBDT} (**s**) is described as follows:

$$f_{GBDT}(\mathbf{s}) = \sum_{k=1}^{K} f_{Tree}{}^{(k)}(\mathbf{s}) \tag{3-27}$$

where K is the number of decision trees constituting the GBDT-predictor.

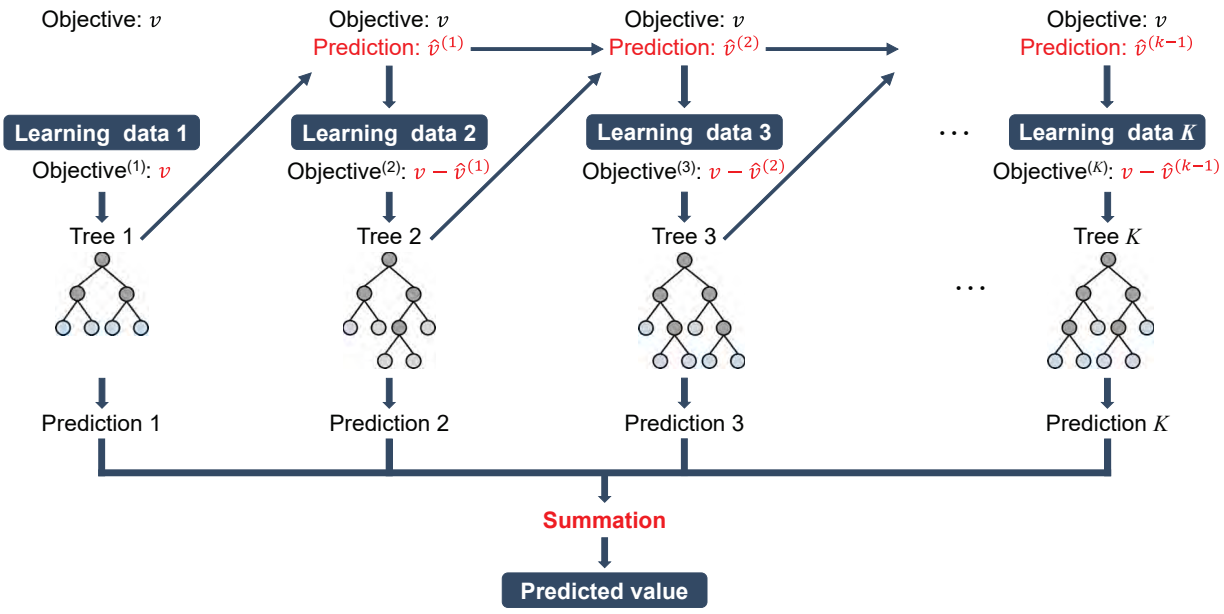


Figure 3-7. Gradient boosting decision tree.

This study uses the framework xgboost [35] to construct the GBDT. The objective function L_{GBDT} in xgboost is formulated as follows.

$$L_{GBDT} = \sum_{j^*=1}^{n^*} l(v_{j^*}, \hat{v}_{j^*}) + \sum_{k=1}^{K} \Omega({f_{Tree}}^{(k)}(\mathbf{s_L}))$$
$$\text{where} \quad l(v_{j^*}, \hat{v}_{j^*}) = \frac{1}{2}\left(v_{j^*} - \hat{v}_{j^*}\right)^2 \tag{3-28}$$
$$\Omega({f_{Tree}}^{(k)}(\mathbf{s_L})) = \gamma T^{(k)} + \frac{1}{2}\lambda \left\|\mathbf{w}^{(k)}\right\|^2 + \alpha \left|\mathbf{w}^{(k)}\right|$$

The loss function l is the error between the objective variable and the predictions, which in the L_{GBDT} is the sum of the errors for all training data. In the regularization term $\Sigma_k \Omega({f_{Tree}}^{(k)}(\mathbf{s_L}))$, γ is a hyperparameter that represents the penalty on the number of leaves $T^{(k)}$ in the k-th decision tree; increasing its value prevents the model from becoming too complex. α and λ are hyperparameters related to the L1 and L2 regularization terms for the leaf weights, respectively, and larger values prevent overlearning. $\mathbf{w}^{(k)} = ({w^{(k)}}_1, \ldots, {w^{(k)}}_T)$ is $T(k)$-dimensional vector whose entries are the values $w^{(k \geq 2)}$ of the leaf nodes in the k-th decision tree (the average value of the training subset contained in the leaf nodes, i.e., the prediction value obtained with ${f_{Tree}}^{(k)}$). Next, the objective function for training the k-th decision tree (i.e., the recurrence formula in Eq. (3-28)) is expressed as follows:

$${L_{GBDT}}^{(k)} = \sum_{j^*=1}^{n^*} l(v_{j^*}, {\hat{v}_{j^*}}^{(k-1)} + {f_{Tree}}^{(k)}(\mathbf{s_L})) + \Omega({f_{Tree}}^{(k)}(\mathbf{s_L})) \tag{3-29}$$

By finding $\mathbf{w}^{(k)}$ such that this ${L_{GBDT}}^{(k)}$ is minimized, the GBDT-predictor $f_{GBDT}(\mathbf{s})$ is constructed. In addition to the regularization term, xgboost also includes the following innovations to suppress overlearning. For example, regularization by shrinkage for the second and subsequent decision trees, sampling of training data and explanatory variables when constructing decision trees, limiting the depth of decision trees, and setting a minimum number of samples when splitting nodes. Among these, the regularization by shrinkage is explained below. In Eq. (3-27), the predictions from the first decision tree are corrected directly with the optimal w obtained in Eq. (3-29); however, there is a high possibility that overlearning will occur if this is not done. Applying a learning rate η ($0 < \eta \leq 1$) to the second and subsequent decision trees gradually compensates for errors in the predictions made by the first decision tree, thereby preventing overlearning. By introducing the learning rate η into Eq. (3-27), the GBDT-predictor can be described as follows:

$$f_{GBDT}(\mathbf{s}) = {f_{Tree}}^{(1)}(\mathbf{s}) + \sum_{k=2}^{K} \eta {f_{Tree}}^{(k)}(\mathbf{s}) \tag{3-30}$$

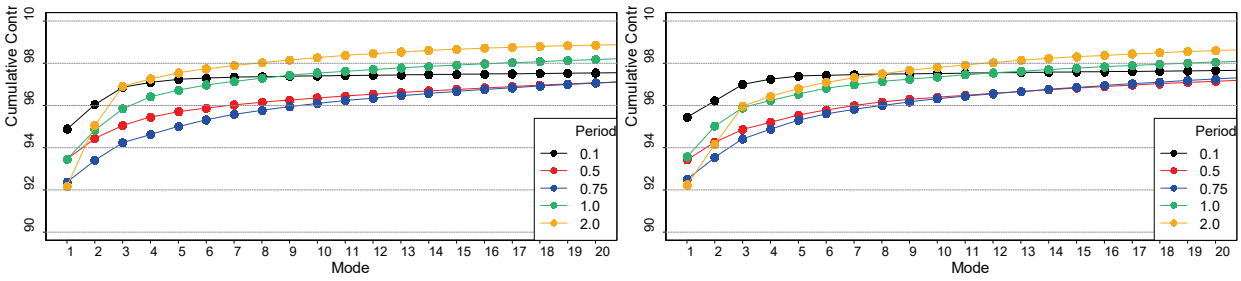
The GBDT has similar advantages to the RF and generally shows higher accuracy than the RF. However, because decision trees in GBDT are serially arranged, model training time is longer, and over-training is more likely to occur than in RF, where decision trees are arranged in parallel. Parameter tuning is also essential to achieve high performance with GBDT, but many hyperparameters make tuning difficult.

3.5 Mode Decomposition of Ground Motion Distributions

As in **Chapter 2**, this chapter presents a numerical example using the distribution of absolute acceleration response spectrum (S_A) of periods 0.1–2.0 s and two horizontal components (FN: Fault-Normal, FP: Fault-Parallel) for 600 cases [36]. The prediction model for ground motion distribution constructed in this study is a surrogate model of the stochastic Green's function method [15], and the target is an earthquake that occurs on a strike-slip fault consisting of a single fault plane.

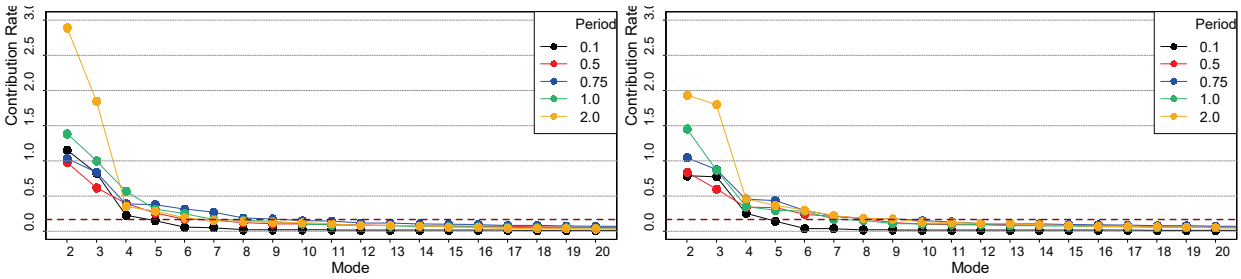
3.5.1 Contribution Ratio of Mode

The mode decomposition method using singular value decomposition is applied to the distributions of S_A for 600 cases by period and component. Fig. 3-8 shows the cumulative ratio of eigenvalues (square of the singular value, which is the diagonal entries of the matrix **D**), and Fig. 3-9 shows the contribution ratio from the 2nd-order mode to the 20th-order mode. First, a common feature of both components is that the contributions of Mode 1 are significantly larger than that of the other. Modes 2–3 also have large contribution in all periods, while Modes 4–6 contributions decrease. The mode contribution ratio exceeds 1/600 (= 0.167 %, the dashed brown line in Fig. 3-9) only for Mode 1 through Modes 4–9; for the higher modes, the information content is less than 1 case. At period 2.0 s, the proportion of Modes 2–3 in both components is larger than in other periods. The reason for this is discussed in **Chapter 4**.



(a) FN component **(b)** FP component

Figure 3-8. Cumulative contribution ratio of the eigenvalue (square of singular value).

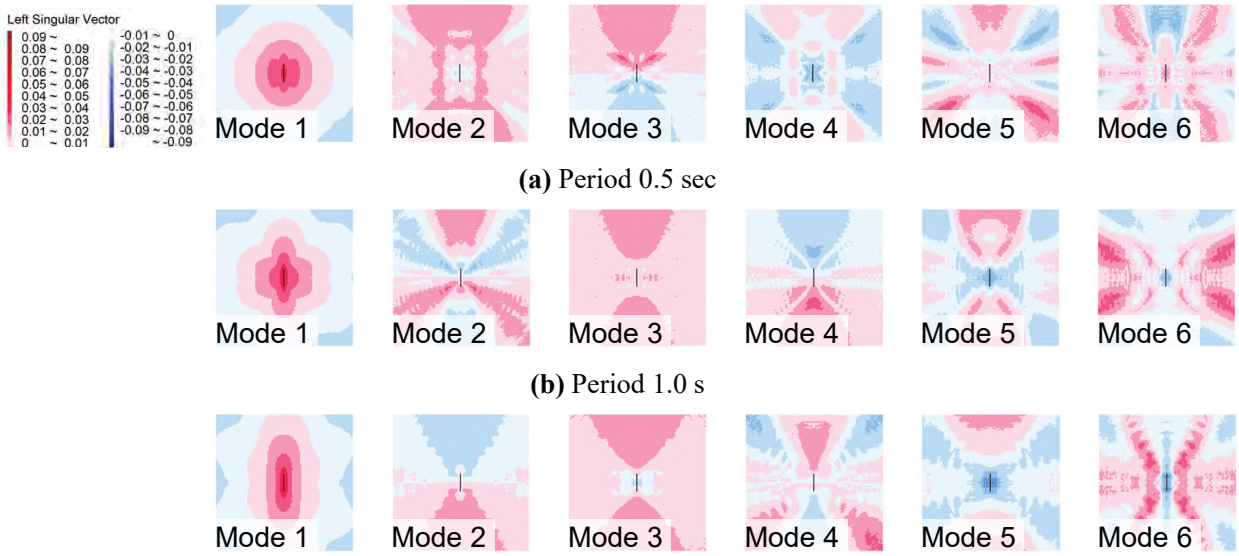


(a) FN component **(b)** FP component

Figure 3-9. Contribution ratio from 2nd to 20th order mode.

3.5.2 Mode Shape

Figure. 3-10 and Fig. 3-11 show the spatial distributions of left singular vectors (the column vector of the matrix **U**), i.e., mode shape, in the FN and FP components, respectively (due to space limitations, only Modes 1–6 with periods 0.5, 1.0, 2.0 s are shown). The mode shape represents the spatial characteristics of ground motion (e.g., distance attenuation of ground motion and expansion of strong ground motion caused by fault rupture propagation). A red point is a positive value, and a blue point is a negative value; a correlation between points with the same (or different) sign is a positive (or negative) correlation. The characteristics of the mode shape and their dominant factors are analyzed in **Chapter 4**.

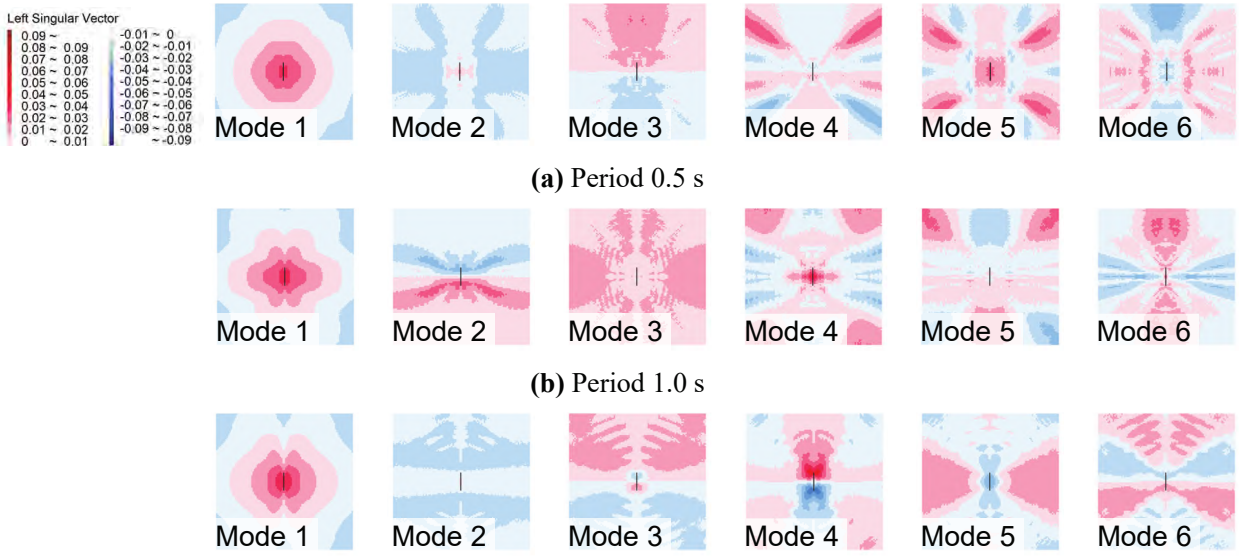


(a) Period 0.5 sec

(b) Period 1.0 s

(c) Period 2.0 s

Figure 3-10. Mode shape in FN component.



(a) Period 0.5 s

(b) Period 1.0 s

(c) Period 2.0 s

Figure 3-11. Mode shape in FP component.

3.5.3 Mean and Standard Deviation of Weight Coefficients

The entries of the right singular vector (the column vector of the matrix **V**) are the weight coefficients that characterize each case. The mean value of the weight coefficients for Mode 1 is 0.0408 ($= N^{-1/2}$), regardless of period or component. The coefficients of variation ranged 1.8–6.4 %, with both components being larger for longer periods. The mean values for Mode 2 and higher are 0, and the standard deviations are 0.0408 ($= N^{-1/2}$).

3.6 Simplified Prediction Model for Ground Motion Distribution based on Mode Decomposition and Machine Learning

3.6.1 Input Data and Hyperparameters

Table. 3-1 summarizes the source parameters (i.e., the features) used in modeling the weight coefficients. Fig. 3-12 shows an image of the location of the hypocenter and asperities for the quantitative variables. Fig. 3-13 shows the images of the classifications of horizontal and vertical arrangement patterns of qualitative variables. These patterns are set up by classifying each case according to the location of large and small asperities and hypocenter. For the horizontal position, let X_L and X_S denote the center of gravity of the large and small asperities, respectively. Similarly, for the vertical position, let Y_L, and Y_S be the centers of gravity of the large and small asperities, respectively.

Horizontal arrangement pattern: four patterns (Fig. 3-13 (a))

- Horizontal pattern 1 (black) : $X_L < X_S$, and the hypocenter is located at the south end of asperity
- Horizontal pattern 2 (red) : $X_L < X_S$, and the hypocenter is located at the north end of asperity
- Horizontal pattern 3 (blue) : $X_L \geq X_S$, and the hypocenter is located at the south end of asperity
- Horizontal pattern 4 (green) : $X_L \geq X_S$, and the hypocenter is located at the north end of asperity

Vertical arrangement pattern: two patterns (Fig. 3-13 (b))

- Vertical pattern 1 (black) : $Y_L < Y_S$
- Vertical pattern 2 (red) : $Y_L \geq Y_S$

For qualitative variables, the MR and SVR were assigned numerical values (0 or 1) by one-hot encoding, while the RF and GBDT were assigned by label encoding according to the pattern numbers shown in Fig. 3-13.

Table 3-1. Features for modeling

Variable Type	Feature		Abbreviation
Quantitative Variable	Seismic moment [dyne*cm] (common logarithmic conversion)		M_0
	Rupture velocity [km/s]		Vr
	Hypocenter (Fig. 3-12 Orange)	Horizontal position [km]	Hypo.H
		Vertical position [km]	Hypo.V
	Large asperity (Fig. 3-12 Blue)	Horizontal position [km]	aspL.H
		Vertical position [km]	aspL.V
	Small asperity (Fig. 3-12 Green)	Horizontal position [km]	aspS.H
		Vertical position [km]	aspS.V
	Ratio between combined asperity area and total rupture area		ratio
Qualitative Variable	Classification by asperities arrangement and hypocenter position (Fig. 3-13)	Horizontal arrangement pattern	Hor.P
		Vertical arrangement pattern	Ver.P

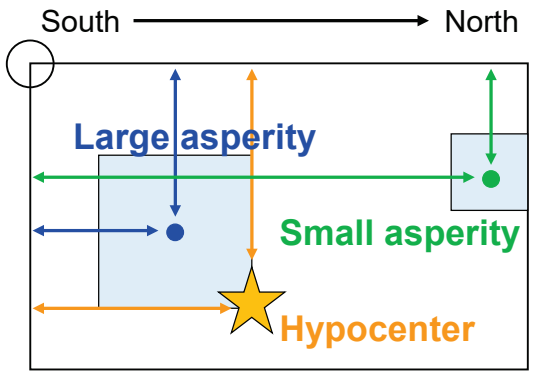
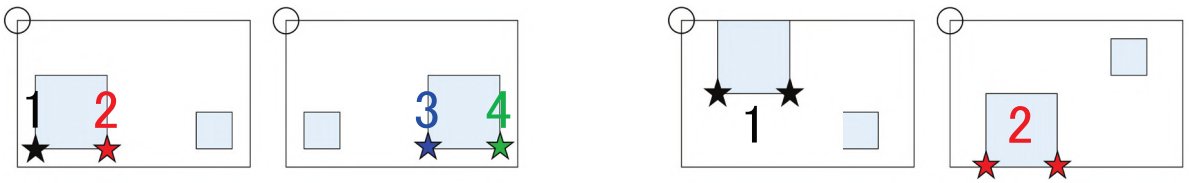


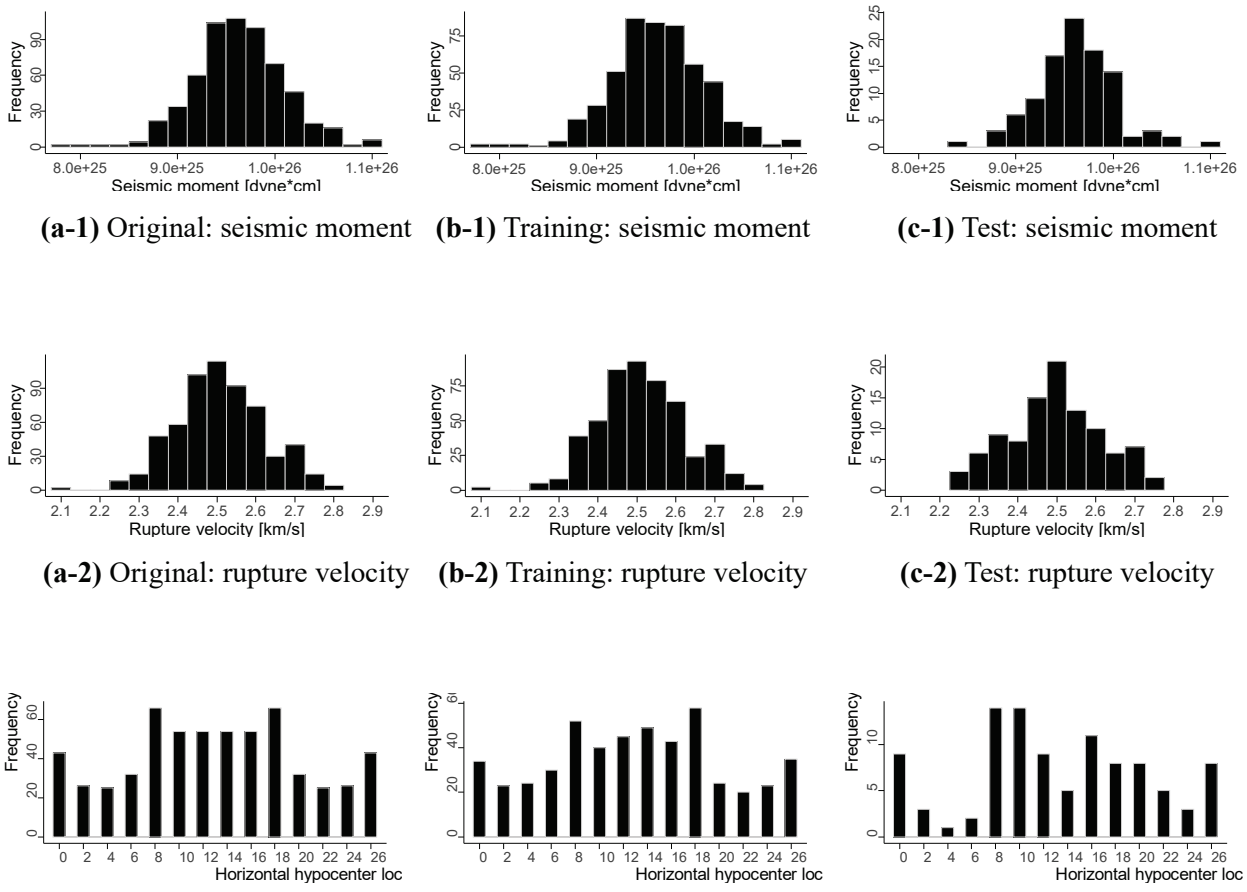
Figure 3-12. Image of the horizontal and vertical location of hypocenter and large and small asperity.



(a) Horizontal pattern classifications **(b)** Vertical pattern classifications

Figure 3-13. Classification by asperities arrangement and hypocenter position.

This study randomly divided the data set of all 600 cases (weight coefficients and source parameters) into 500 cases as training data and 100 cases as test data. Fig. 3-14 shows the seismic moment, rupture propagation velocity, and horizontal hypocenter location for the original (all 600 cases), the training (500 cases) and test (100 cases) data. Compared to the original data (left), the training data (center) and the test data (right) show no bias in the data.



(a-1) Original: seismic moment **(b-1)** Training: seismic moment **(c-1)** Test: seismic moment

(a-2) Original: rupture velocity **(b-2)** Training: rupture velocity **(c-2)** Test: rupture velocity

(a-3) Original: Horizontal hypocenter location **(b-3)** Training: Horizontal hypocenter location **(c-3)** Test: Horizontal hypocenter location

Figure 3-14. Frequency distribution of source parameters in original data (600 cases, left), learning data (500 cases, center) and test data (100 cases, right).

This chapter applied each of the four regression methods (MR, SVR, RF and GBDT) described in ***3.4*** to the training data and constructed predictors of weight coefficient for each period component and mode. The hyperparameters of SVR, RF and GBDT were the same for all five periods, both components and all modes, and were determined exploratively by grid search, subject to minimizing the root mean square error (RMSE) between the weight coefficients and predicted values. Typical hyperparameters for each model are listed below.

SVR: Support Vector Machine

- Regularization parameter C : 10
- Insensitivity parameter ε : 0.05
- Range of influence for training data γ : 0.03

RF: Random Forest

- Number of decision trees K: 500
- Number of explanatory variables in a decision tree:
  Determined by the Function tuneRF() in the program package “randomForest” [37]

GBDT: Gradient Boosting Decision Tree

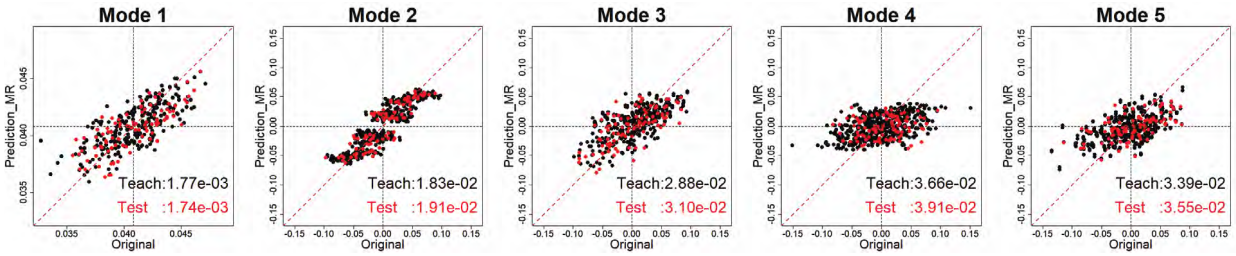
- Number of decision trees K : 1500
- Maximum depth of tree : 4
- Minimum node weights required for branching : 6
- Sampling rate of features in tree :0.8
- Sampling rate of explanatory variables in tree : 0.7
- L1 regularization term α : 0
- L2 regularization term λ : 1
- Learning rate η : 0.01

3.6.2 Weight Coefficient Predictor

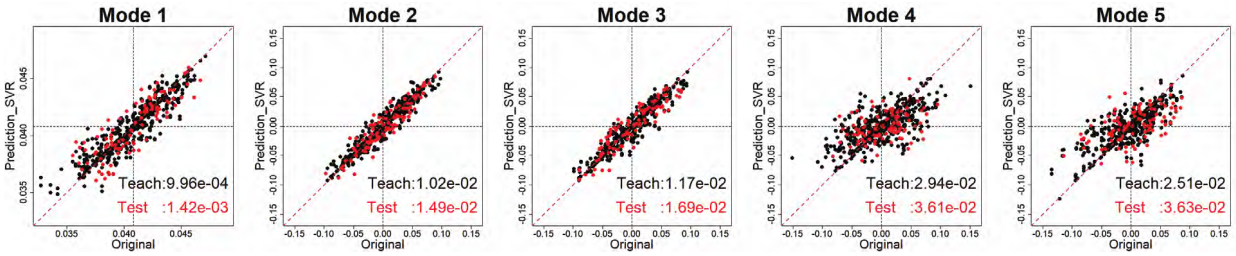
Figure. 3-15 shows the predicted weight coefficients (T = 2.0 s, FN component) by the four predictors (MR, SVR, RF and GBDT). For all three predictors except MR, the predictions for Mode 1–3 are located around the 45-degree line, and the RMSE of the predictions is sufficiently small. Note that, in common with all modes, the prediction tends to be slightly overestimated when the weight coefficient (correct value) is small (left side of the horizontal axis) and slightly underestimated when the weight coefficient is large (right side of the horizontal axis). This is due to the nature of the prediction method itself (the prediction of RF or GBDT: the average or sum of the leaves of the decision trees), and it is possible that the error will decrease as the training data increases.

The weight coefficients for Modes 1–3, which have the largest contributions (Modes 1–3 in order: 92.16%, 2.89%, 1.85%), are accurately modeled by the source parameters. In contrast, some modes above Mode 4 have larger RMSEs, and the range of predictions tended to be narrower relative to the range of correct values (i.e., the explanatory power of the source parameters was low). These trends were common to all four methods and were more pronounced for the higher-order modes. However, the contributions of these modes (Modes 4–6 in order: 0.35%, 0.28%, 0.19%) are sufficiently small to have a small effect on the ground motion distribution. The same trend was observed for the FN component with periods 0.1–1.0 s and the FP component.

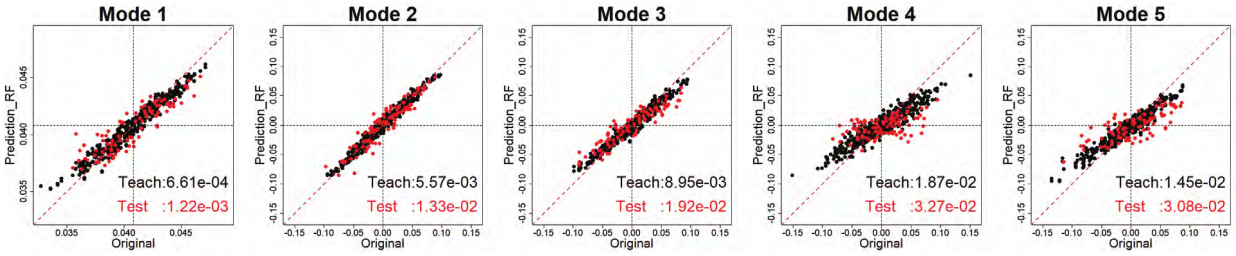
Figure. 3-16 and Fig.3-17 show the RMSEs of the predicted weight coefficients for Modes 1–10 in the FN and FP components, respectively. The RMSEs for the lower-order modes are small, while the RMSEs for the high-order modes are large. The RMSEs of the RF- and GBDT-predictors are often smaller than the MR- and SVR-predictors, regardless of period, component, or mode.



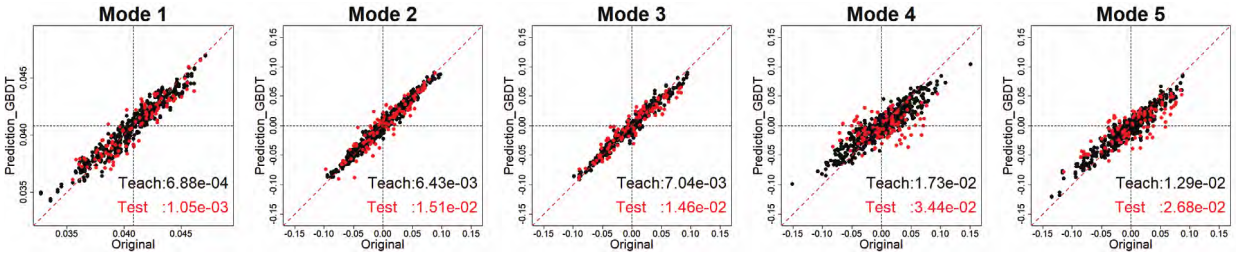
(a) MR-predictor



(b) SVR-predictor



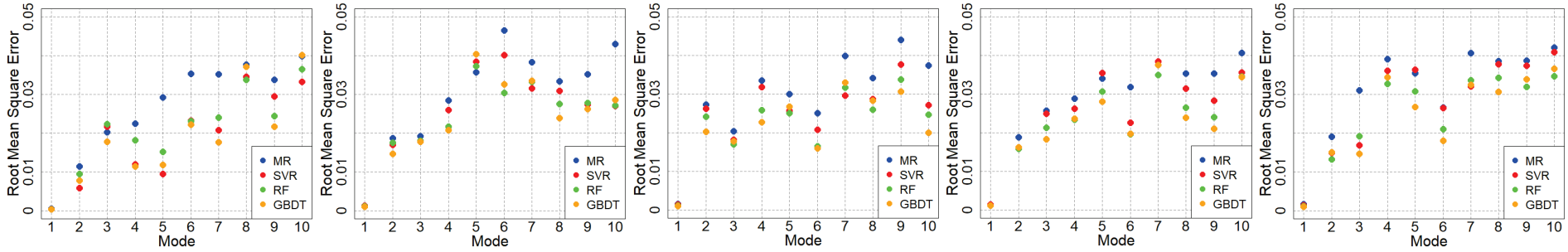
(c) RF-predictor



(d) GBDT-predictor

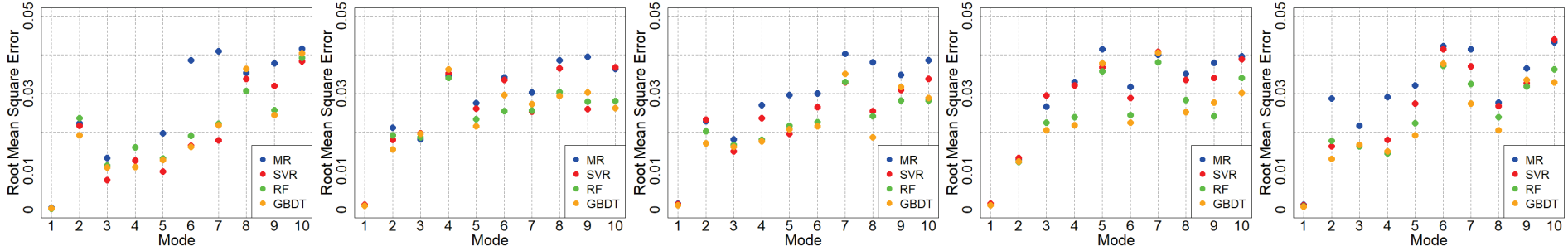
(Black: training data, red: test data, black dashed line: mean of weights coefficients for 600 cases)

Figure 3-15. Scatter plot of weight coefficients (horizontal axis) and predicted values (vertical axis) for 600 cases in period 2.0 s - FN component.



(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

Figure 3-16. RMSE of predicted weight coefficients for 100 test cases in FN component.



(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

Figure 3-17. RMSE of predicted weight coefficients for 100 test cases in FP component.

3.6.3 Consideration of Number of Modes to be Used in Prediction Model

Next, the ground motion distribution is predicted by mode synthesis of the weight coefficients obtained by inputting any source parameters into the weight coefficient predictor. Fig. 3-18 shows the RMSE between the ground motions from the SGM simulation and those predicted by the GBDT-model for each number of modes used (synthesized) in the prediction. The decrease in RMSE with the increase in the number of modes used for prediction is particularly pronounced for Modes 1–3. However, some decrease is also observed for the higher modes. Thus, this study uses up to Mode 15 as a sufficient number of modes for the prediction and constructs a model for the ground motion distribution. The next section provides some examples of prediction models.

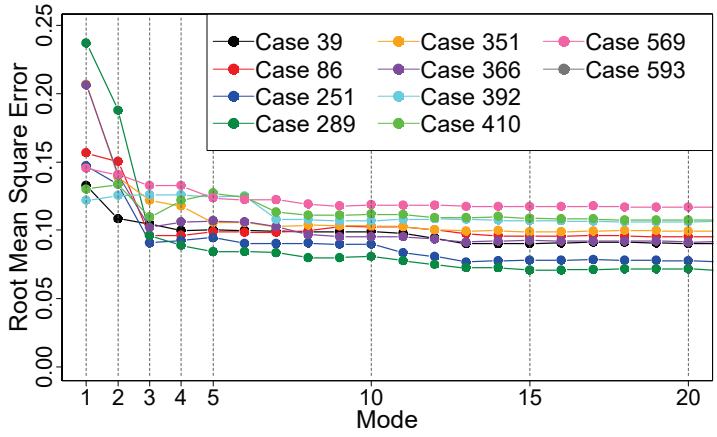


Figure 3-18. RMSE of ground motion distribution (T = 2.0 s, FN component) by number of modes used for prediction in GMPR-model.

3.7 Prediction Performance Evaluation of Prediction Model for Ground Motion Distribution

The prediction performance is evaluated by comparing the ground motion distribution predicted by the prediction model in this study with the distribution calculated by the SGM simulation for the test data. Fig. 3-19 shows the five fault rupture scenarios included in the test data; Fig. 3-20 shows the ground motion distributions of the SGM simulation and predicted by the four prediction models (MR, SVR, RF and GBDT) for FN component in period 2.0 s.

The fault models of Case 86 and Case 289 (Fig. 3-19, left two) show that rupture propagates unilaterally from the southern end of the fault bottom, where the hypocenter is located, to the northern end of the fault top. In the ground motion distribution calculated by SGM simulation (Fig. 3-20 (a)), the strong-motion area, $\log_{10}(S_A\ [cm/s^2]) = 1.2$ (yellow area in the figure) and larger, spreads from the fault to the northern part of the map, reflecting this rupture process. The ground motion distributions predicted by the prediction model (Figs. 3-20 (b)-(e)) also reproduce the spreading of the strong motion zone from the fault to the northern part of the map. Next, attention is paid to the difference between the fault rupture scenarios in these two cases. The fault model of Case 86 has large and small asperities in the direction of rupture progression from the bottom to the top of the fault, while Case 289 has asperities in the direction of rupture progression from the southern to the northern part of the fault. Therefore, in the distributions calculated by SGM simulation, Case 289 shows a stronger impact of forward directivity effect to the north than Case 86. All three prediction models except MR-model reproduce this difference between cases.

In Case 351 (Fig. 3-19 and Fig. 3-20, middle), the ground motion distribution by SGM simulation reflects the effect of the unilateral rupture propagation from the northern to the southern part of the fault, and the strong motion zone spreads from the fault to the south. The distributions by the prediction models also reproduce this spatial trend.

The fault models of Case 410 and Case 569 (Fig. 3-19, right two) show that the hypocenter is located near the center of the fault; the rupture propagates bilaterally from the center of the fault in the north-south direction. The distributions by the SGM simulation (Fig. 3-20 (a)) show that strong motion zone spreads in the north-south direction. The distribution by the prediction models (Figs. 3-20 (b)–(e)) also reproduces the north-south bilateral rupture propagation. Case 410 has a large asperity in the southern part of the fault, and the SGM simulation results show particularly strong ground motion in the southern part of the map. Case 569, in contrast, has a large asperity in the northern part of the fault, and the SGM simulation results are stronger in the northern part of the map. The differences between these cases also appear in the prediction models.

