Reliability assessment of structural dynamic systems subject to earthquake-induced tsunamis

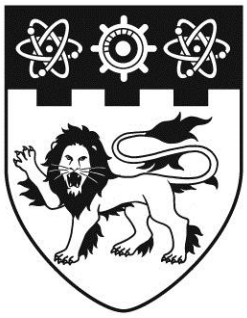
Shao, Zhe

2016

Shao, Z. (2016). Reliability assessment of structural dynamic systems subject to earthquake-induced tsunamis. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/69064

https://doi.org/10.32657/10356/69064



NANYANG TECHNOLOGICAL UNIVERSITY

RELIABILITY ASSESSMENT OF STRUCTURAL DYNAMIC SYSTEMS SUBJECT TO EARTHQUAKE-INDUCED TSUNAMIS

SHAO ZHE

School of Civil and Environmental Engineering

2016

RELIABILITY ASSESSMENT OF STRUCTURAL DYNAMIC SYSTEMS SUBJECT TO EARTHQUAKE-INDUCED TSUNAMIS

SHAO ZHE

School of Civil and Environmental Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

2016

Acknowledgements

First and for most, I am highly indebted to my advisors, Professors Joseph, Cheung Sai Hung and Robert, Tiong Lee Kong, for their guidance and constant support as well as providing insights and ideas on this project. Joseph is one of the smartest and energetic people I have ever met, and has always been there to enlighten me with insightful discussions and advice. Robert has given me invaluable freedom in pursuing my own research interests and offers many suggestions.

I would also like to express my thanks to Dr. Kenneth Macpherson for his in-depth and explicit guidance on the earthquake source models. Many thanks also go to the researchers in different fields: Dr. Pan Wenliang, Dr. Yao Yu and Ms Yao Yao for the implementation of COMCOT, and Dr. Seshu Nimmala on the issue of water part modeling in LS-DYNA. Also special appreciations to my colleagues, Sahil Bensal, Roshan Wahab, Gong Fanming, Queen Suraajini Rajendran for their uncountable help and valuable suggestion. I also appreciate the assistance from Ms. Li Juan and Mr. Suresh Betapudi in teaching me the ArcGIS tool. I would like to thank all the friends who are not listed here but lend me their supports in every possible manner.

Last but not least, my deepest gratitude goes to my family, especially my parents and grandparents for their unconditional trust, love and encouragement.

Table of Contents

Acknowledgements III

Summary VIII

List of Tables X

List of Figures XI

List of Symbols XIV

Chapter 1 Introduction 1

1.1 Motivation 1

1.2 Scope of the study 3

Chapter 2 Literature Review on Tsunami Wave-structure Interaction and Computational Methods 6

2.1 Tsunamigenic earthquake and tsunami earthquake 6

2.1.1 Tsunamigenic earthquake 6

2.1.2 Tsunami earthquake 7

2.2 Earthquake source model and stochastic slip distribution 8

2.2.1 Uniform slip model 9

2.2.2 Stochastic spatial slip distribution 12

2.2.3 Hybrid stochastic k-squared static slip distribution model 13

2.2.4 Temporal slip distribution 14

2.3 Tsunami wave evolution 14

2.3.1 Tsunami generation 15

2.3.2 Tsunami propagation and run-up 17

2.3.3 Numerical methods 19

2.4 Tsunami wave-structure interaction 21

2.4.1 Tsunami force quantification based on experiments and observations 21

2.4.2 Numerical methods 25
2.5 Classical Computational Methods of Reliability Integrals 29
2.5.1 First-order and second-order reliability methods 30
2.5.2 Monte Carlo simulation 30
2.5.3 Markov chain Monte Carlo simulation method 31
2.5.4 Importance sampling 34
2.5.5 Subset Simulation method 35
2.6 Concluding remarks 38
Chapter 3 Proposed Tsunami Wave-structure Interaction Model 41
3.1 General remarks 41
3.2 Tohoku tsunami modelling 41
3.2.1 Tsunami model 42
3.2.2 2011 Tohoku earthquake source parameters 43
3.2.3 Tsunami model results and validation 45
3.3 Structural model setup 46
3.3.1 Structural geometry and material modeling 47
3.3.2 Tsunami force modeling 50
3.3.3 Boundary conditions 51
3.4 Structural performance 52
3.5 Benchmark problem for reliability analysis 54
3.5.1 Stochastic earthquake source model 54
3.5.2 Simplified structural model 56
3.5.3 Benchmark reliability analysis results 58
3.6 Closing remarks 64
Chapter 4 Proposed Adaptive Metamodels for Structural Reliability Analysis 66
4.1 Challenges 66

4.2 Response surface methods 67
4.2.1 Moving least square metamodel 68
4.2.2 Gaussian processes for regression 70
4.2.3 Remarks 73
4.3 Proposed adaptive SS-MLS and SS-GP algorithms for reliability analysis 74
4.4 Illustrative example 77
4.4.1 SS-MLS and SS-GP 77
4.5 Closing remarks 83
Chapter 5 Higher Dimension Metamodel in Reliability Analysis 85
5.1 Sparse Bayesian modeling 85
5.1.1 Sparse Bayesian Inference 86
5.2 Proposed SB-SSMLS for higher order metamodels and higher dimensional random variables 87
5.3 Illustrative example 90
5.3.1 Higher-order SB-SSMLS and SB-SSMGP models 91
5.3.2 Adaptive SB-SSMLS for high dimensional random variables 100
5.4 Closing remarks 108
Chapter 6 Other Enhancements over Metamodel-Based Reliability Analysis 111
6.1 Consistent reliability evaluation using SS-MLS integrated with response conditioning method 111
6.1.1 Response conditioning method 111
6.1.2 Adaptive RCM integrated with SS-MLS algorithm 112
6.1.3 Illustrative example 116
6.2 Subset data selection with multiple criterion in GP regression 119
6.2.1 Subset data selection for Gaussian processes using multi-objective optimization 120

6.2.2 Illustrative example 121
6.3 Closing remarks 126
Chapter 7 Conclusion and Future Works 128
7.1 Conclusion 128
7.2 Future works 131
Reference 133
Appendix A.1 Sample COMCOT Input for the Benchmark Study 145
Appendix A.2 Sample COMCOT Input for Higher Dimensional Case Study 147

Summary

Recent earthquake-induced tsunamis occurred in Padang, 2004 and Tohoku, 2011 brought huge losses of lives and properties. To date, it is neither practical nor possible to accurately predict such hazard and construct perfectly effective countermeasures against them. Therefore, it is of our great interest to quantify the structural risk caused by earthquake-induced tsunamis. Despite continuous advancement in numerical simulation of tsunami and wave-structure interaction, it still remains a computationally challenging task to evaluate the reliability of structural dynamic system, especially when uncertainties related to the system and its modelling are taken into account. The failure of the structure in a tsunami-wave-structural system (the complement of reliability) is defined as any response quantities of the system exceeding specified thresholds during the time when the structure is subjected to dynamic wave impact due to earthquake-induced tsunamis.

In this study, attempts have been made from two perspectives. Firstly, the focus is concentrated on the physical numerical simulation of the tsunami wave-structure interaction. The uncertainty arises from both the stochastic process of earthquake source generation and the structural parameters. Therefore, the tsunami generation, propagation and runup, are carried out, prior to the modelling of the wave-structure interaction. The stochastic earthquake source model is proposed to generate tsunami profile which propagates to shorelines. The numerical analysis of the interaction is performed through LS-DYNA on a high-performance computing system.

Secondly, various approaches based on a novel integration of the Subset Simulation algorithm and two metamodels, *i.e.* modified moving least squares (MLS) response surface approach (Taflanidis & Cheung 2012), as well as Gaussian processes (GP), are proposed to evaluate the reliability of the dynamic system. The metamodels are used to replace the computationally expensive physical numerical simulations. The proposed algorithms follow a logical sequence. First of all, we proposed adaptive SSMLS/SSGP (Chapter 4) in order to assess the reliability of the complex system more accurately. For problems which involve high dimensional variables, the

modified sparse Bayesian inference is proposed and a polynomial up to the 4th order is employed (Chapter 5). On the other hand, the bias (inconsistency) of the metamodels is alleviated to a certain extend by the newly proposed SSMLS-RCM which theoretically can rectify the approximation and converge to the exact result (Chapter 6).

The results showcase the ability of the proposed stochastic earthquake source model and the numerical simulation tools in modelling the complete physical process. Furthermore, the novel integration of the Subset Simulation and metamodels demonstrates potential benefit in estimating failure probability of a highly nonlinear complex system. The effectiveness of the proposed algorithms is discussed by comparing the results with the results obtained from the original Subset Simulation without using the metamodels.

List of Tables

Table 1 List of 10 costliest natural hazards from 1980 to 2014 1

Table 2 List of 10 deadliest tsunamis in the history 2

Table 3 Summary of COMCOT nested grid setting 43

Table 4 Maximum initial water surface height (m) 44

Table 5 Earthquake parameters required in COMCOT and 45

Table 6 Earthquake parameters for the benchmark study 55

Table 7 Location of the 16 slip angles on a fault plane 61

Table 8 Normalized information entropy for the marginal probability distribution of the 16 random variables 63

Table 9 Comparison of the computation efficiency 83

Table 10 Comparison of the computation efficiency for SB-SSMLS algorithms ... 96

Table 11 Fitness of the higher order metamodel compared to benchmark 99

Table 12 Comparison of the computation efficiency for SB-SSMGP algorithms. 100

Table 13 Distributions of the 53 random variables 101

Table 14 Sub-optimized MLS parameters in adaptive SB-SSMLS 102

Table 15 Location of the 36 slip angles on a fault plane 106

Table 16 Computation efficiency for SSMLS-RCM algorithms 118

Table 17 Sample result of non-dominated sorting for MOO-SSGP 121

Table 18 Fitness of the adv-SSGP and MOO-SSGP compared to SS 125

Table 19 Comparison of the computation efficiency 126

List of Figures

Figure 1 Overall flowchart of the study 5

Figure 2 Various types of subduction-zone faulting and examples 7

Figure 3 Schematic cross section of a Décollement 8

Figure 4 The actual fault and two levels of approximation to the slip vector field . 11

Figure 5 Illustration of fault plane and fault parameters 16

Figure 6 Grid system used in COMCOT 21

Figure 7 Multi-layer grid system schematic 21

Figure 8 Conceptual schematic of inundation depth and run-up height 22

Figure 9 Comparison of tsunami bore velocities 24

Figure 10 Schematic fluid-structure interaction domains 25

Figure 11 Monolithic and partitioned approach for the coupled system 26

Figure 12 ALE approach description (Haritos *et al.* 2005) 28

Figure 13 Minami Gamou Wastewater Treatment Plant out-of-plane flexure damage 41

Figure 14 Bathymetry and topology of the three nested grids used in the simulation 43

Figure 15 Maximum inundation map 46

Figure 16 Geometry details of the tsunami-facing wall model 48

Figure 17 *DIF* for concrete material 49

Figure 18 Strain-rate curve of reinforcement bar of with σ_y 460MPa 50

Figure 19 Tsunami wave-structure interaction model setup in LS-DYNA 51

Figure 20 Deformation of the structural wall 53

Figure 21 Time-history curves for wave average pressure and nodal displacement 54

Figure 22 Inundation map generated using proposed earthquake source generator 56

Figure 23 Maximum structural displacement diagram 57

Figure 24 Nodal displacements time-history curve 58

Figure 25 Failure probability estimates of the 30 Subset Simulations and the averaged curve 59

Figure 26 Coefficient of variation of failure probability estimate for benchmark... 60

Figure 27 Kernel density plots of random variables for various intermediate failure levels 62

Figure 28 Correlation matrix of the 16 slip angles 64

Figure 29 Training data selecting process in GP method 77

Figure 30 Failure probability estimates of SS, ori-SSMLS and adv-SSMLS..... 79

Figure 31 Failure probability estimates of SS, ori-SSGP and adv-SSGP with their associated uncertainty levels..... 81

Figure 32 The c.o.v. of failure probability estimator for SS, SSMLS and SSGP.... 82

Figure 33 Sparse Bayesian metamodel framework 89

Figure 34 Failure probability estimates of SS, ori-SSMLS and SB-SSMLS 92

Figure 35 Frequency of the basis terms chosen by sparse Bayesian MLS 93

Figure 36 Failure probability estimates by SS, third order and fourth order SB-SSMLS..... 94

Figure 37 Frequency of the basis terms chosen by SB-SSMLS. 95

Figure 38 The c.o.v. of failure probability estimates for SS, third order and fourth order SB-SSMLS 96

Figure 39 Failure probability estimates of SB-SSGP with various order basis function 97

Figure 40 Failure probability estimates of SS and various orders SB-SSMGP..... 98

Figure 41 The c.o.v. of failure probability estimates for the third and fourth orders SB-SSMGP 99

Figure 42 Failure probability estimates of SS and adaptive SB-SSMLS 102

Figure 43 Correlation matrix of four earthquake parameters 104

Figure 44 Kernel density estimates of earthquake magnitudes for the three intermediate failure levels 104

Figure 45 Kernel density estimates of earthquake depth for the three intermediate failure levels 105

Figure 46 Upper diagonal correlation matrix of the slip angles 106

Figure 47 Upper diagonal correlation matrix of the slip amplitude 107

Figure 48 Correlation matrix of the five structural variables 108

Figure 49 Results of a few typical runs using the RCM 113

Figure 50 Scattergram of the samples selected from the sub-bins 114

Figure 51 An illustration of the interaction between MLS and RCM 116

Figure 52 Failure probability estimates of SS and adaptive RCM estimates 117

Figure 53 The c.o.v. of failure probability estimates for SS and the SSMLS-RCM 118

Figure 54 Scattergram of the samples selected from the sub-bins for SSMLS-RCM 119

Figure 55 Comparison of the correlation coefficients for the direct RCM and SSMLS-RCM 119

Figure 56 Ranking based on the non-domination sorting 124

Figure 57 Failure probability estimates of SS, adv-SSGP, and MOO-SSGP with their associated uncertainty levels 125

Figure 58 The c.o.v. of failure probability estimates for SS, adv-SSGP and MOO-SSGP 126

Figure 59 Schematic structure of the study of the structural reliability analysis ... 131

List of Symbols

M_0	Seismic moment
L, W	Fault length and width
$\vec{\xi}_1$, $\vec{\xi}_2$	Coordinates along fault length and width directions
$\bar{D}$	Ultimate static slip
v_{r}	Rupture velocity
τ_{r}	Rise time
$\Delta\sigma$	Static stress drop
μ	Shear modulus
S	Fault area
k_{x}, k_{z}	Wavenumber in x and z directions
Φ	Phase of the random spectrum
K_{N}	Nyquist wavenumbers
$\Delta L, \Delta W$	Characteristic subfault size
θ, α and λ.	Strike angle, dip angle and rake (slip) angle
u_z	Vertical displacement due to shear dislocation in a half space
P, Q	Water flux
η	Water surface elevation
h	Mean water depth
H	$= \eta + h$. Total water depth
f	Coriolis coefficient
F_{x}, F_{y}	Bottom friction in *x*- and *y*-direction
n	Manning roughness coefficient
Δt	Time step in finite difference scheme
Δx	Grid size in explicit Leap-Frog finite difference scheme
H_{max}	Maximum water depth within the region of the particular grid level
u_{p}	Flow velocity normal to the structural surface
C_D	Drag coefficient
m_i	Debris mass
u_i	Debris impacting velocity

C_m	Added mass coefficient
Ω_f, Ω_s	Fluid domain and structural domain
σ_{ij}^f , σ_{ij}^s	The tensor stress in fluid domain and structural domain, respectively
E	Young's modulus
ν	Poisson ratio
β	Reliability index used in First-order and second-order reliability methods
Φ_{R}	Standard normal cumulative distribution function in first-order and second-order reliability methods
π_i, π_j	Stationary distribution used in Markov chain
δ_i	Coefficient of variation of any intermediate level i
δ	Overall coefficient of variation
$\dot{\varepsilon}$, $\dot{\varepsilon}_s$	Strain rate and static strain rate
m	Intermediate failure levels
n	Total number of numerical evaluations
p_0	Conditional failure probability
n_θ	Spatial dimension (number of random variables)
NS	Number of supporting points
D	Influence radius in moving least square method
c, k	Shape factors in moving least square method
σ_i	Standard deviation of the i-th component supporting points
σ_f	Standard deviation in squared exponential (SE) covariance function
l	Scale length in squared exponential (SE) covariance function
M	Number of metamodel samples for each intermediate failure level
$\boldsymbol{\varepsilon}_i$	Prediction error

Chapter 1 Introduction

1.1 Motivation

Looking back at history, the development of human society is accompanied by natural hazards. To date, no efficient preventions for such catastrophes have been found in view the evidence of enormous economic loss as well as casualties brought by recent hazards as shown in Table 1.

Table 1 List of 10 costliest natural hazards from 1980 to 2014

Rank	Overall loss [US$ billion]	Fatalities	Event	Country	Year
1	210	15,880	Tohoku earthquake	JP	2011
2	125	1,322	Hurricane Katrina	US	2005
3	100	6,430	Kobe earthquake	JP	1995
4	85	84,000	Sichuan earthquake	CN	2008
5	69	210	Hurricane Sandy	US	2012
6	44	61	Northridge earthquake	US	1994
7	43	813	Thailand flood	TH	2011
8	38	170	Hurricane Ike	US	2008
9	30	520	Chile earthquake	CL	2010
10	28	46	Honshu earthquake	JP	2014

Source: Munich Re NatCatSERVICE, 2015

On top of that, tsunami hazard, being one of the deadliest natural catastrophes, shows an increasing occurrence rate worldwide as seen in Table 2 (Geist & Parsons 2011). Even though Japan is considered as a leading country in tackling with tsunamis, the tendency of underestimating the extreme scenario led to the failure of the structure countermeasures such as breakwater and seawalls (Shuto & Fujima 2009) during the 2011 Tohoku event: the maximum potential earthquake was thought as not exceeding Mw 8.2 in that region (Shibata 1998). Similarly, the Sumatra earthquake in 2004 and Chile earthquake in 1960 are also considered as "black swan" events, indicating an unprecedented and unexpected event in human

history at the point when it occurred. It is difficult to predict when and where such a tsunami happens and whether there will be a larger one in the future (Kelleher *et al.* 1973). Therefore, a probabilistic or probability-based analysis provides a rational approach to mitigate the tsunami risk. More importantly, constructing vertical evacuations to high ground is the most crucial strategy to effectively reduce casualty during the tsunami event. Thus, it is of our great interest to quantify the risk to structural dynamic systems due to earthquake-induced tsunamis.

Table 2 List of 10 deadliest tsunamis in the history

Rank	Death toll	Location (Trigger)	Year
1.	350,000	Indonesia (EQ*)	2004
2.	100,000+	Ancient Greece (VC+)	1410 B.C.
3.	100,000	Portugal (EQ)	1755
4.	100,000	Italy (EQ)	1908
5.	40,000	Taiwan (EQ)	1782
6.	36,500	Indonesia (VC)	1883
7.	30,000	Japan (EQ)	1707
8.	26,360	Japan (EQ)	1896
9.	25,674	Chile (EQ)	1868
10.	15,854	Japan (EQ)	2011

* Earthquake

\+ Volcano eruption

Data source: National Geophysical Data Center

The process of an earthquake-induced tsunami interacting with on-shore structures is a multi-physics phenomenon consisting of seismology, fluid dynamics and fluid-structure interactions. Each component can be treated as a complex system and the interaction between them is highly nonlinear.

Some parts of the above problems have been investigated separately by researchers and engineers. For example, Geist (2005) examined the randomness of the spatial slip distribution with its effect on the generated tsunami amplitude. Geist (2005) and Seda & Tuncay (2010) discussed the relationship between tsunami amplitude

and different earthquake source parameters. Besides, tsunami loading experiments are carried out together with numerical analyses to better understand the wave-structure interaction issue. For instance, Nimmala *et al*. (2006) and Lau *et al*. (2011) did detailed bridge modelling to quantify the tsunami wave forces. With uncertainties pervasive in the whole process, tsunami vulnerability and probabilistic tsunami hazard assessment (Koshimura *et al.* 2009; Omira *et al.* 2010) are commonly adopted to quantify the potential risk. This type of analysis mainly focuses on assessing the damage variability and hazard recurrence probability. However, the previous probabilistic studies hardly satisfy the requirement of reliability analysis, which alternatively, studies the ability of a system to perform its required functions within a certain period of time. Given the fact that it is already computationally intensive to run a single structure-tsunami interaction simulation in order to obtain a structural response, a reliability assessment which normally requires thousands of such simulations or more even using the state-of-the-art approach such as Subset Simulation (Au & Beck 2001) can be computationally prohibitive. Thus, this makes the advancement of the stochastic analysis and reliability computation necessary as well as challenging.

1.2 Scope of the study

In this study, the uncertainty arises from both the stochastic earthquake source and the structural parameters. Therefore, the tsunami generation, propagation and runup, are carried out, prior to the modelling of the wave-structure interaction. The structural reliability (or its complement failure probability) under the risk of earthquake-induced tsunamis is examined. The uncertainty parameters come from both the earthquake source generation and structural variables. The stochastic source model couples the hybrid k-squared slip distribution (Gallovič & Brokešová 2004) with the rest of the seismological possible rupture variables. All the earthquake parameters are used to generate tsunami wave profiles which propagate to shorelines. A detailed numerical simulation for the wave-structure interaction is modelled using LS-DYNA. Furthermore, novel hybrid approaches are proposed, which integrates various metamodels, *i.e.* modified moving least squares (MLS) response surface approach (Taflanidis & Cheung 2012), as well as Gaussian

processes (GP) with the Subset Simulation (SS) method (Au & Beck 2001) to evaluate the reliability of the structure which interacts with the dynamic tsunami waves. The structural reliability analysis involves the computation of a high-dimensional probability integral with the integrand involving a function of the solutions of high-dimensional stochastic differential equations. Failure of the structure is defined as any response quantities of the tsunami-wave-structure system exceeding specified thresholds during the time when the structure is subjected to dynamic wave impact due to earthquake-induced tsunamis.