The results show that the ground motion distribution by the prediction model can represent different fault rupture scenarios and roughly reproduce the fault rupture's spatial trend.

Figure. 3-21 shows the scatter plots of the ground motion calculated by SGM simulation and predicted by the prediction model. The MR tends to overestimate the strong ground motion, and its RMSE is larger than the other three prediction models. In contrast, the SVR, RF and GBDT results are located around the 45-degree line, indicating that they can reproduce the strong ground motions.

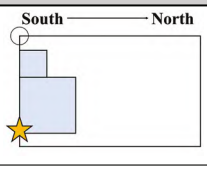
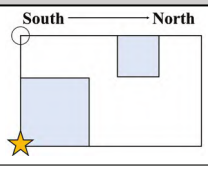
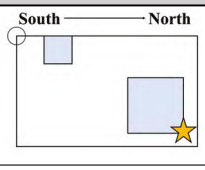
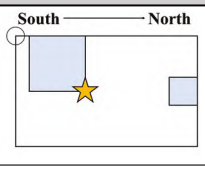
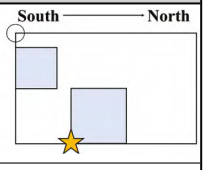
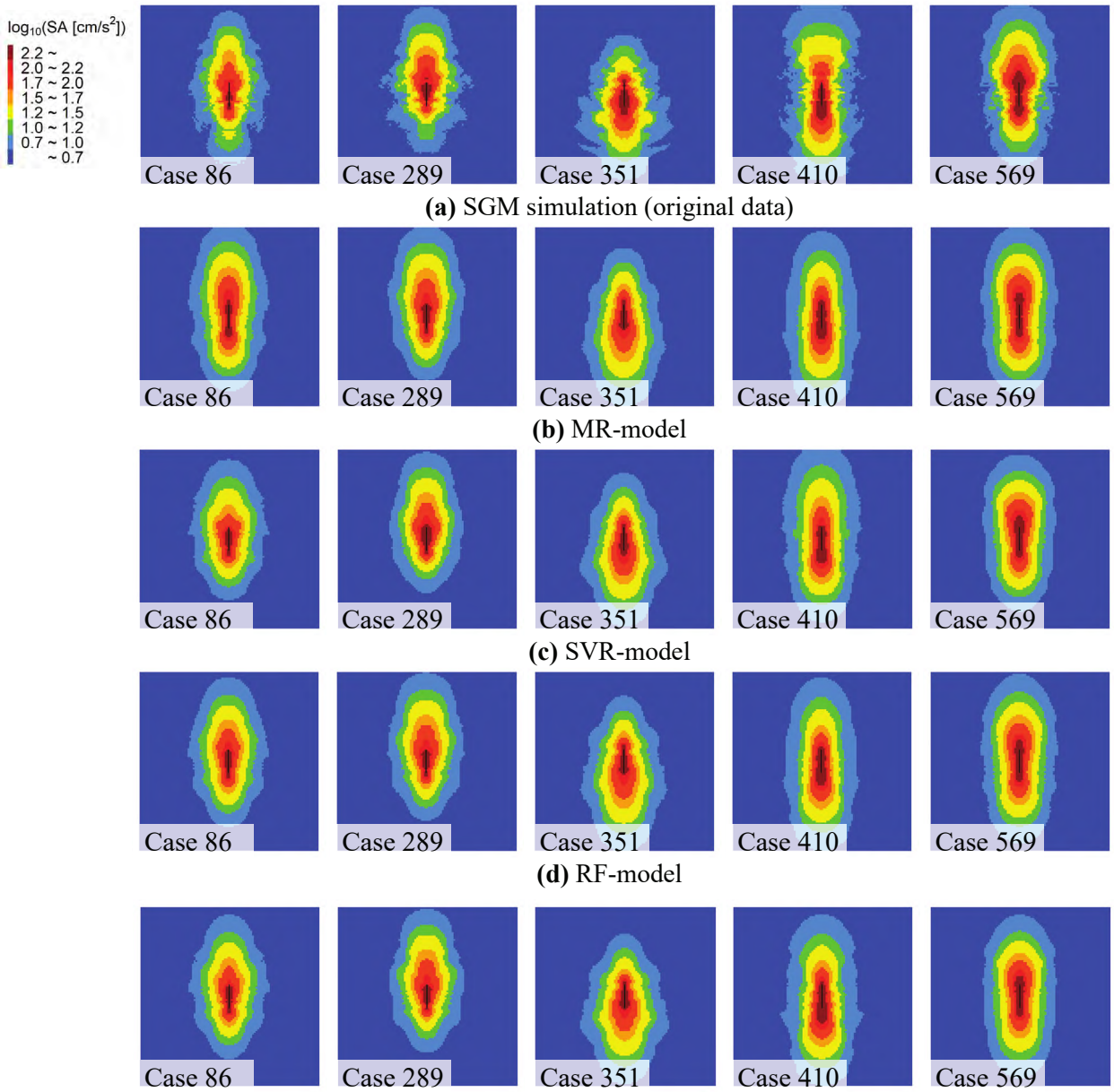
	Case 86	Case 289	Case 351	Case 410	Case 569
Fault model					
M_0 [dyne·cm]	9.53×10^{25}	9.14×10^{25}	9.54×10^{25}	10.0×10^{25}	9.44×10^{25}
Vr [km/s]	2.37	2.50	2.45	2.50	2.56
ratio	0.19	0.33	0.19	0.19	0.24
Hor.P	3	1	4	2	1
Ver.P	2	2	2	1	2

Figure 3-19. Some fault rupture scenarios in test data.



(a) SGM simulation (original data)

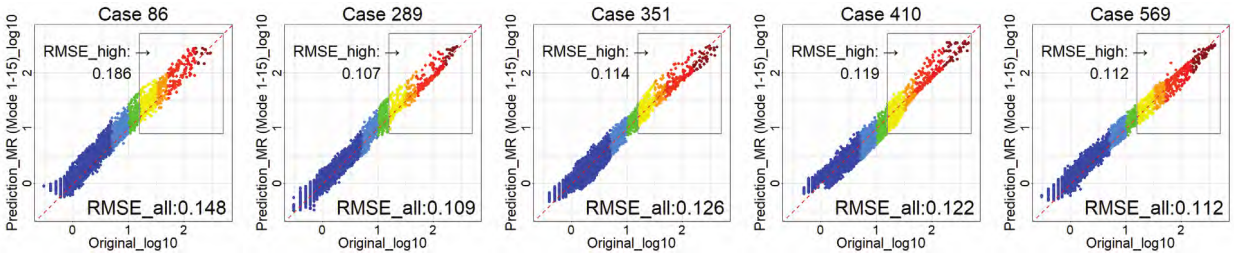
(b) MR-model

(c) SVR-model

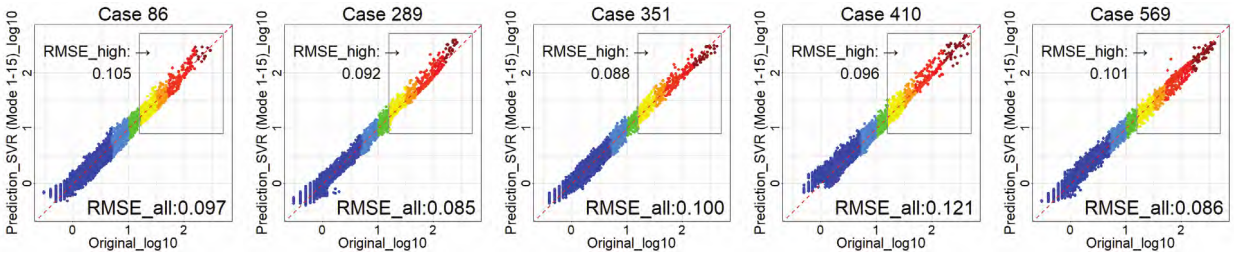
(d) RF-model

(e) GBDT-model

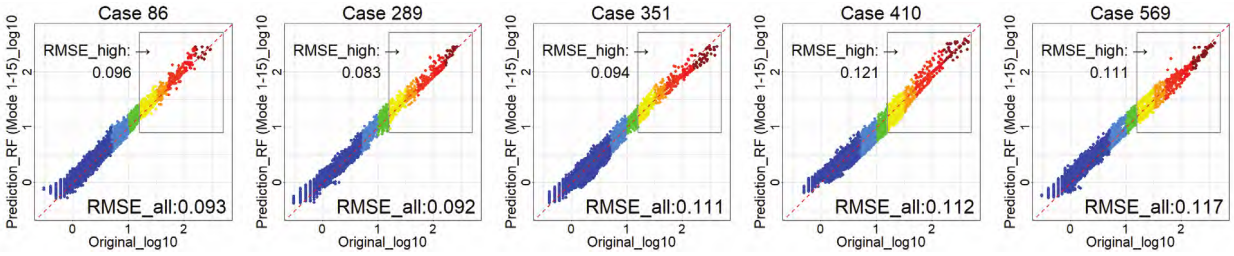
Figure 3-20. Distribution of S_A (T = 2.0 s, FN component).



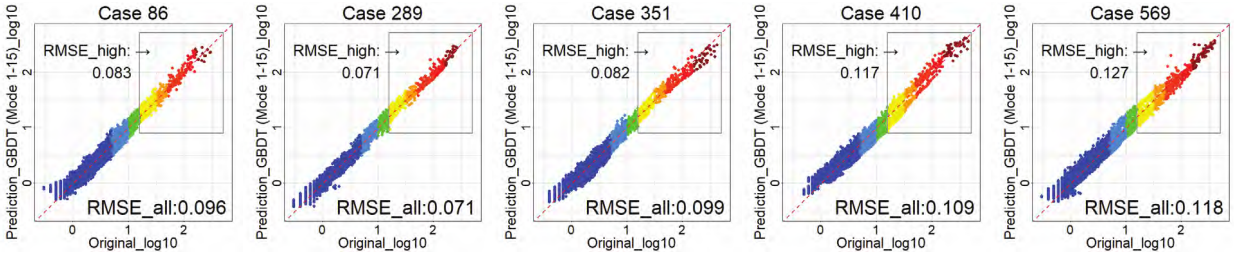
(a) MR-model



(b) SVR-model



(c) RF-model



(d) GBDT-model

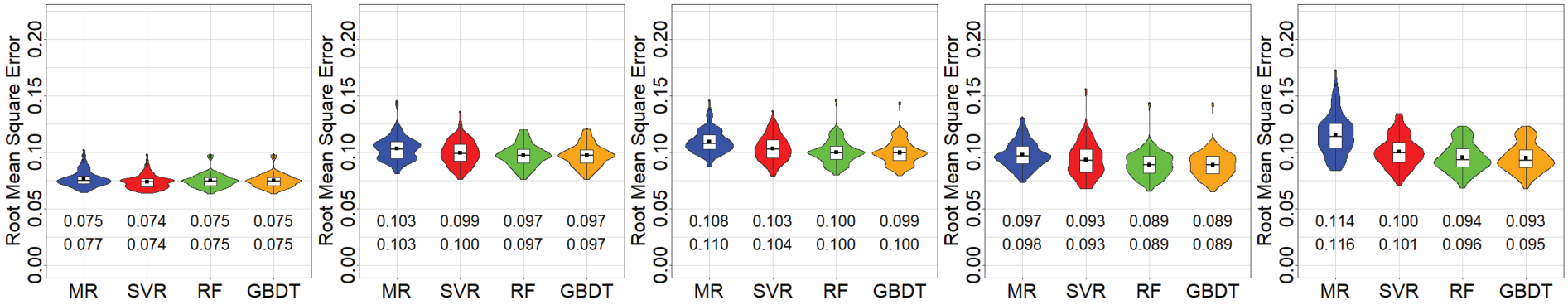
(Color coding follows original S_A value calculated by SGM simulation (a legend in Fig. 3-20). "RMSE_all" is the predicted RMSE for the entire area, and "RMSE_high" is the RMSE for the strong-motion points (gray box; horizontal axis ≥ 1.2).)

Figure 3-21. Scatter plot of S_A (T = 2.0 s, FN component) calculated by SGM simulation (horizontal axis) and predicted values (vertical axis).

Finally, the predicted accuracies of the four prediction models are compared. Fig. 3-22 shows the RMSE distribution of predicted ground motion of FN component in the entire map for the 100 cases of test data. Furthermore, Fig. 3-23 shows the RMSE distribution of FN component limited to the points with strong-motion (periods 0.1–2.0 s in order: $\log_{10}(S_A\ [cm/s^2]) = 2.5$, 2.2, 2, 1.7, and 1.2 and larger) for the test data; the RMSE at period 2.0 s corresponds to the gray box in Fig. 3-21. Fig. 3-24 and Fig. 3-25 also show the RMSE distribution of FP component in the entire map and the strong-motion area, respectively. The model with the highest prediction accuracy is the GBDT-model, which is based on an ensemble method of decision trees. The RF-model, which is also based on decision trees, has the second-highest prediction accuracy. The third most accurate method is the SVR-model, which can handle a wide range of data using a kernel method. The difference between the RF- and SVR-model is small for RMSE in the strong-motion area. Finally, the MR-model has the lowest prediction accuracy because it cannot express nonlinearity.

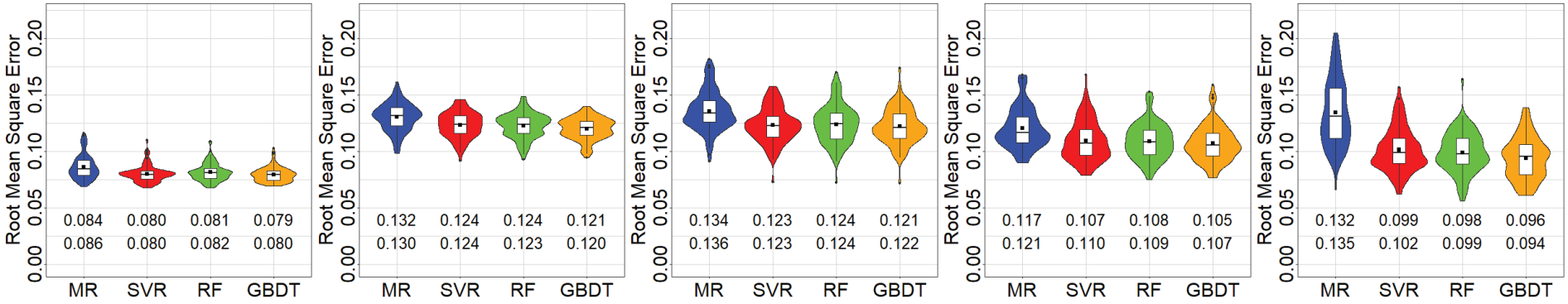
The differences in prediction accuracy among the models are distinct for longer periods when the directivity effect and the radiation pattern are more significant. Since these characteristics are strongly influenced by the combined effects of multiple source parameters, the RF- and GBDT-model, which can reflect the interaction between the factors, are considered to have higher prediction accuracy than the other two models. In contrast, there is no clear differences among the four models for period 0.1 s. This trend may be attributed to the dominant effect of distance attenuation for ground motions in period 0.1 s, which has a relatively simple relationship with the source parameters.

These results show that a simplified prediction model based on mode decomposition and machine learning can predict ground motion distribution reflecting the source characteristics, such as the fault rupture process, by inputting arbitrary fault rupture scenarios (source parameters) into the model. In other words, a surrogate model for a detailed method (in this case, the stochastic Green's function method [15]) could be constructed by combining mode decomposition and machine learning. As the prediction model, the GBDT-model, which shows high prediction accuracy for all periods and components, should be adopted.



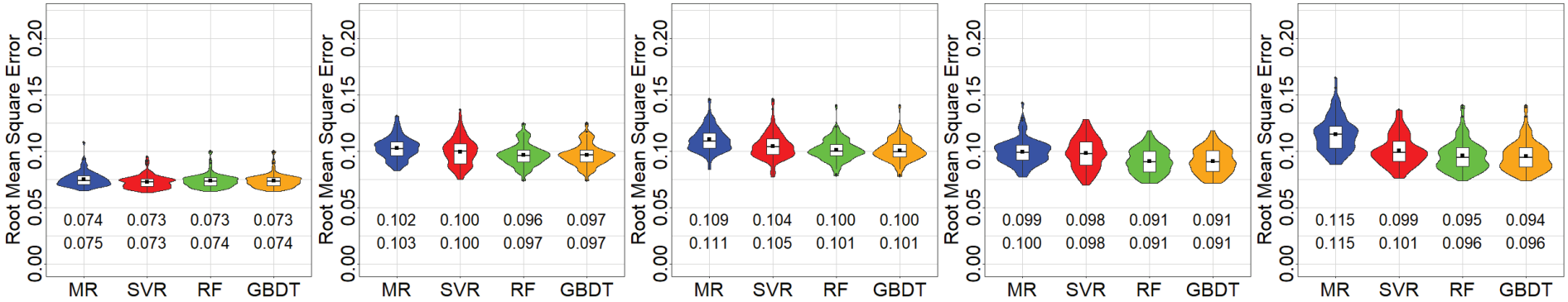
(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

Figure 3-22. RMSE distribution of predicted S_A of FN component in entire map for 100 test cases.



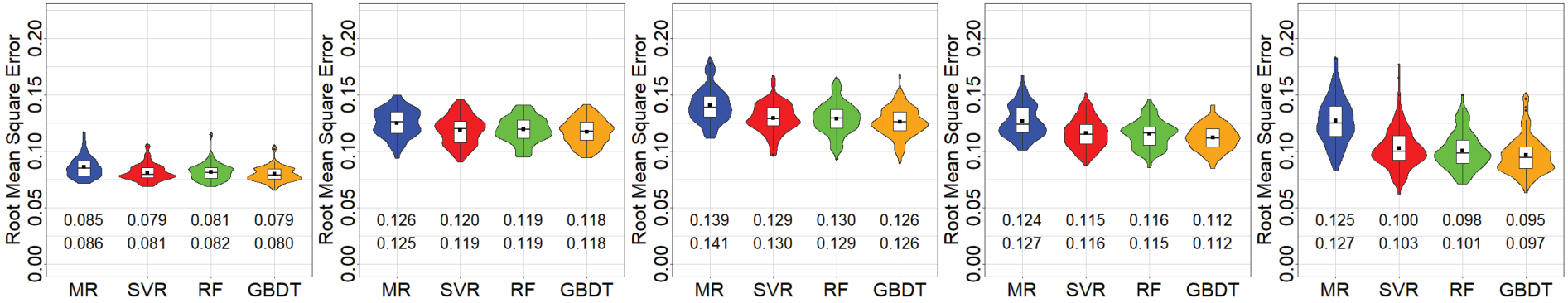
(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

Figure 3-23. RMSE distribution of predicted S_A of FN component in strong-motion area for 100 test cases.



(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

Figure 3-24. RMSE distribution of predicted S_A of FP component in the entire map for 100 test cases.



(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

Figure 3-25. RMSE distribution of predicted S_A of FP component in the strong-motion area for 100 test cases.

3.8 Conclusion

This chapter proposes a simplified prediction model of ground motion distribution based on mode decomposition and machine learning as a surrogate model for a detailed method based on SGM simulation. As a numerical example, a prediction model was constructed for the distribution of the absolute acceleration response spectrum (periods 0.1–2.0 s, FN and FP components) [36] for a strike-slip fault calculated by the stochastic Green's function method [15]. The major conclusions are listed below.

1) Mode decomposition was applied to the distributions of S_A by period and component, then mode shapes, singular values and weight coefficients were obtained. Next, a weight coefficient predictor was constructed using the source parameters as explanatory variables. Multiple regression, support vector regression, random forest, and gradient boosting decision trees were used for the modeling, respectively. By inputting arbitrary fault rupture scenario (source parameters) into the predictor, the three predictors, except MR, could predict the weight coefficients for Modes 1–3, which have high contribution ratios, with high prediction accuracy.

2) By mode synthesis of the predicted weight coefficients, mode shapes, and singular values, the ground motion distribution for arbitrary fault rupture scenario was predicted. The predicted distribution shows spatial characteristics of ground motion corresponding to the fault rupture scenario, which roughly reproduces the ground motion distribution calculated by the detailed method.

3) The differences in prediction accuracy among the four models were not seen in period 0.1 s where the effect of distance attenuation was dominant. However, they were seen in long periods where the directivity effect and radiation pattern were more significant than in short periods. The models with the highest prediction accuracy among the four models were GBDT.

This chapter unifies the hyperparameters for all periods, all components, and all modes for simplicity. Therefore, the prediction accuracy is expected to be improved by tuning the hyperparameters for each mode and adding new feature (i.e., "feature engineering"). The prediction model of ground motion distribution developed in this chapter is based on a single event caused by a single strike-slip fault. Therefore, the future study will construct a prediction model for earthquakes on various fault types, including reverse fault. Furthermore, applying this prediction model to seismic risk assessment will be the subject of future developments.

References in Chapter 3

[1] Fukushima, Y. and Tanaka, T. (1990) A new attenuation relation for peak horizontal acceleration of strong earthquake ground motion in Japan. *Bull. Seism. Soc. Am.*, **80** (4): pp. 757–783. https://doi.org/10.1785/BSSA0800040757

[2] Si, H. and Midorikawa, S. (1999) New attenuation relationships for peak ground acceleration and velocity considering effects of fault type and site condition. *J. Struct. Constr. Eng. Trans. AIJ*, **64** (523): pp. 63–70. https://doi.org/10.3130/aijs.64.63_2 **(in Japanese with English abstract)**

[3] Kanno, T., Narita, A., Morikawa, N., Fujiwara, H. and Fukushima, Y. (2006) A new attenuation relation for strong ground motion in Japan based on recorded data. *Bull. Seism. Soc. Am.*, **96** (3): pp. 879–897. https://doi.org/10.1785/0120050138

[4] Morikawa, N. and Fujiwara, H. (2013) A new ground motion prediction equation for Japan applicable up to M9 mega-earthquake. *JDR*, **8** (5): pp. 878–888. https://doi.org/10.20965/jdr.2013.p0878

[5] Kagawa, T. (2009) A study on empirical ground motion estimation introducing fault rupture propagation effects comparing with ground motions by fault rupture model. *Journal of JSCE A1*, **65** (1): pp. 59–64. https://doi.org/10.2208/jscejseee.65.59 **(in Japanese with English abstract)**

[6] Oji, S., Kanbara, T., Sawada, S. and Iwata, T. (2009) Attenuation relationship by considered the effect of directivity based on equivalent hypocentral distance. *Journal of JSCE A1*, **65** (1): pp. 104–110. https://doi.org/10.2208/jscejseee.65.104 **(in Japanese with English abstract)**

[7] Satoh, T. (2011) Correlation-factor model of rupture directivity and radiation pattern for empirical ground-motion attenuation relations of crustal earthquakes. *J. Struct. Constr. Eng. Trans. AIJ.*, **76** (661): pp. 499–508. https://doi.org/10.3130/aijs.76.499 **(in Japanese with English abstract)**

[8] Abrahamson, N. A., Silva, W. J. and Kamai, R. (2014) Summary of the ASK14 ground motion relation for active crustal regions. *Earthq. Spectra.* **30** (9): pp. 1025–1055. https://doi.org/10.1193/070913EQS198M

[9] Boore, D. M., Stewart, J. P., Seyhan, E. and Atkinson, G. M. (2014) NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthq. Spectra*, **30** (3): pp. 1057–1085. https://doi.org/10.1193/070113EQS184M

[10] Campbell, K. W. and Bozorgnia, Y. (2014) NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthq. Spectra*, **30** (3): pp. 1087–1115. https://doi.org/10.1193/062913EQS175

[11] Chiou, B. S.-J. and Youngs, R. R. (2014) Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthq. Spectra*, **30** (3): pp. 1117–1153. https://doi.org/10.1193/072813EQS219M

[12] Hartzell, S. T. (1977) Earthquake aftershocks as Green's functions. *Geophys. Res. Lett.*, **5** (1): pp. 1–4. https://doi.org/10.1029/GL005i001p00001

[13] Kamae, K., Irikura, K. and Pitarka, A. (1998) A technique for simulation strong ground motion using hybrid Green's Function. *Bull. Seism. Soc. Am.*, **88** (2): pp. 357–367. https://doi.org/10.1785/BSSA0880020357

[14] Irikura, K. and Kamae, K. (1999) Strong ground motions during the 1948 Fukui Earthquake –estimation of broad-band ground motion using a hybrid simulation technique–, *Zishin*, **52** (1): pp. 129–150. https://doi.org/10.4294/zisin1948.52.1_129 **(in Japanese with English abstract)**

[15] Kagawa, T. (2004) Developing a Stochastic Green's function method having more accuracy in long period range to be used in the Hybrid Method. *Journal of JAEE*, **4** (2): pp. 21–32. https://doi.org/10.5610/jaee.4.2_21 **(in Japanese with English abstract)**

[16] Kong, Q., Trugman, T. D, Ross, E. Z., Bianco, J. M., Meade, J. B. and Gerstoft, P. (2019) Machine learning in

seismology: turning data into insights. *Seismological Research Letters*, **90** (1): pp. 3–14. https://doi.org/10.1785/0220180259

[17] Miyamoto, T., Asakawa, T., Kubo, H., Nomura, Y. and Miyamori, Y. (2020) Research trends in machine learning from the perspective of disaster prevention applications. *Artificial Intelligence and Data Science*, **1** (J1): pp. 242–251. https://doi.org/10.11532/jsceiii.1.J1_242 **(in Japanese with English abstract)**

[18] Kerh, T. and Ting, S. B. (2005) Neural network estimation of ground peak acceleration at stations along Taiwan high-speed rail system. *Eng. Appl. Artif. Intell.*, **18** (7): pp. 857–866. https://doi.org/10.1016/j.engappai.2005.02.003

[19] Garcia, S. R., Romo, M. P., and Mayoral, J. M. (2007) Estimation of peak ground accelerations for Mexican subduction zone earthquakes using neural networks. *Geofis. Int.*, **46** (1): pp. 51–63. https://doi.org/10.22201/igeof.00167169p.2007.46.1.2151

[20] Güllü, H. and Erçelebi, E. (2007) A neural network approach for attenuation relationships: an application using strong ground motion data from Turkey. *Eng. Geol.*, **93** (3): pp. 65–81. https://doi.org/10.1016/j.enggeo.2007.05.004

[21] Derras, B., Bard, P. Y., Cotton, F. and Bekkouche, A. (2012) Adapting the Neural Network Approach to PGA Prediction: an Example Based on the KiK-net Data. *Bull. Seism. Soc. Am.*, **102** (4): pp. 1446–1461. https://doi.org/10.1785/0120110088

[22] Derras, B., Bard, P. Y. and Cotton, F. (2014) Towards fully data driven ground-motion prediction models for Europe. *Bull. Earthquake Eng.*, **12** (1): pp. 495–516. https://doi.org/10.1007/s10518-013-9481-0

[23] Derras, B., Bard, P. Y. and Cotton, F. (2016) Site-conditions proxies, ground-motion variability and data-driven GMPEs: insights from NGA-West 2 and RESORCE data sets. *Earthq. Spectra*, **32** (4): pp. 2027–2056. https://doi.org/10.1193/060215EQS082M

[24] Thomas, S., Pillai, G. N., Pal, K. and Jagtap, P. (2016) Prediction of ground motion parameters using randomized ANFIS (RANFIS). *Appl. Soft Comput.*, **40**: pp. 624–634. https://doi.org/10.1016/j.asoc.2015.12.013

[25] Dhanya, J. and Raghukanth, S. T. G. (2018) Ground motion prediction model using artificial neural network. *Pure Appl. Geophys.*, **175**: pp. 1035–1064. https://doi.org/10.1007/s00024-017-1751-3

[26] Derakhshani, A. and Foruzan, A. H. (2019) Predicting the principal strong ground motion parameters: a deep learning approach. *Appl. Soft Comput.*, **80**: pp. 192–201. https://doi.org/10.1016/j.asoc.2019.03.029

[27] Withers, K. B., Moschetti, M. P. and Thompson, E. M. (2020) A machine learning approach to developing ground motion models from simulated ground motions. *Geophys. Res. Lett.*, **47** (6): Article number: e2019GL086690. https://doi.org/10.1029/2019GL086690

[28] Kubo, H., Kunugi, T., Suzuki, W., Suzuki, S. and Aoi, S. (2020) Hybrid predictor for ground-motion intensity with machine learning and conventional ground motion prediction equation. *Sci. Rep.*, **10**: Article No. 11871. https://doi.org/10.1038/s41598-020-68630-x

[29] Nojima, N. Kuse, M. and Duc, L. Q. (2018) Mode decomposition and simulation of strong ground motion distribution using singular value decomposition. *Journal of JAEE*, **18** (2): pp. 95–114. https://doi.org/10.5610/jaee.18.2_95 **(in Japanese with English abstract)**

[30] Lantz, B. (2019) Machine learning with R - third edition. *Packt Publishing ltd*: 458 p. (Translation: Quipu (2021) Tokyo, *Shoeisha Ltd.*: 440 p.) (**in Japanese**)

[31] Bishop, M. C. (2006) Pattern recognition and machine learning. New York, *Springer*: 738 p. (Translation: Motoda, H., Kurita, T., Higuchi, T., Matsumoto, Y. and Murata, N. (2012) Pattern recognition and machine learning – the first. Tokyo, *Maruzen Publishing Co., Ltd.*: 370 p. (**in Japanese**)

[32] Bishop, M. C. (2006) Pattern recognition and machine learning. New York, *Springer*: 738 p. (Translation: Motoda, H., Kurita, T., Higuchi, T., Matsumoto, Y. and Murata, N. (2012) Pattern recognition and machine

learning – the second. Tokyo, *Maruzen Publishing Co., Ltd.*: 450 p. (**in Japanese**)

[33] Takeuchi, I. and Toriyama, M. (2015) Machine learning professional series – support vector machine. Tokyo, *Kodansha Ltd.*: 189 p. (**in Japanese**)

[34] Kadowaki, D, Sakata, R., Hosaka, K. and Hiramatsu, Y. (2019) Winning data analysis techniques with Kaggle. Tokyo, *Gijutsu-Hyoron Co., Ltd.*: 424 p. **(in Japanese)**

[35] Chen, T. and He, T. (2022) Xgboost: extreme gradient boosting. https://cran.r-project.org/web/packages/xgboost/xgboost.pdf

[36] Kagawa, T. (2015) Spatial variability of periodic characteristics among strong ground motions derived from multiple fault rupture scenarios. *Journal of JAEE*, **15** (7): pp. 90–99. https://doi.org/10.5610/jaee.15.7_90 **(in Japanese with English abstract)**

[37] Liaw, A. and Wiener, M. (2002) Classification and Regression by randomForest. *R News*, **2** (3): pp. 18-22.

4 Analysis of Period and Component Dependent Spatial Characteristics of Ground Motion Distributions Using Mode Decomposition and Machine Learning

4.1 Introduction

The research objective is to cover a wide variety of possible ground motion distributions, including those that may occur in the future, in several pre-estimated scenario ground motion distributions. Since the diversity of ground motion distributions depends on the fault rupture scenarios, it is necessary to clarify the relationship between the set values of the fault rupture scenarios (source parameters) and the ground motion distributions. Therefore, this chapter analyzes the effect of variations in the source parameters on the ground motion distributions.

There have been many quantitative evaluations on the effect of different source parameters (e.g., location, area, and intensity of asperities, location of the hypocenter, and rupture velocity) on the spatial distribution of ground motion variability [1–8] (see **Chapter 1** for details). These studies [1–8] focused on the spatial distribution of variability. In contrast, this chapter presents an analysis focusing on the spatial distribution of variation and its spatial correlation structure (hereafter referred to as "spatial characteristics").

A quantitative evaluation method for spatial characteristics of several ground motion distributions is the "mode decomposition method using singular value decomposition" proposed by Nojima et al. [9]. This method is unique and distinctive because it can evaluate the spatial correlation structure for each mode. We have previously used this method to analyze the spatial characteristics of measured seismic intensity distributions (Scenario Earthquake Shaking Maps) for 4–12 cases of fault rupture scenarios [9–11]. However, the limited number of scenarios has limited the diversity of the variability that can be expressed. In addition, because the measured seismic intensity was used as the target, it was impossible to examine the period dependence and component dependence.

This chapter evaluates the spatial characteristics of various ground motion distributions by period and components. As before, the distributions of absolute acceleration response spectrum (S_A) by period and component for 600 cases caused by a strike-slip fault [7] are used for the analysis. In addition, this chapter attempts to clarify the combined effects of multiple source parameter settings on the ground motion distributions by considering the interaction effects among the source parameters and the resulting spatial characteristics for each case. Specifically, some machine learning model's interpretation methods called "explainable AI" is applied to the weight coefficient predictor of each mode (**Chapter 3**) which is modeled on the relationship between the weight coefficients and the source parameters that may affect ground motions. Then, the relationship between the modes and the source parameters is comprehensively analyzed.

Through the above attempts, this chapter aims to grasp the diversity of spatial characteristics expressed by the applied ground motion prediction method (here, the stochastic Green's function method [12]) for a given source parameter. As described in ***2.3***, the ground model of the data in this study has a uniform horizontal stratification structure at all computation points, and the site amplification characteristics are the same at all points [7]. Therefore, the diversity of spatial characteristics for the ground motion distributions expressed in all cases can be considered to reflect only the effects of source rupture within the range of the characterized source model. Accordingly, the ground motion distributions in this chapter are expected to show a clearer relationship between source and spatial characteristics than Scenario Earthquake Shaking Maps in the previous studies [9–11] that reflect complex propagation path characteristics and site

characteristics by considering deep and shallow ground model.

4.2 shows the mode decomposition results again and classifies the characteristics of the mode shape. Next, ***4.3*** outlines “explainable AI,” and ***4.5*** shows the results of interpreting the modes. Finally, ***4.6*** summarizes the conclusions obtained in this chapter.

4.2 Mode Decomposition of Distributions of Absolute Acceleration Response Spectrum

4.2.1 Mode Shape and Cluster Classification

As in **Chapter 3**, the mode decomposition is applied to the distributions of S_A (period: T = 0.1, 0.5, 0.75, 1.0, and 2.0 s, component: Fault-Normal (FN) and Fault-Parallel (FP)) for 600 cases [7] by period and component. Fig. 4-1 and Fig. 4-2 show the spatial distributions of the left singular vectors, which are mode shapes of the FN and FP components, respectively (mode shapes of period 0.5, 1.0, and 2.0 s are restatements of Fig. 3-10 and Fig. 3-11). The mode shape represents the spatial characteristics of ground motion; a correlation between points with the same (or different) sign is a positive (or negative) correlation.

First, feature classification, or similarity evaluation, is performed to capture the general characteristics of the mode shapes of Modes 1–6 (60maps in total) in two horizontal components and 5-period bandwidth. $\mathbf{u}_i$ and $\mathbf{u}_j$ be the mode shapes (unit vectors) of Mode i and Mode j, respectively; the distance d_{ij} (i.e., dissimilarity) based on the cosine similarity S_{ij} between modes is described as follows:

$$d_{ij} = 1 - \left|S_{ij}\right| = 1 - \left|\cos(\mathbf{u}_i, \mathbf{u}_j)\right| = 1 - \left|\frac{\mathbf{u}_i \cdot \mathbf{u}_j}{\left\|\mathbf{u}_i\right\| \cdot \left\|\mathbf{u}_j\right\|}\right| = 1 - \left|\mathbf{u}_i \cdot \mathbf{u}_j\right| \tag{4-1}$$

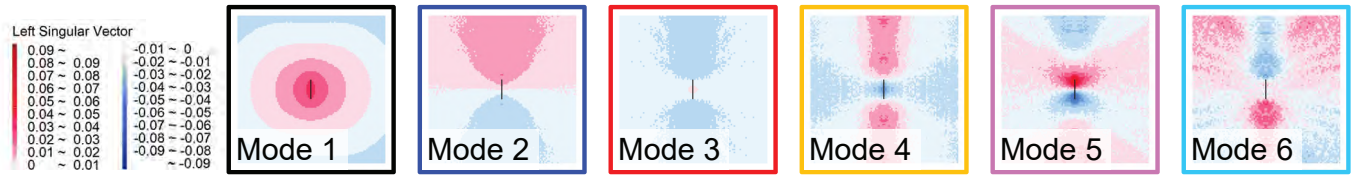
$$0 \le d_{ij} \le 1$$

The hierarchical cluster analysis is used as the analysis method, and the distance between clusters is defined by the group average method.

Figure. 4-3 shows the dendrogram of the hierarchical cluster analysis. For k = 2 clusters, the similarity of mode shapes can be divided into "symmetric modes," whose spatial distribution is symmetric around the perpendicular bisector of the fault, and "reverse modes," whose distribution is reversed. The number of clusters for simply classifying Modes 1–6 is k = 10, and the classification results are shown in Table. 4-1. Cluster numbers are assigned in ascending order from lower-order modes to higher-order modes. Cluster a is hereafter denoted "$C_{a,}$" and the reverse mode is denoted "C_a^*."

The symmetric mode C_1 (black in Fig. 4-1 and Fig. 4-2) contains Mode 1 of all periods and both components, and C_2 (red in Fig. 4-1 and Fig. 4-2)contains "mode with a spatial variation that positively correlated over the almost entire area." In contrast, the reverse mode C_3^* (blue in Fig. 4-1 and Fig. 4-2) contains to "mode that roughly divides the entire area of the map into positive and negative areas." The C_1–C_3^* modes appear in Modes 1–3 for most of all-period and two components. The cumulative contribution of Modes 1–3 is more than 94 % (Fig. 3-8), indicating that these modes almost completely define the ground motion distribution. In some periods and components, the C_4–C_{10} modes do not appear in Table. 4-1, which shows Modes 1–6.

The definition of the distance between clusters is given in Table. 4-2. The validity of the cluster analysis results from each method was compared by the Cohen correlation coefficient [13] (the correlation coefficient between the Cohen matrix representing the distance between clusters and the original distance matrix). The results confirmed that the group average method used in this study is the best definition, taking the maximum valu.



(a) Period 0.1 s

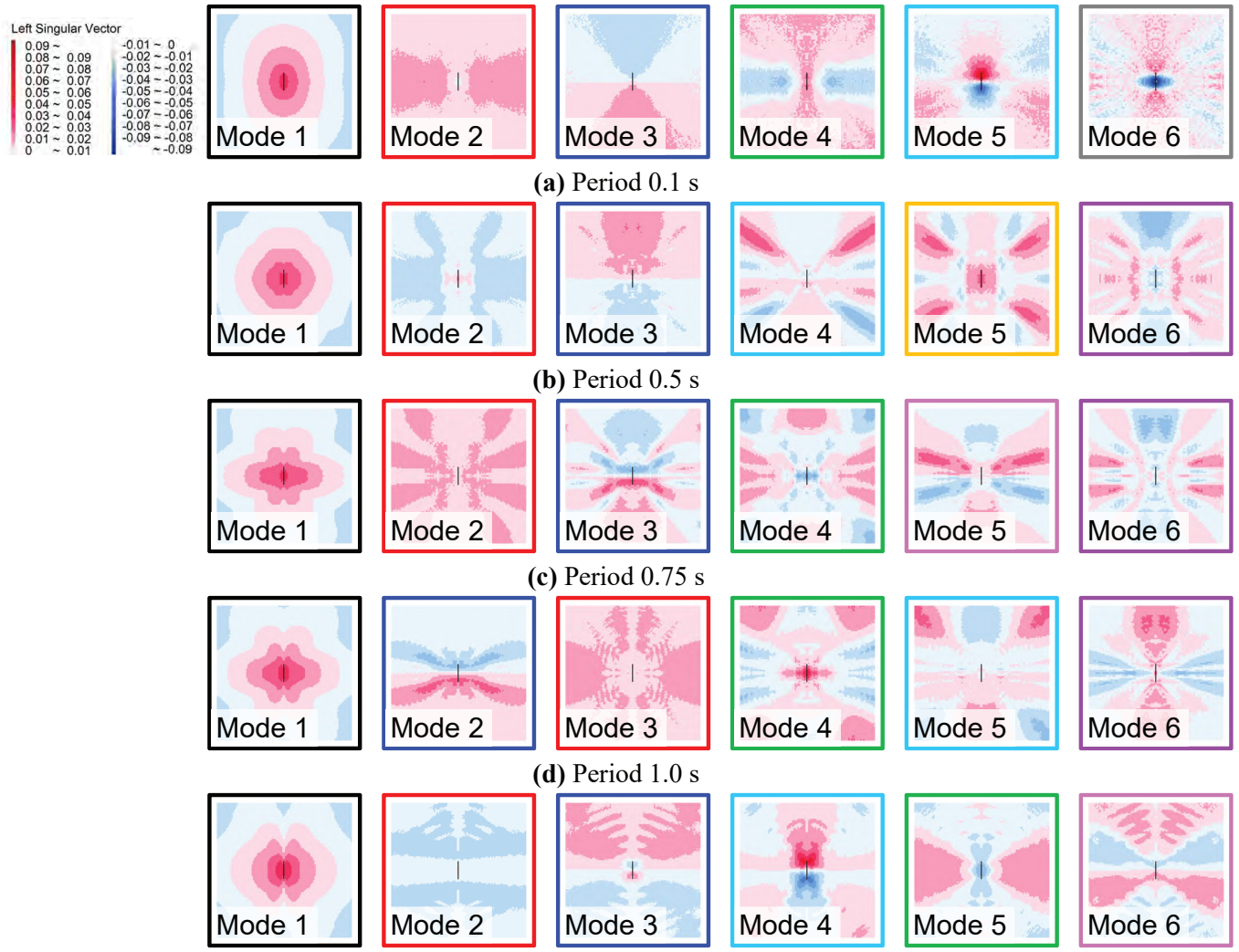
(b) Period 0.5 s

(c) Period 0.75 s

(d) Period 1.0 s

(e) Period 2.0 s

Figure 4-1. Mode shape in FN component. (Color coding follows the classification in Fig. 4-3.)



(a) Period 0.1 s

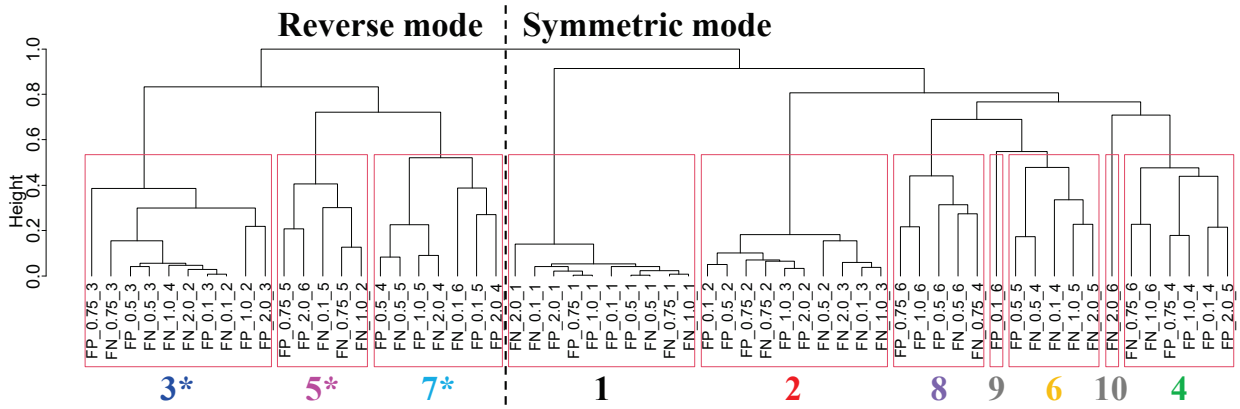
(b) Period 0.5 s

(c) Period 0.75 s

(d) Period 1.0 s

(e) Period 2.0 s

Figure 4-2. Mode shape in FP component. (Color coding follows the classification in Fig. 4-3.)



(The numbers at the bottom indicate the cluster number at k = 10; "*" indicates a reverse mode cluster. For example, "FP_0.75_3" is "Mode 3 in period 0.75 s and FP component".)

Figure 4-3. Dendrogram of mode shapes for Modes 1–6 in FN and FP components and the 5-period band.

Table 4-1. Cluster number (k = 10) of mode shapes in FN and FP components of the 5-period band. ("*" indicates a reverse mode; brackets "()" indicate clusters with fewer than two modes of occurrence in the same component.)

Component	Period [sec]	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
FN	0.1	1	3*	2	6	5*	7*
	0.5	1	2	3*	6	7*	(8)
	0.75	1	2	3*	(8)	5*	(4)
	1.0	1	5*	2	3*	6	(4)
	2.0	1	3*	2	7*	6	(9)
FP	0.1	1	2	3*	4	7*	(10)
	0.5	1	2	3*	7*	(6)	8
	0.75	1	2	3*	4	(5*)	8
	1.0	1	3*	2	4	7*	8
	2.0	1	2	3*	7*	4	(5*)

Table 4-2. Cophenetic correlation coefficient. (Red number is the maximum score.)

Method	Group average method	Nearest neighbor method	Furthest neighbor method	Centroid method	McQuitty's method	Median method	Ward's method
Cophenetic correlation coefficient	0.931	0.742	0.863	0.846	0.925	0.881	0.843

4.2.2 Mode Shape and Weight Coefficient

The ground motion distribution $\mathbf{x}_j$ for Case j ($1 \leq j \leq N = 600$) can be described by a linearly coupled model with mode shape $\mathbf{u}$ and principal component scores dv (= singular value d × weight coefficient v) as follows:

$$\mathbf{x}_j = \sum_{k=1}^{N} d_k v_{kj} \mathbf{u}_k \tag{4-2}$$

Here, the mode shape and singular value are common to all cases, while the values of weight coefficients differ from case to case. Therefore, it can be said that the spatial characteristics of ground motion distributions for 600 cases are defined by the mode shapes common to all cases, and the weight coefficients define the differences among the cases.

The study in this chapter focuses on the weight coefficients for each mode and examines the mode's dominant factors, considering the interaction effects among the source parameters and the resulting spatial characteristics for each case. Considering that the modes are affected by the combined effects of the source parameters, the input-output relationship of the weight coefficient predictor (**Chapter 3**) with all possible source parameters as explanatory variables for influence on the ground motion distribution is clarified. Specifically, "explainable AI" is applied to the weight coefficient predictor based on random forest, which has obtained relatively good results among the prediction models proposed in **Chapter 3**.

4.3 Interpretable Machine Learning "Explainable AI"

4.3.1 Explainable AI

"Explainable AI (XAI)" or "Interpretable Machine Learning (IML)" are interpretation methods for machine learning models [14–16]. This section uses the following four methods of XAI.

- Permutation Feature Importance (PFI)
- Individual Conditional Expectation Plot (ICE)
- Partial Dependence Plot (PD)
- Accumulated Local Effects Plot (ALE)

Each of these features is described below.

4.3.2 Permutation Feature Importance (PFI)

PFI is a method for evaluating the importance of a feature in a prediction model based on the increase in prediction error (decrease in prediction accuracy) when the permutations of a certain feature in the data set are shuffled [14]. A high value of PFI (a large prediction error after shuffling) indicates that the importance of the feature is high. Unlike the correlation coefficient, PFI can evaluate the importance of a feature regardless of the linear system between the predictions and the feature; it allows the macro behavior of the model to be understood.

This chapter adopts the increase in the mean square error (MSE) as the evaluation index for PFI; PFI is defined as the mean increase in MSE over five trials of feature shuffling.

In contrast, PFI has some problems [14]:

- If the model includes features correlated with the features of interest, they will compete with each other for importance and will be underestimated.
- It is unclear how the prediction changes when the feature value changes.

4.3.3 Individual Conditional Expectation Plot (ICE) and Partial Dependence Plot (PD)

ICE is a method for obtaining the change in predictions for a single instance (case) when only a certain feature is changed (other features are fixed) [14, 16]. ICE can represent the heterogeneity of each case, and the relationship between the feature and prediction can be interpreted based on the interaction between the features.

PD is a method for obtaining the average relationship between feature and prediction by averaging the ICEs of all cases in a dataset [14, 16]. Although PD provides stable results due to averaging, it may not represent the relationship between feature and prediction well because the interactions of the features cancel out.

This chapter attempts to provide an integrated interpretation based on both methods by plotting a combination of PD and ICE. Note that these methods independently handle features that originally change in conjunction with other features; thus, care must be taken in interpreting the model in such cases.

4.3.4 Accumulated Local Effects Plot (ALE)

ALE is a method for obtaining the average effect on predictions when only a certain feature value is changed and has similar characteristics to PD [16]. The advantage of ALE compared to PD is that it can express the relationship with the predictions well, even when there are correlated features. The disadvantage is that it cannot be plotted simultaneously as ICE, which provides clues for case-by-case interpretation.

4.4 Modal Interpretation

4.4.1 Modal Interpretation Based on XAI

This chapter applies XAI to the weight coefficient predictor based on random forest in **Chapter 3** and attempts to interpret the modes based on the results. For the analysis, 500 cases of training data are used.

First, PFI is used to determine the importance of features in the model. Next, focusing on the feature with high importance, the relationship between the feature and prediction is evaluated by ICE, PD, and ALE; the mode is interpreted. This chapter basically uses ALE, which can well express the relationship between the feature and the prediction even when the feature is correlated with other features; however, ICE and PD are used when case-by-case heterogeneity is considered. The following paragraphs discuss the dominant factors of the C_1, C_2 and C_3^* modes, which contribute significantly to the ground motion distribution.

Table. 4-3, Fig. 4-4 and Fig. 4-5 show the source parameters (input parameters of the predictor) for analysis (restatements of Table. 3-1, Fig. 3-12 and Fig. 3-13). For details on the horizontal and vertical arrangement patterns, see ***3.6.1***. The asperities area ratio (Fig. 2-2 (c)) is not used for model interpretation due to the large bias of the original data.

Table 4-3. Features for modeling

Variable Type	Feature		Abbreviation
Quantitative Variable	Seismic moment [dyne*cm] (common logarithmic conversion)		M_0
	Rupture velocity [km/s]		Vr
	Hypocenter (Fig. 4-4 Orange)	Horizontal position [km]	Hypo.H
		Vertical position [km]	Hypo.V
	Large asperity (Fig. 4-4 Blue)	Horizontal position [km]	aspL.H
		Vertical position [km]	aspL.V
	Small asperity (Fig. 4-4 Green)	Horizontal position [km]	aspS.H
		Vertical position [km]	aspS.V
	Ratio between combined asperity area and total rupture area		ratio
Qualitative Variable	Classification by asperities arrangement and hypocenter position (Fig. 4-4)	Horizontal arrangement pattern	Hor.P
		Vertical arrangement pattern	Ver.P

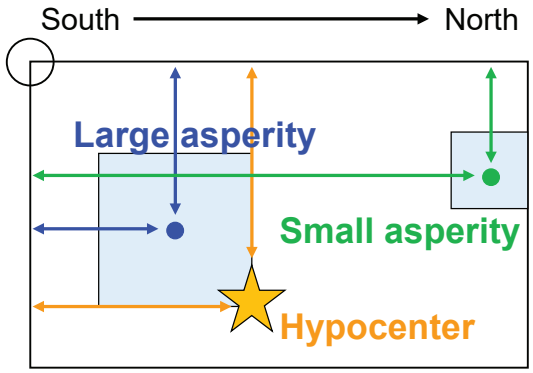
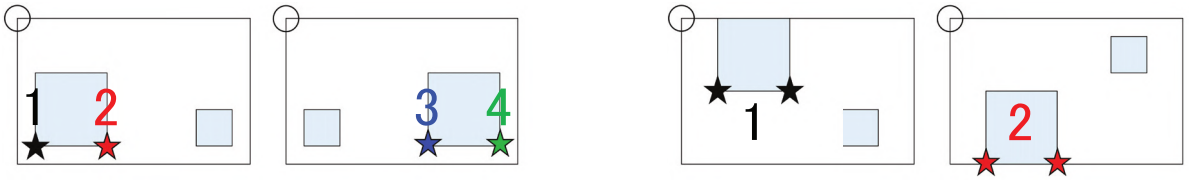


Figure 4-4. Image of the horizontal and vertical location of hypocenter and large and small asperity.



(a) Horizontal pattern classifications **(b)** Vertical pattern classifications

Figure 4-5. Classification by asperities arrangement and hypocenter position.

4.4.1 Cluster C_1: Attenuation Characteristics

The spatial distributions of mode shapes in C_1 (Mode 1 in Fig. 4-1 and Fig. 4-2) are used to examine its characteristics. The spatial distribution of both components is simple, with positive values near the fault and negative values farther away. These distribution trends correspond to the mean distribution for each period (Fig. 2-4 and Fig. 2-5). From the above, C_1 reflects the distribution of deviations for the common logarithmic mean distribution (attenuation characteristics per period and component).

a) PFI in C_1 Mode

XAI is applied to the RF-predictor in C_1 to evaluate the relationship between the source parameters (Table. 4-3) and the weight coefficients. Table. 4-4 shows the importance (PFI) of the source parameters. The PFI is obtained for each model (by period and component: line direction in the table). When the maximum value of PFI for each model is "max," the PFI for the same model in Table. 4-4 is color-coded as follows:

- Red : max × 0.8 ≤ PFI ≤ max
- Blue: max × 0.6 ≤ PFI ≤ max × 0.8

The PFI for each period and component shows that for C_1, the PFI for "rupture velocity (Vr)" or "one of the source parameters of the vertical arrangement of the hypocenter and large and small asperities (Hypo.V, aspL.V, aspS.V, Ver.P)" tend to have high values. On the other hand, the PFI (period T = 2.0 s, FN component) has a high value for "the horizontal position of hypocenter (Hypo.H)". Next, the relationship between these source parameters and weight coefficients is clarified by ALE.

Table 4-4. PFI of source parameters in Cluster 1 represents distance attenuation gradient. (Red (or blue) indicates that the PFI value is 80–100 % (or 60–80 %) of the maximum PFI value for the same period and component.)

Component	Period [s]	M_0	Vr	Hypo.H	Hypo.V	aspL.H	aspL.V	aspS.H	aspS.V	ratio	Hor.P	Ver.P
FN	0.1	4	4	2	12	3	20	2	**44**	1	1	4
	0.5	37	**55**	12	24	11	41	14	**78**	6	3	27
	0.75	32	**262**	16	48	12	49	17	38	9	25	9
	1.0	31	**105**	28	60	17	**133**	19	36	16	13	78
	2.0	32	**182**	**278**	**178**	27	**194**	23	29	9	14	7
FP	0.1	4	5	2	10	2	17	4	**47**	1	1	5
	0.5	35	**57**	16	28	14	37	16	**92**	7	4	30
	0.75	42	**160**	18	40	15	43	23	43	9	10	16
	1.0	31	41	43	**87**	20	**100**	21	53	10	19	**71**
	2.0	24	60	70	**149**	27	**141**	10	22	5	11	21

*** The above values are PFI x 10^8.**

b) Vertical Position of Large Asperity (aspL.V) in Mode 1 (T = 0.1 s, FN Component)

First, as an example of the relationship between the "vertical position of large asperity (aspL.V)" and the weight coefficients, Fig. 4-6 shows the relationship between aspL.V (horizontal axis) and ALE (vertical axis) of Mode 1 (T = 0.1 s, FN. Fig. 4-1 (a)). The black line on the horizontal axis of the figure represents the data distribution of aspL.V for 500 cases of training data.

ALE is positive when aspL.V is shallow (small value) and negative when aspL.V is deep (large value). This trend is the average effect on the predicted value of the feature (aspL.V in this case); when aspL.V is located at shallow, the predicted weight coefficient is larger, indicating stronger ground motion directly above the fault. A similar trend was observed for the "vertical position of small asperity (aspS.V)" (not shown in the figure). Thus, C_1 reflects the effect of the vertical position of the asperities on the strong ground motion.

c) Vertical and Horizontal Position of Hypocenter (Hypo.V, Hypo.H) and Rupture Velocity (Vr) in Mode 1 (T = 2.0 s, FN Component)

Figure. 4-7 shows the results for Mode 1 (T = 2.0 s, FN. Fig. 4-1 (e)). In Fig. 4-7 (a), the ALE corresponding to the "vertical position of hypocenter (Hypo.V)" is negative when Hypo.V is shallow and positive when Hypo.V is deep. This trend seems to reflect that the deeper the hypocenter, the clearer the attenuation characteristics. In Fig. 4-7 (b), the ALE of the "horizontal position of hypocenter (Hypo.H)" is positive when Hypo.H is located near the center of the fault and negative when Hypo.H is located at the edge of the fault. This trend seems to reflect the effect of horizontal bilateral fault rupture propagation. Fig. 4-7 (c) shows that as the "rupture velocity (Vr)" increases, the ALE also increases. This trend reflects that the effect of attenuation characteristics changes with the values of Vr.

d) Vertical Arrangement Pattern (Ver.P) in Mode 1 (T = 1.0 s, FP Component)

Figure. 4-8 for Mode 1(T = 1.0 s, FP. Fig. 4-2 (d)) clearly shows the effect of "vertical arrangement pattern of large and small asperities (Ver.P)". Vertical pattern 2 (red in Fig. 4-5 (b), Ver. P = 2), where the large asperity is located deeper than the small one, has a larger ALE than vertical pattern 1 (black in Fig. 4-5 (b), Ver. P = 1), where the large asperity is located shallower than the small one. This trend reflects that when large and small asperities are placed in the direction of vertical rupture propagation, the effect of the vertical upward forward directivity effect increases and the ground motion directly above the fault becomes stronger.

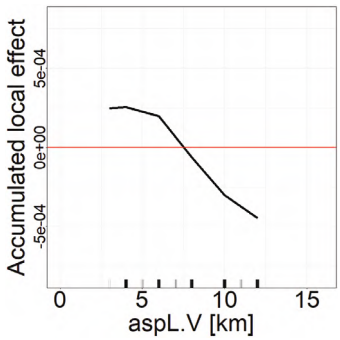
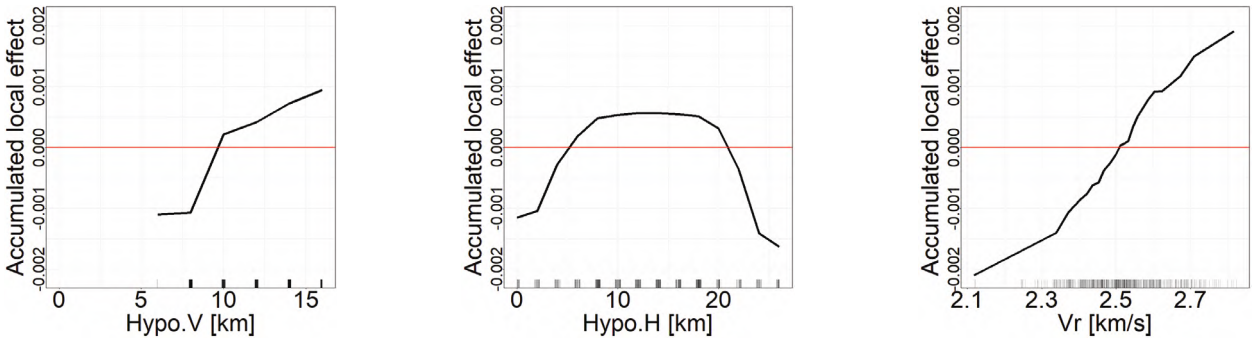


Figure 4-6. Effect of predictions in Cluster 1 and vertical position of large asperity (aspL.V).
(Period 0.1 s, FN component and Mode 1. ALE. Vertical axis: ALE, horizontal axis: parameter.)



(a) Vertical position of hypocenter (Hypo.V)

(b) Horizontal position of hypocenter (Hypo.H)

(c) Rupture velocity (Vr)

Figure 4-7. Effect of predictions in Cluster 1 and source parameters.
(Period 2.0 s, FN component and Mode 1. ALE. Vertical axis: ALE, horizontal axis: parameter.)

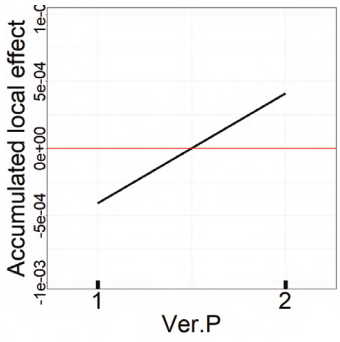


Figure 4-8. Effect of predictions in Cluster 1 and vertical arrangement pattern (Ver.P).
(Period 0.1 s, FP component and Mode 1. ALE. Vertical axis: ALE, horizontal axis: parameter.)

e) Summary in C_1

The dominant factors of C_1 were evaluated comprehensively based on XAI and those shown here. Table. 4-5 shows a summary of the results. The results do not exactly match the red and blue numbers in Table. 4-4. This is because, in addition to the importance of a feature, it also includes an evaluation of the impact of the changes in features on the predictions (the same applies to Tables. 4-6, 4-7 and Tables. 4-8, 4-9 below). It can be observed that C_1 (T = 0.75, 1.0 s) reflect the influence of either “rupture velocity (Vr)” or “horizontal position of hypocenter (Hypo.H)”, while C_1 (T = 2.0 s) reflects both influences. These trends seem to be related to the fact that the longer the period, the more significant the influence of bilateral directivity effects and radiation patterns becomes.

Table 4-5. Dominant source parameters for modes of Cluster 1 (distance attenuation gradient) in the 5-period band and FN and FP components.

Component	Period [s]	Vertical position of hypocenter and asperities (Hypo.V, aspL.V, aspS.V, Ver.P)		Rupture velocity (Vr)	Horizontal position of hypocenter (Hypo.H)
		Gradient of distance attenuation	Vertical forward directivity effect	Gradient of distance attenuation	Bilateral rupture propagation
				Directivity effect + Radiation pattern	
FN	0.1	Asperities (aspL.V, aspS.V): shallow position			
	0.5		Rupture propagation from bottom to top of fault		
	0.75			○	
	1.0		Rupture propagation from bottom to top of fault	○	
	2.0	Hypocenter (Hypo.V): deep location		○	Rupture propagation from near center of fault
FP	0.1	Asperities (aspL.V, aspS.V): shallow position			
	0.5		Rupture propagation from bottom to top of fault		
	0.75			○	
	1.0		Rupture propagation from bottom to top of fault		Rupture propagation from near center of fault
	2.0	Hypocenter (Hypo.V): deep location		○	Rupture propagation from near center of fault

4.4.2 *Cluster C_2: Mode with Positive Correlation in Almost Entire Area*

C_2 is a mode that causes simultaneous changes in ground motion intensity over the entire area and appears as Mode 2 or Mode 3 for both components at all periods (Table. 4-1). The spatial distribution of C_2 (red in Fig. 4-1 and Fig. 4-2) shows that FN component has a larger absolute value in the direction of the fault strike (North-South), while FP component has it in the orthogonal direction to the fault (East-West)., This distribution reflects the influence of bilateral rupture propagation and radiation patterns as in Mode 1 (C_1).

a) PFI in C_2 Mode

Table. 4-6 shows the PFI of the source parameters in C_2 mode. The PFI for "seismic moment (M_0)" is commonly high for C_2 (except for period 2.0 s). The PFI for "rupture velocity (Vr)" and "horizontal position of hypocenter (Hypo.H)" are also high for some periods and components.

Table 4-6. PFI of source parameters in Cluster 2 represents a positive correlation in the whole area. (Red (or blue) indicates that the PFI value is 80–100 % (or 60–80 %) of the maximum PFI value for the same period and component.)

Component	Period [s]	M_0	Vr	Hypo.H	Hypo.V	aspL.H	aspL.V	aspS.H	aspS.V	ratio	Hor.P	Ver.P
FN	0.1	**85**	13	11	18	5	22	5	28	12	40	2
	0.5	**55**	**48**	3	**50**	1	**63**	1	16	2	3	1
	0.75	**74**	**61**	22	14	5	17	4	19	8	9	2
	1.0	**67**	24	31	6	5	8	5	**85**	22	12	3
	2.0	17	46	**86**	8	12	11	6	10	10	18	1
FP	0.1	**90**	16	7	22	3	41	4	33	12	18	2
	0.5	**81**	36	6	46	2	39	2	32	2	6	1
	0.75	**60**	**73**	9	27	3	29	3	18	6	4	2
	1.0	**119**	22	5	13	2	15	3	**93**	39	3	1
	2.0	27	**53**	**77**	15	6	21	2	9	**57**	2	1

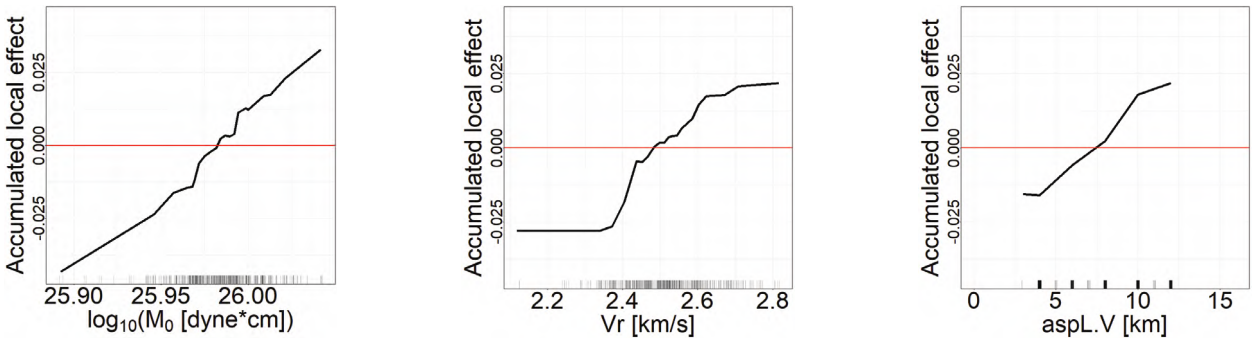
*** The above values are PFI x 10^5.**

b) Seismic Moment (M_0), Rupture Velocity (Vr) and Vertical Location of Large Asperity (aspL.V) in C_2 Mode (T = 0.5 s, FN Component, Mode 2)

Figure. 4-9 (a) for C_2 (T = 0.5 s, FN, Mode 2. Fig. 4-1 (b)) shows the relationship between the "seismic moment (M_0)" and the ALE. Since the spatial distribution of the corresponding mode shape is positive over the entire area, the ground motion intensity is pushed up over the entire area when M_0 is large (positive value of ALE) and pushed down when M_0 is small (negative value of ALE). Figs. 4-9 (b), (c) show the same trend for "rupture velocity (Vr)" and "vertical location of large asperity (aspL.V)" as the above seismic moment (M_0), which seems to reflect the effect on the ground motions in the whole area.

c) Horizontal Position of Hypocenter in C_2 Mode (T = 2.0 s, FN Component, Mode 3)

Figure. 4-10 for C_2 (T = 2.0 s, FN, Mode 3. Fig. 4-1 (e)) shows the relationship between the "horizontal position of hypocenter (Hypo.H)" and the ALE. ALE has the same sign (positive value) as the mode shape when Hypo.H is near the center of the fault and has a different sign (negative value) when Hypo.H is near the edges of the fault. This trend reflects the effect of horizontal bilateral rupture propagation.



(a) Seismic moment (M_0) **(b)** Rupture velocity (Vr) **(c)** Vertical position of large asperity (aspL.V)

Figure 4-9. Effect of predictions in Cluster 2 and source parameters.
(Period 0.5 s, FN component and Mode 2. ALE. Vertical axis: ALE, horizontal axis: parameter.)

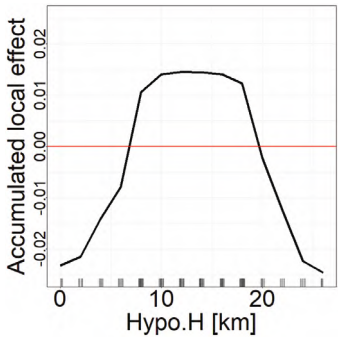


Figure 4-10. Effect of predictions in Cluster 2 and horizontal position of hypocenter (Hypo.H).
(Period 2.0 s, FN component and Mode 3. ALE. Vertical axis: ALE, horizontal axis: parameter.)

d) Summary in C_2

Table. 4-7 summarizes a discussion of the dominant factors of the C_2 mode for all periods and two components based on XAI. In period 1.0 s with FN component (Mode 3 in Fig. 4-1 (d)) and period 2.0 s with both components (FN: Mode 3 in Fig. 4-1 (e), FP: Mode 3 in Fig .4-2 (e)), where the spatial distribution of modes is dominant in the fault strike or orthogonal directions, relationships are found with "rupture velocity (Vr)" and "horizontal position of hypocenter (Hypo.H)". This trend seems to reflect that, as in C_1, the bilateral directivity effect and the influence of the radiation pattern become more pronounced with Vr and Hypo.H.

For C_2 of period 0.75 s (Mode 2 of both components), a larger Vr pushes down the intensity of ground motion in the whole area. On the other hand, Table. 4-5 shows that for C_1 (Mode 1) with the same period, the attenuation characteristics tend to become stronger as Vr increases. Therefore, it is assumed that at period 0.75 s, the mode C_1 represents the global spatial variability, and the C_2 mode, which shows a more complex mode shape than C_1, Adjusts and complements spatial variation in detail.

In the previous studies [9–11] (target: Scenario Shaking Earthquake Maps for 4–12 cases of measured seismic intensity), the C_2 mode rarely appeared because the recipe is basically set to a single seismic moment and rupture velocity. However, in this study, the dataset with a variation in the seismic moment and rupture velocity (Figs. 2-2 (a), (b)) was used; as shown in Fig. 4-9 (a), a uniform spatial variation (C_2 mode) appears in the entire area for all periods with both components. In order to incorporate the uniform spatial variation in the entire area into the spatial characteristics represented by the Scenario Shaking Earthquake Maps, such considerations are necessary for setting the source parameters.

Table 4-7. Dominant source parameters for modes of Cluster 2 (positive correlation in the whole area) in the 5-period band and FN and FP components.

Component	Period [s]	Seismic moment (M_0)	Rupture velocity (Vr)	Horizontal location of hypocenter (Hypo.H, Hor.P)	Vertical location of hypocenter and asperities (Hypo.V, aspL.V, aspS.V)
		Ground motion intensity in whole area	Ground motion intensity in whole area	Bilateral rupture propagation	Ground motion intensity in whole area
			Directivity effect + Radiation pattern		
FN	0.1	○		Rupture propagation from near center of fault	
	0.5	○	○		Hypocenter and asperity (Hypo.V, aspL.V): deep location
	0.75	○	(C_1 Complement)		
	1.0	○		Rupture propagation from near center of fault	Asperity (aspS.V): deep location
	2.0	○	○	Rupture propagation from near center of fault	
FP	0.1	○			
	0.5	○			
	0.75	○	(C_1 Complement)		
	1.0	○			Asperity (aspS.V): deep location
	2.0	○	○	Rupture propagation from near center of fault	

4.4.3 Cluster C_3^*: Mode that Divides Entire Area into Two Positive and Negative Parts

C_3^* is a mode that shows a negative correlation across the perpendicular bisector of the fault and appears as Mode 2 or Mode 3 for both components at all periods (Table. 4-1). The spatial distribution of most of C_3^* (blue in Figs. 4-1 and Fig. 4-2) have large absolute values in the fault strike direction.

a) PFI in C_3^* Mode

Table. 4-8 shows the PFI of the source parameters in C_3^* mode. The PFI for "horizontal location of hypocenter and asperities (Hypo.H, aspL.H, aspS.H, Hor.P)" shows a high value. In the following discussion of the dominant factors of C_3^* mode, in order to analyze the effect of the arrangement of large and small asperities on the ground motion in the direction of rupture propagation, ICE, which can represent the heterogeneity of each case, and PD, which can represent the average change in predicted values, are used. ICE is a sensitivity analysis that visualizes changes in predictions by varying only specific features within the original data range; however, even if there are constraints between features, they are not considered in ICE. Since it is preferable to focus on changes in the original data's neighborhood, the original data predictions are plotted as dots in this study. In addition, ICEs are color-coded corresponding to the horizontal arrangement pattern (Fig. 4-5 (a)) to clarify the characteristics of each case.

Table 4-8. PFI of source parameters in Cluster 3 that represents two divided areas.
(Red (or blue) indicates that the PFI value is 80–100 % (or 60–80 %) of the maximum PFI value for the same period and component.)

Component	Period [s]	M_0	Vr	Hypo.H	Hypo.V	aspL.H	aspL.V	aspS.H	aspS.V	ratio	Hor.P	Ver.P
FN	0.1	1	1	5	1	**230**	1	74	5	0	86	0
	0.5	4	7	14	3	**212**	3	45	6	0	73	1
	0.75	5	16	**132**	3	30	4	43	10	1	11	1
	1.0	7	13	19	5	**91**	5	**101**	29	11	23	2
	2.0	3	7	57	2	25	2	6	3	1	**104**	1
FP	0.1	1	2	5	0	**142**	1	48	2	0	**224**	0
	0.5	5	8	15	2	**141**	3	44	4	1	**100**	1
	0.75	3	4	**143**	2	27	2	38	4	2	7	0
	1.0	2	3	**117**	1	38	2	18	3	1	8	1
	2.0	4	12	40	2	33	3	18	5	6	**121**	0

*** The above values are PFI x 10^5.**

b) C_3^* Mode on Short Period

Figure. 4-11 shows the results of the C_3^* mode for a short period (T = 0.5, FN, Mode 3. Fig. 4-1 (b)). The mode shape is negative values in the southern part of the fault and positive values in the northern part.

Figure. 4-11 (a) shows ICE (black, red, blue, and green line) and PD (yellow line) of the "horizontal position of large asperity (aspL.H)." ICEs and PDs are negative when aspL.H is near the south (small values) and positive when aspL.H is near the north (large values). This trend is the predicted value (ICE: by case, PD: average of all cases) corresponding to the change in the feature (in this case aspL.H). When aspL.H is in the southern part of the fault (negative predicted value), it indicates larger ground motions in the southern part of the map, and when aspL.H is located in the northern part of the fault (positive predicted value), it indicates larger ground motions in the northern part of the map. This trend reflects the bias of the center of gravity of the asperity arrangement in the horizontal direction.

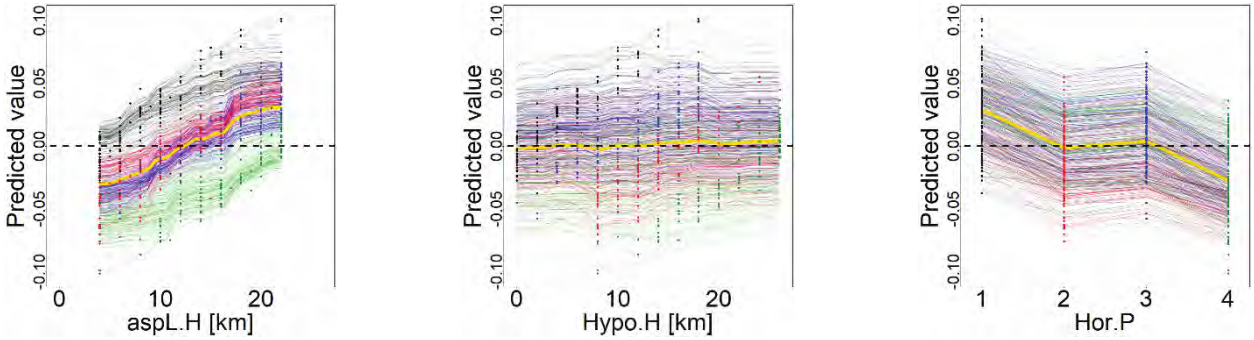
Figure. 4-11 (b) shows the ICEs and the PD of the "horizontal position of the hypocenter (Hypo.H)", but no clear feature can be found in them. However, for the PD of "horizontal arrangement pattern of hypocenter and asperities (Hor.P)" in Fig. 4-11 (c), differences are found in the horizontal patterns 1 and 4 in which the hypocenter is located outside of the large and small asperities. The horizontal pattern 1 (black in Fig. 4-5 (a), Hor.P = 1), where the large asperity is located further south than the small one, is a positive value, and the horizontal pattern 4 (green in Fig. 4-5 (a), Hor.P = 4), where the large one is located further north than the small one, is a negative value. For horizontal patterns 2 and 3 (red and blue in Fig. 4-5 (a), Hor.P = 2, 3), where the hypocenter is located inside the large and small asperities, the PD is almost zero. The ICE also has a maximum value at Hor.P = 1, a minimum value at Hor.P = 4, and intermediate values at Hor.P = 2 and 3. The above results suggest that when the rupture propagates from the end of the fault to the direction where the large and small asperities are located, the ground motion in the direction of propagation is larger.

c) C_3^* Mode on Long Period

Next, Fig. 4-12 shows the results of the C_3^* mode for a long period (T = 2.0, FN, Mode 2. Fig. 4-1 (e)). The spatial distribution of the mode shape is positive values in the southern part of the fault and negative values in the northern part.