The schematic flow of the study on a reliability assessment of structural dynamic systems due to earthquake-induced tsunamis is shown in Figure 1. The thesis is structured in the following way. Chapter 2 firstly describes the state-of-the-art related to the study, including the commonly-used earthquake generation models and classic computational methods for reliability assessment. Then Chapter 3 proposed a series of physical models starting from the earthquake source model to the nonlinear structural response computation (e.g. maximum displacement, inter-story drift ratio, etc.). The benchmark study using the proposed physical process is included at the end of the chapter. It is followed by a newly proposed reliability algorithm in Chapter 4 which generalizes the procedures to evaluate the failure probability of a structural dynamic system under earthquake-induced tsunami risks. In Chapter 5, a novel algorithm is proposed which utilizes response surface models for high-dimensional problems. Chapter 6 discusses two algorithms which serve the purpose of enhancing the performance of the algorithms proposed in Chapters 4 and 5. Finally, Chapter 7 summarizes the results obtained so far and the future works.

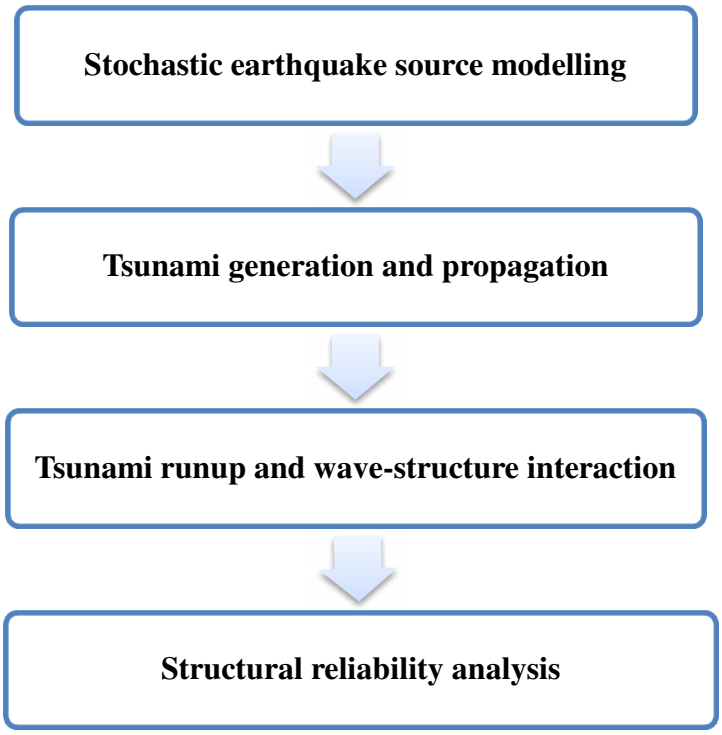


Figure 1 Overall flowchart of the study

Chapter 2 Literature Review on Tsunami Wave-structure Interaction and Computational Methods

2.1 Tsunamigenic earthquake and tsunami earthquake

Earthquake is a result from continuous transformations of tectonic plate movements. In a micro scale, the relative movement of the plates is initiated due to the exceeding of a maximum accumulative stress which a rock can bear on a fault plane. The regions where two or more plates converge are called subduction zones.

2.1.1 Tsunamigenic earthquake

Given the low efficiency in energy transmitting, under-seafloor earthquakes with certain source mechanism and moment magnitude M_w greater than 6.8 can trigger potential tsunamis (Titov 1997). This type of earthquake is referred to as tsunamigenic earthquake. In fact, most of the tsunamigenic earthquakes are from the subduction-zones (Pacheco & Sykes 1992; Geist 1998). There are several types of subduction-zone earthquakes. The majority of the subduction-zone earthquakes are initiated by the relative motion between two plate boundaries, referring to as interplate thrust (Geist 1998). This type of earthquake faulting accounts for more than 90 per cent of the total seismic energy release around the world (Bolt 2005). However, there are possibilities that the earthquake could be in either of the continental crust or oceanic crust, creating back-arc thrust and outer-rise fault, respectively. The latter is responsible for several mega tsunamis, such as 1933 Sanriku, 1990 Mariana and 2011 Tohoku (Lay *et al.* 2011). Figure 2 illustrates various types of subduction-zone faulting.

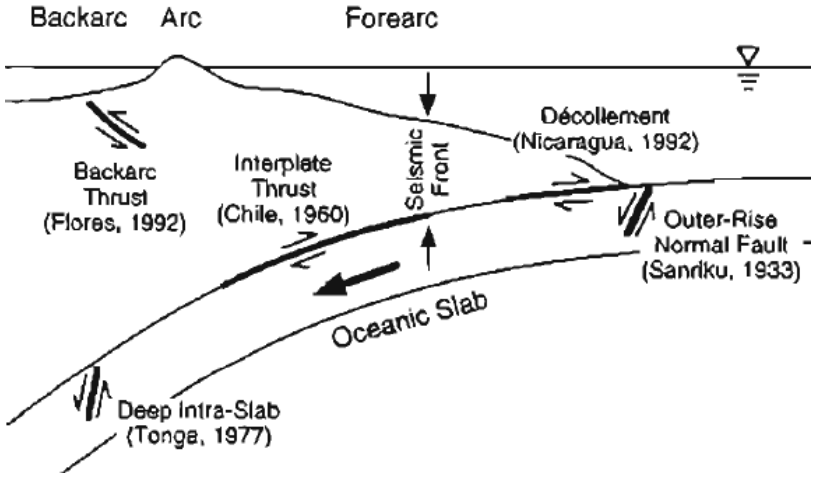


Figure 2 Various types of subduction-zone faulting and examples (Geist 1998)

2.1.2 Tsunami earthquake

On the other hand, there are earthquakes which do not release as much energy in a short duration as tsunamigenic earthquake but generate more tsunamis. It is considered as tsunami earthquakes firstly defined by Kanamori (1972). Tsunami earthquake is a special case of tsunamigenic earthquake. It occurs in shallow interplate region close to the edge of two boundaries of the plates named décollement. In geology, décollement is defined as a process in which some strata partly detach from those underneath and slide over them, causing folding and deformation. A typical décollement process is shown in Figure 3. Kanamori (1972) stated that the tsunami earthquake, despite of its relatively low energy release, contains large effective moment at long period at the order of 100s. This observation indicates a long rupture duration τ_c, comparing to other types of earthquakes. Therefore, the slow rupture velocity v_r is expected, since v_r is inversely proportional to the rupture duration (Aki & Richards 1980; Lay & Wallace 1995). This was also demonstrated by Pelayo and Wiens (1992) and Kanamori and Kikuchi (1993) in their studies on a few damaging tsunami earthquakes. They believed that the rupture velocity of tsunami earthquakes is likely to be limited owing to the low shear wave speed in accreted or subducted sediment near the trench. However, Ihmle (1996) provided another explanation – the low rupture velocity lies in the unique frictional properties of décollement sediment itself and

the shear wave speed in the sediment rock close to the fault zone may be of normal values.

Besides the long rupture duration, another characteristic of tsunami earthquakes is their relatively higher slip amount that could trigger highly nonlinear vertical wave displacement and considerable local tsunamis. Bilek & Lay (1999; 2002) and Seno (2002) suggested that the large dislocation could be the consequence of low rigidity (~10 GPa) of the sediments in a shallow subduction region. This can be explained from the relationship between seismic moment M_0 and fault size A and average slip $\bar{D}$ given by $M_0 = \mu A \bar{D}$ where μ is the shear module (rigidity). Please note that we do not differentiate the type of triggering earthquakes in this study since the scope is to analyse the structural performance given any earthquake which can generate potential tsunamis.

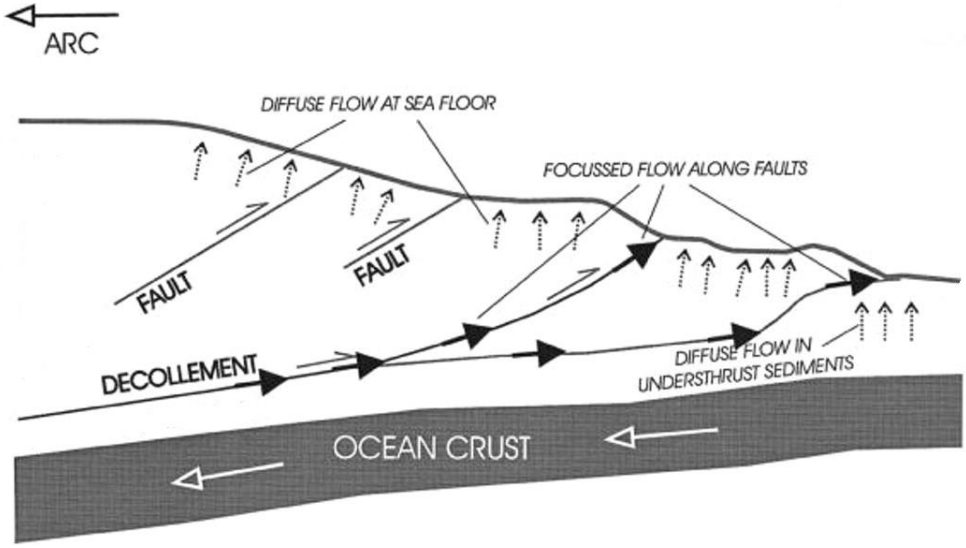


Figure 3 Schematic cross section of a Décollement (Saffer & Bekins 1999)

2.2 Earthquake source model and stochastic slip distribution

Seismic source modelling aims at investigating the kinetic and dynamic properties of the earthquake source by modelling the seismogram as well as other observed data. Two basic models, kinematic and dynamic models are commonly accepted in describing the earthquake source mechanism and dislocation. The concept of the kinematic model was introduced by Volterra (1907). The model quantifies the slip vector on the fault plane using dislocation theory without considering the causing

stress field. Therefore, the slip vector is treated as a function of space and time only. While for the dynamic model, it was first presented in Griffith's work (1921), aiming to further understand the stress change on the fault that causes the fracture. The earthquake can be better understood through the studies of friction laws and associated elasto-dynamic nature of the process. In another word, the dynamic model represents the earthquake source physics. A detailed review of the two models can be found in Madariaga and Olsen (2002). A number of applications have been developed (Kostrov & Das 1989; Madariaga & Olsen 2002) based on these two models. The kinematic model is widely applied in both forward and inverse problems. In solving the forward problem, the kinematic model helps to calculate the ground motion from the earthquake source parameters including the slip vector and geometry of the fault. Many slip distributions are proposed for this purpose, e.g. uniform slip model, k-squared model and other composite models (Zeng *et al.* 1994). Another major application of the kinematic model is to simulate the tsunami and landslide induced by an earthquake. The inverse problem, on the contrast, focuses on reproducing the slip distribution of an earthquake and fitting the result with the observed strong-motion seismograph, teleseismic, GPS and interferometry synthetic aperture radar (InSAR) data. The dynamic model is not employed as frequent as the kinematic model because of its complexity and high dependence on the numerical solution (Hartzell, S. *et al.* 2005). However, this model has become popular in the community recently along with the increasing power of computation technology (Beroza and Mikumo, 1996; Nielsen and Olsen, 2000; Guatteri et al., 2003, 2004; Peyrat and Olsen, 2004). In the following sections, the author presents two kinematic slip distribution models which can be adopted for the tsunami generation process.

2.2.1 Uniform slip model

Similar to other natural systems, an earthquake motion exhibits a source of complexity which till now, is hard to be fully understood and many issues are under dispute (Wyss & Brune 1967; Carlson & Langer 1989; Rice 1993; Madariaga & Cochard 1994; Lavallée *et al.* 2006). The demands of resolution regarding earthquake models are different in analyzing far-field and near-field tsunamis. That

is to say, the propagation distance affects the properties of tsunamis generated by various seismic source parameters (Geist 1998). For a far-field tsunami, the earthquake source model can be treated as point-source or line-source, and it is sufficient to simulate reliable wave amplitude. The source parameter that affects a far-field tsunami the most is the seismic moment M_0 that reflects the total radiated energy in an earthquake event. However, for the near-field case, it is further required to fully represent the earthquake model that depends on a number of source parameters due to the fact that the high frequency portion of the ground motion is dominant in the near field earthquakes. Thus, a finite fault model can be adopted, and both temporal and spatial variations are critical.

After examining the effect on tsunami generation, Geist (1998) and F. Løvholt *et al*. (2012) conclude that the slip vector has the largest influence on tsunami amplitude compared with the geometric parameters, such as rupture area, dip angle and source depth, etc. The uniform slip model, corresponding to a rectangular fault area, also known as the Volterra dislocation model, is a compromise of an actual fault as shown in the middle pane of Figure 4. This kinematic earthquake source model was introduced by Ben-Menahem (1961) and improved by Haskell (1964), in which he assumed a uniform dislocation along the fault. By assuming a rectangular fault with length L in $\vec{\xi}_1$ direction and width W in $\vec{\xi}_2$ direction, the slip function of a point on the fault plane is given as:

$$\left|\Delta u\left(\vec{\xi},\tau\right)\right| = \begin{cases} 0 & \tau < \dfrac{\xi_1}{v_r} \\ \bar{D}\dfrac{\tau - \dfrac{\xi_1}{v_r}}{\tau_r} & \dfrac{\xi_1}{v_r} < \tau < \dfrac{\xi_1}{v_r} + \tau_r \\ \bar{D} & \tau > \dfrac{\xi_1}{v_r} + \tau_r \end{cases} \qquad \xi_1 \in [0, L],\ \xi_2 \in [0, W] \qquad (2.1)$$

where v_r and τ_r are the rupture velocity and rise time, respectively as described above, while $\bar{D}$ is the ultimate static slip value. This function indicates an instantaneous fracture is formed at $\tau = 0$ in W direction and propagates unilaterally

over a length *L* along the $\vec{\xi}_1$ axis. The slip direction is arbitrary but constant over the fault.

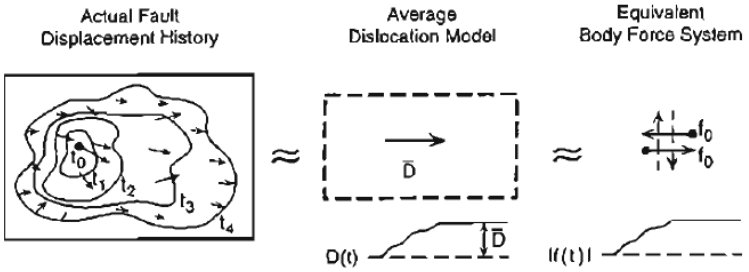


Figure 4 The actual fault and two levels of approximation to the slip vector field (Lay & Wallace 1995)

Conventionally, there are two approaches to evaluate the average slip. The first approach is to compute the average slip amount by assuming either the initial stress τ_0 or final stress τ_1 to be constant (Shimazaki & Nakata 1980). Since the strain drop in Haskell's model is proportional to $\frac{\bar{D}}{L}$ or $\frac{\bar{D}}{W}$, the static stress drop $\Delta\sigma$ can be related to the strain drop. According to Hooke's law, it can be written as

$$\Delta\sigma = C\mu\frac{\bar{D}}{\tilde{L}} \tag{2.2}$$

where C is a nondimensional constant depending on the geometry of fault, and $\tilde{L}$ is the characteristic rupture dimension and is often denoted as $\sqrt{S}$ with *S* being the fault area. The second approach utilizes empirical relations derived from the centroid moment tensor solution of hundreds of earthquakes (Schwartz & Coppersmith 1984). In this method, the slip is directly related to earthquake dimensions *L* and *W*. Two scaling laws, namely, L-model (Scholz 1982) and W-model (Rundle 1989; Romanowicz & Rundle 1993) are proposed, in which the average slip is related to the rupture length and width, respectively. These two models are better fitted in the strike-slip fault type but poorly correlated to subduction-zone earthquake and the reverse-slip type (Scholz 1982; Wells & Coppersmith 1994). Since the installation of the worldwide standard seismic network (WWSSN) was introduced in the 1960s, obtaining higher accuracy ground motion data has become possible. Based on the newly available data, Blaser *et al*.

(2010) shows that the scaling law may follow neither L- nor W-model, rather an A-model in which the slip relies on the total rupture area A.

2.2.2 Stochastic spatial slip distribution

The uniform slip model is widely used in tsunami early warning system, in which the tsunami arrival time is the biggest concern. However, this model lacks of accuracy and may lead to underestimated run-up amplitude compared with observations (Geist 1998; Geist 2005). Studies by Satake (1994) and Piatanesi *et al*. (1996) showed that the nonuniform slip with different spatial distributions highly affect the local tsunami run-up. As mentioned earlier, the slip distribution is derived from the inversion of the seismogram, tsunami waves and geodetic data (Kikuchi & Fukao 1987; Thatcher 1990; Satake & Kanamori 1991; Ozawa *et al.* 2011). The nature of spatial variance is the result of the fact that the rock layers are heterogeneous and the fault geometry is irregular. All of these features point at the stochastic behavior of the slip distribution as the high frequency phases of the ground motion is random in general (Somerville *et al.* 1999).

A large amount of papers have been published with a focus on the stochastic slip model. Haskell (1966) was among the first to employ a kinematic stochastic source model to account for the heterogeneity and the complexity of a strong ground motion. Hanks (1979) and Andrew (1980), with a similar idea, related their models to the widely-observed power laws (Aki 1967) and proposed stochastic source models based on self-similarity and friction functions. Andrew (1980) assumed a scale-invariant stress drop, implying that no matter how large or small the earthquake size is, the stress drop is constant. This assumption is referred to as dynamical similarity in some books (Koyama 1997; Kanamori & Brodsky 2004). Afterwards, Andrew (1981) formulated a stochastic model with random slip velocity function and applied it to the ground motion simulation. Based on Andrew's work, Herrero and Bernard (1994) proposed the k-squared slip distribution model in which the slip amplitude spectrum follows a k^{-2} decay (k being wave number) beyond the corner wavenumber K_c. The radiated wave field of this model is consistent with the widely observed w^{-2} spectral decay in the far-field

displacement. From then on, the k-squared model has been further developed by many scholars (Bernard *et al.* 1996; Mai & Beroza 2002; Gallovič & Brokešová 2004).

2.2.3 Hybrid stochastic k-squared static slip distribution model

Gallovic *et al*. (2004) modified the k-squared slip distribution model and introduced the 2-D corner wave numbers as $k_x^c = \mathrm{K}/L$, $k_z^c = \mathrm{K}/W$, where K is a dimensionless constant. The slip function in terms of wavenumber k_x and k_z is assumed to have spatial spectrum decay following a radically symmetric pattern. Thus, it leads to the expression as shown in Eqn. 2.3.

$$D\left(k_x, k_z\right) = \frac{\Delta\bar{u}LW}{\sqrt{1+\left(\left(\frac{k_x L}{\mathrm{K}}\right)^2+\left(\frac{k_z W}{\mathrm{K}}\right)^2\right)^2}} e^{i\Phi(k_x,k_z)} \tag{2.3}$$

in which Φ is the phase of the random spectrum, L and W are the fault length and width, respectively. The constant K affects the roughness of the generated slip distribution. A detailed discussion regarding the effect of the K value can be found in Gallovic (2004). The random generator for the phase can be either uniform or Gaussian probability distribution. The validation of the k-squared slip model with respect to the earthquake dynamics was done by Jan Burjanek *et al*. (2007) using slip weakening friction law (Hisada 2000).

As discussed in Hisada and Gallovic's studies, the original k-squared model proposed by Bernard *et al*. (1996) used a random slip model and the asperity concentrated on the center of the fault plane. This is in contradiction with the finding of Somerville *et al*. (1999), in which they concluded that larger earthquakes were likely to yield two or more asperities randomly located on the plane. Therefore, a prescribed slip distribution with constant slip for each subfault was suggested to be implemented before adding the stochastic component. This deterministic slip model could be obtained from seismic data inversion or empirical relations. After smoothing out the sudden slip change, the k-squared model is applied in such a way that the phase of random spectrum shall be added to the distribution if the

wavenumber is higher than the threshold, *i.e.* Nyquist wavenumbers K_N, given as $K_N = \sqrt{\left(\frac{1}{\Delta L}\right)^2 + \left(\frac{1}{\Delta W}\right)^2}$ and the characteristic subfault size ΔL and ΔW can be obtained from empirical relations (Mai & Beroza 2002).

2.2.4 Temporal slip distribution

As an earthquake occurs on a finite fault plane with a certain length and width, it consumes a certain amount of time for the rupture to develop and propagate on the fault plane. The temporal variation of the slip vector can be evaluated providing a prior knowledge on the source time function. Basically, the time function is the convolution of the rising time τ_r and rupture duration τ_c. Geist (1998) examined a broad range of rise time from 1s to 100s and found out that increasing rise time has the effect of decreasing the amplitude of a local tsunami as well as the maximum run-up. The rupture velocity, which is inversely proportional to the rupture duration, is comparatively larger than the tsunami propagation speed. Therefore, the effect of the rupture propagation can be neglected and an assumption of instantaneous rupture can give rise to a slightly conservative result (Yamashita & Sato 1974; Kajiura 1982). However, an exception does exist for the tsunami earthquake case, where the rupture velocity can be half of the average value of 2.5 to 3.0km/s (Somerville, Irikura et al. 1999). For example, the dislocation velocity was around 1.0 – 1.5km/s during the Nicaragua tsunami earthquake (Kikuchi & Kanamori 1995).

2.3 Tsunami wave evolution

Theoretically, a tsunami process can be divided into three phases, *i.e.* the formation of initial sea surface disturbance, wave propagation and run-up. Tremendous energy stored in a tsunami wave usually propagates at a speed of 600-900 km/h in deep ocean. As a tsunami moves towards the shoreline, the wave speed drops quickly due to the sea bottom friction, which causes the increase of the wave amplitude. The exchange between gravitational energy and kinetic energy forms the backbone of the tsunami propagation and run-up (Okal 1988).