The PD of the "horizontal position of the hypocenter (Hypo.H)" in Fig. 4-12 (b) is a negative value near the southern end of the fault, 0 at the center, and a positive value near the northern end. In Fig. 4-12 (a), the PD obtained by changing only the "horizontal location of large asperity (aspL.H)" takes zero. The ICE for each case can be classified into two types of cases: one with negative predictions (horizontal patterns 1 and 3: black and blue in Fig. 4-5 (a)) and the other with positive predictions (horizontal patterns 2 and 4: red and green in Fig. 4-5 (a)) when aspL.H is changed from the southern to the northern part of the fault. In horizontal patterns 2 and 3 (red and blue), where the hypocenter is located inside the large and small asperities, the absolute value of the predicted value is small; in other words, the degree of influence is small. In contrast, horizontal patterns 1 and 4 (black and green) strongly influence where the hypocenter is near the fault edge. These results indicate that this mode strongly influences Hypo.H, but the degree of influence depends on the aspL.H of each case.

Next, examine the PD for the “horizontal arrangement pattern (Hor.P)” in Fig. 4-12 (c). The PD is negative values for horizontal patterns 1 and 3 (black and blue in Fig. 4-5 (a), Hor.P = 1, 3), where the hypocenter is located at the southern end of the large asperity, and positive values for horizontal patterns 2 and 4 (red and green in Fig. 4-5 (a), Hor.P = 2, 4), where it is located at the northern end. These trends reflect the horizontal forward directivity effect. In the case of horizontal patterns 1 and 4 (black and green in Fig. 4-5 (a), Hor.P = 1, 4), where the hypocenter is located outside the large and small asperities, the PD and the ICEs have large absolute values. This trend indicates that the forward directivity effect is stronger when large and small asperities are placed in the direction of rupture propagation.

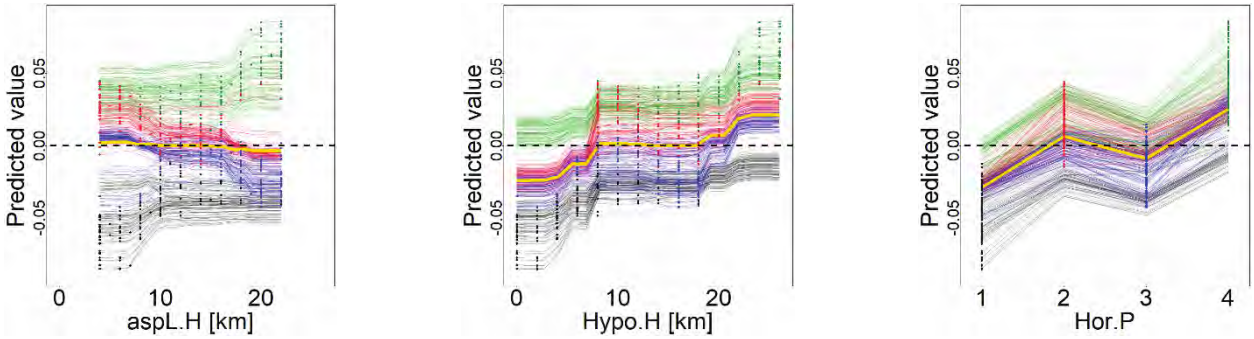


(a) Horizontal position of large asperity (aspL.H)

(b) Horizontal position of hypocenter (Hypo.H)

(c) Horizontal arrangement pattern (Hor.P)

Figure 4-11. Effect of predictions in Cluster 3 and source parameters.
(Period 0.5 s, FP component and Mode 3. ICE (black, red, blue, and green) + PD (yellow).
Vertical axis: Predicted values, horizontal axis: parameter.)



(a) Horizontal position of large asperity (aspL.H)

(b) Horizontal position of hypocenter (Hypo.H)

(c) Horizontal arrangement pattern (Hor.P)

Figure 4-12. Effect of predictions in Cluster 3 and source parameters.
(Period 2.0 s, FN component and Mode 2. ICE (black, red, blue, and green) + PD (yellow).
Vertical axis: Predicted values, horizontal axis: parameter.)

d) Summary in C_3^*

Table. 4-9 summarizes a discussion of the dominant factors of the C_3^* mode for all periods and two components based on XAI. The C_3^* modes are dominated by the effect of the horizontal position of arrangement of asperities for short periods (T = 0.1, 0.5, 0.75, and 1.0 s) and by the effect of forward directivity, which depends on the horizontal position of the hypocenter, for long-period (T = 2.0 s).

In ***3.5.1***, it is pointed out that the contributions of Modes 2–3 at period 2.0 s of both components (Fig. 3-8 and Fig. 3-9) are larger than that of the other periods. This result seems to reflect that the influence of the directivity effect and the radiation pattern is dominant at period 2.0 s, rather than the bias of the seismic moment and the center of gravity of asperities.

Table 4-9. Dominant source parameters for modes of Cluster 3 (two divided areas) in the 5-period band and FN and FP components.

Component	Period [s]	Horizontal location of hypocenter and asperities (Hypo.H, aspL.H, aspS.H, Hor.P)	
		Bias in gravity center of asperities	Horizontal forward directivity effect
FN	0.1	○	Hypocenter (Hor.P): Located outside of large and small asperities
	0.5	○	Hypocenter (Hor.P): Located outside of large and small asperities
	0.75	○	
	1.0	○	
	2.0		Rupture propagation involving large asperities The effect is particularly strong when hypocenter is located outside of the large and small asperities.
FP	0.1	○	Hypocenter (Hor.P): Located outside of large and small asperities
	0.5	○	Hypocenter (Hor.P): Located outside of large and small asperities
	0.75	○	
	1.0	○	
	2.0		Rupture propagation involving large asperities The effect is particularly strong when hypocenter is located outside of the large and small asperities.

4.4.4 Mode Include in Cluster C_4 or C_5^*

From Table. 4-1, the modes in C_4 or C_5^* appear as Mode 4 and higher (except for C_5^* at period 1.0 s of FN component). The spatial distribution of the mode shape (Fig. 4-1 and Fig. 4-2) is "several partitions of the entire map area," but the spatial correlation becomes more complex with higher-order modes. These modes are also attempted to be interpreted based on XAI.

a) Cluster C_5^*: Mode 2 at Period 1.0 s with FN Component

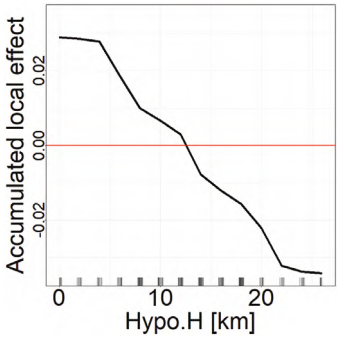
Examine Mode 2 for period 1.0 s of FN component (Fig. 4-1 (d)) in cluster C_5^*. This mode is assumed to represent a significant spatial characteristic because of its high contribution. Positive values in the mode shape are distributed in the map's north and from the fault's south to the southwest and southeast.

Figure. 4-13 shows the relationship between this mode and the source parameters. Figs. 4-13 (a), (b) show that the ALE is a positive value when "horizontal positions of the hypocenter and the large asperity (Hypo.H and aspL.H)" is in the southern part of the fault, and the ALE is a negative value when them is in the northern part. This trend could reflect the expansion of the strong motion area caused by the hypocenter and the large asperity. In addition, the ALE of "horizontal arrangement pattern (Hor.P)" in Fig. 4-13 (c) has positive and negative values for horizontal patterns 1 and 4 (black and green in Fig. 4-5 (a), Hor.P = 1 and 4), where the hypocenter is located outside of the large and small asperities, respectively. This trend reflects that the forward directivity effect is accentuated when large and small asperities are in the direction of rupture progression.

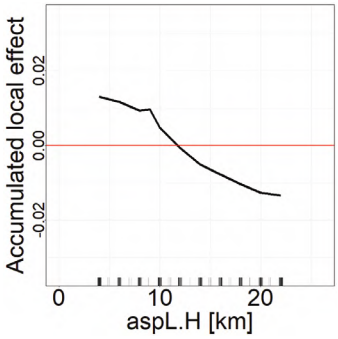
b) Cluster C_4: Mode 4 at Period 1.0 s with FP Component

Finally, evaluate the dominant factors for Mode 4 (T = 1.0 s, FP. Fig. 4-2 (d)) in cluster C_4. Fig. 4-14 shows the ALE of the "horizontal position of the hypocenter (Hypo.H)" and "vertical position of the hypocenter and the large asperity (Hypo.V, aspL.V)". The relationship with these source parameters in this mode is similar to that of Mode 1 (C_1, Fig. 4-2 (d)) of the same period (T = 1.0 s) and the same component (FP) (Table. 4-5). This trend suggests that the details of the global spatial variability represented by the mode shape of Mode 1 are adjusted and complemented by the mode shape of Mode 4, which represents complex spatial variability.

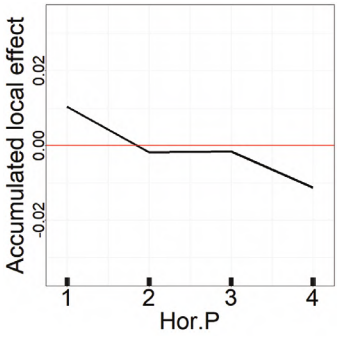
The relationship between the weight coefficients and the source parameters was also examined for other modes above cluster C_4. These modes are assumed to adjust for details of spatial variability not fully represented by the lower-order modes (except for C_5^* at period 1.0 s of FN component).



(a) Horizontal position of hypocenter (Hypo.H)

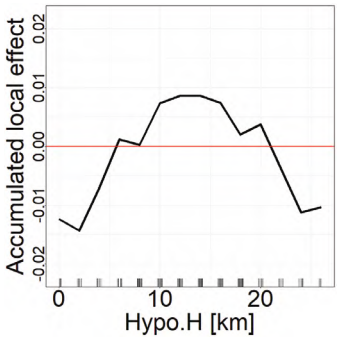


(b) Horizontal position of large asperity (aspL.H)

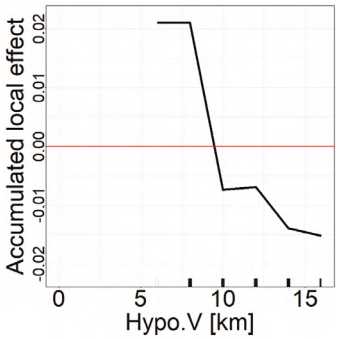


(c) Horizontal arrangement pattern (Hor.P)

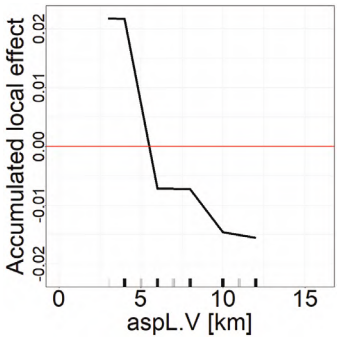
Figure 4-13. Effect of predictions in Cluster 5 and source parameters.
(Period 1.0 s, FN component and Mode 2. ALE. Vertical axis: ALE, horizontal axis: parameter.)



(a) Horizontal position of hypocenter (Hypo.H)



(b) Vertical position of hypocenter (Hypo.V)



(c) Horizontal position of large asperity (aspL.H)

Figure 4-14. Effect of predictions in Cluster 4 and source parameters.
(Period 1.0 s, FP component and Mode 4. ALE. Vertical axis: ALE, horizontal axis: parameter.)

4.4.5 Advantages of Modal Interpretation Based on Explainable AI

This chapter comprehensively analyzes the relationship between modes and all source parameters with possible effects on ground motions by modeling with machine learning regression and evaluating the dominant factors of the mode based on explainable AI (XAI). The advantage of the proposed method is its ability to analyze the combined effects of complex fault rupture scenarios on spatial characteristics. In other words, it is now possible to evaluate the relationship between source parameters and case-specific weight factors, including the arrangement of hypocenter and asperities, as well as the combined effect on the distribution of ground motion.

4.5 Conclusions

This chapter aims to understand the diversity of spatial characteristics expressed by the ground motion prediction method and analyzes the effects of different source parameters on the spatial characteristics by period and component. The analysis covers the distributions of the absolute acceleration response spectrum by period (0.1–2.0 s) and component (FN, FP) for 600 cases caused by a strike-slip fault. These ground motion distributions were calculated by the stochastic Green's function method [12]. Since the site amplification characteristics are the same for all computation points, the spatial characteristics analyzed in this section can be considered to reflect only the effects of source rupture within the setting range of the characterized source model. The major conclusions are listed below.

1) Hierarchical cluster analysis was applied to the spatial distributions of mode shapes obtained by mode decomposition by period and component, and each mode was classified into clusters with clear characteristics. The modes can be classified into two major categories, "symmetric mode" and "reverse mode," with the perpendicular bisector of the fault as the boundary, and Modes 1–6 can be briefly classified into ten categories.

2) For clarifying the relationship between the weight coefficients and the source parameters, an explainable AI (XAI) was applied to the weight coefficient predictor with the source parameters as explanatory variables constructed in **Chapter 3**. Based on the XAI results, the mode's dominant factors were evaluated.

3) Mode 1 mainly represents the attenuation characteristics. Depending on the period and component, the attenuation characteristics and vertical forward directivity effects due to the vertical location of the hypocenter and asperities were observed. For a longer period, the influence of bilateral directivity effects and radiation patterns became more significant and tended to increase in variation depending on the rupture velocity and the horizontal position of the hypocenter.

4) The "mode with positive correlation in almost entire area" indicates that the ground motion intensity varies over the entire area because of the influence of seismic moment and rupture velocity. In addition, this mode also shows the influence of bilateral directivity effect and radiation pattern.

5) The "mode that divides the entire area into two positive and negative parts" is strongly influenced by the center of gravity of the asperity arrangement in the short period and by the horizontal forward directivity effect in the long period.

6) The modes of "positive correlation over the whole area" and the "dividing the entire area into two parts" mostly appear in Mode 2 and Mode 3 for all periods and both components, and their contribution to the spatial structure of the variability is large. In particular, the contribution of these modes is larger in period 2.0 s, when the influence of the directivity effect and the radiation pattern is more significant than in other periods.

The results of this chapter relate to the spatial characteristics represented by the inner fault parameters and the extra fault parameters (e.g., rupture velocity and location of hypocenter) under relatively simple conditions. However, actual earthquakes rarely occur according to the assumed scenario and may be caused by the rupture of a part of the assumed source fault or in conjunction with the surrounding active faults. Therefore, in addition to the inner and extra fault parameters, it is necessary to consider the influence of the outer fault parameters (e.g., geometry and location of the fault) on spatial characteristics in the future. Specifically, we plan to analyze the spatial characteristic of the ground motions for multi-segment earthquakes caused by the coupling of multiple faults for future development.

References in Chapter 4

[1] Nozu, A. and Morikawa, H. (2003) Some characteristics of seismic hazard curve under the predominant influence of active fault. *Journal of JCOSSAR 2003*: pp. 659–664. **(in Japanese with English abstract)**

[2] Okumura, T., Sato, T., Ishii, T., Somerville, P., Graves, R. and Collins, N. (2005) Study on variability of earthquake ground motion due to variation of source parameters. *JSCE Journal of Earthquake Engineering*, **28**: Paper No. P167. https://doi.org/10.11532/proee2005a.28.167 **(in Japanese with English abstract)**

[3] Itoi, T. Midorikawa, S. Kito, J. Miura, H. Uchiyama, Y. and Sakamoto, S. (2009) Variability in response spectra for crustal earthquake using Stochastic Green's Function approach. *Journal of JAEE*, **9** (1): pp. 128–142. https://doi.org/10.5610/jaee.9.1_128 **(in Japanese with English abstract)**

[4] Yamada, M., Senna, S. and Fujiwara, H. (2011) Statistical analysis of ground motions estimated on the basis of a recipe for strong-motion prediction: approach to quantitative evaluation of average and standard deviation of ground motion distribution. *Pure Appl. Tab.Geophys.*, **168**: pp. 141–153. https://doi.org/10.1007/s00024-010-0159-0

[5] Hikita, T., Ikemura, T., Tomozawa, I. and Ishida, H. (2015) Study on prediction method of ground motion response spectrum and variability in prediction results based on fault models. *J. Struct. Constr. Eng. Trans. AIJ*, **80** (715): pp. 1403–1413. https://doi.org/10.3130/aijs.80.1403 **(in Japanese with English abstract)**

[6] Kagawa, T. (2015) A study on spatial and temporal variabilities of strong ground motions based on multiple fault rupture scenarios. *Journal of JSCE A1*, 71 (4): pp. I_191–I_197. https://doi.org/10.2208/jscejseee.71.I_191 **(in Japanese with English abstract)**

[7] Kagawa, T. (2015) Spatial variability of periodic characteristics among strong ground motions derived from multiple fault rupture scenarios. *Journal of JAEE*, 15 (7): pp. 90–99. https://doi.org/10.5610/jaee.15.7_90 **(in Japanese with English abstract)**

[8] Iwaki, A., Morikawa, N., Maeda, T. and Fujiwara, H. (2017) Spatial distribution of ground-motion variability in broad-band ground-motion simulations. *Bull. Seism. Soc. Am.*, **107** (6): pp. 2963–2979. https://doi.org/10.1785/0120170150

[9] Nojima, N. Kuse, M. and Duc, L. Q. (2018) Mode decomposition and simulation of strong ground motion distribution using singular value decomposition. *Journal of JAEE*, **18** (2): pp. 95–114. https://doi.org/10.5610/jaee.18.2_95 **(in Japanese with English abstract)**

[10] Nojima, N. Kuse, M. and Takahashi, Y. (2019) Analysis of spatial variation and correlation of predicted seismic intensity maps for scenario earthquake by used of mode decomposition. *Journal of JAEE*, **19** (1): pp. 121–135. https://doi.org/10.5610/jaee.19.1_121 **(in Japanese with English abstract)**

[11] Nojima, N. and Takahashi, Y. (2020) Effect of source parameter settings on spatial variation and correlation of predicted seismic intensity maps for scenario earthquake. *Report of Tono RIES*, **45**: pp. 53–66. **(in Japanese with English abstract)**

[12] Kagawa, T. (2004) Developing a Stochastic Green's function method having more accuracy in long period range to be used in the Hybrid Method. *Journal of JAEE*, **4** (2): pp. 21–32. https://doi.org/10.5610/jaee.4.2_21 **(in Japanese with English abstract)**

[13] Kin, M. (2007) Data science with R -from basic data analysis to the least methods. Tokyo, *Morikita Publishing Co., Ltd.*, 320 p. **(in Japanese)**

[14] Morishita, K. (2021) Techniques for interpreting machine learning. Tokyo, *Gijutsu-Hyoron Co., Ltd.*, 256 p. **(in Japanese)**

[15] Molnar, C., Bischl, B. and Casalicchio, G. (2018) iml: an R package for interpretable machine learning. *Journal of Open Source Software*, **3** (26), p.786.

[16] Molnar, C. (2019) Interpretable machine learning - a guide for making black box models explainable. https://christophm.github.io/interpretable-ml-book/

5 Diversification of Source Parameter Settings in Recipe for Predicting Strong Ground Motions

5.1 Introduction

The “Recipe for Predicting Strong Ground Motions” [1] has been proposed as a standard for setting up fault rupture scenarios on source faults in Japan. However, the “Scenario Earthquake Shaking Maps (SESMs)” [2], which are set up based on the “recipe,” has a limited number of cases; therefore, SESMs do not cover the various possibilities of ground motion distributions. For this reason, it is important and of great social significance to systematize a method for setting scenario earthquakes that can efficiently cover the diversity of ground motion distributions in terms of disaster countermeasures, designing a city for disaster prevention, and seismic risk assessment. Therefore, this study aims to “diversify ground motion distributions” by expanding the conditions for setting source parameters in the “recipe.”

In **Chapter 4**, this study clarified the mode shapes that define the distribution of ground motions and the source parameters that affect the diversity of the ground motion distributions. Based on the result of **Chapter 4,** in this chapter, the “recipe” setting conditions are extended to represent the diversity of ground motion distributions.

As a numerical example, this study uses 600 cases of fault models that were set with varying values of the “recipe” [3, 4] and their ground motion distributions [5]; this study attempt to reproduce the diversity of spatial characteristics represented by the 600 cases. We attempt to reproduce the diversity of the 600 ground motion distributions by using the ground motion distribution of the fault model conforming to the “recipe” (hereafter referred to as the “recipe case”) and the distribution of the fault model with extended “recipe” setting conditions (hereafter referred to as the “extended case”). Note that in this study, the set values of source parameters for “outer fault parameters (e.g., fault location, fault geometry)” except for seismic moment are fixed, and the set values for a seismic moment and “inner fault parameters (e.g., area and location of asperities, rupture velocity)” and “extract fault parameters (e.g., hypocenter)” are extended.

The data used in this study are characterized source models set up for a strike-slip fault consisting of a single fault plane [5] and the ground motion distributions [5] calculated by the stochastic Green’s function method [6]. In other words, this study aims to cover the diversity of ground motion distributions represented by the stochastic Green’s function method within the setting of the characterized source model for the strike-slip fault by using the recipe cases and the extended cases.

5.2 shows the diversity of the ground motion distributions in the 600 cases by using mode decomposition and cluster analysis. Next, ***5.3*** reveals the spatial characteristics of ground motion distributions and their spread shown by the fault model conforming to the “recipe.” ***5.4*** attempts to diversify the spatial characteristics of scenario earthquakes by extending the “recipe” setting conditions. Finally, ***5.5*** summarizes the conclusions obtained in this chapter.

5.2 Diversity of Ground Motion Distributions

5.2.1 Visualization of Diversity of Ground Motion Distributions

First, the ground motion distribution diversity for the 600 cases is understood. The previous study [7] applied hierarchical cluster analysis and principal component analysis to SESMs to clarify the interrelationships among scenario earthquakes and the influence of source parameters on ground motions, focusing on the distribution shapes. Hierarchical cluster analysis is a classification method that integrates similar cases in a bottom-up approach [8]. In contrast, this study uses non-hierarchical cluster analysis, in which the number of clusters is determined in advance in a top-down approach [8], as a classification method for many cases. This method is outlined in ***5.2.2***.

Figure. 5-1 shows the method for understanding the diversity of the ground motion distribution for the 600 cases. First, mode decomposition is applied to the 600 cases to extract mode forms which are spatial characteristics of the ground motion distribution and principal component scores (PCS; singular value × weight coefficient) which are modal weights. Next, non-hierarchical cluster analysis is applied to the PCSs to cluster the 600 cases based on the similarity of the PCSs (i.e., the similarity of spatial distribution trends represented by the ground motion distribution). Then, using a two-dimensional distribution with mode as the axis, the relative relationships among the clusters are clarified to visualize the diversity of the 600 ground motion distributions.

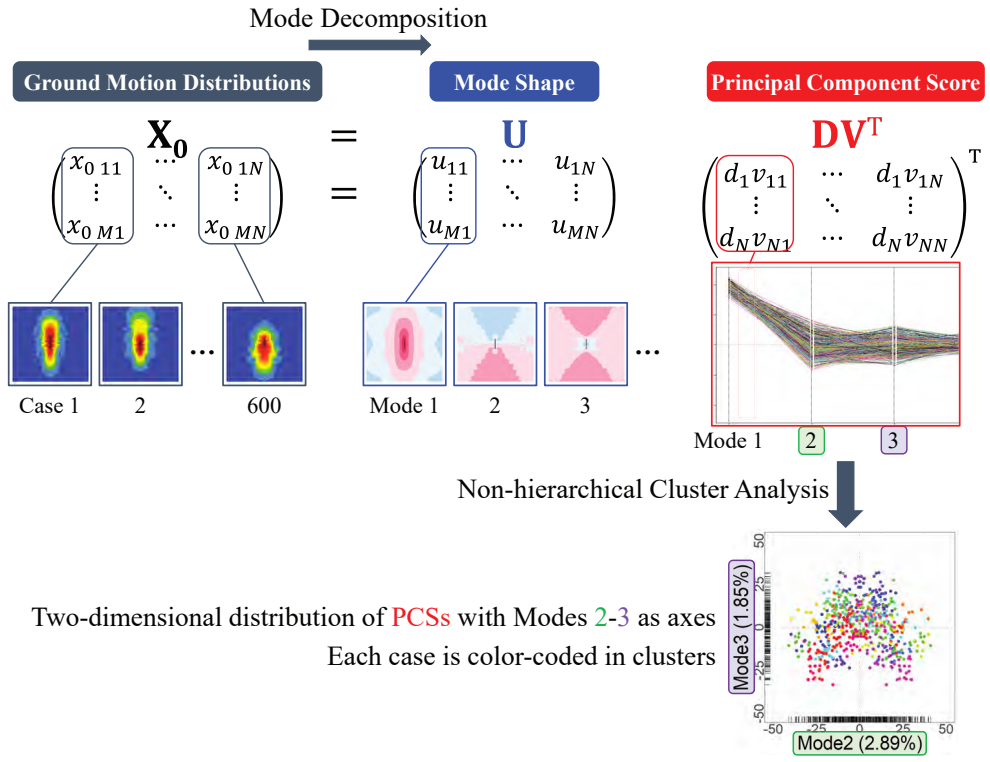


Figure 5-1. Evaluation of the diversity of ground motion distributions using mode decomposition and non-hierarchical cluster analysis.

5.2.2 Non-hierarchical Cluster Analysis

Non-hierarchical cluster analysis is a classification method that forms clusters in a top-down approach by pre-determining the number of partitions k [8]. This analysis method is generally used for big data, and the k-means clustering is representative. This study treats the PCSs of all modes as feature vectors and applies the k-means method to the dissimilarity matrix based on the Euclidean distance between feature vectors. Fig. 5-2 shows an image of the k-means algorithm.

1. Randomly assign the centroids of the initial k clusters.
2. Each element (PCS in each case) is assigned to the nearest centroid, forming k clusters.
3. For each cluster, compute its centroid.
4. Iterate steps 2.–3. until the cases no longer change clusters.

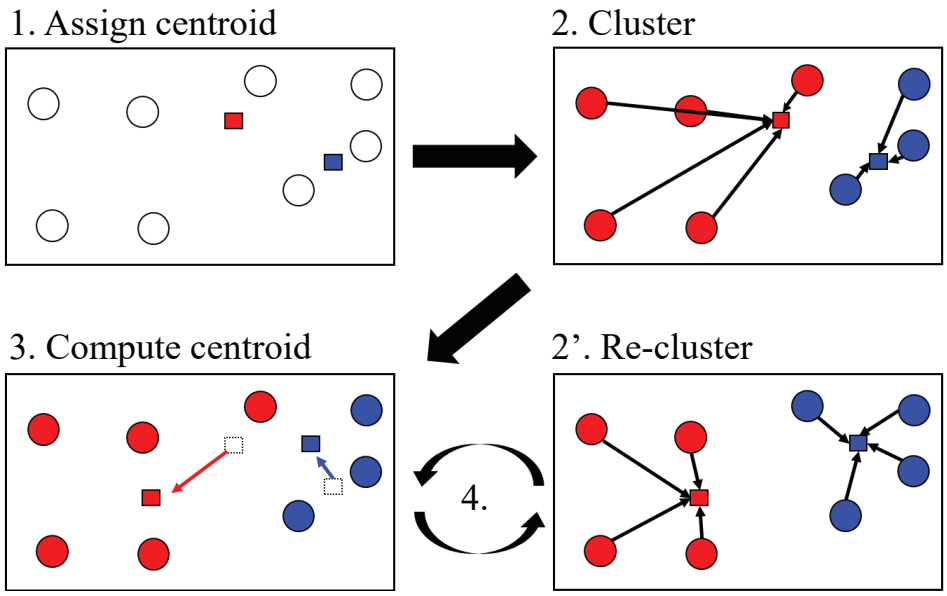


Figure 5-2. K-means algorithm.

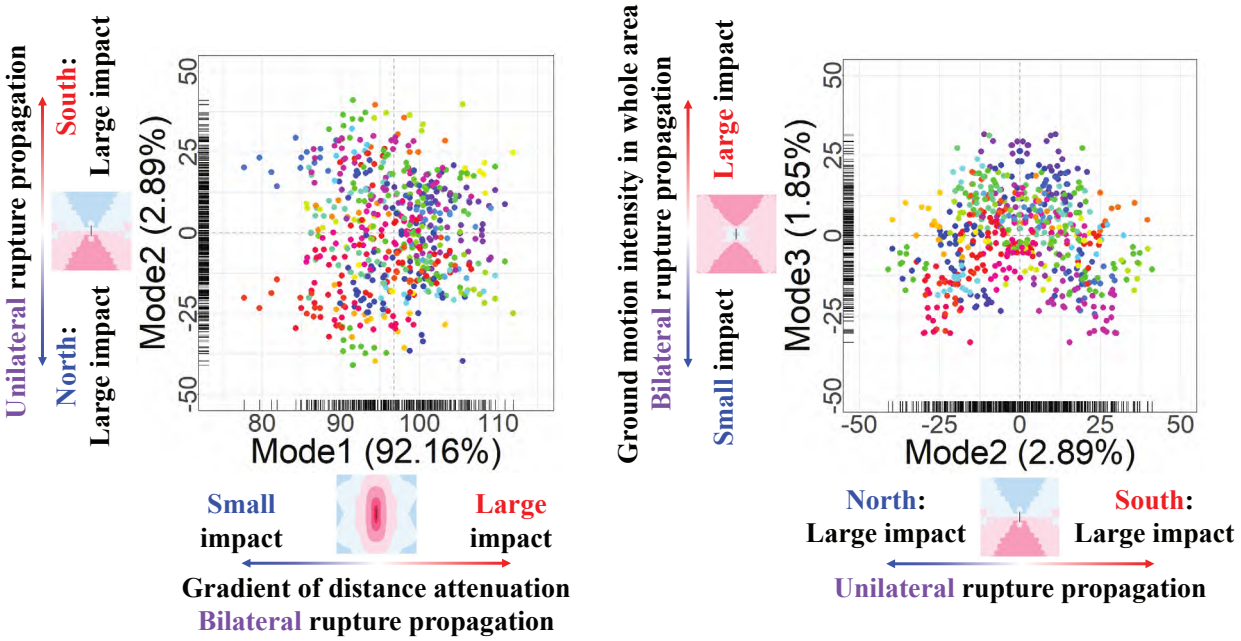
5.2.3 Understanding the Diversity of 600 Ground Motion Distributions

In **Chapter 3**, mode decomposition was applied to the 600 ground motion distributions to obtain the mode shapems and the PCSs; K-means is applied to the PCS, with k = 60 as the number of divisions, assuming that an average of about 10 cases belong to a single cluster. In the following, the analysis results are shown for the distribution of the absolute acceleration response spectrum of the FN component in period 2.0 s (hereafter referred to as S_A (T = 2.0, FN)).

Figure. 5-3 shows the two-dimensional distribution of PCSs with Mode 1-2 and Mode 2-3. Here, the dots represent the PCS for each case, color-coded in clusters classified by non-hierarchical cluster analysis. Each case formss a cluster with its surrounding cases, indicating that the cluster analysis results are good. The two-dimensional distributions represent the diversity of spatial characteristics in the 600 ground motion distributions.

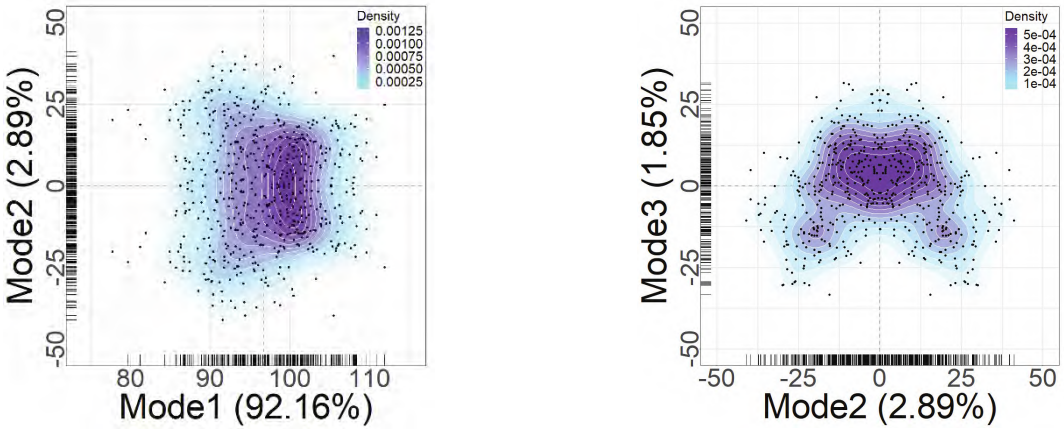
The spatial characteristics of the case where Mode 1 is around 80 and Mode 2 is around 0 in Fig. 5-3 (a), and those where Mode 2 is around 0 and Mode 3 is around -25 in Fig. 5-3 (b) are not represented in the 600 cases. Here, Mode 1 and Mode 3 reflect the effect of bilateral rupture propagation. From Fig. 4-7 (b) and Fig. 4-10 in **Chapter 4**, this effect is stronger when the hypocenter is located at the fault center and weaker when it is located at the fault edge. In contrast, Mode 2 reflects the effect of unilateral rupture propagation. Fig. 4-12 (b) shows that this effect is stronger when the hypocenter is located at the fault edge and weaker when the hypocenter is located at the fault center. Thus, the spatial characteristics in these regions in Fig. 5-3 on Mode 1 or Mode 3 axis appear when the hypocenter is located at the fault edge, whereas on the Mode 2 axis, they appear when the hypocenter is located at the fault center; therefore, these conditions are never satisfied simultaneously, and the spatial characteristics of the region in question are not represented. The above results indicate that there are spatial characteristics in the two-dimensional distribution between some of the modes that are not represented because the setting conditions of the source parameters constrain them. Therefore, when examining the “diversification of ground motion distribution,” it is necessary to extend the “recipe” to consider such diversity that depends on the source parameters’ setting.

Figure. 5-4 shows the kernel density estimation for PCSs in Modes 1-2 and Modes 2-3. The kernel density estimation uses a Gaussian kernel, and the bandwidths are estimated using a bandwidth estimator [9]. The results show that the diversity of spatial characteristics and the density of the 600 cases can be described simultaneously. Therefore, the density analysis results will be one of the indices for determining the spatial characteristics to be expressed preferentially in the expansion of the “recipe” and for comparison with the 600 cases.



(a) PCSs with Mode 1 and Mode 2 **(b)** PCSs with Mode 2 and Mode 3

Figure 5-3. Results of applying non-hierarchical cluster analysis (k = 60) and mode shapes in S_A (T = 2.0, FN). (Dashed line: mean of PCSs, black line on an axis: distribution of PCSs)



(a) Mode 1 and Mode 2 **(b)** Mode 2 and Mode 3

Figure 5-4. Six hundred cases: Kernel density estimation in two-dimensional space of PCS by two modes.

Figure. 5-5 (b) and Fig. 5-6 (b) show the centroid of each cluster. They show the positional relationship among the clusters and simplify the diversity of the ground motion distribution. The following 9 clusters are defined as representative clusters in a Modes 1-2 or Modess 2-3 two-dimensional distribution.

- Red circle: 8 clusters located in the outer periphery
- Blue circle: 1 cluster located in the center

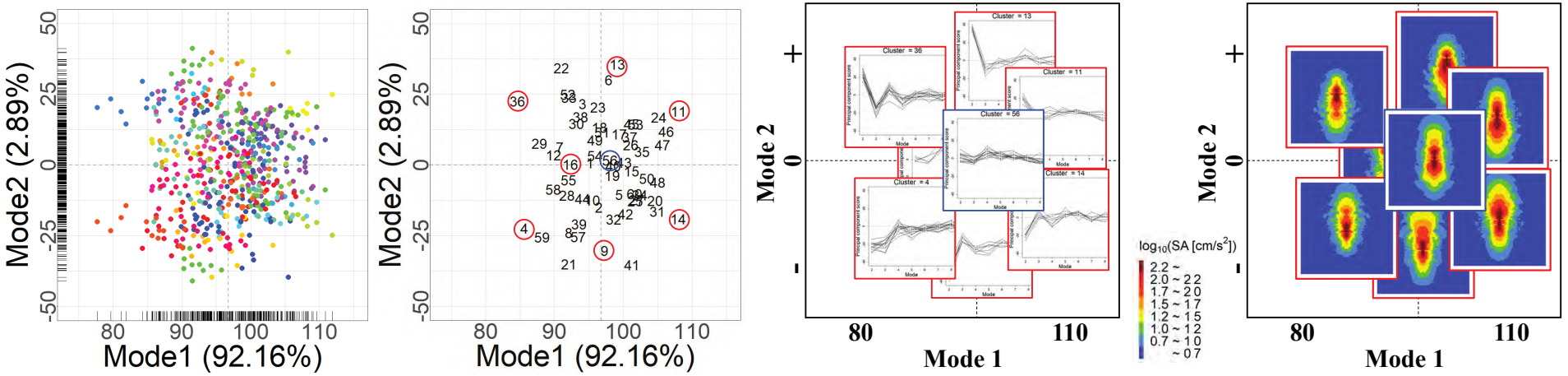
Figure. 5-5 (c) and Fig. 5-6 (c) show the PCSs for all cases belonging to the representative clusters. In Fig. 5-6 (c), the cluster on the right contains the cases with positive values of Mode 2, and the cluster on the left contains the cases with negative values of Mode 2. Furthermore, the upper cluster contains cases with positive values of Mode 3, while the lower cluster contains cases with negative values of Mode 3. In summary, it is found that the cluster analysis integrated the cases showing similar spatial characteristics.

The representative case is defined as the case with the minimum Euclidean distance between the cluster centroid and the cases belonging to the cluster. Fig. 5-5 (d) and Fig. 5-6 (d) show the ground motion distributions for the representative cases.

First, focus on Mode 1 and Mode 2 in Fig. 5-5 (d). The ground motion distribution on the right side of the figure (Mode 1: large) shows high ground motion levels around the fault and strong ground motion ($\log_{10}(S_A\ [cm/s^2]) \geq 1.7$) spreading from the fault to the southern and northern parts of the map. In contrast, on the left side of the figure (Mode 1: small value), the ground motion level around the fault is smaller than that on the right side, and there is no effect of bilateral rupture propagation. Next, on the upper side of the figure (Mode 2: positive), strong ground motions spread from the fault to the southern part of the map, while on the lower side (Mode 2: negative), strong ground motions spread from the fault to the northern part of the map. These trends are consistent with the relationship between mode shapes and PCS showed in Fig. 5-3 (a). The relationship of ground motion distribution between clusters in Mode 2 and Mode 3 in Fig. 5-6 (d) is also consistent with the relationship between mode shape and PCS in Fig. 5-3 (b).

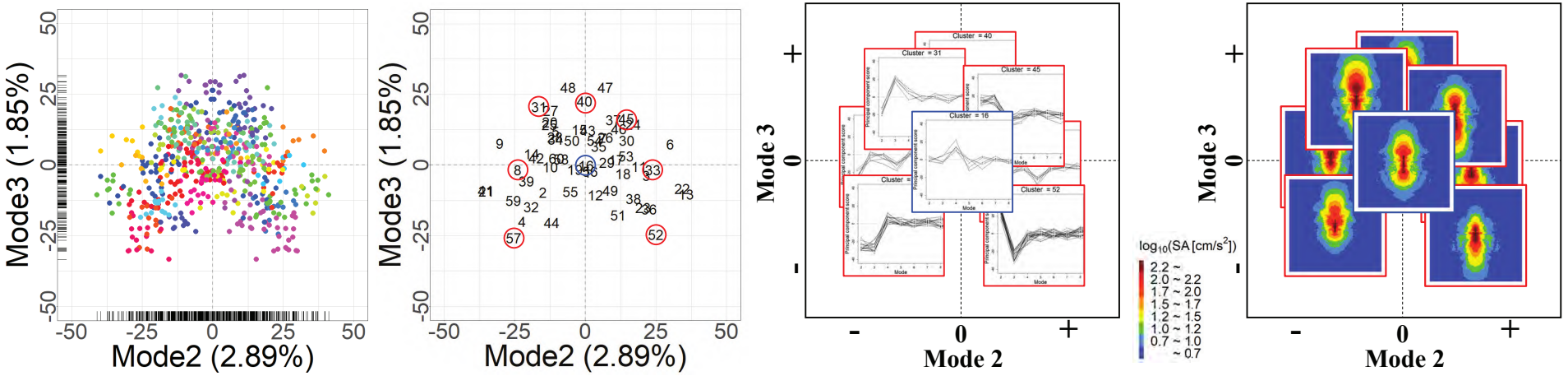
As described above, the ground motion distributions in Fig. 5-5 (d) and Fig. 5-6 (d) reflect the modal characteristics. Hence, by using mode decomposition and non-hierarchical cluster analysis, the diversity of the distribution of ground motions in the 600 cases can be classified and visualized concerning spatial characteristics.

In the following, this study assumes that the 600 fault models that are set with the “recipe” value and its surrounding values and their ground motion distributions [5] cover the diversity of fault models and ground motion distributions; then this study examines the extension of the “recipe.”



(a) 600 cases (Reprint of Fig. 5-3 (a)) **(b)** Centroids of clusters (Number: cluster number) **(c)** PCSs in representative clusters (Red and blue circles in Fig. 5-5 (b)) **(d)** Ground motion Maps in representative clusters

Figure 5-5. Modes 1-2 for 600 cases: Results of applying non-hierarchical cluster analysis (k = 60) and characteristics between clusters.



(a) 600 cases (Reprint of Fig. 5-3 (b)) **(b)** Centroids of clusters (Number: cluster number) **(c)** PCSs in representative clusters (Red and blue circles in Fig. 5-6 (b)) **(d)** Ground motion Maps in representative clusters

Figure 5-6. Modes 2-3 for 600 cases: Results of applying non-hierarchical cluster analysis (k = 60) and characteristics between clusters.

5.3 Fault Models Conforming "Recipe of Predicting Strong Ground Motions" and their Ground Motion Distributions

First, fault models conforming to the setting conditions for the source parameters of the "recipe" (hereafter referred to as "recipe cases") and their ground motion distributions are assumed. Then, the relative positions of the ground motion distributions of the recipe cases to the 600 cases are shown to clarify their diversity. In the following study, the ground motion distribution is estimated by the prediction model using gradient boosting decision trees in **Chapter 3** (modes used for synthesis: Modes 1–15).

5.3.1 Fault Models Conforming "Recipe of Predicting Strong Ground Motions"

Set up fault models for the recipe cases on a transverse strike-slip fault consisting of a single fault plane (length: L = 26 km, width: W = 16 km). Fig. 5-7 shows the fault models for the four recipe cases set based on the source parameters setting conditions in the "recipe" [1, 3, 4] below.

- Seismic moment M_0 : 9.63×10^{25} [dyne・cm]
- Rupture velocity Vr : 2.52 [km/s]
- Asperities area ratio ratio : 22 %
- Stress drop (average) $\Delta\sigma^{*1}$: 2.76 [MPa]
- Stress drop (asperities) $\Delta\sigma_a{}^{*2}$: 12.5 [MPa]
- Arrangement of asperities : A well-balanced arrangement
- Position of hypocenter : Left and right corners of the lower edge of asperities

*1 $\Delta\sigma = 7/16 \times M_0 \times (\pi / LW)^{3/2}$

*2 $\Delta\sigma_a = \Delta\sigma$ / ratio

The four recipe cases are denoted as R1–R4.

5.3.2 Ground Motion Distributions for Recipe Cases

Figure. 5-8 shows the ground motion distributions for recipe cases. The strong ground motions are distributed in the northern (R1, R3) or southern (R2, R4) part of the map.

Figure. 5-9 (a) and Fig. 5-10 (a) show the relative positions of the recipe cases to the 600 cases. These figures plot R1–R4 on Fig. 5-5 (b) and Fig. 5-6 (b). The values for Mode 1 and Mode 3 of the recipe cases are similar to the average of the 600 cases, while Mode 2 has a large absolute value and is positively and negatively split. This result indicates that the gradient of distance attenuation (Mode 1) and the global level of ground motion (Mode 3) for R1–R4 are average, and that these ground motion distributions reflect the effects of unilaterally rupture propagation (Mode 2).

Figure. 5-9 (a) and Fig. 5-10 (a) show the kernel density estimation. Here, the kernel density bandwidths are the same value as in the 600 cases in Fig. 5-4 for comparison with the 600 cases. The bandwidth is the same in the following studies. The recipe cases are located around the regions of high density among the 600 cases. However, the spatial characteristics represented by the recipe cases are very limited, indicating that the diversity of the ground motion distribution cannot be estimated (or represented) from the four recipe cases.

From the above, the ground motion distributions for the recipe case (Fig. 5-8) represent average and limited spatial characteristics for the 600 cases, and do not adequately represent the spread of Mode 1 and Mode 3. Furthermore, the ground motion distribution such that the influence of Mode 2 does not appear (Mode 2 is 0) is also not represented in the recipe cases.

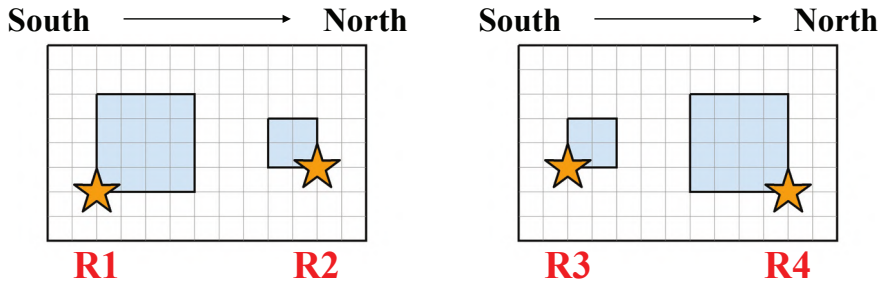


Figure 5-7. Recipe cases: fault models. (Star: hypocenter, blue rectangle: asperity)

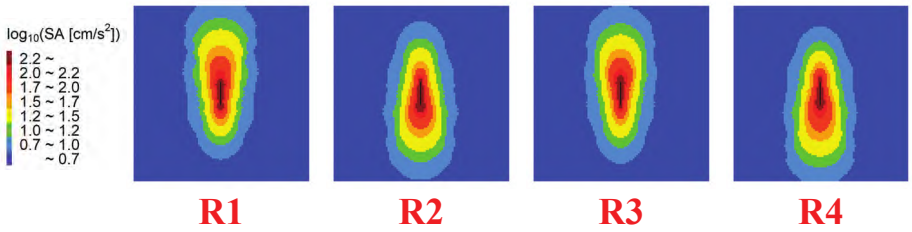
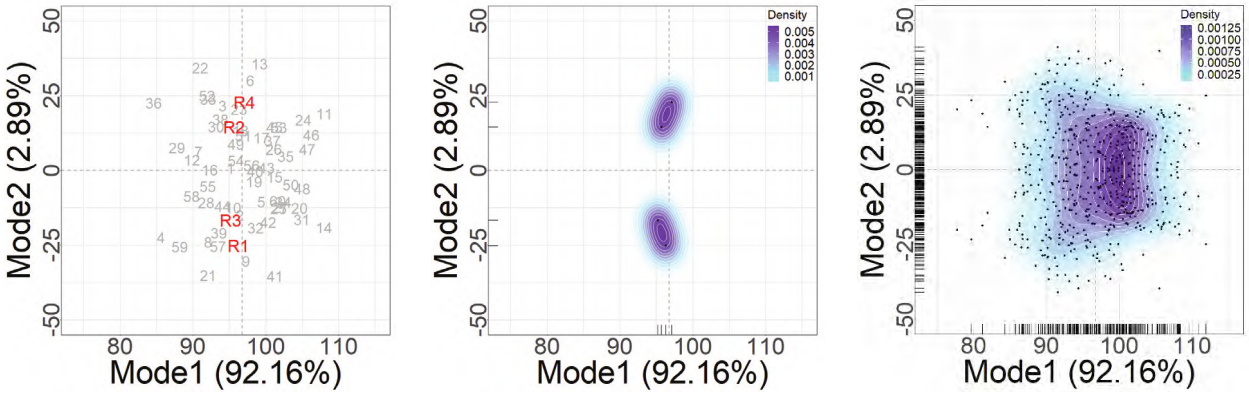


Figure 5-8. Recipe cases: ground motion distributions in S_A (T = 2.0, FN).

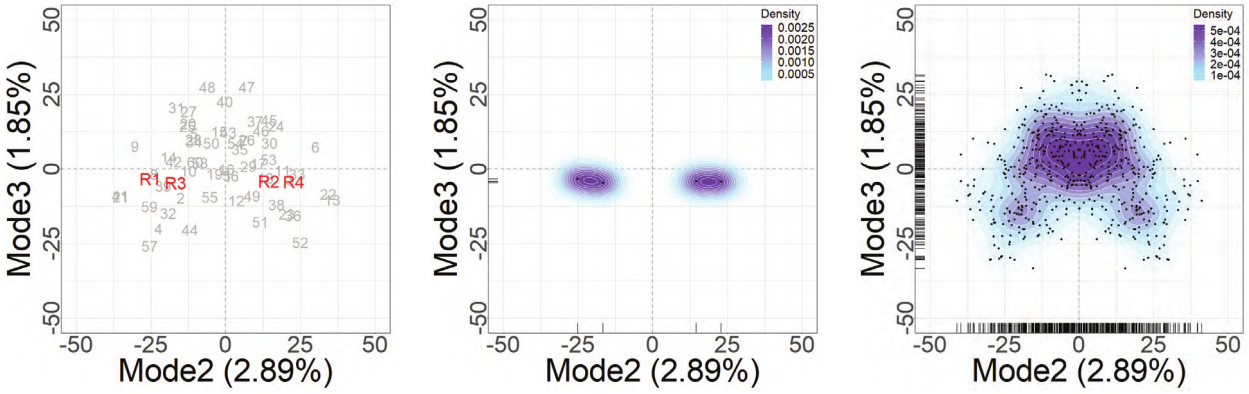


(a) Recipe cases (red) and 600 cases (gray) (Plotted on Fig. 5-5 (b))

(b) Recipe cases: Kernel density estimation

(c) 600 cases: Kernel density estimation (Fig. 5-4 (a))

Figure 5-9. Modes 1-2 for recipe cases.



(a) Recipe cases (red) and 600 cases (gray) (Plotted on Fig. 5-6 (b))

(b) Recipe cases: Kernel density estimation

(c) 600 cases: Kernel density estimation (Fig. 5-4 (b))

Figure 5-10. Modes 2-3 for recipe cases.

5.4 Fault Models Extending "Recipe of Predicting Strong Ground Motions" and its Ground Motion Distributions

By diversifying the "recipe" setting conditions, it attempted to reproduce the diversity of spatial characteristics that could not be expressed in the recipe cases. Based on the analysis results in **Chapter 4**, this study sets fault models with extended setting conditions for source parameters (hereafter referred to as the "extended case"). The source parameters to be extended are described below and listed in Fig. 5-11.

- Horizontal position of hypocenter : H1–H4 (green)

- Rupture velocity (Vr) : Case with large value ; V(L)1–V(L)4 (blue)
  Case with small value ; V(S)1–V(S)4 (blue)

-Seismic moment (M_0) : Case with large value ; M(L)1–M(L)4 (gold)
  Case with small value ; M(S)1–M(S)4 (gold)

- Hypocenter + Rupture velocity (Vr) : Case with large value of Vr ; V(L)H1, V(L)H4 (navy)
  Case with small value of Vr ; V(S)H1, V(S)H4 (navy)

- Hypocenter + Seismic moment (M_0) : Case with large value of M_0 ; M(L)H1, M(L)H4 (brown)
  Case with small value of M_0 ; M(S)H1, M(S)H4 (brown)

- Horizontal arrangement of asperities*: A(H)1–A(H)4 (violet)

- Vertical arrangement pf asperities* : A(V)1–A(V)8 (light blue)

- Corner arrangement of asperities* : A(C)1–A(C)8 (purple)

* The position of the hypocenter also changes with the arrangement of asperities.

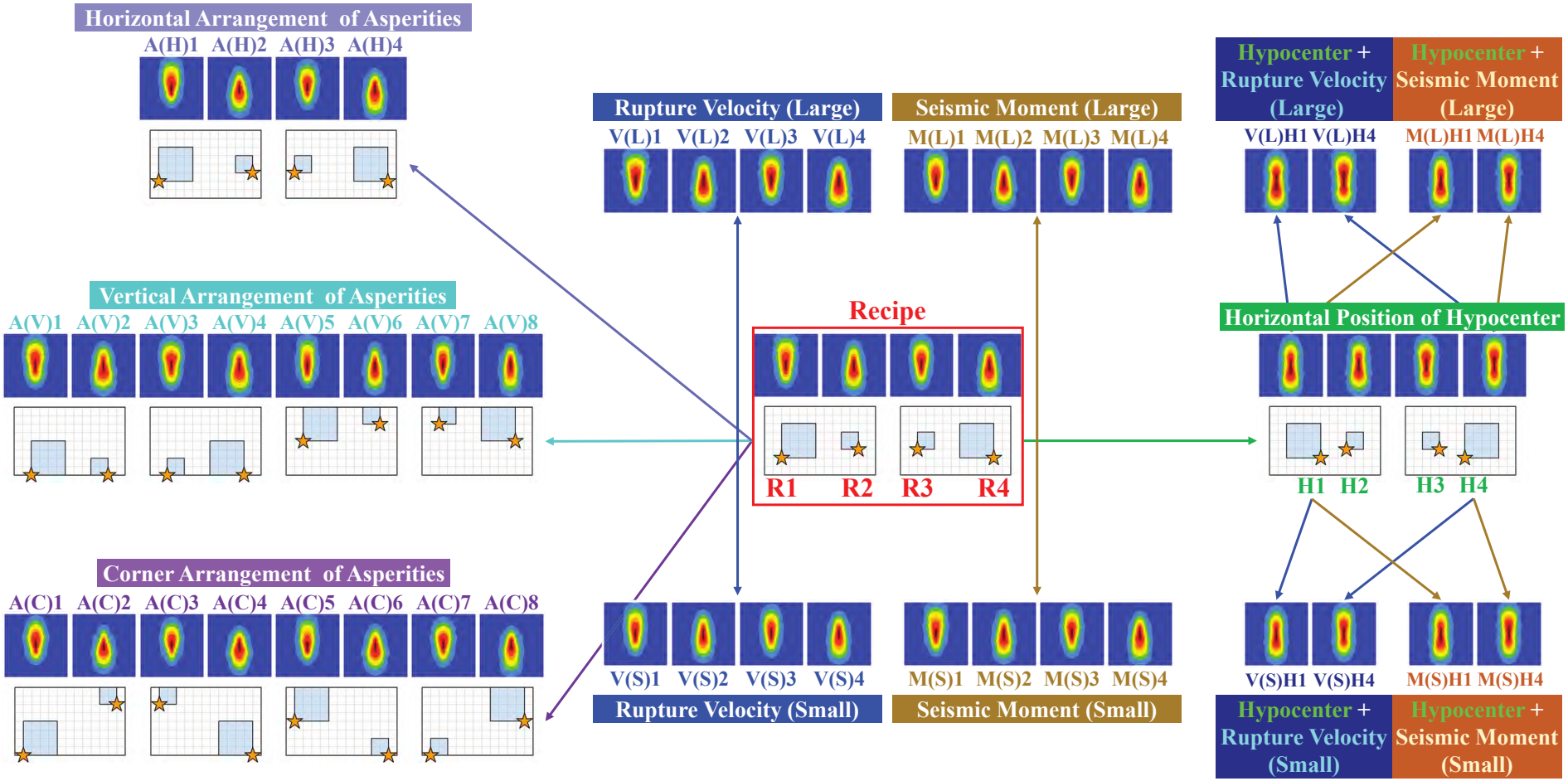


Figure 5-11. Extended conditions for source parameters in the "Recipe."

5.4.1 Cases with Extended Setting Conditions for Hypocenter

From **Chapter 4**, the parameters related to the horizontal position of the hypocenter (Hypo.H, Hor.P) are among the dominant parameters for the spatial characteristics in Modes 1–3 of S_A (T = 2.0, FN). However, in the recipe cases, the hypocenter is set only on the fault edge side of the lower edge of asperity, as shown in Fig. 5-7. Then, as extended cases about the horizontal position of the hypocenter in the "recipe," this study assumes fault models in which the hypocenter is located at the fault central side of the lower edge of asperity (Fig. 5-12, H1–H4). In H1–H4, all parameters (except the hypocenter, e.g., seismic moment and rupture velocity) are the same as in the recipe cases (R1–R4).

Figure. 5-13 shows the ground motion distributions for H1 – H4; these maps reflect the effects of bilateral rupture propagation from the fault to the northern and southern parts of the map.

Figure. 5-14 (a) and Fig. 5-15 (a) show the expression range of the spatial characteristics based on PCSs for the recipe cases (R1–R4: red) and the extended case (H1–H4: green). In H1–H4, Mode 2 is closer to zero than in the recipe cases, indicating that the effect of unilaterally rupture propagation does not appear. The values for Mode 1 and Mode 3 are larger than those of the recipe cases; they indicate that the ground motion levels for H1–H4 are higher than those for the recipe cases, and that the bilateral rupture influences the spread of ground motion.

Figure. 5-14 (b) and Fig. 5-15 (b) show the kernel density estimates for the recipe and extended cases (H1–H4). By setting the extended cases H1–H4 for the horizontal position of the hypocenter, the range of spatial characteristics that can be expressed is broadened; the density distribution for the recipe and extended cases (R1–R4, H1–H4) captures the characteristics of the 600 cases better than the recipe case alone (Fig. 5-10– Fig. 11).

The above results show that by setting H1–H4, it is possible to have a spread for Mode 1 and Mode 3 and also express the spatial characteristics with a small effect of Mode 2.

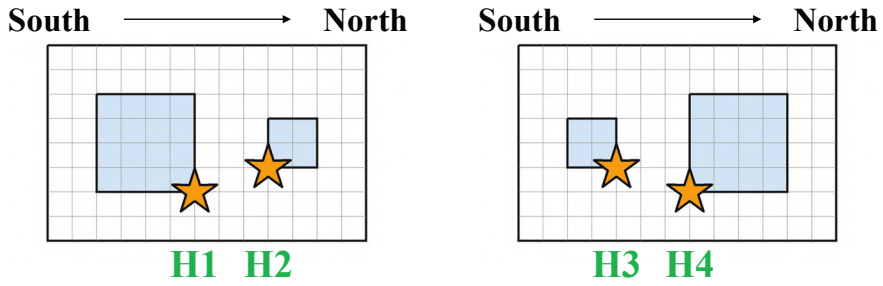


Figure 5-12. Extended cases for hypocenter: fault models.

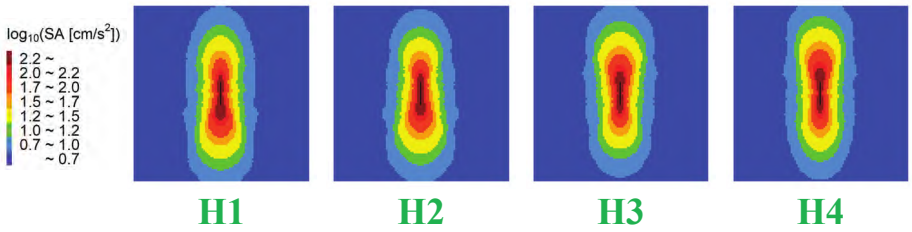
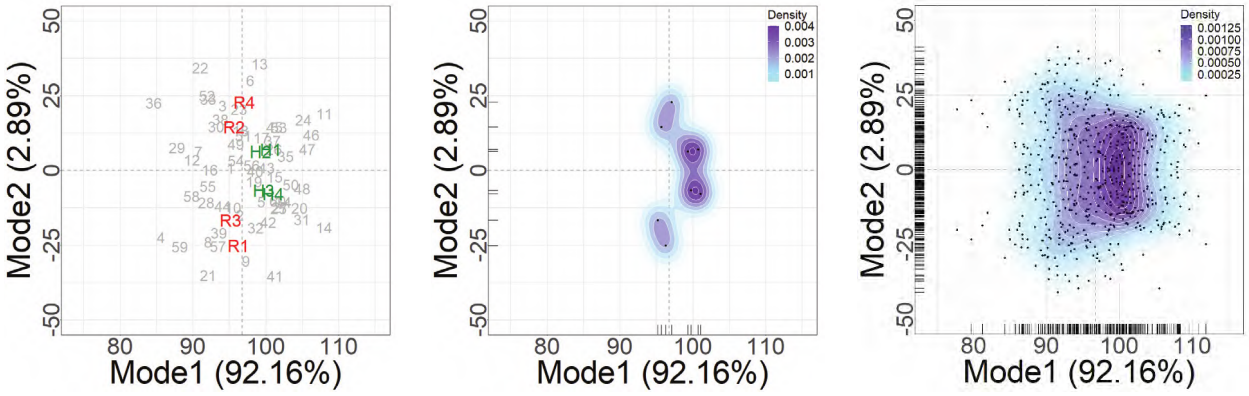


Figure 5-13. Extended cases for hypocenter: ground motion distributions in S_A (T = 2.0, FN).

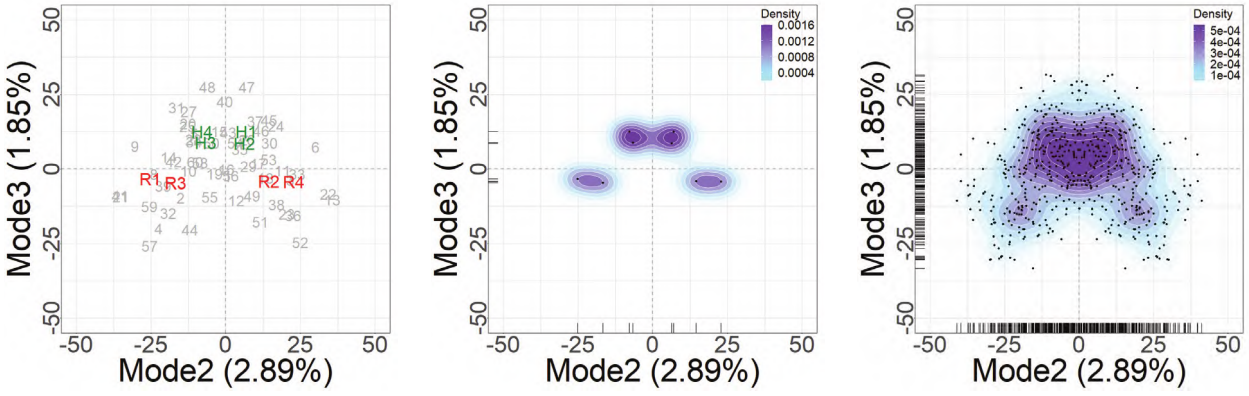


(a) Recipe and extended cases (Plotted on Fig. 5-5 (b))

(b) Recipe and extended cases: Kernel density estimation

(c) 600 cases: Kernel density estimation (Fig. 5-4 (a))

Figure 5-14. Modes 1-2 for extended cases for hypocenter.



(a) Recipe and extended cases (Plotted on Fig. 5-5 (b))

(b) Recipe and extended cases: Kernel density estimation

(c) 600 cases: Kernel density estimation (Fig. 5-4 (a))

Figure 5-15. Modes 2-3 for extended cases for hypocenter.

5.4.2 Cases with Extended Setting Conditions for Rupture Velocity

The “recipe” basically sets only one type of rupture velocity and seismic moment. However, these values are averages based on various observation records and contain variations. Specifically, the set value of rupture velocity in the “recipe” is μ = 2.52 [km/s] [3], [4], and its variation is σ = 0.126 [km/s] [10], [11]. This study assumed fault models with large rupture velocity (V(L)1–V(L)4) and fault models with small one (V(S)1–V(S)4); the rupture velocities of V(L) and V(S) are set as follows:

- V(L)1–V(L)4: $\mu + \sigma$ = 2.646 [km/s]
- V(S)1–V(S)4: $\mu + \sigma$ = 2.394 [km/s]

Figure. 5-16 and Fig. 5-17 show their fault models and ground motion distributions, respectively.

Figure. 5-18 and Fig. 5-19 show the PCSs and their kernel density estimates for the recipe cases (R1–R4: red) and the extended cases (V(L)1–V(L)4, V(S)1–V(S)4: blue). For Mode 1 and Mode 2, the improvement effect of the diversity of the ground motion distribution (hereafter referred to as the “diversifying effect”) by adding the extended case is small. In contrast, for Mode 3, a slight diversifying effect is observed: increasing (/ decreasing) the rupture velocity results in higher (/ lower) global ground motion levels.

The current SESMs basically do not set up fault models that vary the global ground motion level; therefore, when considering diversification of ground motion distribution, it is necessary to set up multiple types of rupture velocities (and seismic moments).

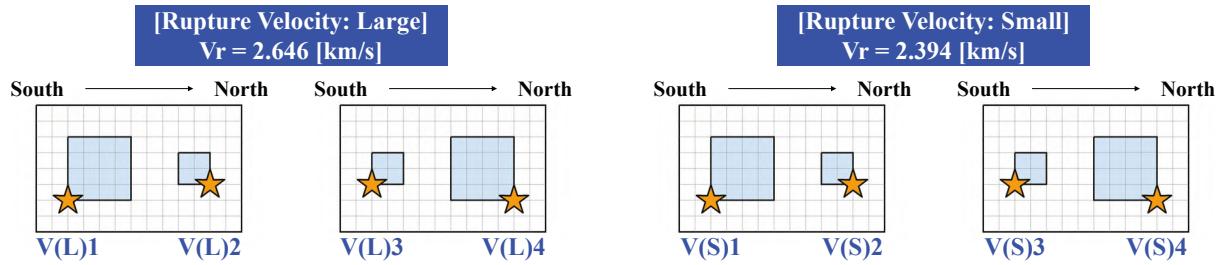


Figure 5-16. Extended cases for rupture velocity: fault models.

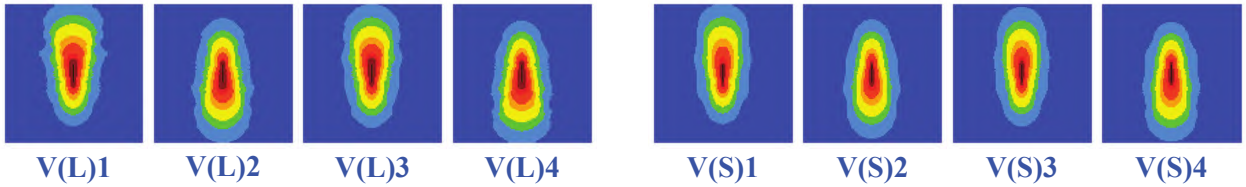
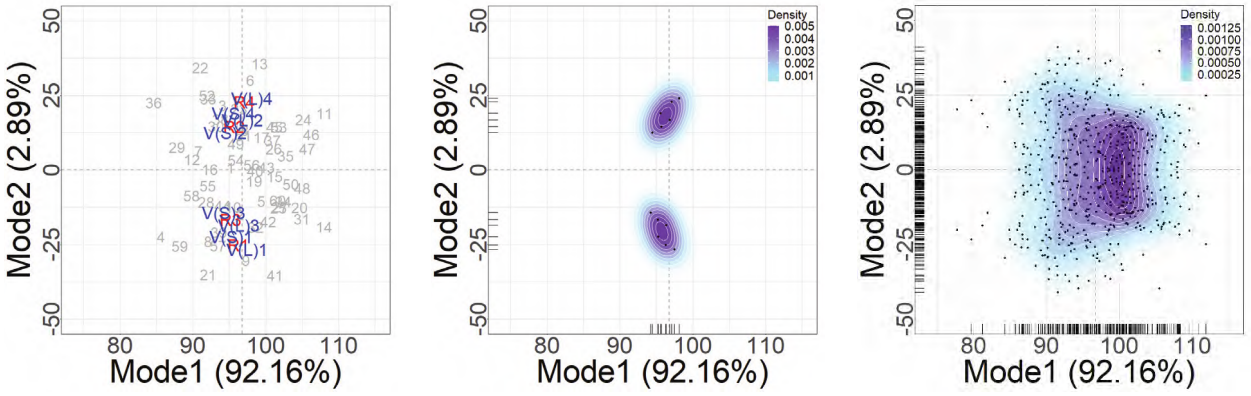


Figure 5-17. Extended cases for rupture velocity: ground motion distributions in S_A (T = 2.0, FN).

s

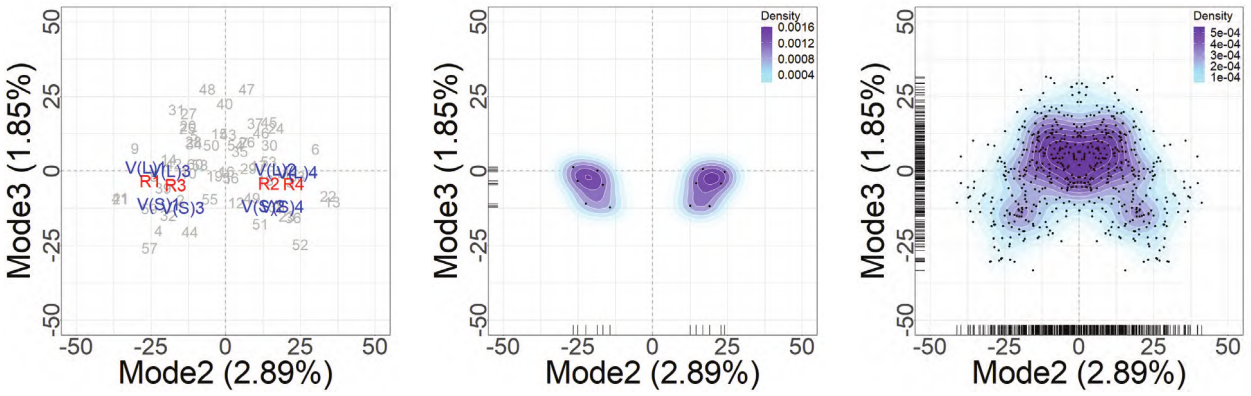


(a) Recipe and extended cases (Plotted on Fig. 5-5 (b))

(b) Recipe and extended cases: Kernel density estimation

(c) 600 cases: Kernel density estimation (Fig. 5-4 (a))

Figure 5-18. Modes 1-2 for extended cases for rupture velocity.



(a) Recipe and extended cases (Plotted on Fig. 5-5 (b))

(b) Recipe and extended cases: Kernel density estimation

(c) 600 cases: Kernel density estimation (Fig. 5-4 (a))

Figure 5-19. Modes 2-3 for extended cases for rupture velocity.

5.4.3 Cases with Simultaneously Extended Setting Conditions for Hypocenter and Rupture Velocity

In ***5.4.1*** and ***5.4.2***, the diversifying effect of the extended cases was clarified by independently varying the horizontal position of the hypocenter and the rupture velocity. Next, the extended cases are set up in which these parameters are varied simultaneously to clarify the diversifying effect.

Fault models (V(L)H1, V(L)H4 and V(S)H1, V(S)H4) are assumed, in which the rupture velocities of H1 and H4 are varied from large or small (Vr = 2.646, 2.394 [km/s]). Fig. 5-20 and Fig. 5-21 show their fault models and ground motion distributions, respectively.

Figure. 5-22 and Fig. 5-23 show the PCSs and their kernel density estimates for the recipe cases (R1–R4: red) and the extended cases (V(L)H1, V(L)H4, V(S)H1, V(S)H4: navy). Focusing on Mode 1 and Mode 3, the PCS is larger than that of the recipe cases (R1–R4); the ground motion distributions (Fig. 5-21) are strongly influenced by the bilateral rupture propagation reflected in these modes. For Mode 2, the absolute value of PCS is smaller than that of the recipe case, indicating that the effect of unilaterally propagating rupture is small.

The PCSs of Mode 1 and Mode 3 for V(L)H1 and V(L)H4 with large rupture velocity are larger than the extended cases where the hypocenter (Figs. 5-14, 15: H1–H4) or the rupture velocity (Figs. 5-18, 19: V(L / S)1–V(L / S)4) are independently varied. Therefore, these extended cases, in which several source parameters are varied simultaneously, have a wider range of spatial characteristics that can be expressed than the case in which only a single parameter is extended.

These results indicate that the expansion condition combining multiple source parameters can express various spatial characteristics more efficiently than the expansion condition for a single parameter. Therefore, when examining the extension of the “recipe,” it is effective to assume extended conditions that consider the interaction of multiple parameters.

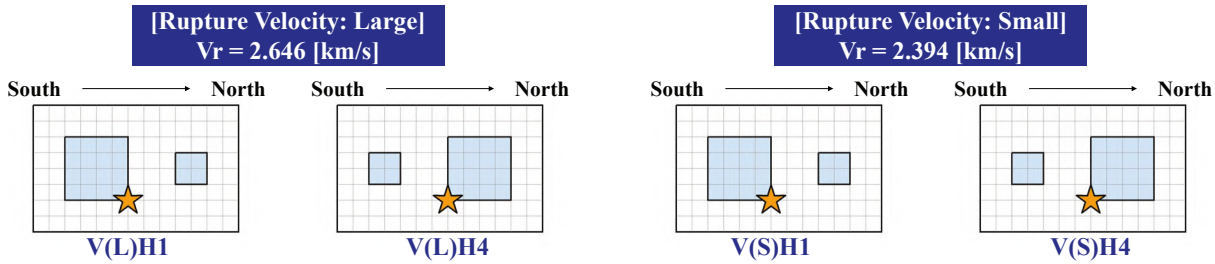
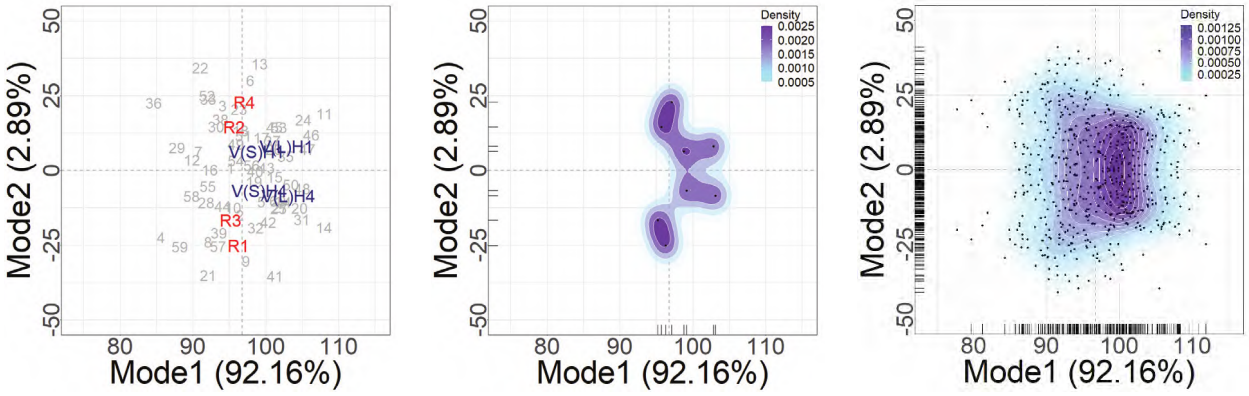


Figure 5-20. Extended cases for hypocenter and rupture velocity: fault models.

Figure 5-21. Extended cases for hypocenter and rupture velocity: ground motion distributions in S_A (T = 2.0, FN).

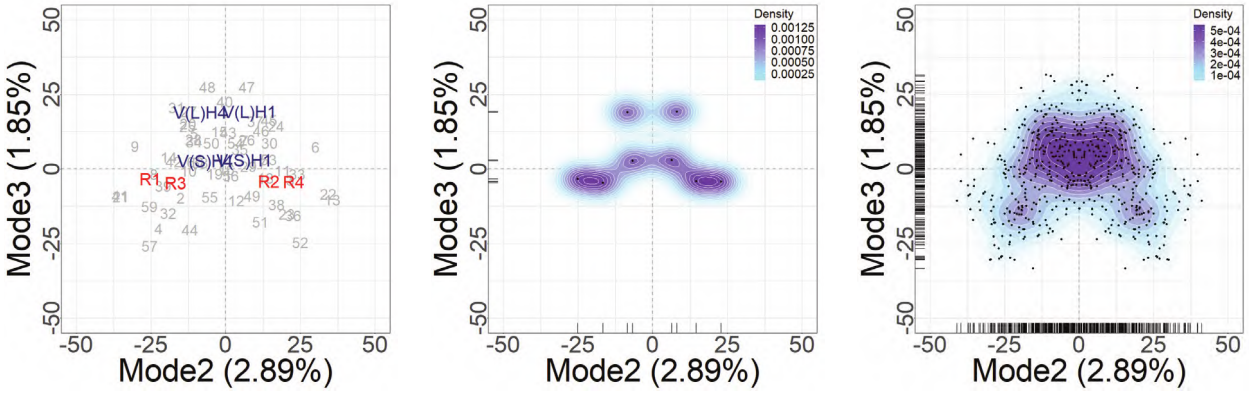


(a) Recipe and extended cases (Plotted on Fig. 5-5 (b))

(b) Recipe and extended cases: Kernel density estimation

(c) 600 cases: Kernel density estimation (Fig. 5-4 (a))

Figure 5-22. Modes 1-2 for extended cases for hypocenter and rupture velocity.