2.3.1 Tsunami generation

In general, volcano eruption, submarine mass failure, asteroid impact and even manmade underwater explosion can trigger a tsunami. However, earthquakes, among all the potential sources, are the leading causes, given the frequency of large earthquakes near a subduction zone.

In seafloor deformation modeling, the elastic fault model is commonly adopted which treats the earth as a homogeneous, isotropic and elastic material (Steketee 1958; Okada 1985). The seafloor displacement depends on several parameters, including epicenter, focal depth, length and width of the fault plane, dislocation or slip, strike angle θ, dip angle α and rake (slip) angle λ. A schematic plot of the fault plane with relevant parameters is shown in Figure 5. The epicenter (marked as $*$ in the figure) is the earth surface projection of its focus or hypocenter, which is the initiation point of a rupture. Focal depth is defined as the vertical measure from the focus to epicenter. The slip vector describes the relative amount of movement and the direction of two sides of the fault moving towards one another. Slip angle λ describes the relative movement of the hanging wall against the footwall. For example, reverse or normal slip type refers to a slip direction of +90 degree or -90 degree, correspondingly. The angle is zero if the slip is left-lateral and 180 degrees if right-lateral slip. Dip angle δ reveals the inclination of the fault measured from the earth surface. For instance, a vertical dip represents vertically formed rupture plane. The strike angle θ is used to specify the orientation of the fault and measured clockwise regarding north. To do so is to keep the dipping always at the right side with respect to the strike direction, so as to eliminate the ambiguity.

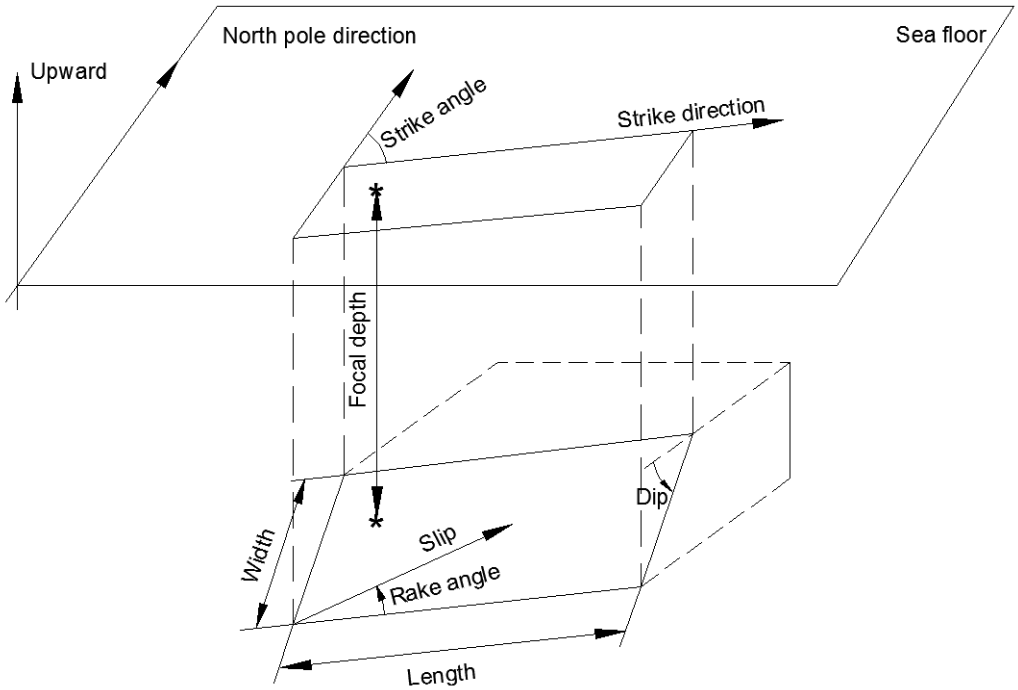


Figure 5 Illustration of fault plane and fault parameters

All the parameters mentioned above are used to calculate the initial water surface displacement. One of the most widely used analytical solutions for the surface deformation in the vertical direction (u_z) due to shear dislocation in a half space is derived by Okada (1985). The calculation of the displacement of a finite rectangular fault with length L and width W are given in Eqn. (2.5) for the strike component and Eqn. (2.7) for the dip component. The final deformation results are represented in a compact form using Chinnery's (1961) notation ||, given as:

$$f(\xi,\eta)\| = f(x,y) - f(x,y-W) - f(x-L,y) + f(x-L,y-W) \quad (2.4)$$

where (ξ, η) and (x, y) represent coordinate systems of fault plane and seafloor, respectively (Okada 1985). For the strike-slip component of the displacement, the normalized vertical displacement is

$$\frac{u_z}{U_1} = -\frac{1}{2\pi}\left[\frac{\bar{d}q}{R(R+\eta)} + \frac{q\sin\delta}{R+\eta} + I_4\sin\delta\right]\Bigg\| \quad (2.5)$$

where

$$\begin{cases} I_4 = \dfrac{\mu}{\lambda+\mu}\dfrac{1}{\cos\delta}\left[\ln\left(R+\bar{d}\right)-\sin\delta\ln\left(R+\eta\right)\right] & \text{for}\cos\delta\neq 0 \\ I_4 = -\dfrac{\mu}{\lambda+\mu}\dfrac{q}{R+\bar{d}} & \text{for}\cos\delta = 0 \end{cases} \tag{2.6}$$

For the dip-slip component of the displacement, the normalized vertical displacement is:

$$\frac{u_z}{U_2} = -\frac{1}{2\pi}\left[\frac{\bar{d}q}{R\left(R+\xi\right)}+\sin\delta\tan^{-1}\frac{\xi\eta}{qR}-I_5\sin\delta\cos\delta\right] \tag{2.7}$$

where

$$\begin{cases} I_5 = \dfrac{\mu}{\lambda+\mu}\dfrac{2}{\cos\delta}\tan^{-1}\left[\dfrac{\eta\left(X+q\cos\delta\right)+X\left(R+X\right)\sin\delta}{\xi\left(R+X\right)\cos\delta}\right] & \text{for}\cos\delta\neq 0 \\ I_5 = -\dfrac{\mu}{\lambda+\mu}\dfrac{\xi\sin\delta}{R+\bar{d}} & \text{for}\cos\delta = 0 \end{cases} \tag{2.8}$$

and in all the equations,

$$\begin{cases} p = y\cos\delta + d\sin\delta \\ q = y\sin\delta - d\cos\delta \\ \bar{d} = \eta\sin\delta - q\cos\delta \\ R^2 = \xi^2+\eta^2+q^2 \\ X^2 = \xi^2+q^2 \end{cases} \tag{2.9}$$

The deformation of the seafloor will be projected onto the mean earth surface, *i.e.* the ellipsoid surface, because the surface deformation derived above is in the plane surface. In general, a rather coarse assumption states that the initial tsunami wave form follows the vertical component of the seafloor displacement due the lack of more information.

2.3.2 Tsunami propagation and run-up

The amplitude of a tsunami generated in deep ocean is much smaller than the ocean depth. Thus the depth-averaged Navier-Stokes equations, referred as linear shallow water equations (SWE) are implemented. The SWE, including continuity and

momentum equations in two orthogonal directions, are sufficient to describe the propagation of a tsunami wave. The equations in spherical coordinates are given as:

$$\begin{cases} \dfrac{\partial \eta}{\partial t} + \dfrac{1}{R\cos\varphi}\left\{\dfrac{\partial P}{\partial \psi} + \dfrac{\partial}{\partial \psi}(\cos\varphi Q)\right\} = -\dfrac{\partial h}{\partial t} \\ \dfrac{\partial P}{\partial t} + \dfrac{gh}{R\cos\varphi}\dfrac{\partial \eta}{\partial \psi} - fQ = 0 \\ \dfrac{\partial Q}{\partial t} + \dfrac{gh}{R}\dfrac{\partial \eta}{\partial \varphi} - fP = 0 \end{cases} \tag{2.10}$$

where η and h refer to the water surface elevation and mean water depth, respectively. The water flux is given by $P = H\overline{u}$, $Q = H\overline{v}$, in which H is the total water depth and $H = \eta + h$. $\overline{u}$ and $\overline{v}$ are velocity components in the orthogonal directions. The Coriolis coefficient $f = \Omega\sin\varphi$ takes into account of the rotation of the earth where Ω is the rotation rate is. Sometimes it is preferred to express the SWE in Cartesian coordinates as follows:

$$\begin{cases} \dfrac{\partial \eta}{\partial t} + \left\{\dfrac{\partial P}{\partial x} + \dfrac{\partial Q}{\partial y}\right\} = -\dfrac{\partial h}{\partial t} \\ \dfrac{\partial P}{\partial t} + gh\dfrac{\partial \eta}{\partial x} - fQ = 0 \\ \dfrac{\partial Q}{\partial t} + gh\dfrac{\partial \eta}{\partial y} - fP = 0 \end{cases} \tag{2.11}$$

The SWE are valid under condition of $\eta / h \ll 1$. Particularly, the water depth should be greater than 50m (Nagano *et al.* 1991; Shuto 1991). When the tsunami wave approaches coastal area, the amplitude of the tsunami becomes larger and the wavelength becomes shorter. The nonlinear effect, *i.e.* convective inertia force and bottom friction terms, cannot be neglected. The following nonlinear shallow water equations (NSWE) are implemented instead. The NSWE in spherical coordinates is given as:

$$\begin{cases} \dfrac{\partial \eta}{\partial t}+\dfrac{1}{R\cos\varphi}\left\{\dfrac{\partial P}{\partial \psi}+\dfrac{\partial}{\partial \psi}(\cos\varphi Q)\right\}=-\dfrac{\partial h}{\partial t} \\ \dfrac{\partial P}{\partial t}+\dfrac{1}{R\cos\varphi}\dfrac{\partial}{\partial \psi}\left\{\dfrac{P^2}{H}\right\}+\dfrac{1}{R}\dfrac{\partial}{\partial \varphi}\left\{\dfrac{PQ}{H}\right\}\dfrac{gH}{R\cos\varphi}\dfrac{\partial \eta}{\partial \psi}-fQ+F_x=0 \\ \dfrac{\partial Q}{\partial t}+\dfrac{1}{R\cos\varphi}\dfrac{\partial}{\partial \psi}\left\{\dfrac{PQ}{H}\right\}+\dfrac{1}{R}\dfrac{\partial}{\partial \varphi}\left\{\dfrac{Q^2}{H}\right\}\dfrac{gH}{R}\dfrac{\partial \eta}{\partial \varphi}+fP+F_y=0 \end{cases} \quad (2.12)$$

and in Cartesian coordinates,

$$\begin{cases} \dfrac{\partial \eta}{\partial t}+\left\{\dfrac{\partial P}{\partial x}+\dfrac{\partial Q}{\partial y}\right\}=-\dfrac{\partial h}{\partial t} \\ \dfrac{\partial P}{\partial t}+\dfrac{\partial}{\partial x}\left\{\dfrac{P^2}{H}\right\}+\dfrac{\partial}{\partial y}\left\{\dfrac{PQ}{H}\right\}gH\dfrac{\partial \eta}{\partial x}+F_x=0 \\ \dfrac{\partial Q}{\partial t}+\dfrac{\partial}{\partial x}\left\{\dfrac{PQ}{H}\right\}+\dfrac{\partial}{\partial y}\left\{\dfrac{Q^2}{H}\right\}+gH\dfrac{\partial \eta}{\partial y}+F_y=0 \end{cases} \quad (2.13)$$

F_x and F_y represent the bottom friction in x- and y-direction and can be defined using Manning's formula as:

$$\begin{aligned} F_x &= \frac{gn^2}{H^{7/3}}P\left(P^2+Q^2\right)^{1/2} \\ F_y &= \frac{gn^2}{H^{7/3}}Q(P^2+Q^2)^{1/2} \end{aligned} \quad (2.14)$$

where n is the Manning roughness coefficient. It should be noted that the SWE and NSWE equations derived above are agreeable with the experimental data under the non-breaking tsunami wave condition. The gap between experimental data and numerical solutions cannot be ignored for breaking waves (Nistor Ioan *et al.* 2010).

2.3.3 Numerical methods

The fast development of 2-D and 3-D tsunami numerical models enables researchers and engineers to better evaluate the tsunami risk. Although for local tsunami events, the simulation time is roughly the same as the tsunami propagation duration, which hardly satisfies the early warning purpose (exceptions exist and efforts are making towards an early warning system by various means (Titov *et al.* 2005), they are valuable from the perspective of investigating the nature of a tsunami and improving the resilience of the society. Several numerical packages are

available, such as Method of Splitting Tsunam (MOST) (Titov & Gonzalez 1997), TUNAMI (Imamura 1995) and Cornell Multi-grid Coupled Tsunami Model (COMCOT) (Wang 2009), etc. The COMCOT, as one of the most widely used numerical models, is adopted in this study to simulate the tsunami generation and propagation. The SWE and NSWE are solved using the explicit Leap-Frog finite difference scheme proposed by Yee (1966). The method updates position $x(t)$ and velocities $v(t)$ at interleaved time points. In 1-D problem, the equations can be given at integer steps for velocity:

$$\begin{cases} x_i = x_{i-1} + v_{i-1/2}\Delta t \\ v_{i+1/2} = v_{i-1/2} + \dot{v}_i \Delta t \\ \dot{v}_i = F(x_i) \end{cases} \tag{2.15}$$

where x_i is position at step i and $v_{i+1/2}$is the velocity at step $i + 1/2$. The time-step Δt must be constant to maintain stability. The volume flux *P* and *Q*, as well as the water surface elevation are staggered in time and space to form a grid system as shown in Figure 6. The COMCOT allows a nested grid system for accurate results. As indicated in Figure 7, the elevation and volume flux in the sub-layer are determined by interpolation from its parent layer. The Courant condition must be satisfied with a time step $\Delta t < \dfrac{\Delta x}{\sqrt{gH_{max}}}$ where H_{max} is the maximum water depth within the region of the particular grid level. The dispersion effect due to the relatively short tsunami wavelength when approaching the coastal area cannot be neglected. Imamura *et al*. (1988) proposed a numerical dispersion term to characterize the physical dispersion by choosing the grid size $\Delta x = \sqrt{4h^2 + gh(\Delta t^2)}$.

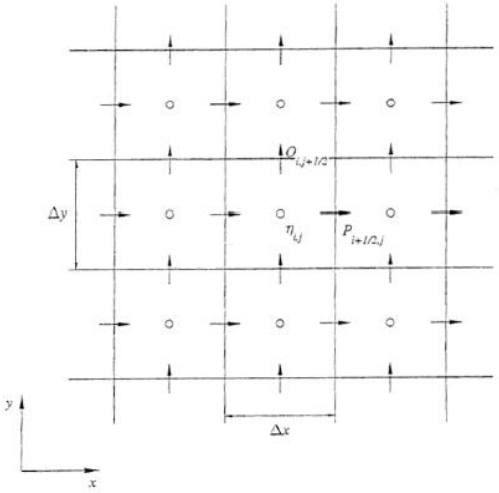


Figure 6 Grid system used in COMCOT (Wang 2009)

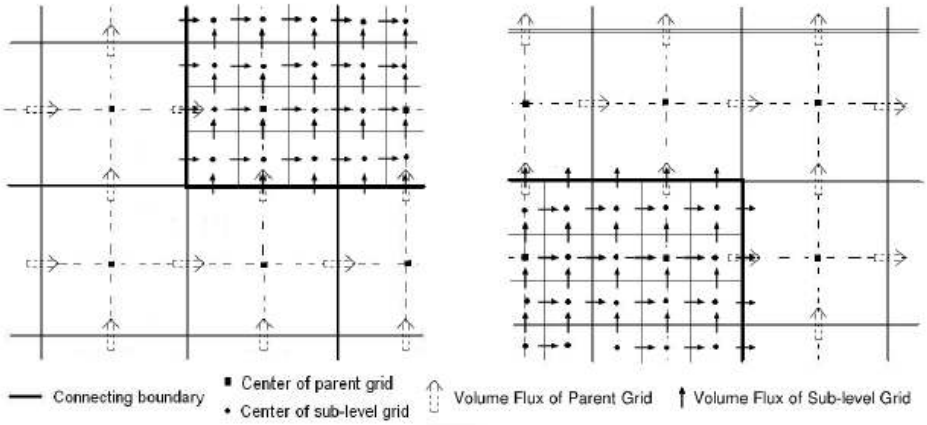


Figure 7 Multi-layer grid system schematic (Wang 2009)

2.4 Tsunami wave-structure interaction

The interaction between an incompressible fluid, e.g. tsunami wave, and a solid structure is a typical multiphysics topic that has a wide range of applications in many fields. The definition of coupled systems is adequate to capture the essence of the fluid structure interaction (FSI) process. Richter (2010) defined it as: "A coupled system S is one in which physically or computationally heterogeneous mechanical components interact dynamically. Let S1 and S2 denote the subsystems. The coupled system S is called one-way, if there is no feedback between the subsystems and two-way, if there are feedbacks between the subsystems". It is extremely difficult, if not impossible, to derive analytical solutions. Therefore, the focus was shifted towards experimental tests and numerical simulations.

2.4.1 Tsunami force quantification based on experiments and observations

The two key parameters that determine the magnitude of tsunami forces are the inundation depth and the flow velocity. The former can be estimated using the maximum water level subtracting the land elevation as indicated in Figure 8. Analytically, the wave loading consists of several types of forces, such as hydrostatic pressure, impingement force, drag force and debris impact, etc. Various empirical equations derived from laboratory tests and field investigations have been

proposed. To some extent, the equations are practical in providing instructions for constructing tsunami-resistant structures.

There are a few construction guidelines quantitatively measuring the tsunami loads acting on the structures, including Federal Emergency Management Agency Coastal Construction Manual (FEMA 55), the City and County of Honolulu Building Code (CCH), the Structural Design Method of Buildings for Tsunami Resistance (SDMBTR), proposed by Okada *et al*. (2005), and Guidelines for Design of Structures for Vertical Evacuation from Tsunamis (FEMA P646). In most of the guidelines, five types of forces are commonly considered (Nistor Ioan, Dan Palermo et al. 2010). They are: 1) hydrostatic force, F_{HS} 2) buoyant force, F_B 3) hydrodynamic (drag) force, F_D 4) surge force, F_S, and 5) impact of debris, F_i. Furthermore, the scouring during the tsunami run-up can yield the foundation deformation and lead to the collapse of the structure, although it is not considered in this paper. Among all the forces, the hydrodynamic and impact forces are the most critical ones as the hydrostatic force is much smaller than the drag and surge forces. The surge force sometimes refers to as impingement of the advancing water bore or the initial bore impact. A brief summary of the five types of force are given below.

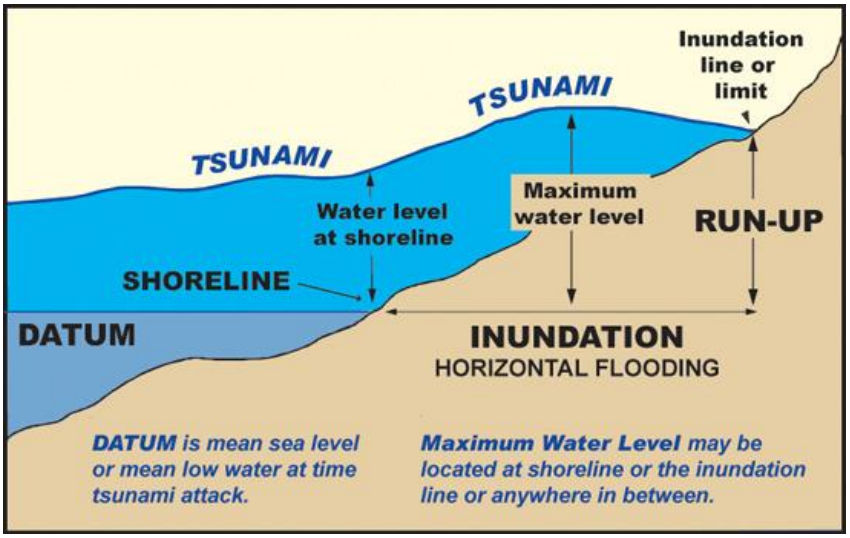


Figure 8 Conceptual schematic of inundation depth and run-up height (*Source*: *http://walrus.wr.usgs.gov/tsunami/srilanka05/measurements.html*)

2.3.1.1 Hydrostatic force

It is generated by static or slow-moving water acting on a vertical surface. The lateral hydrostatic force per unit width proposed by CCH is given as:

$$F_{HS} = \frac{1}{2}\rho g(h + u_p^2 / 2g)^2 \tag{2.15}$$

where u_p is the flow velocity normal to the structural surface. FEMA 55 does not consider the velocity head component appearing in the equation since it is negligible. The hydrostatic force in general is not comparable with the drag force and surge force when the tsunami bore hits the structure. However, the static force cannot be ignored when the tsunami acts like a rapid-rising tide (Dames & Moore 1980; Nistor Ioan, Dan Palermo et al. 2010).

2.3.1.2 Buoyant force

The buoyant force is given by:

$$F_B = \rho g V \tag{2.16}$$

where V denotes the volume of water displaced by the structure components.

2.3.1.3 Hydrodynamic (drag) force

$$F_D = \frac{1}{2}\rho C_D A u^2 \tag{2.17}$$

where C_D is the drag coefficient and A is the projected area normal to the flow direction. The tsunami bore velocity is expressed as $u = h$ in CCH and $u = 2\sqrt{gh}$ in FEMA 55. The difference in the equations results in the discrepancy in the final drag force estimated. In fact, the two velocity estimation proposed in CCH and FEMA P55 are served as the upper and lower bound of other empirical velocity calculations (Kirkoz 1983; Iizuka & Matsutomi 2000; Bryant & Nott 2001) as shown in Figure 9.