(a) Recipe and extended cases (Plotted on Fig. 5-5 (b))

(b) Recipe and extended cases: Kernel density estimation

(c) 600 cases: Kernel density estimation (Fig. 5-4 (a))

Figure 5-23. Modes 2-3 for extended cases for hypocenter and rupture velocity.

5.4.4 Representation of Diversity of Spatial Characteristics Expressed by Ground Motion Distributions of Recipe Cases and Extended Cases

a) Absolute Acceleration Response Spectrum of FN Component in Period 2.0 s

In addition, as shown in Fig. 5-11, this study attempted to the diversification of the ground motion distribution by assuming the following extended cases:

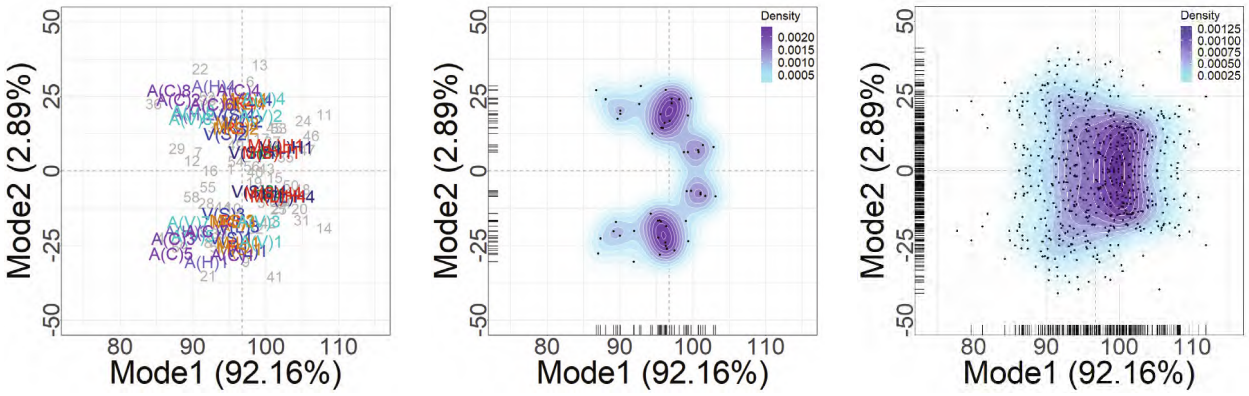
- M(L) and M(S) : Extended case with large and small seismic moments
- M(L)H and M(L)H : Extended case where the hypocenter and the seismic moment are varied simultaneously
- A(H), A(V) and A(C) : Extended cases for asperities arrangement

Here, the set value of seismic moment in the "recipe" is $\mu = 9.63 \times 10^{25}$ [dyne・cm] [3, 4], and its variation is $\sigma = 0.482 \times 10^{25}$ [dyne・cm] [10, 11]. The extended cases with large rupture velocity (M(L) and M(L)H) and fault models with a small one (V(S)1–V(S)4); The seismic moment for the extended cases with large seismic moment (M(L) and M(L)H) and small cases (M(S) and M(S)H) are set as follows:

- M(L)1 – M(L)4, M(L)H1 and M(L)H4 : $\mu + \sigma = 10.1 \times 10^{25}$ [dyne・cm]
- M(S)1 – M(S)4, M(S)H1 and M(S)H4 : $\mu - \sigma = 9.15 \times 10^{25}$ [dyne・cm]

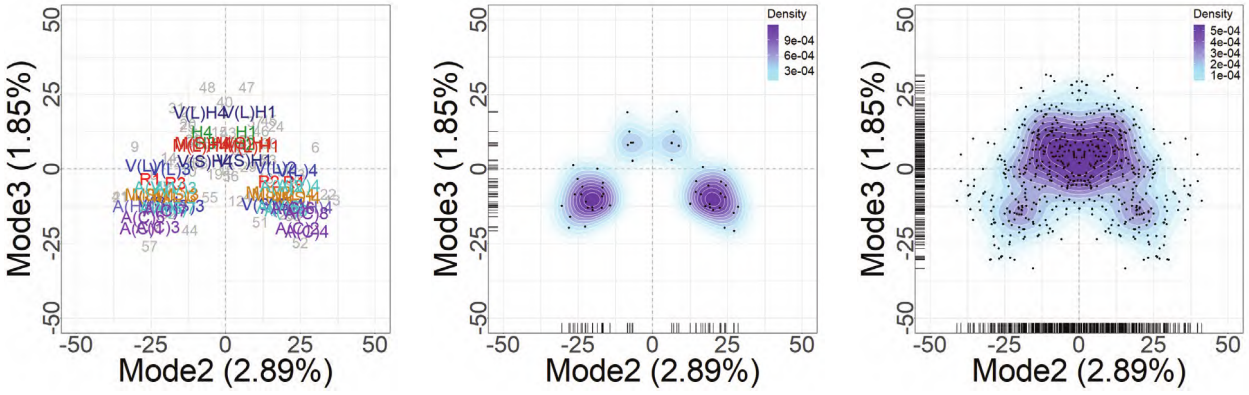
In the extended cases related to seismic moment, the set values of "asperities area ratio" and "stress drop of asperities" related to seismic moment vary simultaneously. Therefore, this study assumed a total of 48 extended cases in addition to the four recipe cases.

Figure. 5-24 and Fig. 5-25 show the diversity of spatial characteristics represented by the 52 recipe and extended cases. These cases are able to reproduce the diversity of spatial characteristics of the 600 cases. However, most of the extended cases are located in the periphery of the recipe cases, and the kernel density shape of the recipe and extended cases does not have the overall spread of the 600 cases; thus, the recipe and extended cases do not efficiently represent the diversity. In addition, some areas are not covered by the recipe and extended cases, such as the area where Mode 2 is 0.



(a) Recipe and extended cases (Plotted on Fig. 5-5 (b)) **(b)** Recipe and extended cases: Kernel density estimation **(c)** 600 cases: Kernel density estimation (Fig. 5-4 (a))

Figure 5-24. Modes 1-2 for recipe cases (4 cases) and all extended cases (48 cases) in S_A (T = 2.0, FN).



(a) Recipe and extended cases (Plotted on Fig. 5-5 (b)) **(b)** Recipe and extended cases: Kernel density estimation **(c)** 600 cases: Kernel density estimation (Fig. 5-4 (a))

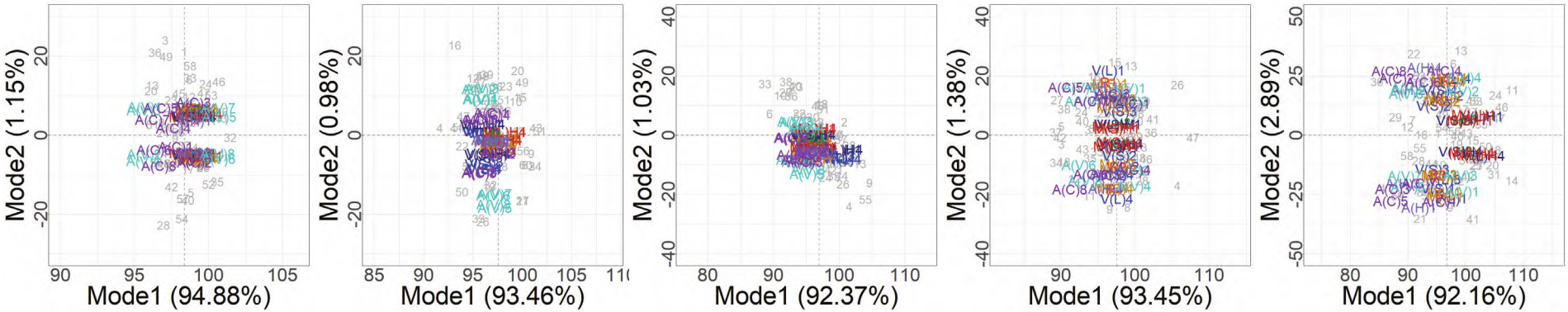
Figure 5-25. Modes 2-3 for recipe cases (4 cases) and all extended cases (48 cases) in S_A (T = 2.0, FN).

b) Absolute Acceleration Response Spectrum of Horizontal Two Components (FN, FP) and Period 0.1–2.0 s

This study attempted to diversify the distribution of S_A (T = 2.0, FN). The same examination was performed for the S_A distributions of other periods (0.1, 0.5, 0.75, and 1.0 s) and component (FP). Fig. 5-26–Fig. 5-29 show the two-dimensional distribution of PCSs for the distributions of S_A (T=0.1 – 2.0, FN and FP) for the recipe and extended cases (52 cases). The recipe and extended cases reproduce some of the diversity of the 600 cases. However, some regions are not fully represented in the extended cases set up in this study, such as Modes 2-3 with FN component in period 0.1 s (Fig. 5-27 (a)) and Modess 2-3 with FP component in period 1.0 s (Fig. 5-27 (a)).

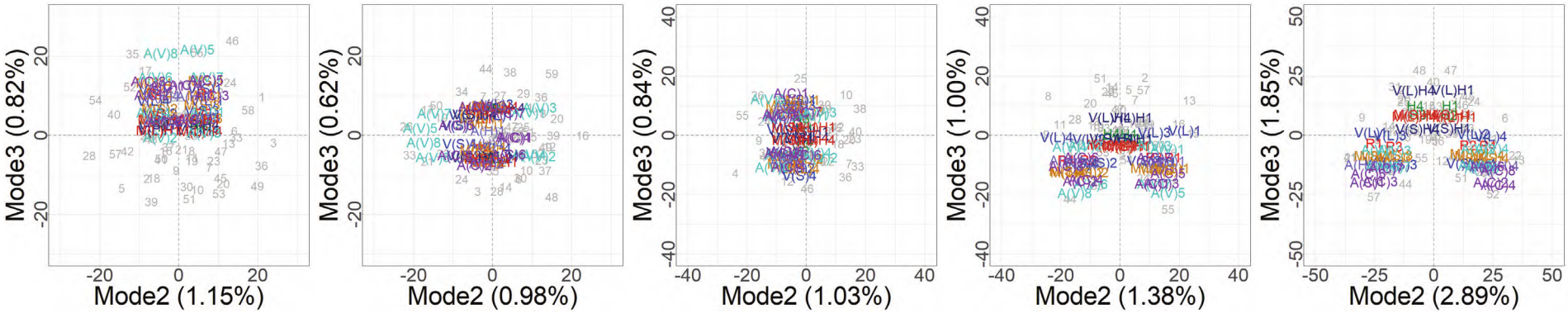
The above results indicate that, by appropriately setting the extended cases, it is possible to represent the diversity of ground motion distribution, regardless of the period or component. However, the modes tend to differ for each period and component, and the dominant source parameters are also different. Therefore, it is necessary to consider the differences for each period and component when assuming extension conditions.

In this study, there were only two types of extended cases in which multiple source parameters were varied simultaneously: the hypocenter and the rupture velocity or seismic moment. Therefore, it is necessary to try various extension conditions, such as extended cases in which the asperity arrangement and the rupture velocity (or seismic moment) are varied simultaneously. This study has 48 extended cases, but considering that the extended cases will be published in addition to the recipe cases, it is desirable to have a relatively small number of extended cases. Therefore, it is necessary to diversify the setting conditions of source parameters while simultaneously eliminating those with little diversifying effect.



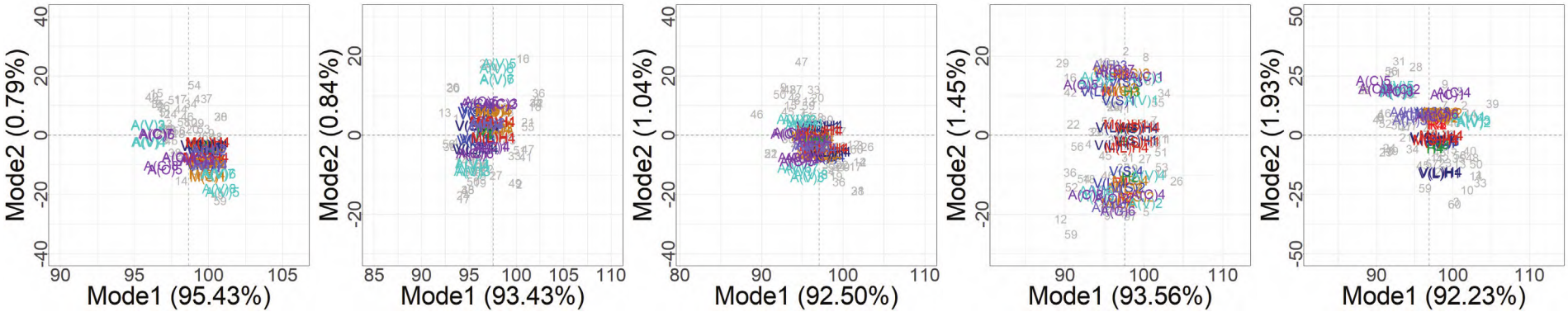
(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s (Fig. 5-24 (a))

Figure 5-26. Modes 1-2 for recipe cases (4 cases) and all extended cases (48 cases) in S_A of FN component.



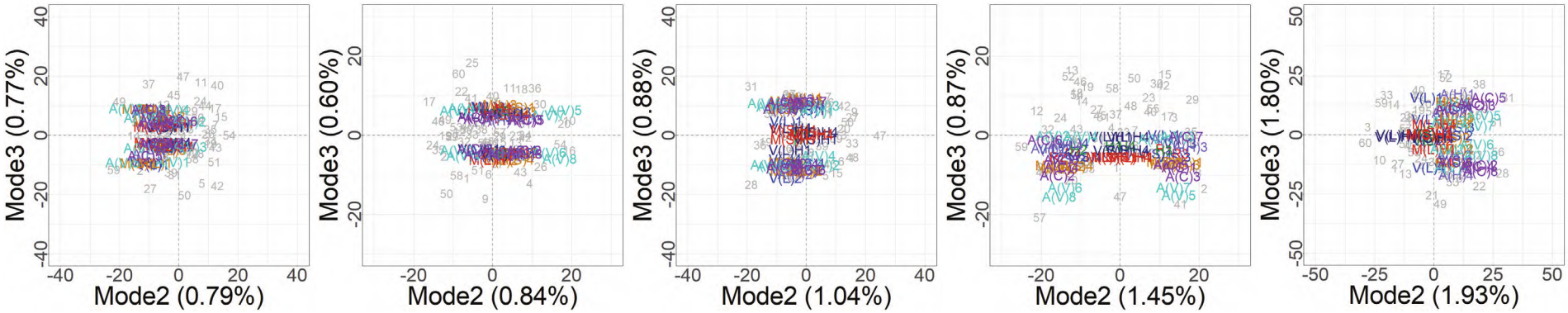
(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s (Fig. 5-25 (a))

Figure 5-27. Modes 2-3 for recipe cases (4 cases) and all extended cases (48 cases) in S_A of FN component.



(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

Figure 5-28. Modes 1-2 for recipe cases (4 cases) and all extended cases (48 cases) in S_A of FP component.



(a) Period 0.1 s **(b)** Period 0.5 s **(c)** Period 0.75 s **(d)** Period 1.0 s **(e)** Period 2.0 s

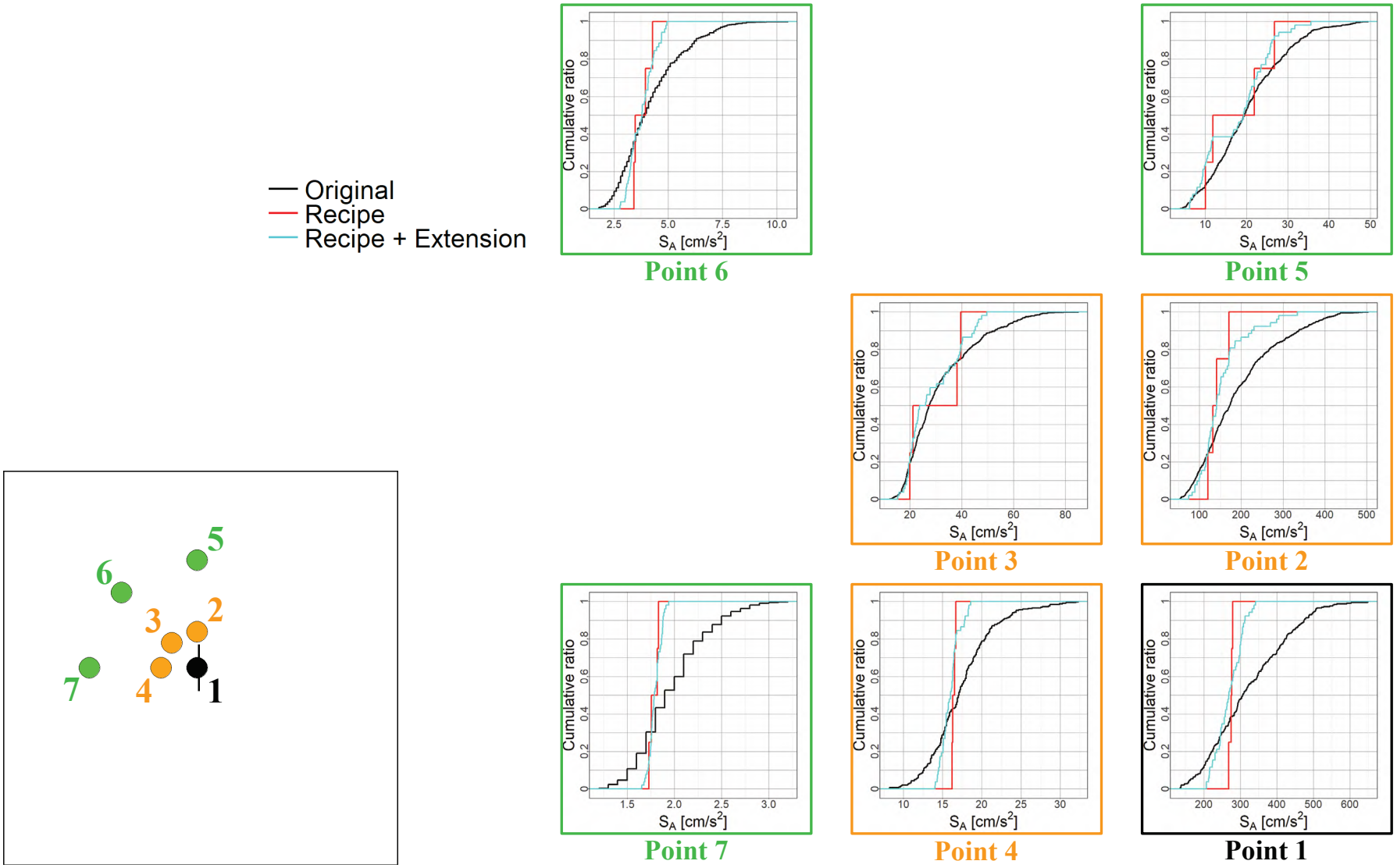
Figure 5-29. Modes 2-3 for recipe cases (4 cases) and all extended cases (48 cases) in S_A of FP component.

5.5 Evaluation of Diversity Focusing on Ground Motion Intensity

Up to this point, this study has evaluated the diversity of "spatial characteristics" of the ground motion distributions represented by the original 600 cases or the recipe cases and the extension cases. In this section, I focus on the "ground motion intensity" itself and evaluate how well the recipe and extension cases reproduce the original 600-case ground motion. In ***5.5.1*** and ***5.5.2***, the ground motion intensity is evaluated for selected points and the entire computational area.

5.5.1 Ground Motion Intensity at Selected Point

Seven points were selected from the computational area, as shown in Fig. 5-30 (a) (the same points as in ***2.6.2***). The recipe cases (red; 4 cases) have extremely narrower ranges of ground motions than the original 600 cases (black) except for Points 3 and 5, which are smaller than the mean of the 60 cases. The distribution shape of the recipe cases and the extension cases (blue; recipe: 4 cases, extension: 48 cases) is smoother and its range is wider than that of the recipe cases alone (red). Points 2, 3, and 5, which are located along and around the strike extension of the fault where the influence of the directivity effect is pronounced (strong ground motions are likely to occur), can roughly represent the range of the target 600 cases (black). The other points do not fully represent the 600-cases range; however, their S_A values are small, except for Point 1. For Point 1, which is directly above the fault, it is possible to represent the S_A of 600 cases by extending the "recipe" setting conditions, such as "set simultaneously large values for magnitude and rupture velocity" or "arrange large and small asperities vertically."



(a) Seven selected points

(b) Cumulative relative frequency graph of S_A

Figure 5-30. Cumulative relative frequency of S_A (T = 2.0 s, FN component) at selected points.

5.5.2 Ground Motion Intensity for Entire Area

Figure 5-31 shows the inverse cumulative distribution of the number of S_A-exposure points (the number of points exposed to the corresponding S_A) over the entire computational area. The gray lines represent each of the 600 cases. The mean and the standard deviation of the 600 cases for each point are μ and σ, respectively (Fig. 2-4 (e)); the black lines represent the S_A distribution of μ (solid line), $\mu \pm \sigma$ (dashed lines) assuming a perfect correlation of the variations. The recipe cases (red; 4 cases) are almost equal (slightly smaller) to the mean (solid black line). The extension cases (blue; 48 cases) are located around the recipe cases (red), and most of the cases fall between μ - σ (lower black dashed line) and μ (solid black line). The above results show that the recipe cases and extended cases cover the ground motion distributions below the mean value of the 600 cases. However, the distribution above the mean value is not fully represented. This result suggests that it is necessary to assume extended "recipe" setting conditions such that stronger ground motions appear.

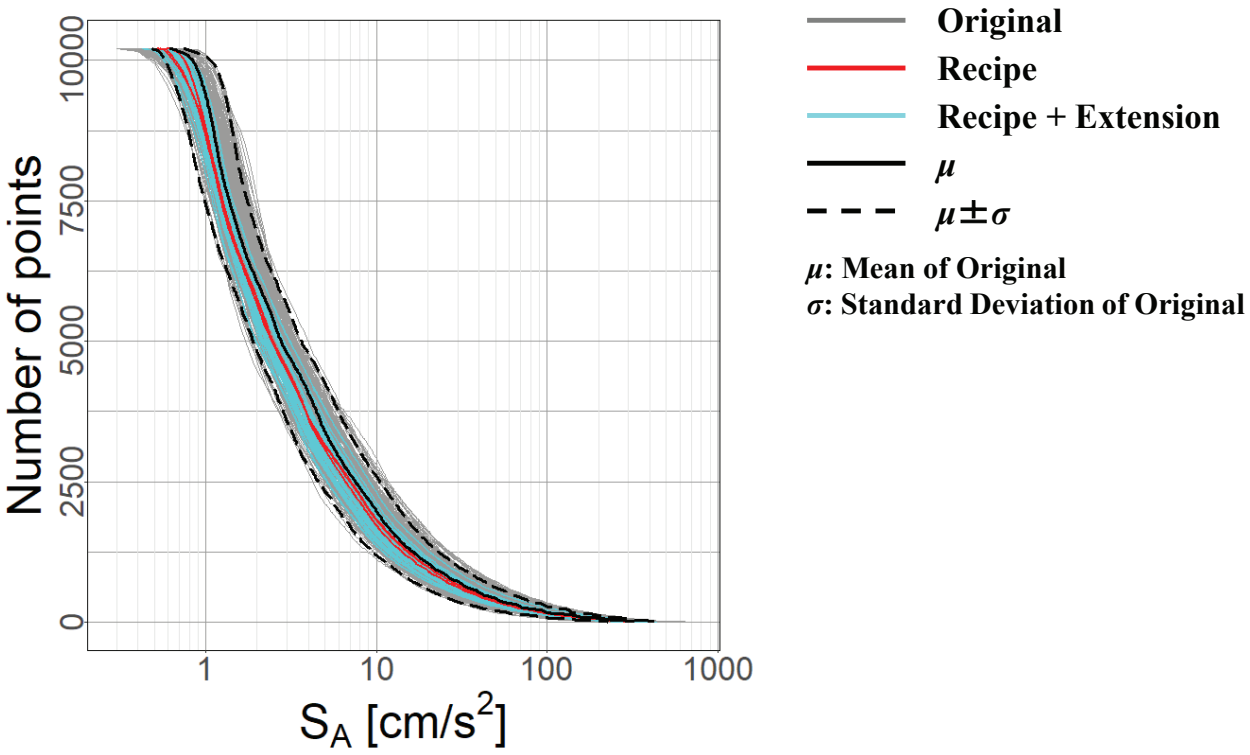


Figure 5-31. Inverse cumulative distribution of the number of S_A (T = 2.0 s, FN component) exposure points over the entire computational area.

5.6 Conclusion

The study in this chapter attempts to diversify the ground motion distribution of scenario earthquakes by extending the setting conditions of the source parameters in the "Recipe for Predicting Strong Ground Motions" [1, 3, 4]. This study targets a characterized source model for a strike-slip fault consisting of a single fault plane [5], and extends the setting conditions for a seismic moment, inner fault parameters and extract fault parameters. The major conclusions are listed below.

1) Mode decomposition was applied to the 600 ground motion distributions [5] to obtain the mode shapes and the principal component scores; non-hierarchical cluster analysis was applied to the PCSs. By arranging the PCSs for Modes 1–3 in two-dimensions, the diversity of the ground motion distributions could be visualized.

2) The fault models complied with the "recipe" setting conditions, and the relative relationship between the ground motion distributions for recipe cases and the 600 cases were clarified. However, the spatial characteristics for the recipe cases were average and limited, and did not cover the diversity of the 600 cases.

3) The cases with extended the "recipe" setting conditions were set up, and an attempt was made to diversify the ground motion distribution using the recipe and extended cases. The source parameters with extended setting conditions are "horizontal position of the hypocenter," "rupture velocity," "seismic moment," and "asperities arrangement." In addition to the extended cases in which these parameters were varied independently, extended cases in which two parameters were varied simultaneously ("hypocenter and rupture velocity" and "hypocenter and seismic moment") were also assumed.

4) Focusing on the "spatial characteristics" and "ground motion intensity" of the ground motion distributions, the diversity of the ground motion distributions represented by the recipe case (4 cases) and the extended case (48 cases) was evaluated. The spatial characteristics that did not appear in the current recipe cases (e.g., differences in ground motion levels and the effects of bilateral rupture propagation) could be expressed by adding the extended cases; thus, the recipe and extended cases allowed for diversification of the seismic motion distribution. This diversifying effect was greater in the case where multiple parameters were combined and extended than in the case where solely a single parameter was extended. At the computational points along and around the strike extension of the fault, the ground motion (S_A) of the recipe and extension case approximately reproduced the range of the 600 cases. The ground motion distributions over the entire computational area of the recipe cases and the extension cases are covered for the distributions below the mean of the 600 cases. However, they do not fully represent the larger-than-average cases.

Yamada et al. [12] proposed a setting method using the LHS (Latin Hypercube Sampling) as an efficient way to set up various characterized source models considering the "recipe" setting values and their variations. Imai et al. [13] proposed a method using copulas to capture the distribution shape of PCSs among modes as a probability distribution, sample PCSs according to this probability distribution, and simulate

ground motion distribution. The sampling using LHS is an efficient method to cover the diversity of source parameters without considering the diversity of spatial characteristics of the ground motion distribution. In contrast, the sampling simulation using copulas is an efficient way to cover various spatial characteristics without directly considering the source parameter settings. Meanwhile, the extension of the “recipe” in this study arbitrarily extends the “recipe” setting conditions, assumes a variety of fault models in a relatively small number of cases, and attempts to cover a variety of ground motion distributions. Future studies will set up and simulate the fault models and ground motion distributions by sampling and simulation using the LHS method [12] and the copula method [13]. Based on these results, the number of cases necessary to efficiently cover the diversity of seismic motion distribution will be judged.

References in Chapter 5

[1] The Earthquake Research Committee, the Headquarters for Earthquake Research Promotion. (2020) Strong ground motion prediction method for earthquakes with specified source faults ("Recipe"). https://www.jishin.go.jp/main/chousa/20_yosokuchizu/recipe.pdf **(in Japanese)**

[2] The National Research Institute for Earth Science and Disaster Resilience. "Japan seismic hazard information station (J-SHIS)." https://www.j-shis.bosai.go.jp/en/

[3] Irikura, K. and Miyake, H. (2001) Prediction of strong ground motions for scenario earthquakes. *J. Geogr*, **110** (6): pp. 849–875. https://doi.org/10.5026/jgeography.110.6_849 **(in Japanese with English abstract)**

[4] Irikura, K., Miyake, H., Iwata, T., Kamae, K., Kawabe, H. and Dalguer, D. L. (2004) Recipe for predicting ground motions from future large earthquakes. *Proc. of 13WCEE*: Paper No. 1371.

[5] Kagawa, T. (2015) Spatial variability of periodic characteristics among strong ground motions derived from multiple fault rupture scenarios. *Journal of JAEE*, **15** (7): pp. 90–99. https://doi.org/10.5610/jaee.15.7_90 **(in Japanese with English abstract)**

[6] Kagawa, T. (2004) Developing a Stochastic Green's function method having more accuracy in long period range to be used in the Hybrid Method. *Journal of JAEE*, **4** (2): pp. 21–32. https://doi.org/10.5610/jaee.4.2_21 **(in Japanese with English abstract)**

[7] Yoshida, K., Nojima, N. and Takahashi, Y. (2020) Classification of characteristics of multiple predicted seismic intensity distributions for a scenario earthquake. *Journal of JSCE A1*, **76** (4): pp. I_420–I_429. https://doi.org/10.2208/jscejseee.76.4_I_420 **(in Japanese with English abstract)**

[8] Kin, M. (2007) Data science with R -from basic data analysis to the least methods. Tokyo, *Morikita Publishing Co., Ltd.*, 320 p. **(in Japanese)**

[9] Ripley, B. D., Venables, W. N., Bates, D. M., Hornik, K., Gebhardt, A. and Firth, D. (2022) MASS: Support functions and datasets for Venables and Ripley's MASS. https://cran.r-project.org/web/packages/MASS/MASS.pdf

[10] Somerville, P. G., Irikura, K., Graves, R., Sawada, S., Wald, D., Abrahamson, N., Iwasaki, Y., Kagawa, T., Smith, N. and Kowada, A. (1999) Characterizing crustal earthquake slip models for the prediction of strong ground motion. Seismological Research Letters, 70 (1): pp. 59–80. https://doi.org/10.1785/gssrl.70.1.59

[11] Kagawa, T., Irikura, K. and Somerville, P. G. (2004) Differences in ground motion and fault rupture process between the surface and buried rupture earthquakes. EPS, 56: pp. 3-14. https://doi.org/10.1186/BF03352486

[12] Yamada, M., Senna, S. and Fujiwara, H. (2007) Statistical analysis of predicted ground motions on the basis of a recipe for strong-motion prediction. *Journal of JAEE*, 7 (1): pp. 43–60. https://doi.org/10.5610/jaee.7.43 **(in Japanese with English abstract)**

[13] Imai, R., Kasui, N., Iwaki, A. and Fujiwara, H. (2020) A sample generation of scenario earthquake shaking maps via modal decomposition and empirical copula. *Proc. of the 15th Annual Meeting of JAEE*: Paper No. S15-18.

6 Analysis of Spatial Characteristics of Tsunami Inundation Depth Caused by the Nankai Megathrust Earthquakes

6.1 Introduction

An earthquake with submarine slip has the potential to cause tsunamis and other disasters simultaneously. Therefore, it is important to analyze these multi-hazards and conduct risk assessments from the viewpoint of disaster prevention. The studies in **Chapters 2–5** have applied the "decomposition, synthesis, classification, regression, and prediction" method to the ground motion distributions; then, the studies have performed analyses related to seismic risk assessment, such as analysis of the spatial characteristics of ground motion distributions and simulation of ground motion distributions. The study in **Chapter 6** aims to clarify the usefulness of these methods, i.e., their applicability to multi-hazards, as a stepping stone to multi-hazard risk assessment. Specifically, the spatial characteristics of the tsunami inundation depth distributions are analyzed using the "decomposition" method.

The tsunami caused by the Nankai trough megathrust earthquake has the potential to invade wide areas based around the Pacific coast and cause severe damage. For example, many key energy-related facilities, such as thermal power plants, nuclear power plants, and liquefied natural gas (LNG) terminals, are located in coastal areas; thus, tsunami damage to these facilities could lead to long-term supply disruptions of electricity and city gas. Therefore, the key to rational tsunami countermeasures is proper tsunami damage estimation and risk assessment. The working group on megathrust earthquake models for the Nankai trough [1] established in the central disaster management council has assumed and published the distributions of tsunami inundation depth caused by the earthquakes expected to result in the largest tsunami damage. However, the occurrence frequency is very infrequent [2]. The tsunami source models for 11 cases were set up in this assumption, and the assumed area covers a wide region from Kyushu to Kanto.

As in seismic analysis, handling uncertainty in fault models is an important issue in tsunami analysis, and many studies have been carried out as follows.

- Imamura et al. [3] : Evaluation of the effects of fault length and strike differences on the spatial distribution of tsunami water levels and wave heights.
- Kurita et al. [4] : Quantitative evaluation of the variation of computed water levels in tsunami analysis using tsunami trace heights in the Tohoku region.
- Yasuda et al. [5] : Quantitative evaluation of the uncertainty of tsunami height using the amount of slip on the fault plane as a probability quantity.
- Kotani et al. [6] : Quantifying the maximum tsunami height uncertainty by simulating tsunami heights using response surfaces.
- Kotani et al. [7] : Analysis of site-to-site correlation of tsunami heights by principal component analysis.

The study in this chapter applies the decomposition method to the inundation depth distributions and the physical implications of the spatial characteristics of inundation depth reflected in the base forms obtained from the decomposition by relating it to the tsunami source fault model. Specifically, the spatial characteristics are analyzed for the inundation depth distributions of 11 cases published by The working group on megathrust earthquake models for the Nankai trough [1]. In addition to the singular value decomposition used in **Chapter 4** (target: ground motion distribution), a non-negative matrix factorization is used as the analysis method.

6.2 describes applying singular valued decomposition and non-negative matrix factorization to tsunami inundation depth distributions. ***6.3*** outlines the tsunami inundation depth distributions caused by the Nankai megathrust earthquakes. Next, ***6.4*** shows the results of applying singular value decomposition, and ***6.5*** and **6.6** show the results of non-negative matrix factorization. **6.7** compares the results of both methods. Finally, ***6.8*** summarizes the conclusions obtained in this chapter.

6.2 Singular Value Decomposition and Non-negative Matrix Factorization of Tsunami Inundation Depth Distributions

6.2.1 Singular Value Decomposition and Non-negative Matrix Factorization

Table. 6-1 summarizes the contents of singular value decomposition (SVD) and non-negative matrix factorization (NMF) [8, 9]. The greatest advantage of SVD is that the orthogonal bases are obtained as eigenvectors of mutually uncorrelated modes; this allows the spatial correlation of each mode to be represented as an independent component. The feature space (mode shape) of SVD represents the variation structure of the entire space. In contrast, NMF is characterized by the approximate representation of a non-negative matrix using only the addition of non-negative bases; this facilitates interpretation and semantics in the decomposition and synthesis of non-negative physical quantities. The NMF feature space (non-negative basis space) represents the variation structure of the local space.

Table 6-1. Singular value decomposition and non-negative matrix factorization.

Method	**Singular Value Decomposition (SVD)**	**Non-negative Matrix Factorization (NMF)**
Data Type	All data, regardless of whether the sign is positive or negative ($M \times N$ matrix)	Non-negative data only ($M \times N$ non-negative matrix)
Basis Number (Factorization Rank)	N ($<<M$)	k ($2 \leq k \leq \min\{M, N\} = N$) * k is an optional integer
Decomposed Basis	Orthonormal basis (Each mode is independent) Basis reflects the variation structure in the entire space	Non-negative basis (Symplectic basis) Basis reflects the variation structure in the local space
Synthesis of Bases	Original matrix is represented by addition and subtraction of orthogonal modes - Synthesis of all N modes : perfect reproduction - Synthesis of low-order modes : low-dimensional approximation	Original non-negative matrix approximated by only addition of non-negative basis (k-dimensional low-dimensional approximation) * The approximation accuracy generally improves as the factorization rank k increases.
Order of Basis Number	Eigenvalues in ascending order - Mode 1 has maximum weight (the greatest importance) - Higher-order modes have smaller weights (lower importance)	Factorization rank is optional (No difference in relative importance for each basis)
Advantage	Orthogonal modes allow each mode to be evaluated independently	Non-negative bases addition allows easy interpretation of decomposition and synthesis results
Disadvantage	Negative values in the bases can make interpretation of decomposition and synthesis results difficult depending on the physical quantity	- Target data is only non-negative values - Need to set appropriate factorization rank k

6.2.2 Matrix Representation of Tsunami Inundation Depth Matrix

Let x (≥ 0) denote a variable representing the inundation depth, and the inundation depth distribution (M: total number of computation sites) is represented by a column vector $\mathbf{x} = (x_1, \ldots, x_M)^{\mathrm{T}}$. The inundation depth distributions for N cases with different parameter settings are described by the $M \times N$ ($M >> N$) matrix $\mathbf{X}$ as follows:

$$\mathbf{X} = \left(\mathbf{x}_1, \ldots, \mathbf{x}_N\right) = \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{M1} & \cdots & x_{MN} \end{pmatrix} \tag{6-1}$$

where x_{ij} is the inundation depth at site i in Case j.

6.2.3 Singular Value Decomposition of Inundation Depth Distributions

The application of SVD to the inundation depth distributions is the same as the ground motion distributions in ***3.3***; thus, only a brief flow is described here. The matrix $\mathbf{X}$ of the inundation depth distributions is normalized by using the mean μ_G and standard deviation σ_G of all entries of $\mathbf{X}$, which is defined as the matrix $\mathbf{X_0}$. Applying SVD to $\mathbf{X_0}$ yields three matrices $\mathbf{U_0}$, $\mathbf{D_0}$, and $\mathbf{V_0}$.

$$\mathbf{X_0} = \mathbf{U_0 D_0 V_0}^{\mathrm{T}} = \begin{pmatrix} \mathbf{u}_{01} & \cdots & \mathbf{u}_{0N} \end{pmatrix} \begin{pmatrix} d_{01} & & 0 \\ & \ddots & \\ 0 & & d_{0N} \end{pmatrix} \begin{pmatrix} \mathbf{v}_{01} & \cdots & \mathbf{v}_{0N} \end{pmatrix}^{\mathrm{T}} \tag{6-2}$$

The decomposition results of SVD, $\mathbf{U_0}$, $\mathbf{D_0}$, and $\mathbf{V_0}$, are uniquely determined. The computation of $\mathbf{U_0 D_0 V_0}^{\mathrm{T}}$ in Eq. (6-2) cancels out the sign arbitrariness in the singular vectors $\mathbf{u_0}$ and $\mathbf{v_0}$. Please refer to ***3.3*** for details of SVD, including the implications of $\mathbf{U_0}$, $\mathbf{D_0}$, and $\mathbf{V_0}$. It should be noted that in ***3.3***, it is denoted as $\mathbf{X_0} = \mathbf{UDV}^{\mathrm{T}}$.