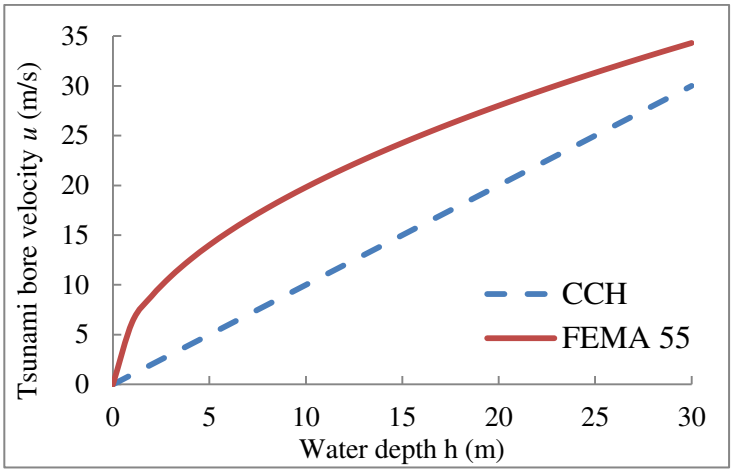


Figure 9 Comparison of tsunami bore velocities

2.3.1.4 Surge force

The surge force recommended in CCH is given by:

$$F_S = 4.5\rho g h^2 \tag{2.18}$$

Yeh & Robertson (2005) states that the value of 4.5 given in CCH is conservative and may overestimate the magnitude of the surge force. In FEMA P646, the surge force is quantified as 1.5 times of the hydrodynamic force listed in Eqn. (2.17).

2.3.1.5 Debris impact force

The debris impact force is consistent in both FEMA P55 and CCH as follows:

$$F_i = m_i \frac{u_i}{\Delta t} \tag{2.19}$$

where m_i is the debris mass, u_i denotes the debris impacting velocity, and Δt is the interaction duration between the debris and the structure. The determination of Δt relies on the impacting materials and the differences between FEMA P55 and CCH reside in the empirical impacting duration estimation. In FEMA P646, the impact force is given as:

$$F_i = C_m u_i \sqrt{mk} \tag{2.20}$$

where C_m is the added mass coefficient and m and k are the mass and the effective stiffness of the debris, respectively.

2.4.2 Numerical methods

A laboratory test is straightforward and provides reliable results. It is widely used to investigate the basic aspects of fluid-structure interaction. However, experiments are often limited in scale and the boundary conditions, which make it difficult to simulate complex problems. Therefore, numerical methods are mostly adopted. Numerous advanced algorithms were developed, aiming at solving the interaction problems for the coupled system.

2.4.2.1 General notes

As mentioned earlier, the FSI problem involves two or more subsystems. We can consider a computational domain Ω with external boundary Γ. A prototype problem may contain a fluid domain Ω_f, a structural domain Ω_s, and a common boundary Γ_s defined as $\Gamma_s = \Omega_f \cap \Omega_s$. The concept is illustrated in Figure 10. The FSI problem is formulated on moving domains, indicating both the fluid domain and structural domain are changing with time although the total computational domain remains unchanged, *i.e.* $\Omega = \Omega_f(t) \cup \Omega_s(t)$.

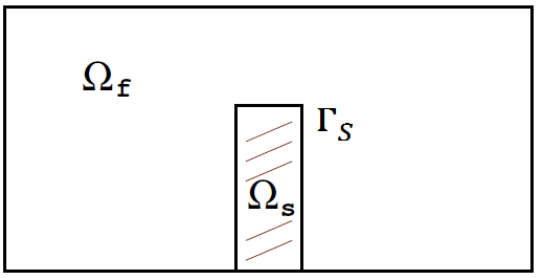
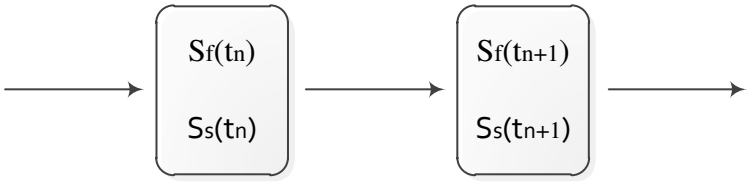


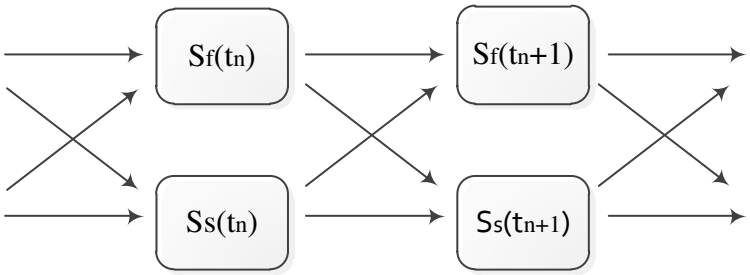
Figure 10 Schematic fluid-structure interaction domains

Numerical methods for the FSI are grouped into two categories globally, namely monolithic and partitioned methods. In the first approach, the fluid and structural domains are solved simultaneously under an integrated mathematical framework. The interaction on the boundary Γ_s is computed in an implicit manner. The monolithic approach theoretically can achieve a better accuracy but requires a large amount of computational resources. Besides, it is difficult to maintain a robust numerical code. Nevertheless, one can decouple the interaction process and formulate the problem using the partitioned approach. As indicated from its name,

the partitioned approach solves the two domains separately using different algorithms. The interaction thus is evaluated explicitly. The most desirable advantage of this method is that there are many well-established algorithms for a complex system in both fluid and structural domains. However, the tradeoff is that a small time step resulting from the stability consideration requires a large number of iterations. Another difficulty is to keep track of the interface location since the two domains use entirely different algorithms. Both of the approaches are demonstrated in Figure 11 below. S_f and S_s denote the solutions for the fluid and structure, respectively. Most of the available numerical packages dealing with FSI adopt partitioned algorithms.



(a) Monolithic approach



(b) Partitioned approach

Figure 11 Monolithic and partitioned approach for the coupled system

2.4.2.2 Governing equations of continuum mechanics

Generally speaking, the governing equations for both fluid domain and structural domain follow the D'Alembert's principle which gives:

$$\rho \dot{v}_i - \sigma_{ij,j} + f_i = 0 \tag{2.21}$$

where $\dot{v}_i$ is the time derivative of velocity field given by $\dot{v}_i = \frac{\partial^2 u_i}{\partial t^2}$ (u_i being the displacement), $\sigma_{ij,j}$ is the internal stress tensor and f_i denotes the body forces and often written as ρg_i. The first two terms in Eqn. (2.22) have different expressions in fluid and structural domains.

In the fluid domain, according to the conservation of mass and momentum in fluid dynamics, the inertia term can be written as:

$$\dot{v}_i = \frac{Dv_i}{Dt} = \frac{\partial v_i}{\partial t} + v_j \nabla v_i \tag{2.22}$$

Further assumption of incompressible Newtonian fluid leads the tensor stress into the following form,

$$\sigma_{ij}^f = -p\delta_{ij} + \tau_{ij} \tag{2.23}$$

where $\tau_{ij} = \mu(\nabla v_i + \nabla v_j)$ and μ is the fluid dynamic viscosity. However, the same term is represented in a different form for structural domain. For example, the structural stress tensor following Hook's law is given by:

$$\sigma_{ij}^s = \lambda\delta_{ij}\varepsilon_{kk} + 2G\varepsilon_{ij} \tag{2.24}$$

where $\varepsilon_{ij} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right)$, $G = \frac{E}{2(1+\nu)}$, $\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}$ (E and ν are Young's modulus and Poisson ratio, respectively) when assuming a linear elastic material.

The structural domain and fluid domain problems are often described with the Eulerian and the Lagrangian formation, respectively. The nodes in the Lagrangian meshing coincide with the structure element nodes, and the displacement of the structure corresponds to the moving of the Lagrangian meshing. The boundary conditions are easy to impose, because the edge of the Lagrangian description represents the total structural domain. This feature leads to the wide application of the Lagrangian formulation in solving structural problems. However, the large deformation of the structure leads to severe distortion of the meshing, thus causing instability and unrealistic results. Therefore, for fluid domain that often experiences sudden and sharp deformation, it hardly uses the Lagrangian description. Instead, the Eulerian formulation is generally utilized to describe the fluid domain. The fluid domain under the Eulerian description deforms and moves within a space-fixed meshing. In such a case, a convection term is introduced to account for the fluid material moving in and out from one element to adjacent elements in each time step.

2.4.2.3 ALE formulation

Both the Lagrangian and the Eulerian description have limitations when dealing with FSI problems. First of all, all-Lagrangian meshing is inappropriate due to the considerable displacement of the fluid material. Secondly, if the fluid domain is described by the Eulerian formulation while, the structural domain the Lagrangian meshing, a large amount of fluid material advection may present at the interaction boundaries, which causes inaccurate numerical results and a smaller time step, thus a longer computation time is required. Arbitrary Lagrangian Eulerian (ALE) formulation is designed to tackle this problem. Unlike the pure Eulerian formulation where the meshing is fixed, the ALE meshing allows a certain degree of movement as the simulation progresses. The advection of the fluid material still exists but much less than the Eulerian description, indicating potentially more accurate results. In fact, the Eulerian formulation is a special case of the ALE in which the moving velocity of the meshing is zero. Different approaches used to formulate the problem are compared in Figure 12.

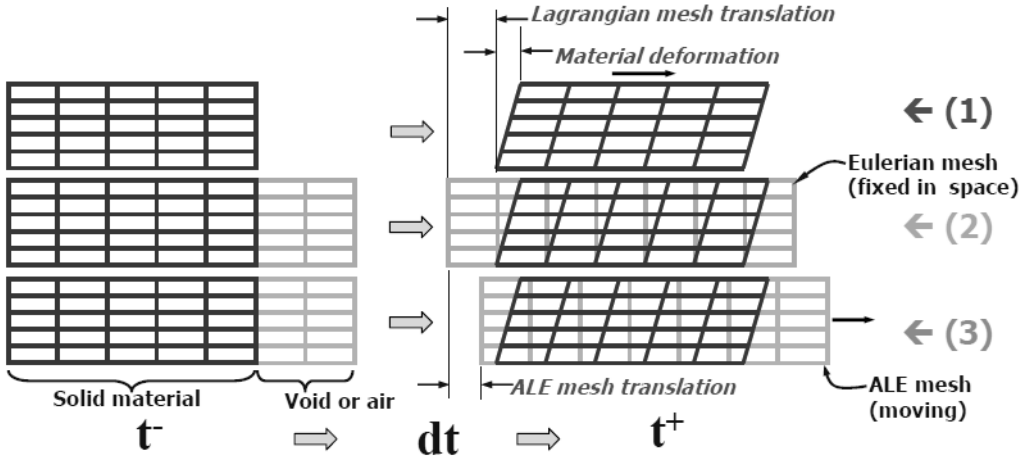


Figure 12 ALE approach description (Haritos *et al.* 2005)

The governing equation of the ALE formulation is similar to Eqn. (2.22) except for the introduction of the convection term. The equation is expressed as follows:

$$\rho\dot{v}_i + \rho(v_j - v_j^m)v_{i,j} - \sigma_{ij,j} + f_i = 0 \qquad (2.25)$$

where v_j and v_j^m represent, respectively, fluid velocity and meshing velocity. If $v_j^m = 0$, we obtain again, the Eulerian description as the convection of meshing is

diminished. On the other hand, the Lagrangian formulation can be derived if $v_j = v_j^m$. Consequently, $v_j - v_j^m$ stands for the relative movement between the fluid and the ALE meshing.

2.5 Classical Computational Methods of Reliability Integrals

Uncertainty and randomness are inherent in engineering problems and the deviation from the "most likely" value is unavoidable. Structure performance can vary dramatically given a slight change of the input variables and initial conditions. In present study, additional challenge is added to provide the relatively small occurrence rate of tsunami hazards. Structural reliability provides a rational probabilistic measure to quantify the structural system performance when the structural system interacts with the dynamic tsunami waves. The failure probability (a compliment to reliability) can be written in terms of a limit state function $L(\boldsymbol{\theta})$ given by:

$$P_F = P\left(L(\boldsymbol{\theta})<0\right) = \int I_F\left(L(\boldsymbol{\theta})<0\right) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{2.26}$$

where $\boldsymbol{\theta} \in \mathbb{R}^n$ represents the vector of uncertain variables with joint probability density function (PDF) $p(\boldsymbol{\theta})$ of the random variables. The indicator function $I_F = 1$, if $L(\boldsymbol{\theta})<0$ and $I_F = 0$ otherwise.

The structural reliability can be estimated by two measures: safety index and failure probability (Choi *et al.* 2006). A general solution of the integral is intricate and sometimes impossible. Hence, approximation and simulation methods are implemented. In the following section, we firstly discuss the first-order and second-order reliability methods (FORM/SORM). The ordinary Monte Carlo simulation (MCS) is quickly reviewed since it is a well-known standard procedure with applications in almost all the disciplines. An introduction of Monte Carlo Markov chain (MCMC) method is presented followed by Metropolis-Hasting (M-H) algorithm and a detailed discussion of Subset Simulation method including the modified M-H algorithm.

2.5.1 First-order and second-order reliability methods

The FORM was first proposed by Hasofer and Lind (1974) on the basis of the first-order second moment method (FOSM) (Cornell 1969) with an idea of most probable point (MPP) on the linearized limit state function $L(X) = 0$. After transformation from the original correlated random variable space into an independent standard normal space U, the reliability index β is written as:

$$\beta = \min_{u \in F} (u^{\mathrm{T}} u)^{1/2} \tag{2.27}$$

where F denotes the failure domain and u be standard normal random variables. The reliability index represents the algebraic distance between the origin and the linearized limit state surface. Thus, the failure probability can be obtained from the linearized limit state surface at MPP u^*:

$$P_F \cong \Phi_{\mathrm{R}}(-\beta) \tag{2.28}$$

where Φ_{R} is the standard normal cumulative distribution function. u^* actually is the point on the limit state surface closest to the origin and derivation the of u^* requires an iterative algorithm (Ang & Tang 1984; Choi, Grandhi et al. 2006).

In the case where the first-order linearization of the limit state surface is inadequate to capture the nonlinearity of the limit state function $L(\boldsymbol{\theta})$, the second-order reliability method (SORM) is used to enhance the accuracy of the approximation by expanding the Taylor Series up to the second order.

2.5.2 Monte Carlo simulation

The Monte Carlo simulation (MCS) is a probabilistic approach that simulates possible scenarios using a large number of random numbers. It can be applied to different problems and give a complete approximation of the distribution. Each parameter of the system is computed by using random samples drawn from its predefined distribution. In reliability problems, MCS calculates the expectation of the indicator function I_{F} by computing the ratio of the failure samples over the total sample size. The MCS estimator for P_F can be expressed as:

$$P_F = P(F) = \frac{1}{N} \sum_{k=1}^{N} I(\theta_k \in F) = \frac{N_F}{N} \tag{2.29}$$

where θ_k is independent, identically distributed (i.i.d.) input variable drawn from its PDF $p(\boldsymbol{\theta})$. N_F is the number of samples lying in the failure region F. The accuracy of the MCS is measured by its variance given as:

$$Var(P_F) = Var(\frac{N_F}{N}) = \frac{P_F(1-P_F)}{N} \tag{2.30}$$

Since the samples are independent and identically distributed, the estimation is unbiased. The coefficient of variation (c.o.v.) δ can be measured by calculating the ratio of the standard deviation to the mean of P_F:

$$\delta = \frac{\sqrt{Var(P_F)}}{P_F} = \sqrt{\frac{1-P_F}{NP_F}} \tag{2.31}$$

The MCS is the most robust stochastic simulation method which can be used to evaluate the integral appearing in Eqn. (2.26). However, when estimating small failure probability (*i.e.*, P_F <<1), MCS is inefficiency as it requires at least $1/P_F$ samples simulated according to $p(\boldsymbol{\theta})$.

2.5.3 Markov chain Monte Carlo simulation method

The MCMC was invented soon after the ordinary MCS by Metropolis *et al*. (1953) when studying the phase equilibrium equation. It was applied to Bayesian inference problems after 80s with the usage of Gibbs sampler (Geman & Geman 1984) which sometimes refers to as data augmentation after Tanner & Wong (1987), though the theory of the MCMC was not fully understood till early 90s (Gelfand & Smith 1990; Geyer & Thompson 1992). Ordinary MCS generates samples which are independently and identically distributed whereas the MCMC proposes stationary and reversible samples according to the Markov chain. This considers as the most different feature between the two, and in fact MCS can be considered as a special case of MCMC.

The Markov chain Monte Carlo (MCMC) simulation method tends to generate failure samples more efficiently than MCS since the proposal distribution can be chosen to do local random walk, instead of the i.i.d. sample. Essentially, the MCMC is an integration of Monte Carlo and Markov chain process. Therefore a brief introduction of Markov chain is necessary before reviewing the properties of the

MCMC method. Let X_1, X_2…denote a sequence of random variables, the set of variables are Markov chain if the conditional probability of x being in present state X_{n+1} only depends on its previous state and irrelevant to its future or past states. Formally,

$$P\left(X_{n+1}=x|X_1=x_1,X_2=x_2,\ldots,X_n=x_n\right)=P\left(X_{n+1}=x|X_n=x_n\right) \quad (2.32)$$

The conditional probability of X_{n+1} given X_n is called transitional probability. In practical MCMC problem, the states are uncountable. Therefore, the initial probability and transitional probability are treated as unconditional probability distribution and conditional probability distribution, respectively. Two important features of Markov chain in developing MCMC are introduced below:

- A probability distribution π on X is a stationary distribution or invariant distribution if

$$\sum_{i\in X}\pi_i P\left(X_{n+1}=j|X_n=i\right)=\pi_j \quad (2.33)$$

  where $\pi_i=P(X_n=i)$.

- The transitional distribution of a Markov chain is reversible with respect to an initial distribution if

$$\pi_i P\left(X_{n+1}=j|X_n=i\right)=\pi_j P\left(X_{n+1}=i|X_n=j\right) \quad (2.34)$$

  The reversibility property of Markov chain, also called detailed balance for the transitional distribution, implies stationarity, but not vice versa.

The basic idea of the MCMC is to construct a conditional distribution, e.g. $q(\bullet|F_i)$ for drawing samples that can be considered as the next state of the Markov chain. Eventually the stationary distribution π of the Markov chain is, in fact, the target distribution to generate sample from (Au & Beck 2001). Many algorithms were developed, including Metropolis-Hasting (M-H), Gibbs sampling, slice sampling, Hamiltonian Monte Carlo (HMC), and many others (Andrieu *et al.* 2003).

Metropolis-Hastings algorithm

The M-H method is the most widely applied MCMC algorithm. It is mainly used to (1) calculate the expectation of complicated distributions and (2) generate samples

from an unknown distribution or a distribution known up to a normalizing constant which typically appears in Bayesian posterior distribution. The M-H algorithm is a two-stage process. First of all, a candidate sample is generated following a proposed conditional distribution. Secondly, the candidate is either accepted or rejected as the next state of the sample with prescribed criteria. Specifically speaking, the M-H method does the following:

- Generate candidate sample θ^* having conditional probability density given θ_k as $q(\theta^*|\theta_k)$. Suppose the target distribution is given as $p(\cdot)$, then the acceptance ratio is:

$$r^* = \frac{p(\theta^*)q(\theta_k \mid \theta^*)}{p(\theta_k)q(\theta^* \mid \theta_k)} \tag{2.35}$$

- Accept the candidate sample θ_{k+1} with probability $A = \min\left\{1, r^*\right\}$ and reject it and keep the current sample θ_k as the next sample (*i.e.* $\theta_{k+1} = \theta_k$) with probability of $1 - A$.

The first step can be considered as a local random walk around the current sample θ_k. The second step acts as "pushing" the proposal distribution towards the region where the probability density of the target distribution is higher. Therefore, given the satisfaction of the conditions for Markov chain (Gelman & Shirley 2011), the simulated sample shall converge to the target distribution after a certain number of iterations though the initial samples may reside far from the region with high probability center of the target distribution.

Gibbs sampler

As mentioned earlier, the proposed samples in the M-H algorithm may be rejected which leads to extra computation. An alternative is the Gibbs sampler which always accepts the proposed samples (Geman & Geman 1984; Gilks 2005). In other words, the acceptance ratio for Gibbs sampler is 1. However, one needs to provide an analytical expression of the conditional distribution of each variable given the rest can be drawn readily from random samples. Equivalently, given a D-dimensional target distribution $p(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_D]^{\mathrm{T}}$, one must provide the conditional probability $p(\boldsymbol{\theta}_i \big| \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \ldots, \boldsymbol{\theta}_D) = p(\boldsymbol{\theta}_i \big| \boldsymbol{\theta}_j),\ j \neq i$. The

implementation of the Gibbs sampler is almost the same as the M-H algorithm. In fact, sampling from the full conditional distribution in the Gibbs sampler is a special case of the M-H technique. Thus, the acceptance ratio now can be written as:

$$r_i = \frac{p(\theta_i^* \,|\, \{\boldsymbol{\theta}\}_{\neq i})\, p(\theta_i \,|\, \{\boldsymbol{\theta}\}_{\neq i})}{p(\theta_i \,|\, \{\boldsymbol{\theta}\}_{\neq i})\, p(\theta_i^* \,|\, \{\boldsymbol{\theta}\}_{\neq i})} = 1 \tag{2.36}$$

where $p(\theta_i \,|\, \{\boldsymbol{\theta}\}_{\neq i})$ is both the target and proposal PDF, and $\{\boldsymbol{\theta}\}_{\neq i} = [\theta_1, \theta_2, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_D]^{\mathrm{T}}$. Therefore the proposed sample will always be accepted. One drawback of the Gibbs sampler is the high correlation between the samples due to the conditional probability used to generate new samples.

2.5.4 Importance sampling

The importance sampling is a variance-reduction version of the original MC method. The basic idea of the method is to produce samples efficiently in the important region so as to reduce the variance of estimator. In reliability aspect, it means to generate samples that lie in the failure region more frequently (Engelund & Rackwitz 1993; Katafygiotis & Zuev 2008). Specifically, the failure probability is rewritten as:

$$\begin{aligned} P_F &= \int I_F(\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &= \int \frac{I_F(\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} q(\boldsymbol{\theta})\, d\boldsymbol{\theta} = E\left(\frac{I_F\, p}{q}\right) \end{aligned} \tag{2.37}$$

Theoretically, the failure probability is estimated by the sample mean estimator similar to the one in MCS, which is given as:

$$P_F = \frac{1}{N} \sum_{k=1}^{N} \frac{I_F(\boldsymbol{\theta}^{(k)})\, p(\boldsymbol{\theta}^{(k)})}{q(\boldsymbol{\theta}^{(k)})} \tag{2.38}$$

in which $q(\boldsymbol{\theta})$ denotes the important sampling density (ISD). One can also see that MCS is a special case of importance sampling where $q(\cdot) = p(\cdot)$. The most critical challenge in importance sampling is to construct ISD so that $\boldsymbol{\theta}^{(k)} \sim q(\cdot)$. The optimal ISD is given by the conditional PDF: $q_{opt}(\boldsymbol{\theta}) = I_F(\boldsymbol{\theta})\, p(\boldsymbol{\theta}) / P_F = p(\boldsymbol{\theta} \,|\, F)$. However, this is practically infeasible since the optimal ISD requires samples from the failure region which is to be estimated.

In general, there are two ways to provide ISD which covers the important regions in the failure domain. The first one is to characterize the region using design points (Shinozuka 1983). These points contain the highest probability density compared to the rest of the points. However, this method requires a considerable amount of time in solving a constrained optimization problem, especially in the case with a multi-modal probabilistic distribution (Melchers 1989; Liu & Der Kiureghian 1991). Alternatively, the kernel density estimator which does not require design points was adopted in many studies (Bucher 1988; Ang *et al.* 1992). It makes use of the pre-samples lying in the failure region to form a mixture of probability distribution. Thus, the more representative the samples, the better the kernel density estimator. Theoretically, the kernel density can take up any shape according to the samples thus it is possible to propose the distribution similar to the optimal ISD. The drawback of the kernel density estimator is that it suffers the *curse of dimensionality* as the performance deteriorates exponentially with the increasing number of random variables (Au & Beck 2003).

2.5.5 Subset Simulation method

The MCS method can be adopted to evaluate the integral appearing in Eqn. (2.26). Conversely, as mentioned earlier, the problem arises as MCS requires about $10/P_F$ number of simulations to capture small probability events (e.g. $P_F < 10^{-3}$ with associated c.o.v. of P_F estimator of 30%). The Subset Simulation (SS) method proposed by Au & Beck (2001) reduces the computation burden significantly by evaluating a sequence of intermediate failure events F_i such that $F_1 \supset F_2 \supset \cdots \supset F_m = F$. Thus, the computation of rare failure event probability is converted to the problem of assessing several relatively frequent events which are conditioned on a previous failure event. In other words, the computation of the rare event problem is equivalent to computing a product of conditional probabilities $P(F_i | F_{i-1})$ given as:

$$
\begin{aligned}
P_F &= P(F_m) \\
&= P(F_m|F_{m-1})\, P\left(\bigcap_{i=1}^{m-1} F_i\right) \\
&= P(F_1) \prod_{i=2}^{m} P(F_i|F_{i-1})
\end{aligned}
\tag{2.39}
$$

and

$$
P(F_1) \cong \frac{1}{N} \sum_{k=1}^{N} I_{F_1}(\boldsymbol{\theta}_k \in F_1) \tag{2.40}
$$

where the initial failure probability $P(F_1)$ can be readily estimated more accurately using direct MCS using the samples $\boldsymbol{\theta}_k$, $k = 1,\ldots, N$, simulated according to $p(\boldsymbol{\theta})$ since the failure probability for the first level is relatively high. The conditional probabilities $P(F_i|F_{i\text{-}1})$ are derived using samples generated by the procedures described as follows:

$$
P(F_i \mid F_{i-1}) \approx \frac{1}{N} \sum_{k=1}^{N} I(\theta_k \in F_i) \tag{2.41}
$$

where θ_k follows $p(\theta \mid F_{i-1})$, $k = 1,\ldots,N$.

The general steps of the component-based M-H algorithm (modified M-H, MMH) proposed by Au and Beck (2001) are summarized as follows:

- Let $j = 1,2,\ldots,n$ denote the number of uncertain parameters (random variables) and for each parameter group generate a candidate sample $\boldsymbol{\xi}_{k+1}^{j}$ following a pre-selected proposal PDF $q_j\left(\bullet \middle| \boldsymbol{\theta}_k^j\right)$ where $\boldsymbol{\theta}_k^j$ is the j-th parameter group of the k-th sample following $P(\boldsymbol{\theta}|F_{i\text{-}1})$. Compute the acceptance ratio as:

$$
r_{k+1}^{j} = \frac{p^j\left(\xi_{k+1}^{j}\right) q_j\left(\theta_k^j \middle| \xi_{k+1}^{j}\right)}{p^j\left(\theta_k^j\right) q_j\left(\xi_{k+1}^{j} \middle| \theta_k^j\right)} \tag{2.42}
$$

  Set $\boldsymbol{\theta}_{k+1}^{j} = \boldsymbol{\xi}_{k+1}^{j}$ with probability of $\min\left\{1, r_{k+1}^{j}\right\}$ and $\boldsymbol{\theta}_{k+1}^{j} = \boldsymbol{\theta}_k^j$ with a probability of $1 - \min\left\{1, r_{k+1}^{j}\right\}$.

- Accept $\boldsymbol{\theta}_{k+1}^{j}$ as the next sample $\boldsymbol{\theta}_{k+1}$ following $P(\boldsymbol{\theta}|F_{i\text{-}1})$ if $\boldsymbol{\theta}_{k+1} \in F_i$, otherwise reject it and set $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$.

The greatest improvement of MMH over M-H is that some component groups of Markov chain samples in MMH are able to change which avoid high correlation between current and next samples. This feature is more critical when a large number of random variables are involved. The c.o.v. δ_i of any intermediate level $P(F_i|F_{i\text{-}1})$ derived by Au and Beck (2001) is given as:

$$\delta_i = \sqrt{\frac{(1-P_i)}{NP_i}}(1+\gamma_i) \tag{2.43}$$

where

$$\gamma_i = 2\sum_{k=1}^{N/N_c-1}\left(1-\frac{kN_c}{N}\right)\rho_i(k) \tag{2.44}$$

and

$$\rho_i(k) = R_i(k)/R_i(0) \tag{2.45}$$

where ρ_i is the correlation coefficient of the stationary sequence at lag k. The covariance sequence is estimated by:

$$R_i(k) \approx \tilde{R}_i(k) = \left(\frac{1}{N-kN_c}\sum_{j=1}^{N_c}\sum_{l=1}^{N/N_c-k} I_{jl}^{(i)} I_{j,l+k}^{(i)}\right) - \tilde{P}_i^2 \tag{2.46}$$

where N_c, $I_{jl}^{(i)}$ and $I_{j,l+k}^{(i)}$ are defined as in Au and Beck (2001). Further, if the intermediate failure probabilities $P(F_i|F_{i\text{-}1})$ are assumed to be uncorrelated, the overall c.o.v. of the P_F estimator is:

$$\delta = \sqrt{\sum \delta_i^2} \tag{2.47}$$

In practice, using samples distributed according to $p(\boldsymbol{\theta}|F_{i\text{-}1})$ and the value of the limit state function $L(\boldsymbol{\theta})$ corresponding to each sample, the intermediate threshold b_i corresponding to F_i, $i = 1, 2\ldots m$, can be chosen adaptively such that the conditional probability $P(F_i|F_{i\text{-}1}) = p_0$. A study done by Zuev *et al*. (2012) has suggested a conditional probability p_0 between [0.1, 0.3] and the acceptance ratio between 30% and 50% to achieve an optimal performance for SS.

2.6 Concluding remarks

This chapter reviews the physics of tsunami generating mechanism, as well as the propagation and run-up. The computational method of the fluid-structure interaction is also introduced. Lastly, several computational methods of reliability assessment are discussed.

The majority of tsunamis are triggered by under seafloor earthquakes. Both kinematic and dynamic models are conducted by researchers in order to better understand the underlying mechanism. The two models are not isolated from each other. In fact, they are interconnected as stated by Kanamori and Brodsky (2004) as follows:

$$G_c = \frac{1}{2}\Delta\sigma D_c \tag{2.48}$$

where G_c is the fracture energy required to create new crack surfaces, $\Delta\sigma$ refers to the static stress drop, and D_c is the critical slip-weakening distance.

Although the constant slip model has been used for the past several decades, the variation in the slip distribution produces a more accurate tsunami model given the fact that frictions at different locations of the rupture surface are heterogeneous. With the advancement in computational simulation, stochastic slip distribution models are frequently adopted. The hybrid stochastic k-squared static slip distribution model can preserve the original shape of the deterministic slip distribution while at the same time being more realistic compared with the constant slip model.

The risk assessment component comes into the picture when the tsunami profile, characterized into different types of forces, smashes onshore structures. The deformation of the structures can be splendid and therefore leaves a challenging task for risk mitigation in the aftermath. The wave-structure interaction involved is a multiphysics process. The ALE approach which combines the best features of Eulerian and Lagrangian descriptions is the most suitable technique for modeling

the interaction process. An imaginary mesh with velocity varying between the one of fluid particle and stationary keep the deformation in manageable level.

The FORM and SORM methods are still among the most popular techniques in structural reliability computations due to their simplicity. However, a major issue with FORM and SORM is that the linearized and second order limit state functions cannot be adopted to represent complex systems since the true limit state function cannot be explicitly expressed.

The MCS does not require any mathematical simplifications and considers the entire probability domain for simulation. The statistical features are well studied. As one of the most common numerical stochastic approximation algorithms used worldwide, it can solve a wide range of problems involving any types of response functions explicitly or implicitly. As stated earlier, a potential issue with MCS is that the number of simulations required is inversely proportional to the associated failure probability. This leads to a considerable often prohibitive amount of computational time. Furthermore, the convergence rate of MCS is rather slow ($1/\sqrt{N}$).

The MCMC method is currently the most general technique in generating samples from any type of posterior probability distributions. The popularity of the MCMC method revitalized the Bayesian statistics. Two of the well-known MCMC algorithms are M-H and Gibbs sampler. The M-H algorithm is more general given the fact that Gibbs sampler can be slow in exploring the space when the dependence of the component groups is strong (Gilks & Roberts 1996). The biggest short-coming of MCMC is that the samples are correlated due to the use of Markov chains in the proposal step. Importance sampling alleviates the correlation issue by incorporating an adaptive kernel density estimator since the samples are independent and identically distributed. However, constructing an accurate important sampling density (ISD) is challenging, especially in high dimensional case.

The basic idea of the Subset Simulation is straightforward: to partition the rare probability event into a sequence of more frequent failure events. The MCMC-based algorithms are adopted in each intermediate failure level of the SS except for the MCS sampling in the first level. Instead of using the original M-H algorithm, MMH is proposed, which to some extent eliminates the inapplicability of M-H in higher dimensional problems. However, it adds another layer of complexity in order to check if the proposed samples are located within the failure space. More often than not, the numerical computation required in the last step can be time demanding.

Chapter 3 Proposed Tsunami Wave-structure Interaction Model

3.1 General remarks

The structural response under tsunami wave impact is illustrated using Minami Gamou Wastewater Treatment Plant, a three-story concrete structure located at Sendai coastline. The actual structure undergone severe damage and failed through out-of-plane flexure due to the tsunami impact as shown in Figure 13. In this study, the structure model is simplified with properly defined boundary conditions. The maximum displacement during the wave-structure interaction is considered as the output variables denoted as $z(\theta)$ where θ denotes the input variables, e.g. wave velocity and wave height.

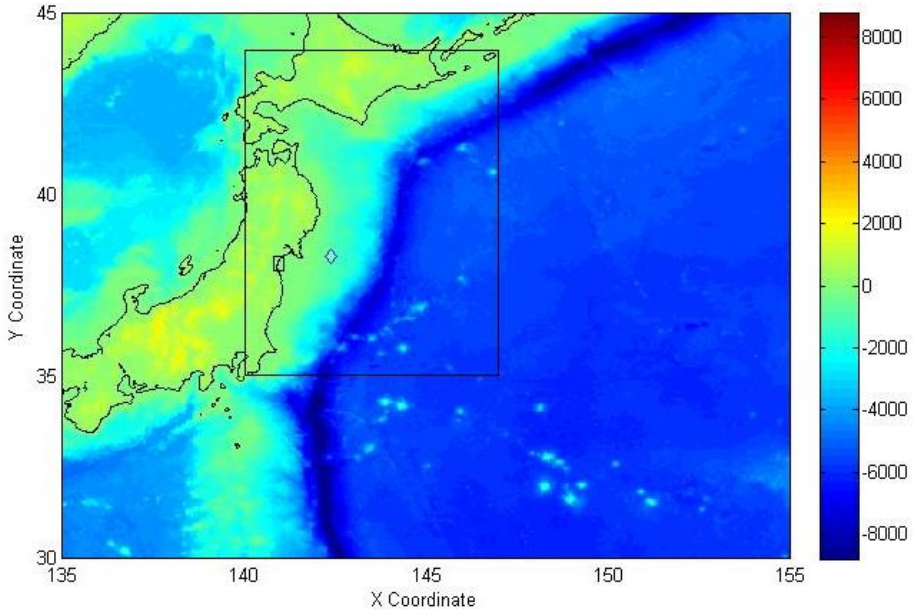
Figure 13 Minami Gamou Wastewater Treatment Plant out-of-plane flexure damage (a) exterior and (b) interior (Carden *et al.* 2012)

3.2 Tohoku tsunami modelling

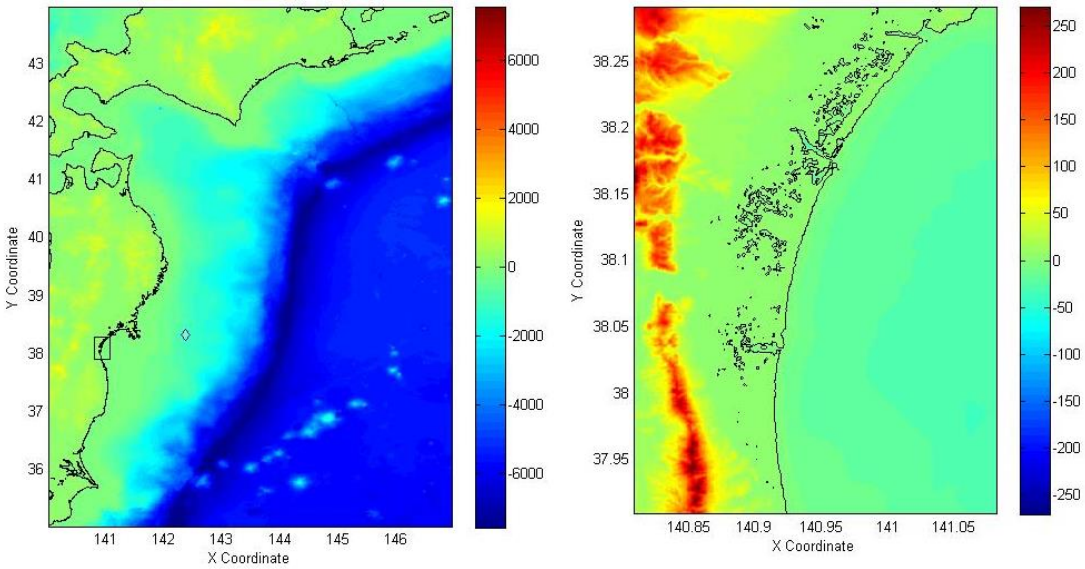
After the Tohoku tsunami in 2011, numerous studies have been conducted in order to investigate both the earthquake source mechanism (Fujii & Satake 2007; Ozawa, Nishimura et al. 2011; Atsushi Nozu 2013; Breanyn *et al.* 2013), and local & far-field tsunami propagation with energy dissipation (Ozaki 2011; Grilli *et al.* 2012). Site surveys for the structural damage evaluation were also carried out under groups of experts (Chock *et al.* 2012; Yim *et al.* 2012). The existing studies mentioned above are utilized to quantify the earthquake source parameters. The generated tsunami profiles are compared with the observed data.

3.2.1 Tsunami model

The earthquake source parameters are input into the numerical software, COMCOT, to generate tsunami amplitudes and volume flux. The Tohoku tsunami event in 2011 is modelled to evaluate the adequacy of the bathymetric and topographic data used in this study. Three levels of nested grids are used so as to accurately simulate the tsunami run-up. The resolution of the first layer (layer01) is 2 arc-minutes (≈3,700m) with nested grids of 0.5 arc-minutes (≈1,000m) and 0.05 arc-minutes (≈100m), respectively, for the sub-layers (layer02 and layer03). Figure 14 shows the three nested grids in the computational domain and the epicenter of the Mw 9.1 Tohoku earthquake is denoted in diamond shape in layer01 and layer02. The bathymetric data for the first layer is obtained from ETOPO2 of National Geophysical Data Center (NGDC). The second layer dataset is provided by GEBCO_08 grid data. In order to acquire higher resolution topological data for Sendai region, Shuttle Radar Topography Mission (SRTM 90) data covering digital elevation model above mean sea level, is combined with the interpolated GEBCO_08 grid data using ArcGIS. The detailed information about the nested grid data is provided in Table 3.



(a) Layer 01



(b) Layer 02 (c) Layer 03

Figure 14 Bathymetry and topology of the three nested grids used in the simulation

Table 3 Summary of COMCOT nested grid setting

	Layer01	Layer02	Layer03
Grid cells (Lon ×Lat)	601*451	836 * 1076	330*460
Lat. Range [deg]	30.000-45.000	35.021-43.979	37.909-38.291
Lon. Range [deg]	135.000-155.000	140.021-146.979	140.800-141.083
Grid size	2.0	0.5	0.05
Parent grid	-	01	02
Grid ratio	-	4	10
Time step	1.00	0.25	0.025
Coordinates	Spherical	Spherical	Spherical
Governing equation	Linear SWE	Linear SWE	Nonlinear SWE
Manning coefficient	-	-	0.035

3.2.2 2011 Tohoku earthquake source parameters

As mentioned above, a number of studies are published regarding the 2011 Tohoku earthquake source mechanism. Four source mechanism models, developed by Hayes (2011), Shao *et al*. (2011), Fujii *et al*. (2011), and Wei *et al*. (2012), are chosen as candidate models. Initial water surface displacement and generated maximum inundation maps in Sendai Plain are compared for all the models. In

addition, Global Centroid Moment Tensor (GCMT) solution (http://www.globalcmt.org) is provided, which represents the constant slip model. Shao *et al*. and Hayes included the time variation with the resultant total rupture duration spanning to 170 sec and 240 sec, respectively. The time variation affects the initial water surface profile as shown in Table 4, and the inundation afterwards. Keeping all other parameters constant, an instantaneous slip model would trigger a higher initial water amplitude since the energy is released at the same time, *i.e.* at t = 0. Hayes' solution has a longer rise time which leads the energy released with a longer duration, thus causing a lower maximum initial surface height compared to Shao's time-variant source model that activates an initial water as high as 14.5 meters at t = 60s. The initial surface profiles for the instantaneous slip models vary from 6.7 meters to 13.6 meters.

Table 4 Maximum initial water surface height (m)

	Hayes	Shao *et al*.	Fujii *et al*.	Wei *et al*.	GCMT
t = 0s	1.9209	3.0616	13.6192	6.6766	11.6194
t = 60s	9.0425	14.4667	-	-	-
t = 120s	6.4717	12.0385	-	-	-
t = 180s	4.3766	9.7381	-	-	-

For both instantaneous and transient seafloor motion, the fault plane can be divided into several subfault segments. By assigning various slip amplitudes and the starting time to each subfault, a non-uniform transient seafloor rupture motion can be generated. The number of subfault is a delicate parameter to be decided. On one hand, according to Geist (2005), the initial profile of the generated tsunami will be underestimated if the size of the subfaults is larger than the focal depth of that particular event. On the other hand, even though there is no restriction for more subfaults, the influence of subfault size vanishes if it is less than the focal depth.

A comparison of the maximum inundation maps from different earthquake source models suggests that both Shao's and GCMT models are consistent with the observations. Please note that other models are more conservative than these two only in Sendai plain. It is not equivalent to arrive at the conclusion that the others

are not accurate in general. On the contrary, other models provide better or same results in other regions along the eastern coast of Japan. This is also stressed in Breanyn (2013). Table 5 provides the parameters required in COMCOT and an example of GCMT solution for the Tohoku earthquake source parameters.