In applying SVD to the inundation depth distribution, the lower limit of the inundation depth x [m] is set to 0.1 m, and the entry of the matrix $\mathbf{X}$ is $\log_{10} \max\{x, 0.1\}$.

6.2.4 Non-negative Matrix Factorization of Inundation Depth Distributions

The non-negative matrix $\mathbf{X}$ is approximately decomposed into the non-negative matrices $\mathbf{W}$ and $\mathbf{H}$ by applying NMF [8, 9].

$$\mathbf{X} \simeq \mathbf{WH} = \begin{pmatrix} w_{11} & \cdots & w_{1k} \\ \vdots & \ddots & \vdots \\ w_{M1} & \cdots & w_{Mk} \end{pmatrix} \begin{pmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{k1} & \cdots & h_{kN} \end{pmatrix} \tag{6-3}$$

where k is the number of adopted bases satisfying $k \leq \min\{M, N\} = N$, and the number of basis l goes from 1 to k. $\mathbf{W}$ and $\mathbf{H}$ in Eq. (6-3) have the following meanings.

- Matrix $\mathbf{W}$: $\mathbf{W}$ is the $M \times k$ basis matrix with no negative entries. The basis vector $\mathbf{w}_l = (w_{1l}, \ldots,$

$w_{il}, \ldots, w_{Ml})^{\mathrm{T}}$, the l-th column of **W**, represents the basic form that defines the spatial characteristics in Basis l. The map of the basis vector visualizes the spatial characteristics at grid cells.

- Matrix **H**: **H** is the $k \times N$ coefficient matrix with no negative entries. The coefficient vector $\mathbf{h}_j = (h_{1j}, \ldots, h_{lj}, \ldots, h_{kj})^{\mathrm{T}}$, the j-th column of **H**, represents the weights of each basis in Case j.

The non-negative linear combination **WH** represents the low-rank approximation of the matrix **X**. The non-negative constraint of NMF tends to make **H** sparse. NMF bases (**W** and **H**) are easy to interpret because of the above "addition of non-negative bases" and "sparsity of **H**."

In NMF, the approximation accuracy of the matrix **X** generally improves as the factorization rank k increases. However, it is necessary to exploratively decide an appropriate factorization rank k to extract and evaluate spatial characteristics of **X**.

The NMF algorithms are described below. First, $\mathbf{W_0}$ and $\mathbf{H_0}$, the initial matrices of W and H, are defined by non-negative values. Then, **W** and **H** in Eq. (6-3) that approximate **X** are obtained by repeatedly applying the NMF update equations. The update equations are defined as an optimization problem to minimize Kullback-Leibler (KL) divergence D as follows [10]:

$$h_{lj} \leftarrow h_{lj} \frac{\sum_{i=1}^{M} w_{il} x_{ij} \big/ (\mathbf{WH})_{ij}}{\sum_{i=1}^{N} w_{il}} \tag{6-4}$$

$$w_{il} \leftarrow w_{il} \frac{\sum_{j=1}^{N} h_{lj} x_{ij} \big/ (\mathbf{WH})_{ij}}{\sum_{j=1}^{N} h_{lj}} \tag{6-5}$$

$$D = KL(\mathbf{X} \,||\, \mathbf{WH}) = \sum_{i,j}^{M,N} \left[x_{ij} \log \frac{x_{ij}}{(\mathbf{WH})_{ij}} - x_{ij} + (\mathbf{WH})_{ij} \right] \tag{6-6}$$

The approximation accuracy of NMF also depends on the initial values of **W** and **H**, i.e., the initial matrix $\mathbf{W_0}$ and $\mathbf{H_0}$. Therefore, two NMF initialization methods are used in this study: the standard method using uniform random numbers and the method using non-negative double singular value decomposition (NNDSVD) [11].

6.2.5 NMF: The Initialization Method Using Uniform Random Numbers

In the initialization method using uniform random numbers, each entry of $\mathbf{W_0}$ and $\mathbf{H_0}$ is given a uniform random number which is continuously distributed in the interval [0, max{**X**}] (max{**X**} is the maximum value of all entries of **X**). Since $\mathbf{W_0}$ and $\mathbf{H_0}$ are given different values for each trial, their convergence result, **W** and **H**, are also not uniquely determined. Therefore, selecting the optimal **W** and **H** from multiple trials is generally adopted in uniform random initialization. However, the selection result is also one of the decomposition results, and it is necessary to note that the characteristics of the decomposition result possibly differ depending on a set of uniform random numbers.

6.2.6 NMF: The Initialization Method Using Non-negative Double Singular Value Decomposition

The initialization method using NNDSVD [11] is described below. First, SVD is applied to the matrix **X**. It should be noted that the SVD technique (***6.2.2***, Eq. (6-2))applies SVD to the standardized matrix $\mathbf{X_0}$, while NNDSVD applies SVD to **X**.

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}} = \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_N \end{pmatrix} \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_N \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_N \end{pmatrix}^{\mathrm{T}} \tag{6-7}$$

Next, the singular value vectors $\mathbf{u}_l$ and $\mathbf{v}_l$, the l-th column vectors in the matrices **U** and **V**, are separated into positive entries $\mathbf{u}_{+l}$ and $\mathbf{v}_{+l}$, and negative entries $\mathbf{u}_{-l}$ and $\mathbf{v}_{-l}$.

$$\mathbf{u}_{+l} = \max\{\mathbf{u}_l, \mathbf{0}\}, \quad \mathbf{u}_{-l} = \min\{\mathbf{u}_l, \mathbf{0}\} \tag{6-8}$$

$$\mathbf{v}_{+l} = \max\{\mathbf{v}_l, \mathbf{0}\}, \quad \mathbf{v}_{-l} = \min\{\mathbf{v}_l, \mathbf{0}\} \tag{6-9}$$

Where $\mathbf{u}_{+l}$ and $\mathbf{u}_{-l}$ are M-dimensional, and $\mathbf{v}_{+l}$ and $\mathbf{v}_{-l}$ are N-dimensional. **0** is a zero vector (M-dimensional in Eq. (6-8) and N-dimensional in Eq. (6-9)). max{} and min{} shall be applied to each vector entry. The products of the norm in the singular vector by sign are denoted by σ_{+l} and σ_{-l}.

$$\sigma_{+l} = \left\|\mathbf{u}_{+l}\right\| \cdot \left\|\mathbf{v}_{+l}\right\| \tag{6-10}$$

$$\sigma_{-l} = \left\|\mathbf{u}_{-l}\right\| \cdot \left\|\mathbf{v}_{-l}\right\| \tag{6-11}$$

The matrices **W** and **H** consist of non-negative bases; thus, NNDSVD extracts non-negative bases (diagonal basis) $\mathbf{u'}_l$ and $\mathbf{v'}_l$ from the singular vectors (orthonormal bases) $\mathbf{u}_l$ and $\mathbf{v}_l$. $\mathbf{u'}_l$ and $\mathbf{v'}_l$ are defined as follows:

$$(\mathbf{u'}_l, \mathbf{v'}_l, \sigma'_l) = \begin{cases} (\mathbf{u}_{+l}, \mathbf{v}_{+l}, \sigma_{+l}) & \text{if } \sigma_{+l} \geq \sigma_{-l} \\ (-\mathbf{u}_{-l}, -\mathbf{v}_{-l}, \sigma_{-l}) & \text{otherwise} \end{cases} \tag{6-12}$$

The l-th column vector of the initial matrix $\mathbf{W_0}$ is denoted by $\mathbf{w}_{0l}$, and the l-th row vector of the initial matrix

$\mathbf{H_0}$ is denoted by $\mathbf{h_{0(}}_{l)}$; then, these are defined by using $\mathbf{u'}_l$ and $\mathbf{v'}_l$ as follows:

$$\mathbf{w_{0}}_l = \sqrt{d_l \sigma'_l} \cdot \frac{\mathbf{u'}_l}{\left\|\mathbf{u'}_l\right\|} \tag{6-13}$$

$$\mathbf{h_{0(}}_{l)} = \sqrt{d_l \sigma'_l} \cdot \frac{\mathbf{v'}_l}{\left\|\mathbf{v'}_l\right\|} \tag{6-14}$$

where the basis number of NNDSVD corresponds to the mode number of SVD.

The initial matrices $\mathbf{W_0}$ and $\mathbf{H_0}$ are set up through the above computational procedure. As mentioned above, the decomposition result of SVD is essentially uniquely determined (except for the arbitrariness of the sign); hence, unlike the initialization using uniform random numbers in ***6.2.5***, the NNDSVD initial matrices $\mathbf{W_0}$ and $\mathbf{H_0}$ and its convergence results $\mathbf{W}$ and $\mathbf{H}$ are also uniquely determined. In addition, the zero entries in $\mathbf{W_0}$ and $\mathbf{H_0}$ that are defined in Eqs. (6-8) and (6-9) are also retained as zero entries in the convergence results $\mathbf{W}$ and $\mathbf{H}$ after applying the update equations (Eqs. (6-4) and (6-5)). Therefore, NNDSVD has the potential to produce higher sparsity of decomposition results than uniform random numbers.

6.3 Tsunami Inundation Depth Distributions Cased by the Nankai Megathrust Earthquakes

This study uses data from the second report published by the working group report on megathrust earthquake models for the Nankai trough [2].

6.3.1 Tsunami Source Model

The working group report on megathrust earthquake models for the Nankai trough [1] examined 11 cases (N = 11) of tsunami source models that cause the largest class of tsunamis among the M_W = 9 class megathrust earthquakes (Fig. 6-1 and Table. 6-2) [2]. The five cases (Cases 1–5) are the base cases, and the other six cases (Cases 6–11) are the derived cases. The base cases have one large slip zone (LSZ) and one super large slip zone (SLSZ) in the tsunami source model, respectively; all five base cases cover the entire Nankai Trough. The hypocenters are set up at 20 km depth near the center of the LSZ. In Case 2, the hypocenter is set off the Shionomisaki in the Kii Peninsula based on the past earthquakes (the 1944 Tonankai earthquake and the 1946 Syowa Nankai earthquake). The derived cases include two cases (Case 6, 7) with splay faults in the slip zone and four cases (Case 8 – 11) with two LSZ and two SLSZ. The hypocenters are set at the same locations as in the base cases that are the target of derivation.

6.3.2 Inundation Depth Distribution

The inundation depth is evaluated on a 10 m mesh scale in the coastal and upland areas, and the value is computed in units of 0.01 m. The tsunami source model is a model in which the rupture propagates from the hypocenter. The tide level is set as the high tide level, and obstacles such as houses are expressed as roughness coefficients. For sites with levees (including sluice gates), two conditions are considered: “conditions under which levees cease to function due to ground motion” and “conditions under which the levees cease to function due to tsunami overtopping the levees.” This study uses the inundation depth distributions under the latter condition.

The target area is from Kyusyu to Kanto, and the number of meshes with which inundation depths are commonly evaluated in all 11 cases (N = 11) is about 6.64 million. This study analyzed M = 6644 meshes (NMF: M = 6641, excluding the three meshes evaluated as 0.00 m in all 11 cases), from which one mesh was selected for every 1000 meshes.

Figure. 6-2 shows the inundation depth distributions for 11 cases. The inundation depths are larger along the Pacific coast, mainly in the vicinity of LSZ, including SLSZ (hereafter collectively referred to as “LSZ”). Fig. 6-3 shows the mean and standard deviation of inundation depth distribution. The mean and standard deviation show large values along the Pacific coast and smaller values toward the inland seas.

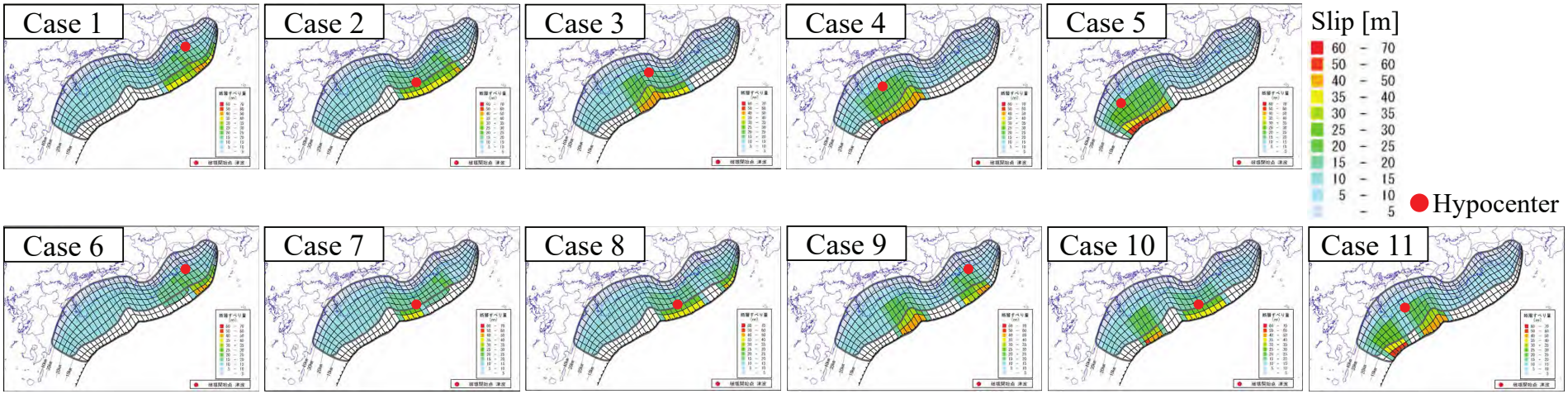


Figure 6-1. Tsunami source models for the Nankai megathrust earthquakes. (Note: Adapted from [2] and partially edited)

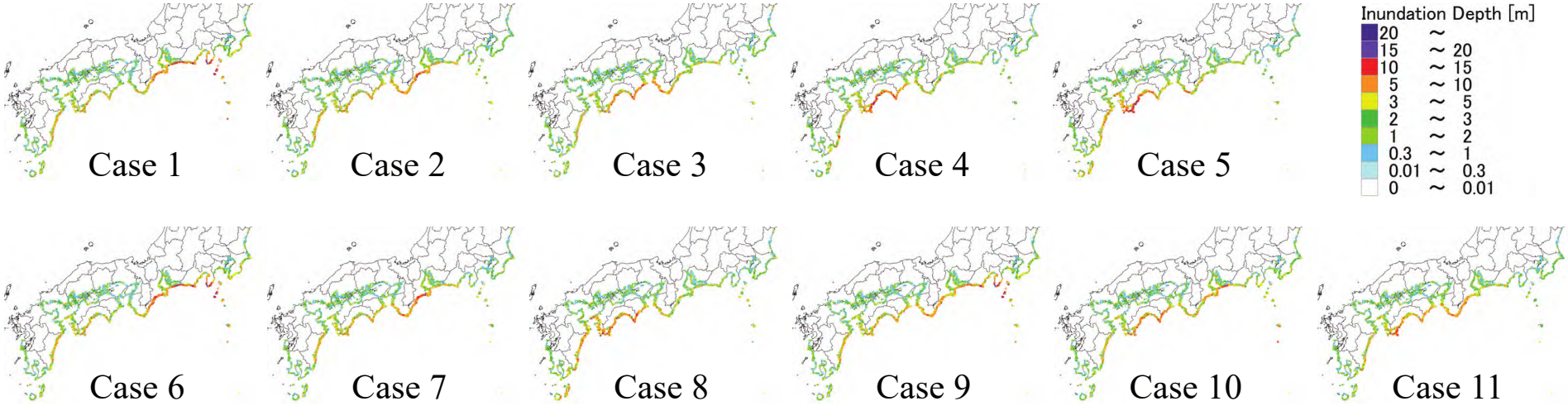
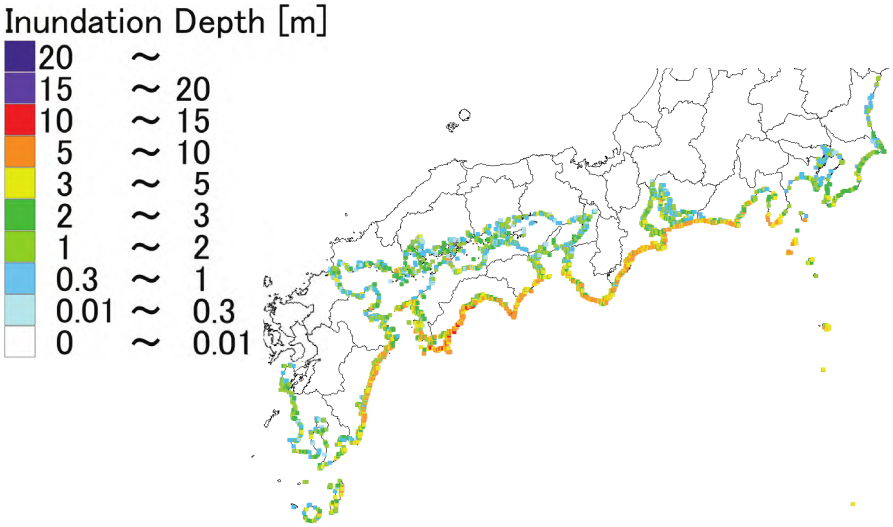


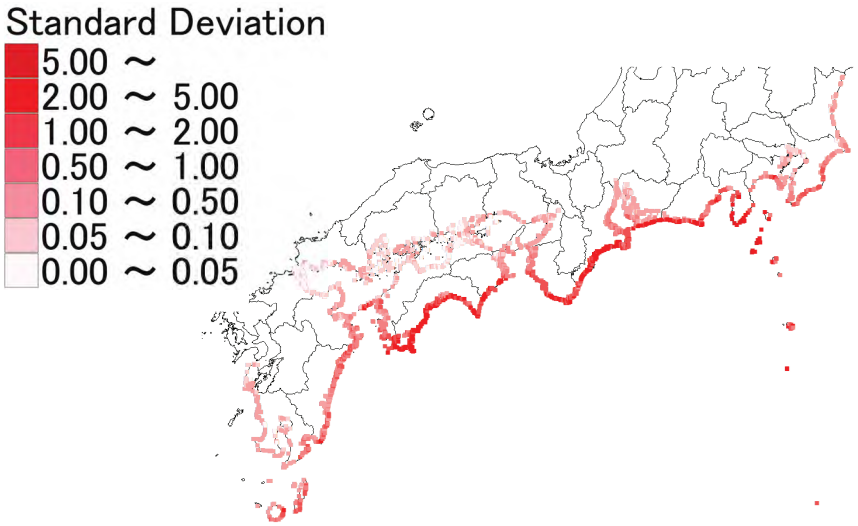
Figure 6-2. Tsunami inundation depth distributions caused by the Nankai megathrust earthquakes. (Source data: [1])

Table 6-2. Parameter of tsunami source model for the Nankai megathrust earthquakes. (Source data: [2])

Type	Case	Large slip zone (LSZ), Super Large slip zone (SLSZ)		M_0 [Nm]	M_W	Average Slip [m]	Basic case targeted for derivation (Location of hypocenter)
		Location	Arrangement pattern				
Basis	1	The Suruga Bay – Off the Kii Peninsula	LSZ + SLSZ	6.1×10^{22}	9.1	10.3	
	2	Off the Kii Peninsula	LSZ + SLSZ	6.3×10^{22}	9.1	10.7	
	3	Off the Kii Peninsula – Off Shikoku	LSZ + SLSZ	6.4×10^{22}	9.1	10.9	
	4	Off Shikoku	LSZ + SLSZ	6.4×10^{22}	9.1	10.8	
	5	Off Shikoku – Off Kyushu	LSZ + SLSZ	6.3×10^{22}	9.1	10.7	
Derivation	6	The Suruga Bay – Off the Kii Peninsula	LSZ + (SLSZ, Splay fault)	5.4×10^{22}	9.1	9.0	Case 1 (Case 1)
	7	Off the Kii Peninsula	LSZ + (SLSZ, Splay fault)	5.3×10^{22}	9.1	8.8	Case 2 (Case 2)
	8	The Suruga Bay – Off the east coast of Aichi Off the south coast of Mie – Off Tokushima	LSZ + SLSZ LSZ + SLSZ	6.2×10^{22}	9.1	10.4	Case 1, 2 (Case2)
	9	Off Aichi – Off Mie Off the Muroto cape	LSZ + SLSZ LSZ + SLSZ	6.7×10^{22}	9.1	11.3	Case 1, 3 (Case 2)
	10	Off the south coast of Mie – Off Tokushima Off the Ashizuri cape	LSZ + SLSZ LSZ + SLSZ	6.4×10^{22}	9.1	10.8	Case 2, 4 (Case 2)
	11	Off the Muroto cape Sea of Hyuga	LSZ + SLSZ LSZ + SLSZ	6.6×10^{22}	9.1	11.2	Case 4, 5 (Case 4)



(a) Mean



(b) Standard deviation

Figure 6-3. Distribution maps of mean and standard deviation of tsunami inundation depth for 11 cases.

6.4 Analysis of Spatial Characteristics of Inundation Depth Using SVD

This section presents the results of applying SVD to the 11 cases of inundation depth distributions.

6.4.1 Contribution Ratio of Mode

Figure. 6-4–Fig. 6-6 show the singular value, contribution ratio, and cumulative contribution ratio, respectively. Mode 1 is the most significant, with a contribution ratio of 84.62 %. Mode 2 and higher modes have smaller but constant contributions than Mode 1, with the cumulative contribution ratio exceeding 98% for Mode 6.

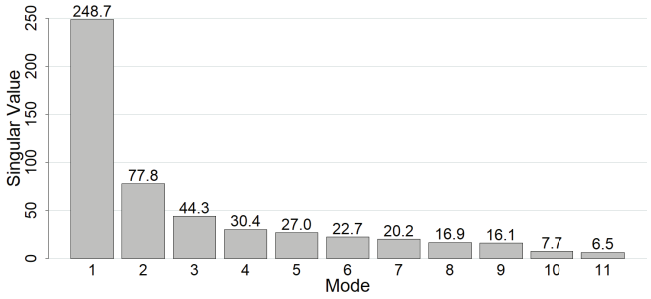


Figure 6-4. Singular value.

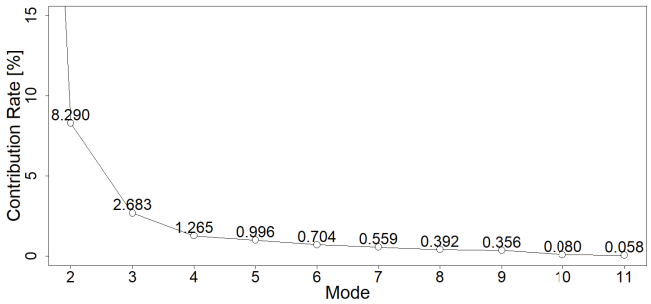


Figure 6-5. Contribution ratio.

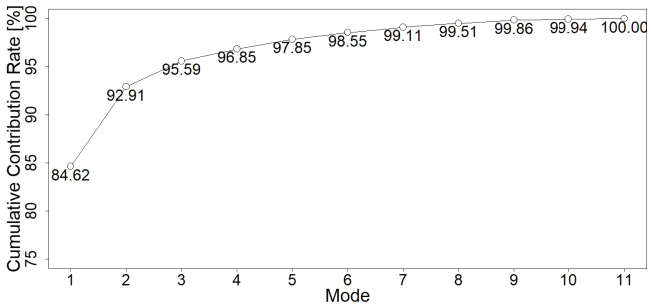


Figure 6-6. Cumulative contribution ratio.

6.4.2 Mode Shape and Spatial Characteristics

Figure. 6-7 shows the left singular vector **u** that represents the mode shape of inundation depth, and Fig. 6-8 shows the right singular vector **v** that represents the case-specific weight for each mode (only Modes 1–6 are shown). The relationship between the characteristics of each mode and the inundation depth distributions is discussed below, relating to Figs. 6-8, 6-9 to the tsunami source model (Fig. 6-1). The results are summarized in Table. 6-3 below.

a) Mode 1

The left singular vector has positive values along the Pacific coast and negative values toward the inland seas, similar to the mean inundation depth distribution (Fig. 6-3 (a)). The right singular vector has almost the same for all cases (mean: 0.3014, standard deviation: 0.0070). Thus, Mode 1 represents the average characteristics of the inundation depth distribution for the 11 cases.

b) Mode 2

The left singular vector has positive values in the east of the Kii Peninsula and negative values in the west. The right singular vector has positive values in Cases 1 and 2 (basis cases) and their derivatives cases where the LSZ is located east of the Kii Peninsula. Therefore, Mode 2 reflects the arrangement condition of LSZ in the east and west, with the Kii Peninsula as the boundary.

c) Modes 3–11

Modes 3 and 4 also reflect the arrangement condition of LSZ. Modes 5–11 also seem to have spatial characteristics that reflect the arrangement condition of LSZ; however, the higher-order modes have more complicated mode shapes, which may play a role as a tuning term for the details of each case.

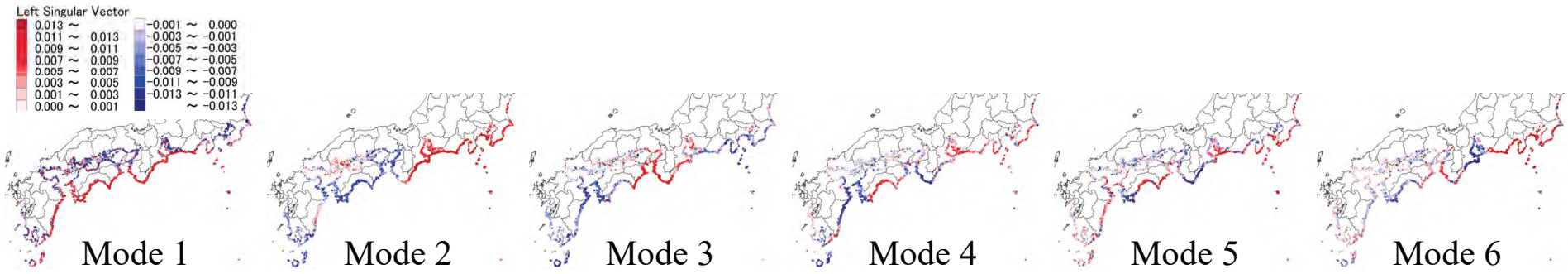


Figure 6-7. Distribution of left singular vector (mode shape).

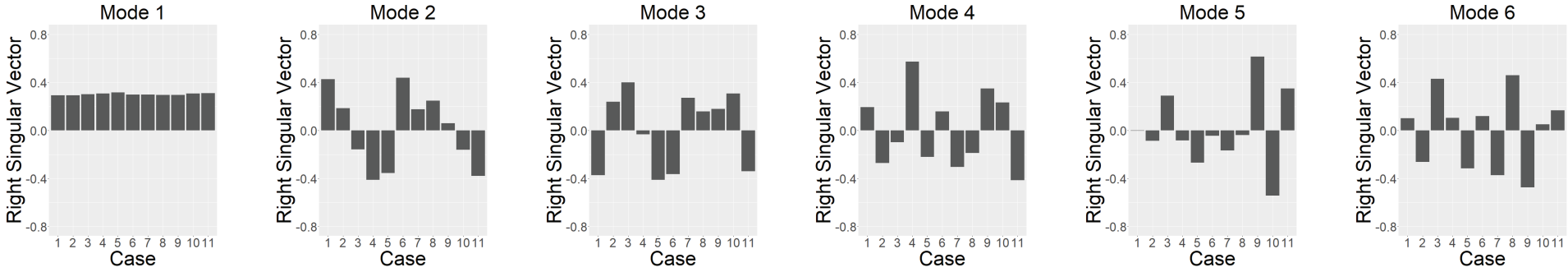


Figure 6-8. Right singular vector (weight coefficient).

6.5 Analysis of Spatial Characteristics of Inundation Depth Using NMF (Initialization Method: Uniform Random Numbers)

This section presents the results of applying NMF using uniform random numbers as the initialization method (hereafter referred to as "NMF-R").

6.5.1 NMF (Initialization Method: Uniform Random Numbers): Variation in Approximation Accuracy with Different Initial Values and Factorization Ranks

In NMF, the approximation accuracy of the matrix **WH** to the matrix **X** varies with the factorization rank. In NMF-R, the initial matrices $\mathbf{W_0}$ and $\mathbf{H_0}$ are different in every trial, so the convergence result **WH** is not uniquely determined; the approximation accuracy in NMF-R also varies in every trial. The factorization rank k is set to $2 \leq k \leq 11$ ($= N$), and convergence calculation was performed with 20 different uniform random numbers in each k. Fig. 6-9 shows the variation of the approximation accuracy in terms of KL divergence of **WH** to **X**; the red point and numbers represent the mean of the KL divergence in 20 trials. The KL divergence depends on each factorization rank. The effect of improvement in approximation accuracy with increasing factorization rank k (hereafter referred to as "accuracy improvement effect") becomes smaller between k = 6 and 7 (decrease in the KL divergence between k = 4–5–6–7–8: 775, 668, 293, and 221, respectively).

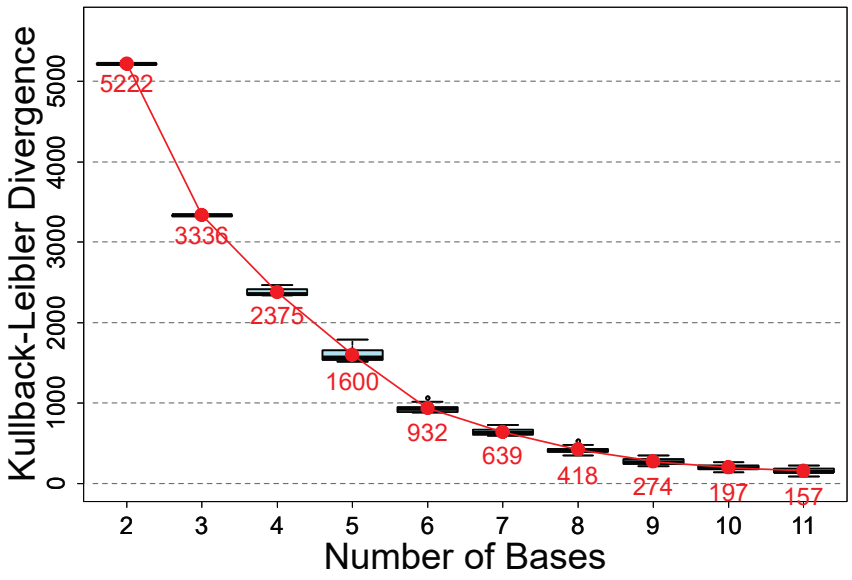


Figure 6-9. Boxplot of KL divergence with different initial values for each factorization rank in NMF-R. (Number of trials for each factorization rank: 20, red point and number: mean of 20 trials)

6.5.2 NMF (Initialization Method: Uniform Random Numbers): Basis Space and Spatial Characteristics

The results of NMF-R for the factorization rank k = 5–7 (k = 6 just before the accuracy improvement effect becomes small, and before and after k = 6) were compared (figures are omitted). Focusing on the base cases (Cases 1–5), k = 6 can be approximated with two bases, while k = 5 and 7 require three; therefore, the simpler k = 6 was adopted in this study.

Figure. 6-10 shows the distributions of the basis vector **w** in NMF-R, and Fig. 6-11 shows the coefficient vector **h**. The dominant factors in the spatial distribution of the basis vectors are discussed below, relating the coefficient vectors to the tsunami source models(Fig. 6-1). Finally, the results are summarized in Tab. 6-3, together with the SVD results.

a) Basis 1

The basis vector has large values along the pacific coast from Aichi prefecture to Shizuoka prefecture. The coefficients of Basis 1 have relatively large values in Case 1 (base case) and some of its derivatives (Cases 6 and 8) where the LSZ is located in the vicinity of such areas. Therefore, Basis 1 reflects the arrangement condition of LSZ in Aichi–Shizuoka.

b) Basis 2

The basis vector has large values along the pacific coast from Mie prefecture to Aichi prefecture. The coefficients of Basis 2 have relatively large values in Cases 1 and 2 (base cases) and some of their derivative (Cases 6, 7, and 9) where the LSZ is located in the vicinity of such areas. Therefore, Basis 1 reflects the arrangement condition of LSZ in Mie–Aichi.

c) Bases 3–6

Bases 3–6 also reflect the spatial characteristics related to the arrangement condition of LSZ.

From the coefficient vectors, it can be read that Cases 1–5 (base case) are approximated mainly by two bases, and Cases 6–11 (derived case) are approximated by the bases that constitute the derived base cases. However, the number of zero entries in the matrix **H** is not large, and the sparsity of **H,** a feature of NMF, is not fully demonstrated. In other words, the approximate model **WH** in NMF-R is complicated by many bases, and the semantics of the decomposition and synthesis is unclear.

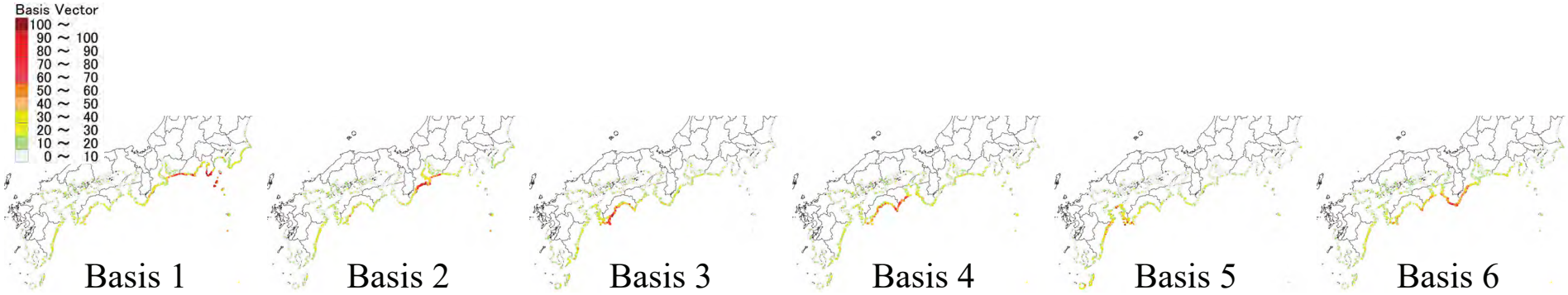


Figure 6-10. Distribution of basis vector in NMF-R ($k = 4$).

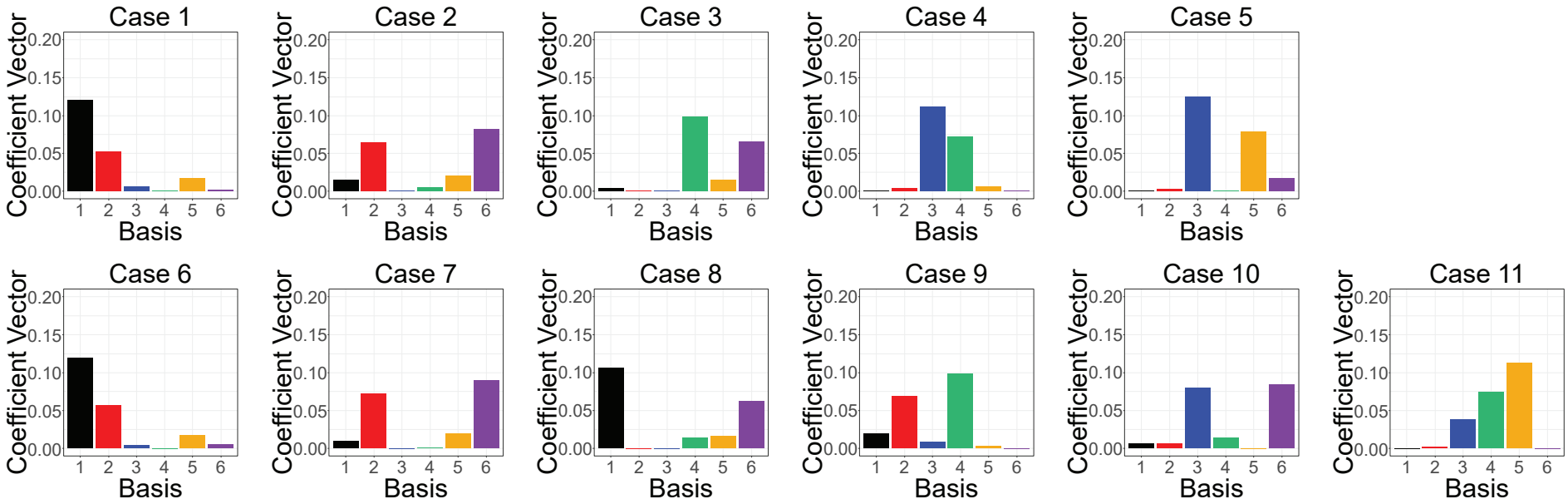


Figure 6-11. Coefficient vector in NMF-R ($k = 4$).

6.6 Analysis of Spatial Characteristics of Inundation Depth Using NMF (Initialization Method: NNDSVD)

This section presents the results of applying NMF using NNDSVD as the initialization method (hereafter referred to as "NMF-NNDSVD").

6.6.1 Orthonormal Basis of Inundation Depth Distribution based on SVD

Figure. 6-12 shows the distributions of the left singular vector obtained by applying SVD to the inundation depth distributions **X**, and Fig. 6-13 shows the right singular vectors (only Modes 1–6 are shown). In MNF-NNDSVD, the matrix $\mathbf{W_0}$ is constructed based on the left singular vectors (Fig. 6-12) according to Eq. (6-13), and the matrix $\mathbf{H_0}$ is constructed based on the right singular vectors (Fig. 6-13) according to Eq. (6-14).

In ***6.4***, SVD was applied to the log-transformed and standardized inundation depth distributions. However, NMF-NNDSVD does not perform these transformations because the non-negative value condition is no longer satisfied by these transformations. Therefore, the singular vector of NMF-NNDSVD (Fig. 6-12 and Fig. 6-13) differs from the singular vectors of SVD (Fig. 6-7 and Fig. 6-8), but the general trends are similar.