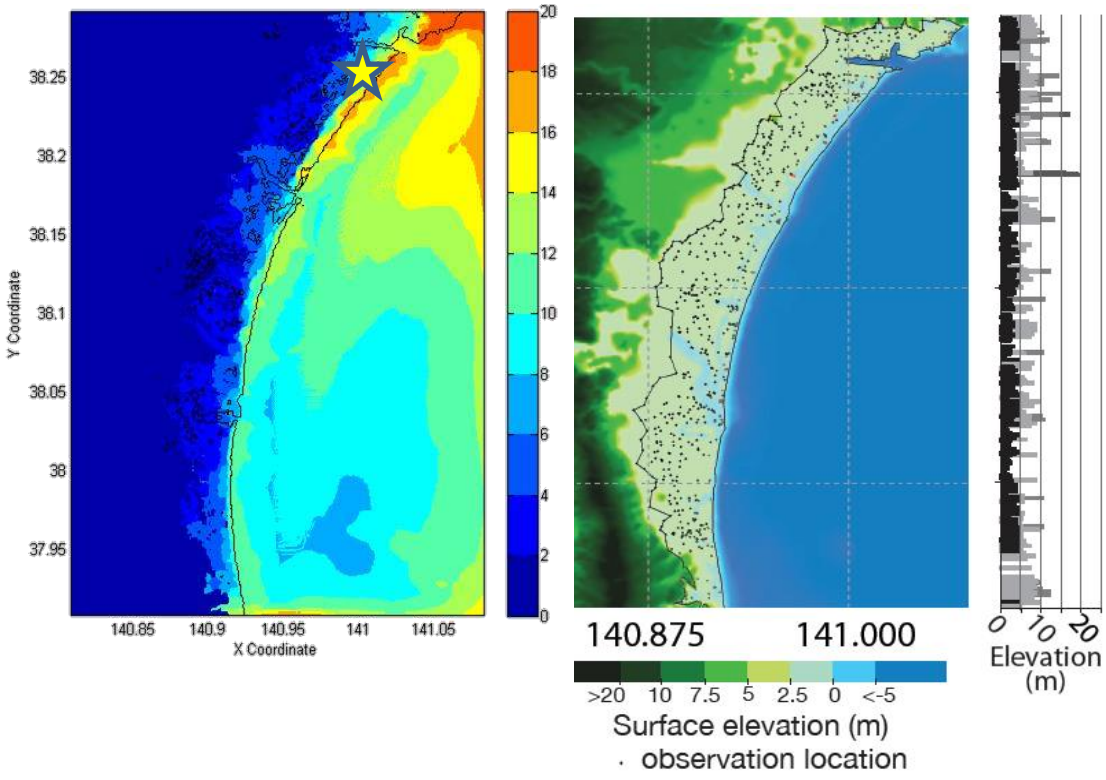
Table 5 Earthquake parameters required in COMCOT and 2011 Tohoku earthquake parameters from Harvard GCMT

Parameter	COMCOT Range	Tohoku Earthquake
Epicenter (Lat, Lon) [deg]	(-90 – +90, 0 – 360)	38.32, 142.37
Strike direction (θ) [deg]	(0 – 360)	203.00
Dip angle (α) [deg]	(0 – 90)	10.00
Rake angle (λ) [deg]	(-180 – 180)	88.00
Focal depth [m]	N.A.	24400
Length of fault [m]	N.A.	300000
Width of fault [m]	N.A.	150000
Dislocation (slip) [m]	N.A.	30

Note: N.A. indicates no limits for the corresponding parameter in COMCOT

3.2.3 Tsunami model results and validation

The earthquake source model developed by Shao *et al*. is implemented to further compute the wave amplitude and velocity at Sendai plain. Figure 15(a) shows the maximum inundation map for layer03, which corresponds to the Sendai plain. The yellow star in the figure indicates the location of Gamou wastewater treatment plant. The total wall-clock time in the simulation is three hours. As shown from the figure, higher water amplitude is observed on the northeast corner of the interested area at a wave height of 20 meters and above. The inundation depth decreases with the further distance from the shoreline. A comparison is made with post-tsunami survey done by Breanyn *et al*. (2013) as shown in Figure 15. The water height is measured from the mean sea level for both cases. The simulation result of the maximum inundation depth matches the observation, although the elevation at each point slightly differs, mainly due to the inaccuracy of the interpolated topological data and numerical error.



(a) (b)

Figure 15 Maximum inundation map (a) COMCOT simulation (b) Post-tsunami survey observations and inundation lines (Breanyn, Gusman et al. 2013)

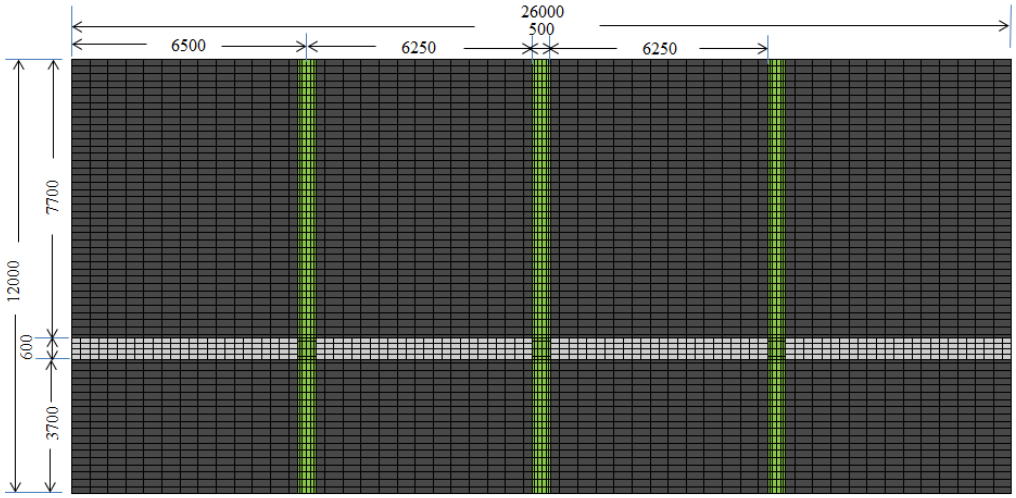
The water depth is obtained by subtracting the elevation of 5 meters at Gamou wastewater treatment plant. Therefore, the tsunami wave height responsible for the impact loading is 4.6 meters. This wave height is used in the numerical simulation of the tsunami-wave-structure interaction.

3.3 Structural model setup

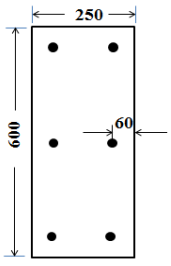
The structural model in this case study is simplified since the out-of-plane deflection occurred only on the side of the ocean-facing wall with a two-story high-bay as shown in Figure 13. The tsunami wave is represented by its amplitude and velocity derived from COMCOT.

3.3.1 Structural geometry and material modeling

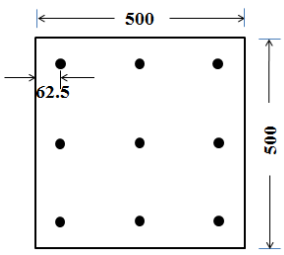
The thickness of the wall is assumed to be 400mm.The details of the structural dimension is demonstrated in Figure 16. The unit is in millimeter. All the values are derived from the site survey pictures for both the exterior and interior parts of the structure. The height of the wall structure is equivalent to a common three-story building. Beams are located at one third of the total height of the wall. The roof-high beam is neglected since the structural wall is fixed at the four sides. Eight #25 (diameter of 25.4mm) steel rebars for the column and six for the beam are arranged as longitudinal reinforcement. The #19 rebars are used for both longitudinal and horizontal reinforcement for the wall. The clear distance between two adjacent rebars in the concrete wall is 500mm and 200mm, respectively, as shown in Figure 16(d).



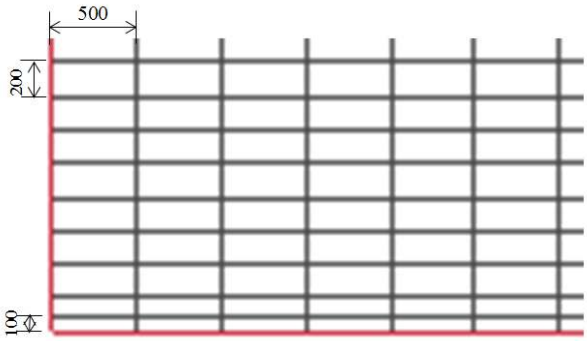
(a) concrete wall with columns and beams at clear height of 3700mm



(b) reinforcement details of the beam



(c) reinforcement details of the column



(d) reinforced grid of wall (corner portion of the wall for illustration only)

Figure 16 Geometry details of the tsunami-facing wall model

The column element size is 200mm × 125mm × 125mm and further divided into 200mm × 62.5mm × 62.5mm locally, to ease the modeling of the longitudinal reinforcement. The beam element size is 250mm × 125mm × 150mm. The wall element has a dimension of 500mm × 200mm × 200mm. A total number of 30,000 elements are used in this model.

3.3.1.1 Concrete material modeling

The concrete material for this study has a density of 2400kg/m^3, Young's modulus of 27800MPa and Poisson's ratio of 0.2. There are several concrete constitutive models in LS-DYNA. Material type MAT_72R3, namely, CONCRETE DAMAGE_REL3 (Schwer & Malvar 2005; Magallanes *et al.* 2010), is adopted in this study with the following reasons. First of all, it is a plastic-based model using three-invariant failure surface formulation, *i.e.* maximum failure surface, residual failure surface and yield failure surface. Therefore it can simulate the nonlinear behavior of the concrete material. Secondly, the user only needs to provide the unconfined compression strength f_c' and concrete mass density, which simplifies the modeling procedure and minimizes the potential numerical uncertainties. The rest of the parameters in MAT_72R3 are self-generated by the program, which provides a good average fit to various material characterization tests. The unconfined compressive strength f_c' for the concrete is 34.5MPa in this case study. Lastly, the MAT_72R3 allows the strain-rate effect, which is important in impact analysis since the concrete static strength can be doubled or tripled in terms of the Dynamic Increase Factor (*DIF*). For compression strength, *DIF* is expressed as:

$$DIF = \begin{cases} (\dot{\varepsilon} / \dot{\varepsilon}_s)^{1.026\alpha_s} & \dot{\varepsilon} \le 30s^{-1} \\ \gamma_s (\dot{\varepsilon} / \dot{\varepsilon}_s)^{1/3} & \dot{\varepsilon} \ge 30s^{-1} \end{cases} \tag{3.1}$$

where $\dot{\varepsilon}$ is the strain rate, $\dot{\varepsilon}_s$ is the static strain rate, $\gamma_s = e^{6.156\alpha_s - 2}$ with $\alpha_s = 1/(5 + 9 f_c / f_{c0})$ in which $f_{c0} = 10\text{MPa}$. The strain rate table adopted for this study is shown below based on Malvar *et al*. (1998). The negative value in abscissa denotes concrete in tension.

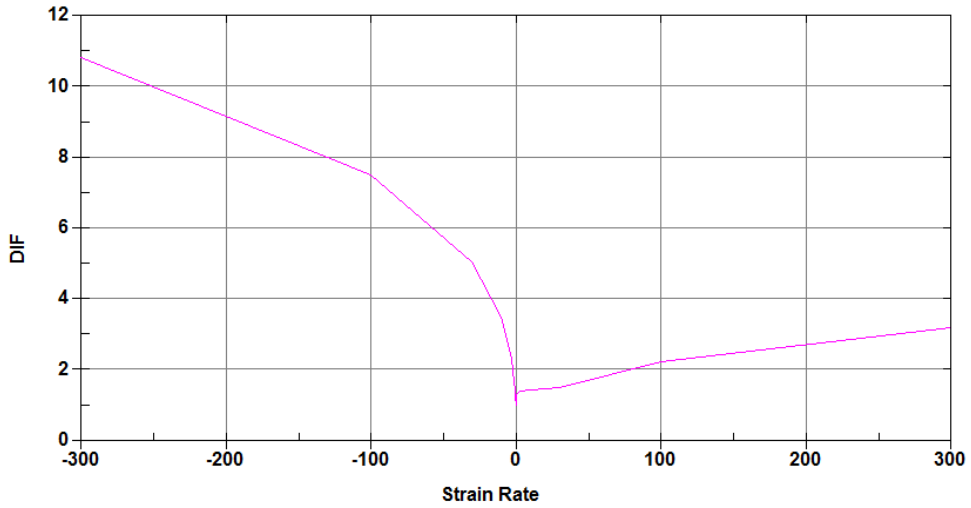


Figure 17 *DIF* for concrete material

3.3.1.2 Reinforcement material modeling

Most of the tensile strength of a concrete structure is provided by the embedded reinforcement bar. The MAT_72R3 concrete model does not consider the reinforcement. Therefore, a separate material model for the reinforcement is required. Again, various material models are capable to represent the characteristic of steel in LS-DYNA. In this study, a common choice of MAT_03, PLASTIC_KINEMATIC card is applied. A bilinear curve is defined to represent the plastic behavior and isotropic, kinematic hardening effect of the steel material. The mass density of the reinforcement is 7800kg/m^3, Young's modulus E_s = 200GPa and Poisson's ratio equals to 0.3. The stress-strain relationship for the initial loading is shown in Figure 18 below.

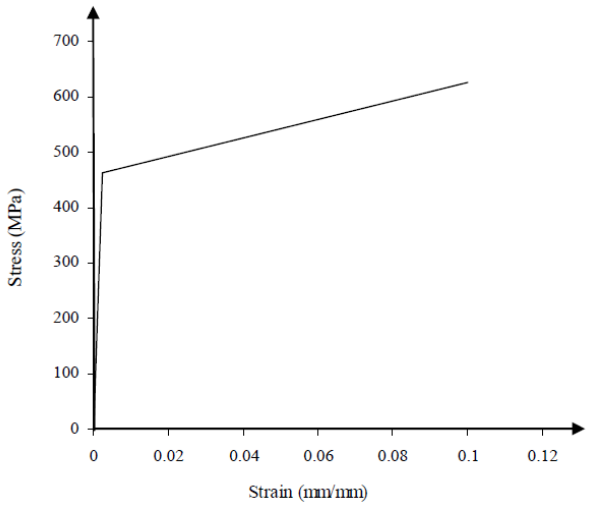


Figure 18 Strain-rate curve of reinforcement bar of with σ_y 460MPa

3.3.2 Tsunami force modeling

In order to acquire the structural response under tsunami loadings, a two-step loading pattern is considered. The gravity is firstly treated as a body force acting on all the material, including water, concrete and reinforcement. The quasi-static gravity loading takes effect within 0.6s. Then, the wave loading characterized by its height and speed crashes the structural wall. The water height variation can be modeled by activating different groups of water part elements in non-overlapping time intervals. The minimum water element size decides the highest resolution of the water part. The water element size used for this case study is 1m × 1m × 0.5m. Owing to the large water element size, it is not possible to trace the velocity changes within 1m. Therefore, a constant velocity is assumed throughout the simulated period. This assumption is valid since the total simulated time is 3 sec. The depth-averaged water velocity components u and v are derived based on the simulated water flux P and Q in x and y directions, respectively. The velocity components at a particular grid i are given as

$$u_i = \frac{P_i}{H_i}, v_i = \frac{Q_i}{H_i} \tag{3.2}$$

where H_i is the water depth at the i-th grid. The resultant velocity U is defined as the square root of the sum of squares of the two velocity components. In this case study, the resultant wave velocity is 5.5m/s.

3.3.3 Boundary conditions

Boundary conditions are highly important for numerical modeling. As shown in Figure 19, the red box displays the water part, and the blue box the "void" region which is going to be occupied by water during the simulation. The structure stands in the void region with a distance of 2.0 meters from the front of the water. The boundary conditions for the two boxes are specified as follows: 1) the eight nodes of the water part (red) are fixed in all directions, 2) four edges of the water part in the z direction are constrained in the x and y directions, 3) four edges along the x direction of both water and void parts are constrained in the y and z directions, 4) bottom surface, *i.e.* x-y plane at the bottom, is constrained in the z direction, 5) all the nodes in the x-z plane are constrained in the y direction, 6) nonreflecting boundary is assigned to the y-z plane surface at the back of the structure to avoid the bounce of tsunami wave which is unrealistic. With the boundary conditions modeled above, the wave can flow towards the structure without leakage and flood out of the fluid domain freely. The structural boundary condition is rather straightforward as mentioned previously: four sides of the structural wall are fixed.

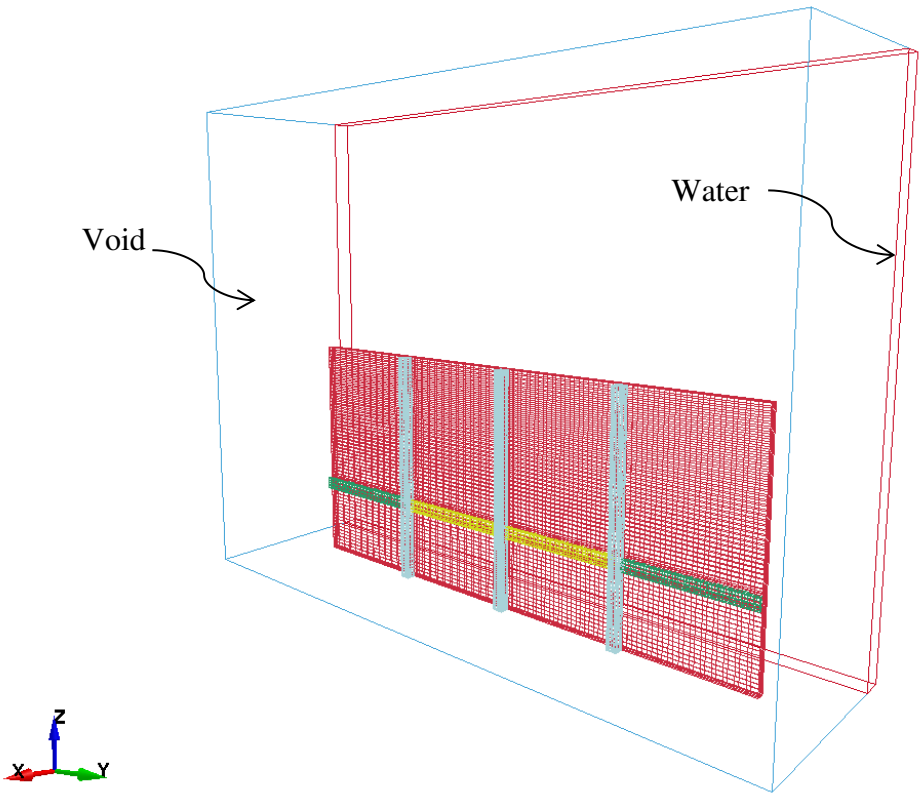
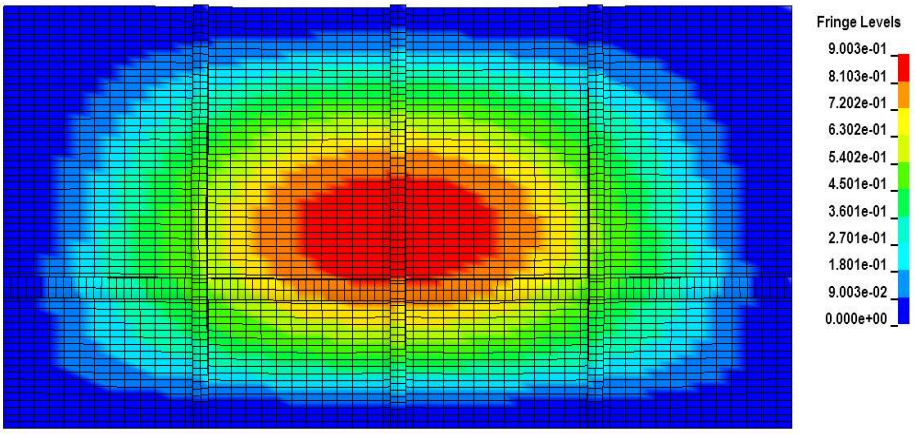


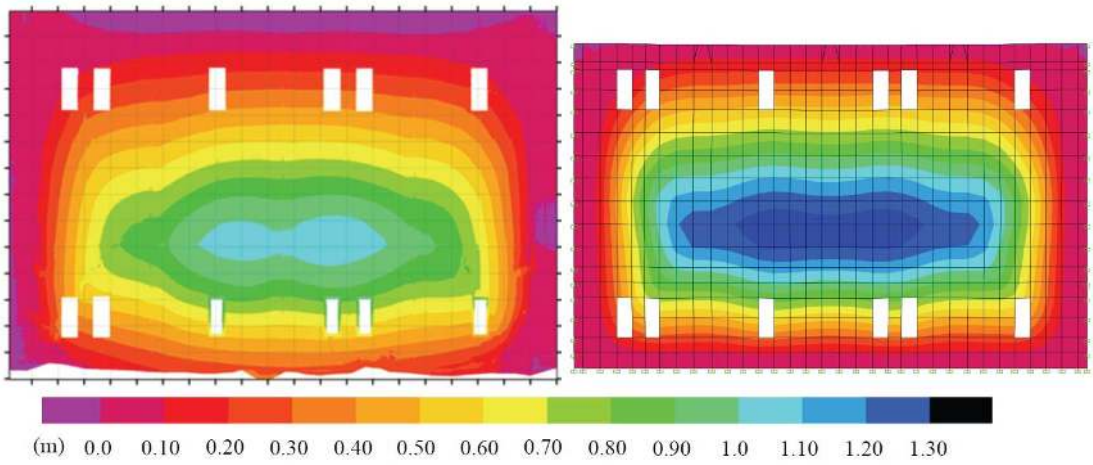
Figure 19 Tsunami wave-structure interaction model setup in LS-DYNA

3.4 Structural performance

With the properly defined wave loadings and boundary conditions, the resultant x-direction deflection of the structural wall is shown in Figure 20(a). The maximum displacement of 0.9m occurs near the center at time of t = 5.0s. The result is compared with the observation obtained from LiDAR scan in Figure 20(b) and static FEM analysis done by Carden *et al*. (2012) in Figure 20(c). The wave pressure distributions used in Carden's analysis is based on the empirical equation derived by Robertson *et al*. (2010). It can be seen that the magnitude of the highly plastic wall displacement shown in Figure 20(a) exhibits a similar pattern and magnitude compared to the scanned image (Figure 20(b)), *i.e.*, the maximum displacement located below the geometry center of the wall with a magnitude of 1.0 – 1.1 meters. The simulated maximum deflection of the current study is less than the ones from the observation by 0.1m - 0.2m. Note that the maximum displacement of 0.9m using LS-DYNA is based on the tsunami velocity and height of 5.5m/s and 4.9m, respectively. These two values estimated by Carden et al. (2012) are 6.5m/s and 6m respectively, according to the footage and site survey. The wave profile differences mainly contribute to the difference in the maximum displacement. In general, the position of the maximum displacement and the deformation region agree well with both LiDAR result and Carden's analysis.



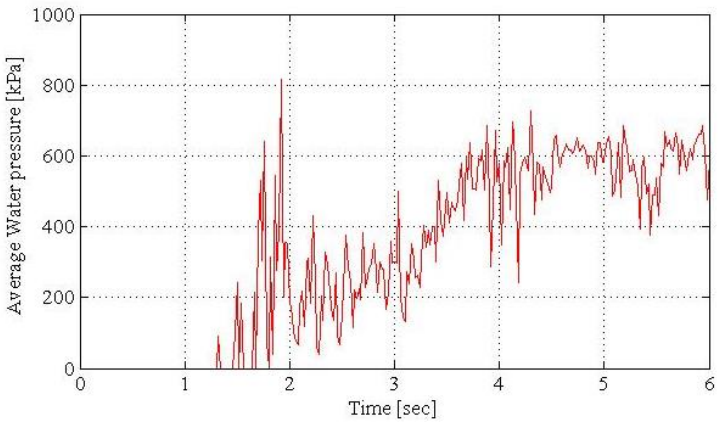
(a)



(b) (c)

Figure 20 Deformation of the structural wall (a) FEM tsunami impact analysis from this study (b) a triangulated LIDAR scan showing out-of-plane deformation (c) a FEM analysis of damaged wall (Carden, Chock et al. 2012)

The time-history results of the pressure variation and displacements at selected node are presented in Figure 21. The average pressure is measured from the central bottom of the tsunami-facing wall. The maximum resulting pressure distribution represents an applied load of 900kN/m of the width of the wall. The selected node is placed at the joint of the middle column and beam (Figure 16a). From the preliminary result, it is noticeable that the largest pressure appears during the initial impact between the tsunami bore and the structural wall. However, a monolithic increasing curve is observed for the x-direction displacement.



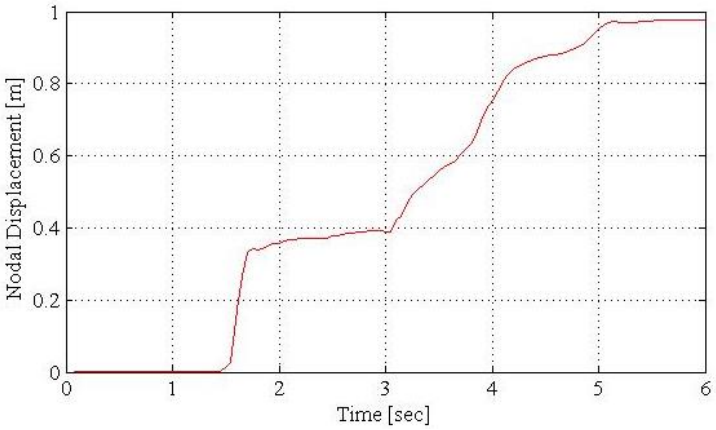


Figure 21 Time-history curves for wave average pressure (up) and nodal displacement (bottom)

The total computational time is 5.2 hours. The computational platform used for the wave-structure interaction is a four-processor computer, Intel Core i7-2600. The minimum time step Δt, controlled by the smallest element size in the computational domain, is 2.7×10^{-5} second on average.

3.5 Benchmark problem for reliability analysis

The numerical simulation of the tsunami and wave-structure interaction demonstrated in the previous sections indicates that the designed simulation processes along with the numerical tools are sufficient to precisely model the tsunami wave-structure interaction problem. Thus, it can be adopted to carry out the reliability analysis of on-shore structures subject to earthquake-induced tsunami waves.

3.5.1 Stochastic earthquake source model

For this study, a total of 16 random variables, *i.e.* rake angles from 16 subfaults, with a truncated normal distribution of mean 90 and standard deviation 30 are generated. This is consistent with the observation of historical earthquakes in Tohoku region (Yamamoto & Hori 2004). Other earthquake source parameters are fixed to keep the number of free variables in a relatively low dimension. This synthetic earthquake model only represents a small portion of all the possible tsunamigenic earthquake scenarios due to the assumed ranges of the earthquake source parameters. The earthquake moment magnitude M_w is set to be 8.3, which

corresponds to the seismic moment of $M_0 \approx 3.5 \times 10^{28}$ dyne-cm. The fault length and width are approximately equal to 230km and 90km, respectively, based on the empirical relations $\log_{10}L = a+bM_w$ (a, b being constant coefficients) from Blaster *et al*. (2010). The correlation length and width are 45km and 18km, respectively, according to Mai & Beroza (2002). The number of subfaults are six for both along-strike and along-dip directions, resulting in a total number of 36 subfaults on the fault plane. The 2011 Tohoku Earthquake parameters are used for the dip angle, strike angle, and fault depth. The slip amplitudes, modelled by the hybrid k-squared model, are fixed as well. Due to the fact that slip amplitudes are confined at the edges of the fault plane, 16 non-zero amplitudes out of 36 subfaults are obtained. The detailed information of the earthquake model parameters are compiled in Table 6. A sample input to COMCOT can be found in Appendix A.1.

Table 6 Earthquake parameters for the benchmark study

Earthquake parameter	Value used
Magnitude [Mw]	8.3
Epicenter [deg]	38.1(latitude) & 143.2 (longitude)
Strike direction (θ) [deg]	199.0
Dip angle (α) [deg]	10.0
Rake angle (λ) [deg]	X~*N*(90,30)* with truncated interval [-180, 180]
Focal depth [m]	24400
Fault Length[m]	10^(-2.37+0.57*Mw)**
Fault Width [m]	10^(-1.86+0.46*Mw)**
Dislocation (slip) [m]	$\overline{D} = \frac{M_0}{\mu A}$ & hybrid k-squared slip model***
Fault discretization	6 × 6

* Yamamoto & Hori, 2004;

** Blaster, 2010;

*** Gallovic, 2004.

To demonstrate the validity of the proposed stochastic earthquake source model, especially the hybrid k-squared slip distribution model, a set of earthquake parameters are generated with an earthquake magnitude of 9.1, and the strike angle,

dip angle, slip angle are the same as those in the 2011 Tohoku earthquake event. The simulated inundation map is shown in Figure 22. It can be seen that the inundation boundary is slightly larger than the site survey result in Figure 15b. The inundation height close to the coastal region is similar to the one in Figure 15b. More specifically, the inundation distribution is consistent with the actual inundation map, *i.e.*, the inundation height is ranged between 12m – 18m in the north and 6m – 12m in the south. Thus, for the purpose of this study, the proposed stochastic earthquake source generator is adequate to produce credible future events for reliability analysis purpose.

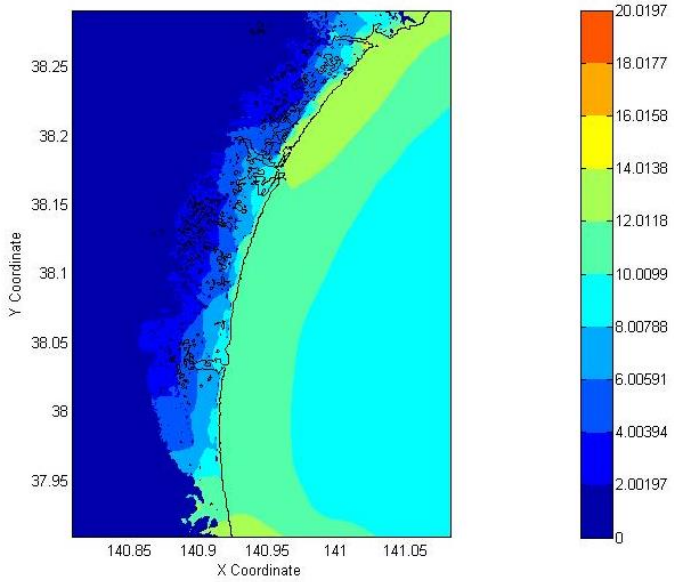


Figure 22 Inundation map generated using proposed earthquake source generator

3.5.2 Simplified structural model

As mentioned earlier, the overall computational time of the entire tsunami wave-structure interaction process, including tsunami propagation and run-up simulation, easily exceeds six hours. Since the focus of the study is to propose efficient algorithms which can reduce the total number of simulations required, it is logical to simplify the numerical model to reduce the individual computational cost. For illustrative purpose, the structural model is simplified into a structural wall without column and beam components. The geometry of the model is scaled to half of the original one in order to reduce total number of elements. The water element size used for this benchmark is 0.25m × 0.25m × 0.25m. The refined resolution allows a

minimum water height difference to be 0.20m. The velocity and tsunami wave height are extracted from the previous COMCOT simulation with constant values assumed during the impact. A total wave-structure interaction of 3.0 seconds is simulated using LS-DYNA.

The total numerical simulation time has dropped from 5.2 hours to 3 hours using the same computational environment. The structural displacement relative to its original position is plotted in Figure 23 using a random sample generated from the stochastic source model described in the last section. The wave height and velocity are 2.25m and 5.0m/s, respectively. The fringe level indicates the displacement in meter. The resultant displacement time-history from the node 20318 is recorded as shown in Figure 24. The maximum value from the record is chosen as the maximum displacement of the structure.

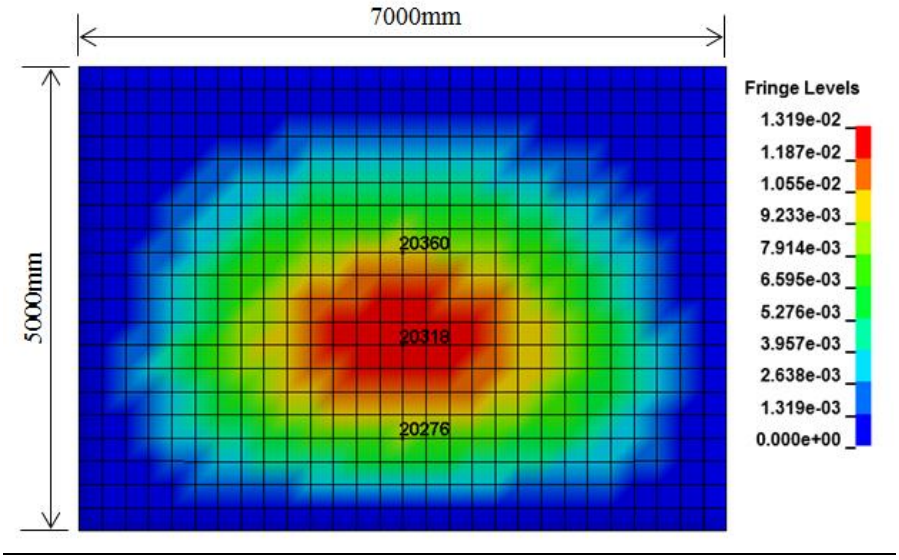


Figure 23 Maximum structural displacement diagram

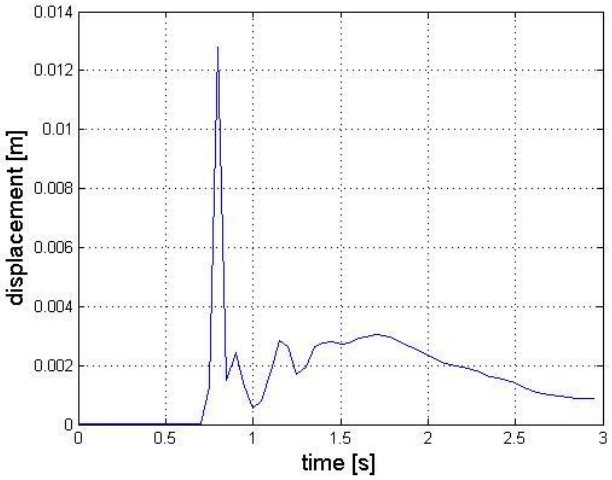


Figure 24 Nodal displacements time-history curve

3.5.3 Benchmark reliability analysis results

In this benchmark study, the original Subset Simulation (SS) is used. Four intermediate failure levels (m = 4) are calculated with N = 500 samples of $\boldsymbol{\theta}$ generated for each level. The sample are selected such that $P(F_i|F_{i-1}) = p_0$ is approximately 0.1. Thus, the total number of numerical evaluations is $n = N\times[1 + (m - 1)(1 - p_0)] = 2000$. Normal distribution is used as the proposal PDF $q_j\left(\bullet\middle|\boldsymbol{\theta}_k^j\right)$. Thus, the acceptance ratio becomes the acceptance ratio between the target distribution with the proposed and the original sample $r_{k+1}^j = p^j\left(\xi_{k+1}^j\right)\Big/ p^j\left(\boldsymbol{\theta}_k^j\right)$. The computational time for the SS is approximately 98 hours with parallel computing using 64 logical processors on a work station with AMD Opteron 6282 SE @ 2.60GHz and a RAM of 96 GB. A total of 50 independent runs were carried out in order to obtain the average failure probability as well as estimating the c.o.v. of the SS results.

The four intermediate failure threshold levels are estimated to be b_1 = 1.115cm, b_2 = 3.133cm, b_3 = 4.821cm and b_4 = 6.068cm on average, which correspond to the maximum structural displacements considering all the nodes and the whole duration. These values will be used later as benchmark for evaluating the accuracy of the results obtained by the proposed algorithms. The mean exceedance probability

curve as well as the failure probability curves obtained in the 30 independent SS runs is presented in Figure 25. In order to investigate the accuracy and bias of the Subset Simulation, the results are compared with the original MCS with 10,000 samples. It can be seen that the results by the two methods agree well up to the second intermediate failure level ($P_F = 10^{-2}$). The discrepancies beyond the second intermediate failure level are greater as the accuracy in estimating failure probability decreases due to the insufficient number of samples in the MCS method. According to Eqn. (2.31), the c.o.v. at the second level is approximately 10% with 10,000 samples, which is an reliable estimate. However, the c.o.v. of the MCS estimator rises to 30% and 100% for the third and fourth intermediate failure levels, respectively.

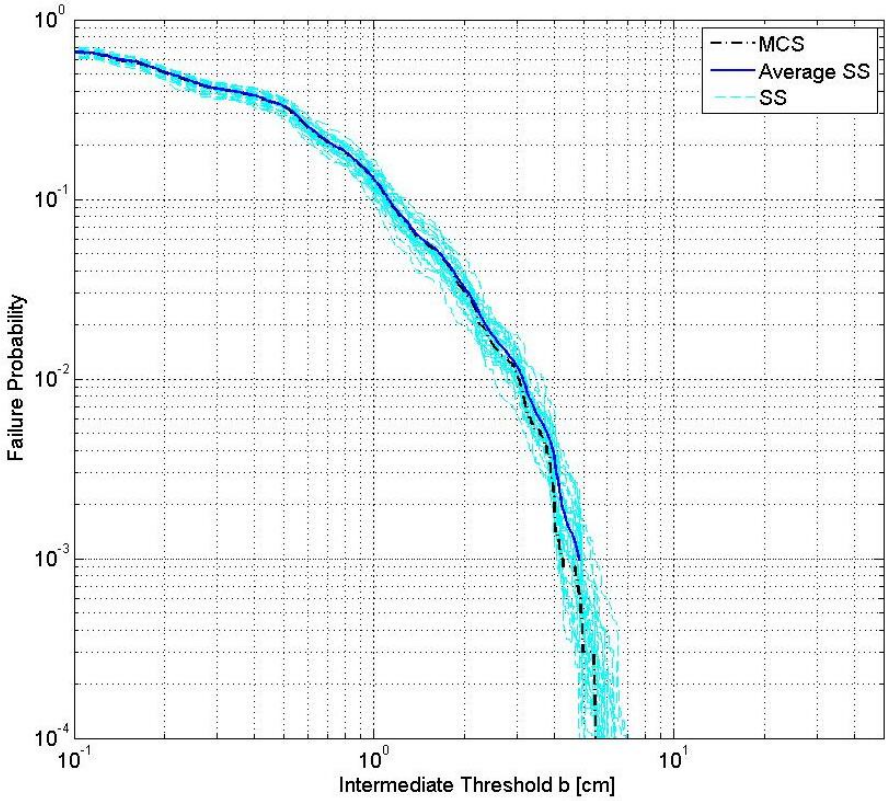


Figure 25 Failure probability estimates of the 30 Subset Simulations and the averaged curve

The variability of the Subset Simulation estimator is examined using the sample c.o.v. over the 30 independent runs. It is compared with the lower bound c.o.v. and theoretical c.o.v. of MCS using Eqn. (2.43) and Eqn. (2.31), respectively. The comparison is plotted in Figure 26. The c.o.v. estimation of MCS at a given

intermediate failure level is assumed to have the same number of stochastic samples as for the SS method so as to keep the computational effort comparable. As seen from the figure, the uncorrelated c.o.v. (dashed line) lies close to the sample c.o.v. and hence the correlation between the conditional failure probabilities of consecutive levels in SS is small. According to the trend of the c.o.v., it is clear that the efficiency of the SS method outweighs the MCS with decreasing failure probabilities. Thus, the results from the SS method are valid to be used as the benchmark for the following chapters against which the proposed algorithms are compared.

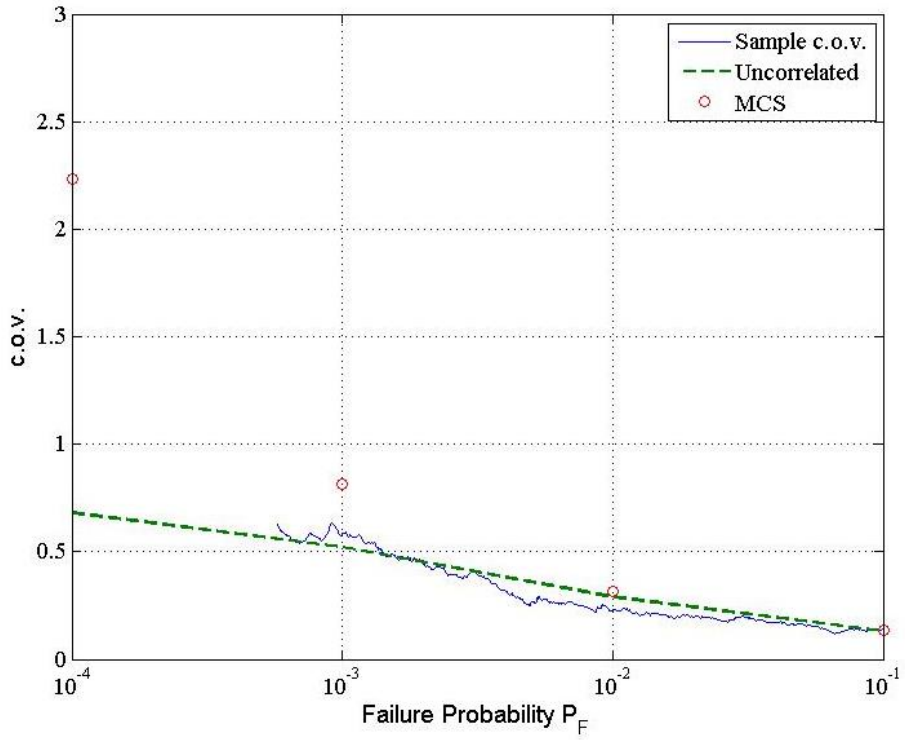


Figure 26 Coefficient of variation of failure probability estimate for benchmark

Besides the overall performance of the reliability algorithm, the evolution of the marginal probability distribution with respect to the intermediate failure levels is also of interest since it can provide insight of the physical process. The relative position of the 16 variables on an imaginary fault plane is shown in Table 7. Four types of lines indicating the kernel density estimates of the four intermediate failure levels is shown in Figure 27: solid line for the 1st level, dashed line for the 2nd level, dot-dashed line for the 3rd level, and dotted line for the last level. The x- and y-axes represent the rake angle and probability density function of a subfault,

respectively. Note that the probability density functions for the first intermediate failure level of all the random variables are almost identical because they follow $N(90, 30)$. The probability density functions estimated using conditional failure samples and kernel density start to deform from the second failure level onwards and each with its own way. The more the discrepancy, the greater sensitivity it contains. According to the figure, the majority of the model parameters share a similar shape of the probability density functions among themselves for the first two to three intermediate failure levels, except for V2, V3, V6 and V7 in which clear trends of moving leftward or rightward can be observed. Furthermore, the estimated marginal probability density function of the fourth failure level has more peaks than the previous failure levels. Information entropy which quantitatively measures the relative change between different failure levels of the marginal proposal distribution is adopted. The results are compiled in Table 8. Apparently, larger information entropy ultimately indicates greater influence on the stochastic sampling process. The data has been normalized to 0 – 1 range for easier comparison. It agrees with the observation from Figure 27 that V6 and V7 have relatively largest information entropy among the 16 random variables for the first and second intermediate failure level, respectively. Furthermore, the V11 is the most sensitive one in the third failure level.