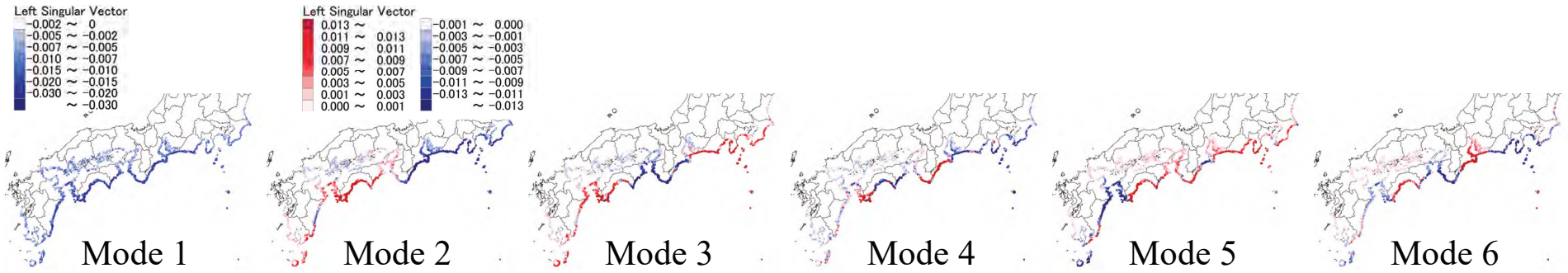


Figure 6-12. Distribution of left singular vector in SVD for NMD-NNDSVD.

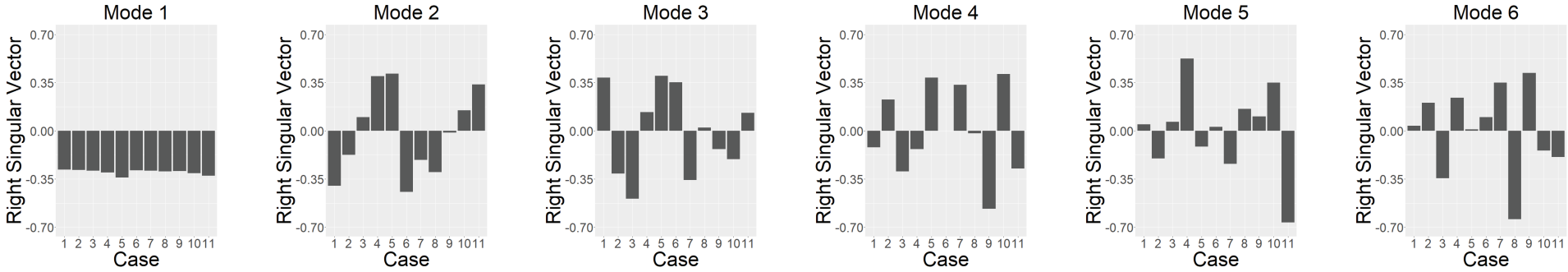


Figure 6-13. Right singular vector in SVD for NMD-NNDSVD.

6.6.2 NMF (Initialization Method: NNDSVD): Variation in Approximation Accuracy with Different Initial Values and Factorization Ranks

NMF-NNDSVD with the factorization rank k set to $2 \leq k \leq 11$ was applied to the inundation depth distributions, respectively; Fig. 6-14 shows the approximation accuracy of **WH** in NMF-NNDSVD. Compared with the approximation accuracy of NMF-R (black dashed line in Fig. 6-9 and Fig. 6-14), NMF-R has higher approximation accuracy, and this trend becomes more pronounced as the number of bases increases. This trend is discussed in ***6.7***, based on the characteristics of the bases.

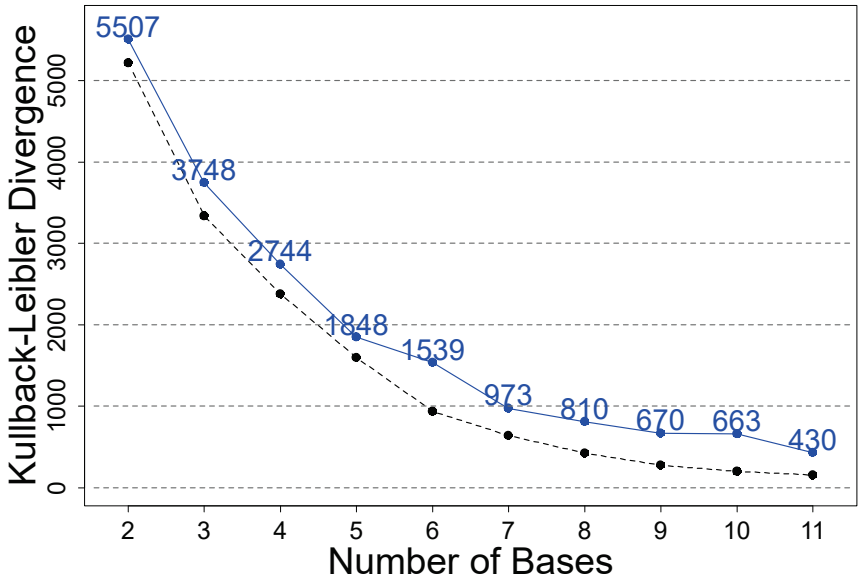


Figure 6-14. KL divergence for each factorization rank in NMF-NNDSVD. (Blue points, line, and number: NMF-NNDSVD, black points, dashed line, and number: NMF-R)

6.6.3 NMF (Initialization Method: NNDSVD): Basis Space and Spatial Characteristics

The factorization rank $k = 6$ was adopted in NMF-NNDSVD for comparison with NMF-R. Fig. 6-15 shows the distributions of the basis vector **w** in NMF-NNDSVD, and Fig. 6-16 shows the coefficient vector **h**. The dominant factors in the spatial characteristics of the inundation depth are analyzed from the basis and coefficient vectors. The results are summarized in Tab. 6-3, together with the SVD and NMF-R.

a) Basis 1

The basis vector has large values on almost the Pacific coast, especially in and around Shikoku (mainly in Kochi Prefecture), and the values become smaller toward the inland seas. This distribution trend is also observed in the mean inundation depth distribution in Fig. 6-2. The coefficient vectors indicate that Basis 1 contributes to all 11 cases. In particular, the coefficients for Cases 4 and 5 (bases cases) and their derivatives (Cases 10 and 11), where the LSZ is located in the vicinity of Kochi Prefecture, are larger than those for the other cases. Therefore, Basis 1 reflects the scale of inundation depth in the whole area (especially in western Shikoku).

Basis 1 contributes to all cases, whereas Bases 2–6 contribute to different cases depending on the basis. That difference suggests that Basis 1 covers the entire area, while Bases 2–6 represents each case's local space.

b) Basis 2

The basis vector has large values in the eastern Kii Peninsula and to the east of it. The coefficients of Basis 1 contribute to Cases 1 and 2 (base cases) and their cases (Cases 6–9) where the LSZ is located in the vicinity of such areas. Thus, Basis 2 reflects the arrangement condition of LSZ in the eastern Kii Peninsula and to the east.

c) Bases 3–6

Basis 3 – 6 also reflect the arrangement condition of LSZ.

Each case is approximated by Basis 1 and other bases. The coefficient value in Basis 1 expresses the level of inundation depth over the entire area; Bases 2–6 express the local spatial characteristics of the inundation depth for each case (the local areas with particularly large inundation depth values in the whole area). The coefficient vectors of the matrix **H** contain 2–4 zero entries each, which improves the sparsity. Thus, it can be read that a small number of bases approximates each case.

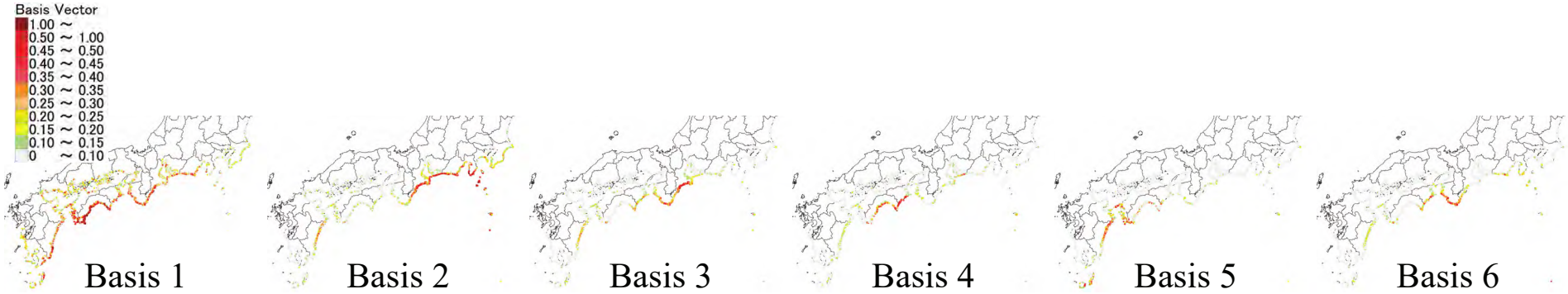


Figure 6-15. Distribution of basis vector in NMF-NNDSVD ($k = 4$).

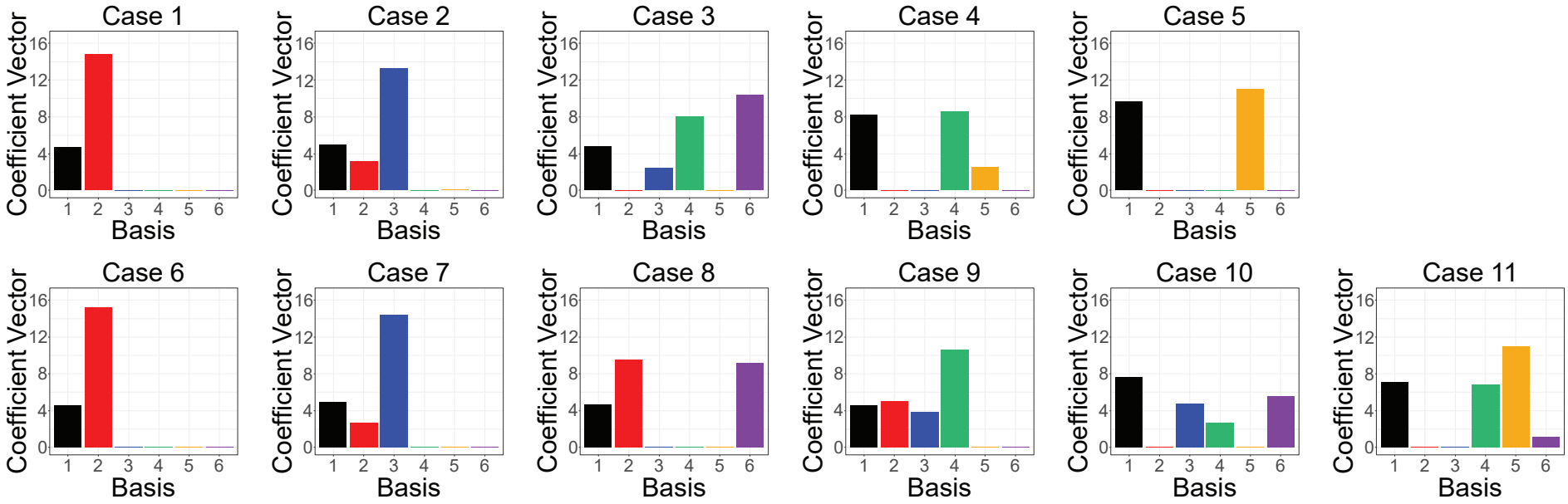


Figure 6-16. Coefficient vector in NMF-NNDSVD ($k = 4$).

6.7 Comparison of Decomposition Results in SVD and NMF

Table. 6-3 summarizes the analysis results of the influencing factors on the spatial characteristics of the inundation depth using SVD and NMF (NMF-R and NMF-NNDSVD). The order of modes in SVD and the order of bases in NMF-NNDSVD correspond to the scale of singular values, as shown in Eqs. (6-2) and (6-7). In contrast, the bases of NMF-R yield different results depending on the trials, and the order of the bases is arbitrary.

First, comparisons are made for decomposition bases. For SVD, the lower-order modes are highly important and closely related to the arrangement of LSZ in the tsunami source models. The mode shapes reflect the distinctive spatial characteristics of the entire inundation depth distributions. However, the weight of a mode decreases in importance with higher-order modes, and the mode shapes become more complex and difficult to interpret. For NMF, the weight of a basis represented by the coefficient vector characterizes each case; there is no relative difference in their importance. For NMF-R, both bases are related to the arrangement of LSZ, and their base spaces represent the local spatial characteristics. For NMF-NNDSVD, in contrast, Basis 1 represents a spatial distribution covering the entire area, contributing to all cases. Bases 2–6 reflect the local spatial characteristics associated with the arrangement of LSZ in the tsunami source model. Furthermore, NMF-NNDSVD has a high sparsity of the coefficient matrix **H** due to Basis 1 covering the entire area and Bases 2–6 characterizing each case, which facilitates the interpretation of the decomposition results. NMF-R does not have a basis that covers the entire area and contributes to all cases, such as Mode 1 for SVD and Basis 1 for NMF-NNDSVD.

Next, the stability of the decomposition results is compared. Since the SVD result is uniquely determined, the NMF-NNDSVD result using it as the initial values is also uniquely determined; thus, these techniques give stable results. In contrast, NMF-R results differ from trial to trial because uniform random numbers set the initial values.

Finally, the accuracy of the low-dimensional approximations by NMF are compared. As shown in Fig. 6-14, the residuals based on the KL divergence for NMF-R (black in the figure) were smaller than that for NMF-NNDSVD (blue in the figure). This trend is due to the complex contribution of the basis to each case in NMF-R, as shown in Fig. 6-11; this indicates that approximation accuracy is a trade-off for sparsity.

Table 6-3. Analysis results of factors influencing the inundation depth distribution using SVD and NMF

	Singular Value Decomposition
Mode 1	Average inundation depth distribution * This mode contributes to all cases
Mode 2	Arrangement condition of large slip zone in east and west, with the Kii peninsula as the boundary
Mode 3	Arrangement condition of large slip zone on the Kii peninsula and both sides
Mode 4	Arrangement condition of large slip zone in the four-part split target areas (Off Kyushu, off Shikoku, off the Kii peninsula, off Aichi – Shizuoka)
Mode 5 – Mode 11	Arrangement condition of large slip zone in the split target areas * Left singular vector: Split by mode order in the strike direction of the fault * Higher-order mode: Fine adjustment term for each case

	Non-negative Matrix Factorization (Factorization rank k = 6)	
	NMF-R (Initialization method: Uniform random)	**NMF-NNDSVD (Initialization method: Non-negative double singular value decomposition)**
Basis 1	Arrangement condition of large slip zone in Aichi – Shizuoka	The scale of inundation depth distribution over the entire target area * This basis contributes to all cases
Basis 2	Arrangement condition of large slip zone in Mie – Aichi	Arrangement condition of large slip zone in the eastern part of the Kii Peninsula and east of it
Basis 3	Arrangement condition of large slip zone in western Shikoku	Arrangement condition of large slip zone on the Kii peninsula
Basis 4	Arrangement condition of large slip zone in eastern Shikoku	Arrangement condition of large slip zone in eastern Shikoku
Basis 5	Arrangement condition of large slip zone in Kyushu – western Shikoku	Arrangement condition of large slip zone in Kyushu – western Shikoku
Basis 6	Arrangement condition of large slip zone in the Kii peninsula	Arrangement condition of large slip zone in the western part of the Kii peninsula

6.8 Conclusions

In this chapter, “data mining” and “machine learning” methods for analyzing ground motion distribution are applied to the tsunami inundation depth distributions; the usefulness of their methods for multi-hazards was clarified. Specifically, the spatial characteristics of the inundation depth distributions were analyzed using singular value decomposition and non-negative matrix factorization. As a numerical example, the tsunami inundation depth distributions for 11 cases caused by the Nankai trough megathrust earthquake were used. The major conclusions are listed below.

1) In SVD, the mode shapes represent the variation structure of the entire area. Mode 1 reflects the average characteristics of the inundation depth distribution. Mode 2 and higher modes reflect the arrangement condition of large slip zone, including super large one.

2) In NMF using uniform random numbers as the initialization method, the bases represent the local area’s variation structure. The bases reflect the arrangement condition of a large slip zone. Each base contributes to a case where the arrangement is closely related.

3) In NMF using NNDSVD as the initialization method, Basis 1 contributes to all cases, and its distribution of the basis vector represents the spatial distribution that covers the entire area. Bases 2–6 reflect the local spatial characteristics associated with the arrangement condition of a large slip zone.

This study used the tsunami inundation distributions for only 11 cases. For future developments, the analysis will be performed for many inundation depth distributions with more diversity in terms of setting conditions of slip angle and slip amount. The Tsunami Hazard Station (J-THIS) [12] established by the national research institute for earth science and disaster prevention has provided the tsunami inundation depth distributions for 2720 cases of the characterized fault models with different combinations of earthquake area and slip amount. These data will be used to analyze the spatial characteristics of inundation depth. Furthermore, we plan to construct a prediction model of tsunami inundation depth distribution based on “decomposition and synthesis” and “machine learning modeling.”

References in Chapter 6

[1] The central disaster management council, working group on megathrust earthquake models for the Nankai trough. https://www.bousai.go.jp/jishin/nankai/model/index.html **(in Japanese)**

[2] The central disaster management council (2012) Working group report on megathrust earthquake models for the Nankai trough (second report). https://www.bousai.go.jp/jishin/nankai/taisaku/pdf/20120829_2nd_report01.pdf **(in Japanese)**

[3] Imamura, F. and Shuto, N. (1989) Effect of fault parameter estimation error on tsunami height. *Proc. of Coastal Eng., JSCE*, **36**: pp. 178 – 182. https://doi.org/10.2208/proce1989.36.178 **(in Japanese)**

[4] Kurita, T., Matsuyama, M. and Uchino, D. (2013) Uncertainty of numerical tsunami simulation evaluated in comparison with the field survey results of the 2011 Tohoku earthquake tsunami. *Journal of JSCE B2*, 69 (2): pp. I_216 – I_220. https://doi.org/10.2208/kaigan.69.I_216 **(in Japanese with English abstract)**

[5] Yasuda, T., Maruyama, T., Goda, K., Mori, N. and Mase, H. (2015) Uncertainty analysis of Nankai trough earthquake tsunamis using stochastic source model. *Journal of JSCE B2*, **71** (2): pp. I_295 – I_300. https://doi.org/10.2208/kaigan.71.I_295 **(in Japanese with English abstract)**

[6] Kotani, T., Takase, S., Moriguchi, S., Terada, K, Fukutani, Y., Nojima, K. and Sakuraba, M. (2016) Numerical-analysis-aided probabilistic tsunami hazard evaluation using response surface. *Journal of JSCE A2*, **72** (1): pp. 58 – 69. **(in Japanese with English abstract)**

[7] Kotani, T., Moriguchi, S., Terada, K., Takase, S., Otake, Y., Fukutani, Y. and Sakuraba, M. (2018) A tsunami risk analysis considering correlations among coastal cities. *Proc. of 7th Asia Conference on Earthquake Engineering*, Paper No. ACEE0095.

[8] Lee, D. and Seung, H. (1999) Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**: pp. 788 – 791.

[9] Kameoka, H. (2012) Non-negative matrix factorization. *Journal of SICE*, **51** (9): pp. 835 – 844. https://doi.org/10.11499/sicejl.51.835 **(in Japanese)**

[10] Brunet, J. P., Tamayo, P., Golub, T. R. and Mesirov, J. P. (2004) Metagenes and molecular pattern discovery using matrix factorization. *PANS*, **101** (12): pp. 4164 – 4169. https://doi.org/10.1073/pnas.0308531101

[11] Boutsidis, C. and Gallopoulos, E. (2008) SVD based initialization: A head start for non-negative matrix factorization. *Pattern Recognition*, **41**: pp. 1350 – 1362. https://doi.org/10.1016/j.patcog.2007.09.010

[12] The National Research Institute for Earth Science and Disaster Resilience. “Japan tsunami hazard information station (J-THIS).” https://www.j-this.bosai.go.jp/ **(in Japanese)**

7 Conclusions

7.1 Results and Future Developments of each Chapter

This study examined “Diversification of Ground Motion Distributions,” which aims to cover various earthquakes, including the future earthquake in the Scenario Earthquake Shaking Maps (SESMs) [1], and “Simulation of Ground Motion Distribution,” which is a part of advanced use of SESMs (or already computed ground motion distribution data). In particular, the former aims to achieve the above objective by diversifying the set values of source parameters in the “strong ground motion prediction method for earthquakes with specified source faults (“recipe”)” [2], which is the standard setting criterion for scenario earthquakes. The following is a list of results and future developments in **Chapters 2–6**.

7.1.1 Chapter 2: Simulation of Strong Ground Motion Distributions of Two Horizontal Components Base on Singular Value Decomposition Analysis

- Objective: Simulation of ground motion distributions

a) Results of Chapter 2

The study in **Chapter 2** extends the simulation method for single-component ground motion distributions proposed in the previous study [3] to two horizontal components ground motion distributions. The simulation methods in the previous study [3] and **Chapter 2** can generate many ground motion distributions that preserve the spatial variation structure of the original ground motion distribution.

By applying singular value decomposition analysis to the ground motion distributions of two horizontal components by period, the modal forms that reflect the mutual spatial variability structure of two components and the principal component scores that represent the case-dependent modal weights were extracted for each mode. Next, the matrices of principal component scores for the two components are replaced by matrices that reproduce the cross-covariance structure between the two components; the ground motion distributions of the two components are generated. Four simulation methods were used: "Method A: treating the two components independently as uncorrelated," "Method B: considering correlation by correlation simulation," "Method C: making the two components perfectly correlated," and "Method D: randomly giving the difference between the two components." The results showed that Method B adequately reproduced the ground motion distribution of two components and their interactive spatial variation structure.

b) Future Developments of Chapter 2

The simulation method proposed in **Chapter 2** can simultaneously simulate the spatial distribution of two physical indicators. For future developments, this method should be applied to the simultaneous simulation of ground motion distributions in different periodic bands caused by the same earthquake is the subject of future work.

7.1.2 Chapter 3: Development of Simplified Prediction Model for Ground Motion Distribution

- Objective: Simulation of ground motion distributions

a) Results of Chapter 3

The study in **Chapter 3** proposed a simplified prediction model of ground motion distribution based on mode decomposition and machine learning as a surrogate model for a detailed based on strong ground motion simulation. As a numerical example, a prediction model was constructed for the distribution of absolute acceleration response spectrum (S_A, periods 0.1–2.0 s, FN- and FP-component) [4, 5] for a strike-slip fault calculated by the stochastic Green's function method [6].

The mode decomposition was applied to the distributions of S_A by period and component, then modal form, singular value and weight coefficients were obtained. Next, a weight coefficient predictor was constructed using the source parameters as explanatory variables. The modeling used "multiple linear regression," "support vector regression," "random forest," and "gradient boosting decision trees (GBDT)," respectively. Finally, the ground motion distribution for any fault rupture scenario was predicted by mode synthesis of the predicted weight coefficients, modal forms, and singular values.

The predicted distribution shows spatial characteristics of ground motion corresponding to the fault rupture scenario, which roughly reproduces the ground motion distribution calculated by the detailed method. The differences in prediction accuracy among the four models were not seen in period 0.1 s, where the effect of distance attenuation was dominant. However, they were seen at long periods, where the directivity effect and radiation pattern were more significant than at short periods. The models with the highest prediction accuracy among the four models were GBDT.

b) Future Developments of Chapter 3

The prediction model of ground motion distribution developed in **Chapter 3** is based on a single event caused by a single strike-slip fault. Therefore, the future study will construct a prediction model for earthquakes on various fault types, including reverse fault. Furthermore, applying this prediction model to seismic risk assessment will be the subject of future developments.

7.1.3 Chapter 4: Analysis of Period and Component Dependent Spatial Characteristics of Ground Motion Distributions Using Mode Decomposition and Machine Learning

- Objective: Diversification of ground motion distributions

a) Results of Chapter 4

The study in **Chapter 4** analyzed the period and component dependent spatial characteristics of ground motion distributions. This chapter revealed the combined effects of multiple source parameter settings on the ground motion distributions. Specifically, "explainable AI," some machine learning model's interpretation methods, was applied to the weight coefficient predictor of each mode (**Chapter 3**); the relationship between the weight coefficients and the source parameters was comprehensively analyzed.

The spatial characteristics that contributed significantly to the ground motion distributions were "the attenuation characteristics," "the spatial structure with a uniform variation of ground motion intensity and positive correlation over the entire area," and "the spatial structure with a negative correlation which represents the difference of ground motion intensity between the north and south of the map." The attenuation characteristics were related to the vertical arrangement of hypocenter and asperities. The spatial structure represents a positive correlation in the whole area was dominated by the effect of the seismic moment in the short periods and by the effect of rupture velocity in the long periods. The spatial structure representing two divided areas was dominated by the effect of the horizontal center of gravity of asperities in the short periods and by the effect of the horizontal position of the hypocenter in the long periods. The bilateral and unilateral directivity effects and the radiation patterns strongly affect the spatial characteristics for long periods.

b) Future Developments of Chapter 4

The results of **Chapter 4** relate to the spatial characteristics represented by the inner fault parameters and the extra fault parameters (e.g., rupture velocity and location of hypocenter) under relatively simple conditions. However, actual earthquakes rarely occur according to the assumed scenario and may be caused by the rupture of a part of the assumed source fault or in conjunction with the surrounding active faults. Therefore, in addition to the inner and extra fault parameters, it is necessary to consider the influence of the outer fault parameters (e.g., geometry and location of the fault) on spatial characteristics in the future. Specifically, we plan to analyze the spatial characteristic of the ground motions for multi-segment earthquakes caused by the coupling of multiple faults for future development.

7.1.4 Chapter 5: Diversification of Source Parameters in Recipe for Predicting Strong Ground Motions

- Objective: Diversification of ground motion distributions

a) Results of Chapter 5

The study in **Chapter 5** tried to cover the diversity of ground motion distribution efficiently and examines the "Recipe for Predicting Strong Ground Motions [2]," which is the standard setting criterion for the characterized source model for SESMs in Japan. Specifically, the spatial characteristics of SESMs are diversified by the "recipe cases" with the fault models set up in accordance with the "recipe" and the "extended cases" with fault models set up by extending the "recipe" setting conditions. Here, the target "diversity of spatial characteristics" refers to the spatial characteristics represented by the ground motion distributions for 600 cases, which were set by Kagawa [5, 6] with variations in the standard values of the "recipe."

Non-hierarchical cluster analysis was applied to the PCSs of the 600 ground motion distributions; by arranging the PCSs for Modes 1–3 in two-dimensions, the diversity of the ground motion distributions could be visualized. After plotting the recipe cases on a two-dimensional distribution of PCSs for the 600 cases, the spatial characteristics for the recipe cases were average and limited, and did not cover the diversity of the 600 cases. The combination of the recipe cases and the extended cases allowed the representation of a wide diversity of ground motion distributions, including spatial characteristics that do not appear in the recipe cases alone. This diversifying effect was greater for the extended case with multiple parameters extended simultaneously than for the extended case with solely a single parameter extended. At the computational points along and around the strike extension of the fault, the ground motion (S_A) of the recipe cases and the extension case approximately reproduced the range of the 600 cases. The ground motion distributions over the entire computational area of the recipe cases and the extension cases were covered for the distributions below the mean of the 600 cases, but did not fully represent the larger-than-average cases.

b) Future Developments of Chapter 5

Yamada et al. [7–9] proposed a setting method using the LHS (Latin Hypercube Sampling) as an efficient way to set up various characterized source models considering the "recipe" setting values and their variations. Imai et al. [10] proposed a method using copulas to capture the distribution shape of PCSs among modes as a probability distribution, sample PCSs according to this probability distribution, and simulate ground motion distribution. The sampling using LHS is an efficient method to cover the diversity of source parameters without considering the diversity of spatial characteristics of the ground motion distribution. In contrast, the sampling simulation using copulas is an efficient way to cover various spatial characteristics without directly considering the source parameter settings. Future studies will set up and simulate the fault models and ground motion distributions by sampling and simulation using the LHS method [7–9] and the copula method [10]. Based on these results, the number of cases necessary to efficiently cover the diversity of seismic motion distribution will be judged.

7.1.5 Chapter 6: Analysis of Spatial Characteristics of Tsunami Inundation Depth Caused by the Nankai Megathrust Earthquakes

- Objective: Identifying the applicability and usefulness of data mining and machine learning methods used in this study to multi-hazards

a) Results of Chapter 6

The study in **Chapter 6** clarifies the applicability and usefulness of data mining and machine learning methods used in this study to multi-hazards. Specifically, the spatial characteristics of the tsunami inundation depth distributions are analyzed using the "decomposition" method (SVD: Singular Value Decomposition, NMF: Non-negative Matrix Factorization). As a numerical example, the tsunami inundation depth distributions for 11 cases caused by the Nankai trough megathrust earthquake [11] were used.

In SVD, the modal forms represent the variation structure of the entire area. Mode 1 reflects the average characteristics of the inundation depth distribution. Mode 2 and higher modes reflect the arrangement condition of large slip zone, including super large one. In NMF using uniform random numbers as the initialization method, the bases represent the local area's variation structure. The bases reflect the arrangement condition of a large slip zone. Each base contributes to a case where the arrangement is closely related. In NMF using NNDSVD as the initialization method, Basis 1 contributes to all cases, and its distribution of the basis vector represents the spatial distribution that covers the entire area. Basis 2–6 reflect the local spatial characteristics associated with the arrangement condition of a large slip zone.

The results of these analyses reveal the applicability of multi-hazard in the analytical methods proposed in this study.

b) Future Developments of Chapter 6

The study in **Chapter 6** used the tsunami inundation distributions for only 11 cases. For future developments, the analysis will be performed for many inundation depth distributions with more diversity in terms of setting conditions of slip angle and slip amount. The Tsunami Hazard Station (J-THIS) [12] established by the national research institute for earth science and disaster prevention has provided the tsunami inundation depth distributions for 2720 cases of the characterized fault models with different combinations of earthquake area and slip amount. These data will be used to analyze the spatial characteristics of inundation depth. Furthermore, we plan to construct a prediction model of tsunami inundation depth distribution based on "decomposition and synthesis" and "machine learning modeling."

7.2 Reach and Advantages/Disadvantages of This Study

I repeated the research summary and issues in each chapter. Next, I summarize how far these results have reached the ultimate goal of this study. In addition, the advantages and disadvantages of the research and the proposed method are summarized.

First, the proposed method's overall advantage is that it can generate ground motion distributions and evaluate spatial characteristics in relation to physical interpretations by using “data mining and machine learning.” As new methods are developed in the future, various analyses are expected to be possible. In the following, I describe the degree of achievement and advantages and disadvantages of the two major themes, “simulation of ground motion distributions” and “diversification of ground motion distributions,” for the ultimate research goals.

7.2.1 Simulation of Ground Motion Distributions

The “simulation methods for ground motion distribution using computed data” in **Chapters 2** and **3** have fully achieved the research goal of showing examples of the effective use of computed data. Various ground motion distribution (or tsunami inundation depth distribution) data have been simulated with the remarkable development of computer resources in recent years. With the method proposed in this study, these data can be efficiently used for practical purposes such as risk assessment.

In contrast to these advantages, a disadvantage is that the proposed method’s simulation results (or accuracy) depend on the quality and quantity of the original computed data. Thus, the original dataW’s diversity is the key to solving this problem. Suppose I can propose an “extended recipe version” that covers the diversity of ground motion distributions by completing the “diversification of ground motion distributions (extension of “recipe” setting conditions)” in this study. In that case, this problem can be solved to some extent.

Another disadvantage is that the simulation results are results within the diversity (or variability) of the original data (i.e., interpolation concerning the original data), not completely unknown results (i.e., extrapolation). In particular, the “ground motion distribution prediction model” in **Chapter 3**, due to the nature of its prediction algorithm (machine learning), cannot predict data outside the range of the training data (more precisely, the accuracy of the prediction results cannot be guaranteed).

7.2.2 Diversification of Ground Motion Distributions

In **Chapter 4**, a method is proposed to evaluate “spatial characteristics of ground motion distributions,” which is important when considering “diversification of ground motion distributions.” Previous methods for evaluating spatial characteristics have only applied mode decomposition to ground motion distributions [3]. In this paper, the method is developed by modeling the weight coefficients obtained from the mode decomposition by the source parameters using machine learning, and evaluating the characteristics of the

mode based on an explainable AI (XAI). The advantage of this method is that it can consider the interaction of source parameters in the evaluation. However, the evaluation results are based on visualizing and quantifying the input-output relationship of the model and do not explain the weight coefficients themselves. Therefore, the evaluation results depend on various factors involved in model construction, such as the accuracy of the model's predictions and the appropriateness of the explanatory variables. Furthermore, the above characteristics are disadvantages of this evaluation method itself. Therefore, when using this method, it is important to consider the above drawbacks and interpret the results based on the mechanical characteristics of the model.

Chapter 5, "diversification of ground motion distributions (extension of "recipe" setting conditions)," is a basic study of this theme. The content of this paper was consistently focused on the 600 ground motion distributions caused by "the strike-slip fault consisting of a single fault plane" located in "uniform horizontally stratified ground" [4, 5]. An attempt was made to reproduce the diversity of the 600 cases using the recipe cases and the extended cases. In this paper, while the diversity of spatial characteristics of the ground motion distributions (absolute acceleration response spectrum (period 2.0 s, fault-normal component)) for the 600 cases can be roughly reproduced; however, there is still room for reconsideration regarding the setting of the expansion conditions. Note that this is based on only the 600 cases of strike-slip fault; additional studies for a wide variety of faults are needed to complete the study. In addition, further analysis and consideration of complex ground structures and other conditions that more closely resemble real-world conditions should be conducted in the future. Finally, it is necessary to demonstrate the usefulness of the "extended version of the recipe" by performing analyses on the active fault that actually caused the earthquake and its ground motion distribution. After these analyses and discussions, this study is complete.

The completion of this study is expected to encourage the export of scenario seismic hazard maps (or "recipe") to other countries (or the introduction of such systems in other countries). As described in **Chapter 1**, many countries, including the U.S., have published only the predicted ground motion distribution based on a GMPE as a hazard map for a scenario earthquake. In contrast, there were no examples, such as Japan's SESMs, where source fault models that can consider the detailed rupture process scenarios of the fault are assumed, and the ground motion distributions are predicted using a detailed method. As a reason for this, the United States Geological Survey (USGS) [13] stated the following.

> "our approach is to generally show the average effect because it is difficult to justify a particular choice of hypocenter or to show the results for every possible hypocentral location." [14]

The completion of this study will lead to the creation of scenario ground motion distribution maps that efficiently cover the uncertainties in the source parameters, and is expected to provide a clue to solving this problem.

References in Chapter 7

[1] The National Research Institute for Earth Science and Disaster Resilience. “Japan seismic hazard information station (J-HIS).” https://www.j-shis.bosai.go.jp/en/

[2] The Earthquake Research Committee, the Headquarters for Earthquake Research Promotion. (2020) Strong ground motion prediction method for earthquakes with specified source faults (“Recipe”). https://www.jishin.go.jp/main/chousa/20_yosokuchizu/recipe.pdf **(in Japanese)**

[3] Nojima, N. Kuse, M. and Duc, L. Q. (2018) Mode decomposition and simulation of strong ground motion distribution using singular value decomposition. *Journal of JAEE*, **18** (2): pp. 95–114. https://doi.org/10.5610/jaee.18.2_95 **(in Japanese with English abstract)**

[4] Kagawa, T. (2015) A study on spatial and temporal variabilities of strong ground motions based on multiple fault rupture scenarios. *Journal of JSCE A1*, 71 (4): pp. I_191–I_197. https://doi.org/10.2208/jscejseee.71.I_191 **(in Japanese with English abstract)**

[5] Kagawa, T. (2015) Spatial variability of periodic characteristics among strong ground motions derived from multiple fault rupture scenarios. *Journal of JAEE*, 15 (7): pp. 90–99. https://doi.org/10.5610/jaee.15.7_90 **(in Japanese with English abstract)**

[6] Kagawa, T. (2004) Developing a Stochastic Green’s function method having more accuracy in long period range to be used in the Hybrid Method. *Journal of JAEE*, **4** (2): pp. 21–32. https://doi.org/10.5610/jaee.4.2_21 **(in Japanese with English abstract)**

[7] Yamada, M., Senna, S. and Fujiwara, H. (2007) Statistical analysis of predicted ground motions on the basis of a recipe for strong-motion prediction. *Journal of JAEE*, 7 (1): pp. 43–60. https://doi.org/10.5610/jaee.7.43 **(in Japanese with English abstract)**

[8] Yamada, M., Senna, S. and Fujiwara, H. (2007) statistical analysis of predicted ground motions on the basis of a recipe for strong-motion prediction for -dip-slip fault and strike slip fault-. *JSCE Journal of Earthquake Engineering*, **29**: pp. 104–113. https://doi.org/10.11532/proee2005a.29.104 **(in Japanese with English abstract)**

[9] Yamada, M., Senna, S. and Fujiwara, H. (2011) Statistical analysis of ground motions estimated on the basis of a recipe for strong-motion prediction: approach to quantitative evaluation of average and standard deviation of ground motion distribution. *Pure Appl. Geophys.*, **168**: pp. 141–153. https://doi.org/10.1007/s00024-010-0159-0

[10] Imai, R., Kasui, N., Iwaki, A. and Fujiwara, H. (2020) A sample generation of scenario earthquake shaking maps via modal decomposition and empirical copula. *JAEE Annual Meeting 2020*. **(in Japanese)**

[11] The central disaster management council (2012) Working group report on megathrust earthquake models for the Nankai trough (second report). https://www.bousai.go.jp/jishin/nankai/taisaku/pdf/20120829_2nd_report01.pdf **(in Japanese)**

[12] The National Research Institute for Earth Science and Disaster Resilience. “Japan tsunami hazard information station (J-THIS).” https://www.j-this.bosai.go.jp/ **(in Japanese)**

[13] The United States Geological Survey. https://www.usgs.gov/

[14] Worden, C. B., Thompson, E. M., Hearne, M. G., and Wald, D. J. (2020) ShakeMap Manual Online: technical manual, user’s guide, and software guide. U. S. Geological Survey. https://doi.org/10.5066/F7D21VPQ.