Table 7 Location of the 16 slip angles on a fault plane

V1	V2	V3	V4
V5	V6	V7	V8
V9	V10	V11	V12
V13	V14	V15	V16

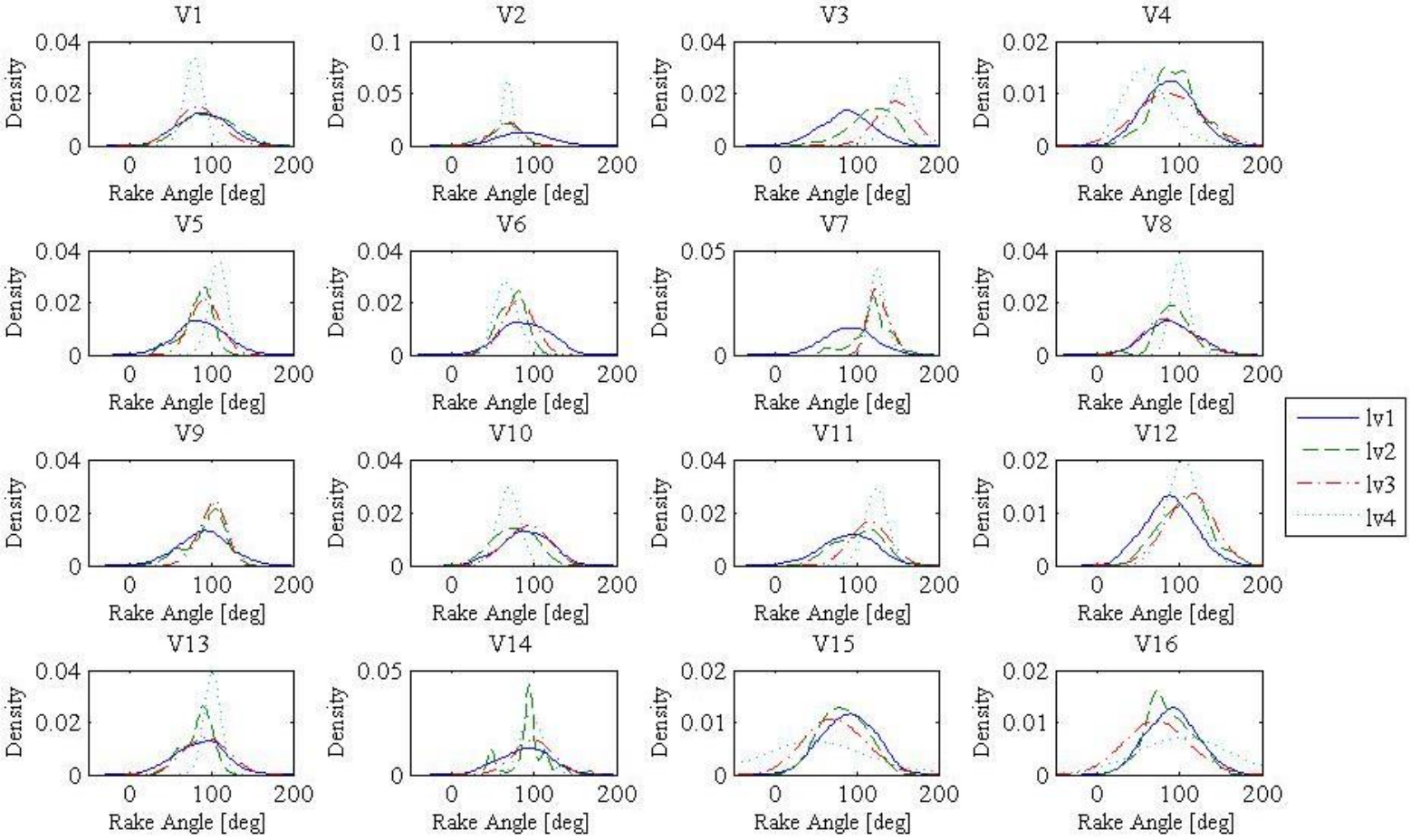


Figure 27 Kernel density plots of random variables for various intermediate failure levels

Table 8 Normalized information entropy for the marginal probability distribution of the 16 random variables

	Level 1-2	Level 2-3	Level 3-4
V1	0.06	0.20	0.46
V2	0.82	0.33	0.18
V3	0.25	0.25	0.20
V4	0.00	0.02	0.04
V5	0.26	0.00	0.21
V6	**1.00**	0.98	0.13
V7	0.21	**1.00**	0.84
V8	0.04	0.33	0.15
V9	0.19	0.16	0.25
V10	0.01	0.71	0.24
V11	0.04	0.38	**1.00**
V12	0.05	0.35	0.01
V13	0.06	0.18	0.15
V14	0.05	0.39	0.05
V15	0.08	0.11	0.00
V16	0.02	0.36	0.14

For the 16 slip angles, a correlation matrix is showcased in Figure 28. In the upper diagonal of the matrix, positive correlations are displayed in blue and negative correlations in red colour. The colour intensity and the size of the circle are proportional to the correlation coefficients indicated in the lower diagonal of the matrix. The legend colour shows the correlation coefficients and the corresponding colours. The paired variables of slip angle with absolute higher correlation value can be found between V3-V13, V6-V14, V6-V16, and V7-V11. Note that the actual interaction of these variables on a fault plane is much more complex and requires further investigation.

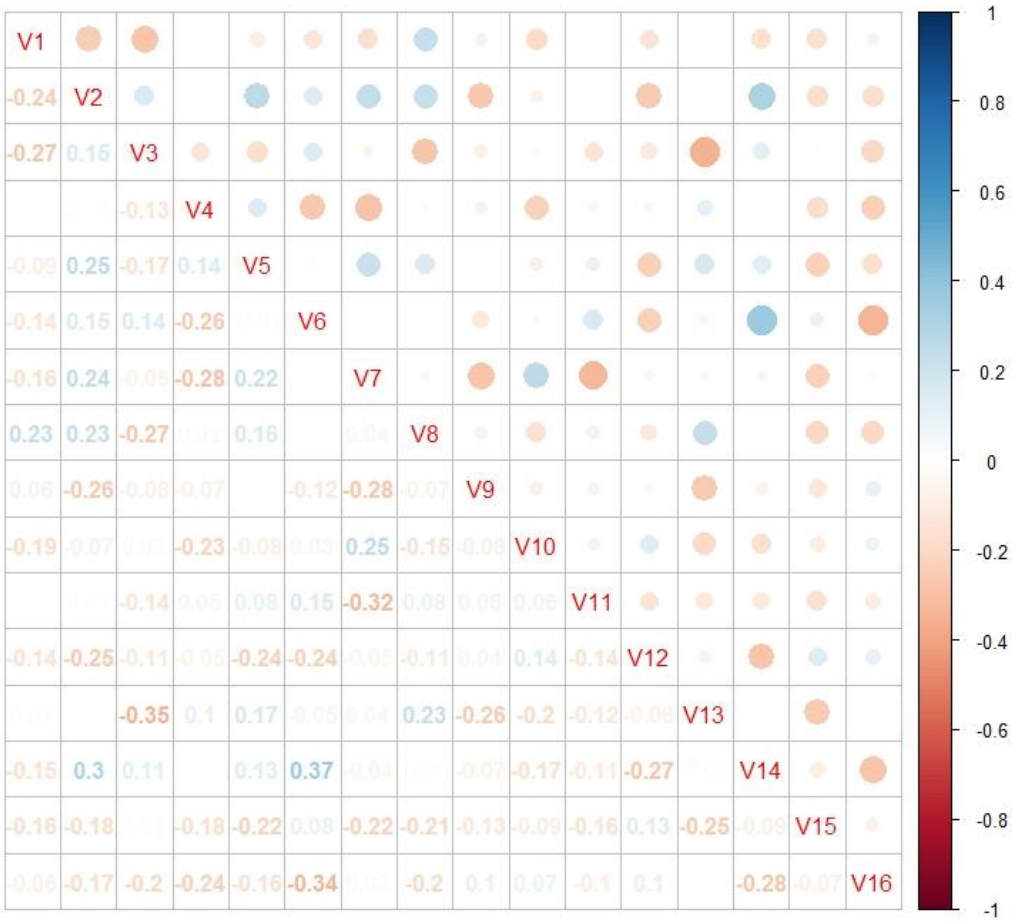


Figure 28 Correlation matrix of the 16 slip angles

3.6 Closing remarks

The process of a tsunami generation and wave-structure interaction is presented in this chapter. An accurate tsunami profile at any location can be obtained by providing reliable bathymetric dataset with appropriately defined earthquake source parameters. Nonlinear structural response under transient and sustained tsunami wave loading is modeled using LS-DYNA. The results lay the foundation for the structural reliability analysis and provide insights of the structural performance under tsunami wave loading.

A benchmark reliability study using a simplified structural model is constructed by the original SS method. The maximum structural displacements correspond to various intermediate failure probabilities have been carried out. The resultant failure

probability and variation of the SS method are compared with the ones obtained from MCS. A parametric study is presented in order to better understand the statistical relationship among the random parameters. Apart from this, there are a few points worth mentioning.

i. Although the hybrid k-squared stochastic slip model is used, a lower subfault number *i.e.*, 6×6 with large earthquake amplitude, *i.e.*, M_w 9.1, result in stable slip amplitude since no corner frequency exceeds the threshold for any of the subfaults.

ii. The Tukey window (Tukey 1967), also known as the tapered cosine window is adopted to "seal" the fault boundary as well as to smooth the dislocation in order to avoid spectral holes caused by the sudden change. A common choice of the parameter α in the Turkey window ranging from 0.1 to 0.5. An α value of 0.2 is selected and proven to be suitable for this study.

iii. The parametric study concentrates on the evolution of the marginal proposal distribution. The interaction among the variables, such as correlation, though calculated, can hardly be used to explain the physics of the tsunami earthquake mechanism.

Chapter 4 Proposed Adaptive Metamodels for Structural Reliability Analysis

4.1 Challenges

The case study demonstrated in Chapter 3 indicates that the designed simulation processes along with the numerical tools are sufficient to precisely model the tsunami-wave-structure interaction problem. A benchmark reliability study of on-shore structures subject to earthquake-induced tsunami is also illustrated using the original Subset Simulation (SS) method. 16 random variables are involved in the benchmark problem. A general evaluation framework using the SS method can be summarised as follows:

1) Generate stochastic earthquake source parameters following prescribed distributions;
2) Obtain tsunami wave profiles from the numerical model;
3) Evaluate the structure responses (e.g. maximum displacement, acceleration, inter-story drift, etc.) performance during the wave-structure interaction using LS-DYNA;
4) Repeat steps (2) and (3) using original MCS to generate N (e.g. 500) samples;
5) Use MCMC-based algorithms (e.g. modified Metroplis-Hasting) to further generate conditional failure samples which require running steps (2) and (3) and computing the limit state function $L(\theta)$ and thus, the failure probability of the structure.

The Steps (4) to (5) represent the core of reliability assessment in this study. More importantly, the most computational demanding part of the reliability analysis in this work is the computation of $L(\boldsymbol{\theta})$ which requires numerous computationally intensive dynamic analyses as stated in Section 3.4. On top of that, when estimating small failure probability (*i.e.*, $P_F << 1$) with a coefficient of variation (c.o.v.) of, for example, 30% in its estimator, the problem arises as the MCS requires about $10/P_F$ samples simulated according to $p(\boldsymbol{\theta})$, which in turn, requires about $10/P_F$ tsunami-wave-structural interaction dynamic analyses. For instance, at least 10,000 of stochastic samples are required in order to get an acceptable result (e.g. c.o.v. of 30%

in P_F estimator) with failure probability of 10^{-3}. This is equivalent to compute the above-mentioned physical process 10,000 times. Even using the Subset Simulation algorithm to obtain a c.o.v. of about 24.5% in P_F estimator, a total number of at least 1,500 stochastic samples are required, assuming N = 500 samples for each intermediate failure level $P(F_i|F_{i-1}) = p_0 = 0.1$. Even though the computational time required by the original Subset Simulation algorithm can be reduced to 90 – 100 hours (*i.e.* 3.8 – 4.2 days) with the help of a HPC system, it is still far from being acceptable for a practical engineering problem such as the one investigated in this study. Not to mention that the assessment is based on a simplified structural model used in the benchmark study.

4.2 Response surface methods

The most computational demanding part of the reliability analysis in this work is due to numerous computations of limit state function $L(\boldsymbol{\theta})$ which requires a large number of computationally intensive dynamic analyses. The key to solving this type of problem is to construct an approximation model to replace the original true limit state function. By doing so, the number of computational intensive numerical simulation realizations is reduced though a certain number is still required in order to estimate the function.

Many response surface models (RSM), also refer to as metamodels, are valid candidates to partially replace the original physical model for the reliability analysis. Some of the most popular metamodels are as moving least square response surface approximation (MLS), Gaussian processes (GP) (Rasmussen & Williams 2006), artificial neural network (ANN) and stochastic response surface methods such as stochastic polynomial chaos (Ghanem & Spanos 1990; Xiu & Karniadakis 2002) and stochastic collocation methods (Babuska *et al.* 2007; Nobile *et al.* 2008), to name just a few. In the following sections, the author will briefly introduce the moving least square response surface approximation. The most recent development by Taflanidis & Cheung (2012), is presented to exhibit a better solution for highly nonlinear systems. At last but not the least, the GP approximation model is described.

4.2.1 Moving least square metamodel

The MLS has a wide range of applications in meshless methods and image processing, etc. It was introduced by Lancaster *et al*. (1981) based on the Shepard scheme (Shepard 1968). Let $\boldsymbol{\theta} = [\theta_1 \ldots \theta_{n_\theta}]^T \in R^{n_\theta}$ be the free variables, the goal is to define a function $\hat{f}(\boldsymbol{\theta})$ to approximate the system response, given a new set of independent variables. The formulation of MLS is similar to the global least squares method except that the coefficient vector in the MLS is a function of $\boldsymbol{\theta}_i$ since the interpolated points are weighted by their distance to the support points. The expression is given below:

$$\hat{f}(\boldsymbol{\theta}) = \sum_{i=1}^{NB} b_i(\boldsymbol{\theta}) a_i\{\boldsymbol{\theta}\} = \mathbf{b}(\boldsymbol{\theta})^T \mathbf{a}\{\boldsymbol{\theta}\} \tag{4.1}$$

where $\mathbf{b}(\boldsymbol{\theta})$ and $\mathbf{a}\{\boldsymbol{\theta}\}$ are the vector of basis functions and the corresponding coefficients, respectively. *NB* denotes the total number of elements in $\mathbf{b}(\boldsymbol{\theta})$ and $\mathbf{a}\{\boldsymbol{\theta}\}$. A common choice of a second order polynomial basis functions (i.e. $\mathbf{b}(\boldsymbol{\theta}) = [1 \ \theta_1 \ldots \theta_{n_\theta} \ \theta_1^2 \ \theta_1\theta_2 \ldots \theta_{n_\theta}^2]^T$) would give *NB* as follows:

$$NB = \frac{n_\theta(n_\theta + 3) + 2}{2} \tag{4.2}$$

where n_θ is the spatial dimension (Fries & Matthies 2003). The unknown coefficients $\mathbf{a}\{\boldsymbol{\theta}\}$ can be determined by minimizing the weighted sum of square error $J_R\{\boldsymbol{\theta}\}$ at supporting points. Unlike the global least squares method that evaluates the model without considering the relative importance of information near the interpolated point, the MLS metamodel involves a weight function $w\{\boldsymbol{\theta}\}$ which helps to construct the model in a local approximation manner. The idea is to compute the weight function for any arbitrary fixed point $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ and compare it with the entire variable field by measuring the distance between the interpolated point and the supporting points $\{\theta_I; I = 1, \ldots, NS\}$ expressed as $d(\boldsymbol{\theta}; \boldsymbol{\theta}_I) = \|\boldsymbol{\theta} - \boldsymbol{\theta}_I\|$

where $\|\cdot\|$ denotes a norm in $\mathbb{R}^{n_\theta}$. The weighted sum of square error $J_R\{\boldsymbol{\theta}\}$ for MLS is thus expressed as

$$\begin{aligned} J_R\{\boldsymbol{\theta}\} &= \sum_{I=1}^{NS} w\{\boldsymbol{\theta}\}\left[\boldsymbol{b}(\boldsymbol{\theta}_I)^T \boldsymbol{a}\{\boldsymbol{\theta}_I\} - f(\boldsymbol{\theta}_I)\right]^2 \\ &= [\mathbf{Ba}\{\boldsymbol{\theta}\} - \mathbf{F}]^T \mathbf{W}\{\boldsymbol{\theta}\}[\mathbf{Ba}\{\boldsymbol{\theta}\} - \mathbf{F}] \end{aligned} \tag{4.3}$$

$$\text{where } \begin{cases} \mathbf{B} &= [\mathbf{b}(\boldsymbol{\theta}_1)\ldots\ \mathbf{b}(\boldsymbol{\theta}_{\mathrm{NS}})]^{\mathrm{T}} \\ \mathbf{F} &= [f(\boldsymbol{\theta}_1)\ldots f(\boldsymbol{\theta}_{\mathrm{NS}})]^{\mathrm{T}} \\ \mathbf{W}\{\boldsymbol{\theta}\} &= \mathrm{diag}[w(d(\boldsymbol{\theta};\ \boldsymbol{\theta}_1))\ldots\ \mathrm{w}(d(\boldsymbol{\theta};\ \boldsymbol{\theta}_{NS}))] \end{cases}$$

The minimization of Eqn. (4.3) yields the following solution:

$$\boldsymbol{a}\{\boldsymbol{\theta}\} = \mathbf{M}^{-1}\{\boldsymbol{\theta}\}\mathbf{G}\{\boldsymbol{\theta}\}\mathbf{F} \tag{4.4}$$

where $\mathbf{M} = \mathbf{B}^{\mathbf{T}}\mathbf{W}\{\boldsymbol{\theta}\}\mathbf{B}$ and $\mathbf{G}\{\boldsymbol{\theta}\} = \mathbf{B}^{\mathbf{T}}\mathbf{W}\{\boldsymbol{\theta}\}$. Thus the approximation function is expressed as

$$\hat{f}(\boldsymbol{\theta}) = \mathbf{b}^{\boldsymbol{T}}(\boldsymbol{\theta})\mathbf{M}^{-1}\{\boldsymbol{\theta}\}\mathbf{G}\{\boldsymbol{\theta}\}\mathbf{F} \tag{4.5}$$

In principle, any function which has a maximum at d = 0 and vanishes outside the influence domain is capable to be used as weight function. Many weight functions have been proposed, e.g. in forms of Gaussian (Levin 1998), exponential, etc. A popular choice of the weight function $w(d(\boldsymbol{\theta};\boldsymbol{\theta}_i))$ is expressed by:

$$w(d) = \begin{cases} \dfrac{e^{-\left(\frac{d}{cD}\right)^{2k}} - e^{-\left(\frac{1}{c}\right)^{2k}}}{1 - e^{-\left(\frac{1}{c}\right)^{2k}}} & \text{if } d < D \\ 0 & \text{if otherwise} \end{cases} \tag{4.6}$$

in which c and k are the shape factors. The distance $d(\boldsymbol{\theta};\ \boldsymbol{\theta}_I)$ between the interpolated point and the supporting points can be measured by computing the weighted quadratic norm denoted as

$$d(\boldsymbol{\theta};\boldsymbol{\theta}_I) = \sqrt{\sum_1^m \left(\theta_i - \theta_{i,I}\right)^2 v_i^2} \tag{4.7}$$

where v_i in the above equation represents the relative weight of each variable $\boldsymbol{\theta}_i$. The weight function for a specific supporting point is non-zero only if the distance

of that point to the interpolated point is within an influence domain. Please refer to Taflanidis & Cheung (2012) for more details. The influence radius D has to be chosen such that enough supporting points are included and yet vanish after the prescribed threshold.

The weight selection highly affects the efficiency of the interpolation function. Traditional MLS response surface approximation method does not consider how the response surface needs to be created or established so as to benefit the stochastic sampling from some target probability distribution, for example, simulating samples according to the target distribution $\pi(\boldsymbol{\theta})$ taking the following form:

$$\pi(\boldsymbol{\theta}) = \frac{z(\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int_{\Theta} z(\boldsymbol{\theta})\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto z(\boldsymbol{\theta})\, p(\boldsymbol{\theta}) \tag{4.8}$$

where $z(\boldsymbol{\theta})$ is some function of the system output which needs to be obtained by running a computational intensive numerical analysis and $p(\boldsymbol{\theta})$ is some probability density function. Taflanidis and Cheung (2012), based on Taflanidis' previous finding (Taflanidis 2009), proposed a relative information entropy to quantify the discrepancies between the marginal target distribution $\pi(\theta_i)$ and the marginal proposal distribution $q(\theta_i)$:

$$D_{re}(\pi(\theta_i) \,||\, q(\theta_i)) = \int_{\Theta_i} \pi(\theta_i) \log\left(\frac{\pi(\theta_i)}{q(\theta_i)}\right) d\theta_i \tag{4.9}$$

The greater the value, the more sensitive a particular parameter is, thus, more important in generating samples following the target distribution. The weights v_i, integrates the relative information entropy and the normalized standard deviation σ_i, into the following expression:

$$v_i = \frac{D_{re}(\pi(\theta_i) \,||\, q(\theta_i))}{\sigma_i} \tag{4.10}$$

where σ_i is the standard deviation of the i-th component supporting points.

4.2.2 Gaussian processes for regression

The Gaussian processes (GP) is defined as a collection of random variables with every point in some input space following Gaussian distribution (Rasmussen &

Williams 2006). The basic theory of Gaussian processes can be dated back to 1940's (Wiener 1949). It has been widely used in spatial statistics field under the name of *kriging* method (Matheron 1973; Ripley 1991). O'Hagan *et al.* (1978) generalized the GP and presented it in a general regression context. The idea of Gaussian processes regression in machine learning was firstly proposed by Williams and Rasmussen (1996). However, it was not commonly used till late 1990's when computational technology facilitated the application of GP in large data set (Tang *et al.* 2010). A brief introduction of GP in regression is given in the following section.

Let us consider *N* realization of responses $f_i = f(\boldsymbol{x}_i),\ i = 1, 2, \dots N$, computed from real-time numerical simulations. Each response corresponds to *D*-dimensional training dataset $\boldsymbol{X} = (\boldsymbol{x}_1^D, \dots, \boldsymbol{x}_N^D)$. A GP regression model considers the deterministic output $f(\boldsymbol{x})$ as a random process which has a prior distribution following with a Gaussian stochastic process. This can be written as:

$$y \sim GP\big(m(x), k(x, x')\big) \tag{4.11}$$

where m(x) and $k(x, x')$ represent the mean function and the covariance function, respectively. One common choice of the mean function is a linear combination of elementary basis functions, which can be written as follows:

$$m(x) = \sum_{j=0}^{k} \beta_j m_j(x) = M(x)\boldsymbol{\beta} \tag{4.12}$$

where $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots \beta_k]^{\mathrm{T}}$ is the coefficient vector, and $M(x) = [m_1(x), \dots m_k(x)]$ represents the basis function.

The covariance function $k(x, x')$ describes the similarity between the two data points according to their nearness and has to be positive semidefinite. Theoretically, any arbitrary positive definite matrix can be adopted as a covariance function (Bailer-Jones 2002). The covariance function is categorized as stationary and non-stationary, in general, depending on whether it is affected by the translations. A widely-adopted stationary covariance function is squared exponential (SE) covariance function which is given as:

$$k_{SE}(x,x') = \sigma_f^2 \exp(-\frac{(x-x')^2}{2l^2}) \tag{4.13}$$

where σ_f and l are called *hyperparameters* following the nomenclature in the neural network method. They define the amplitude and characteristic length-scale, respectively. Other types of stationary covariance functions include rational quadratic (QR), the Matérn class, and piecewise polynomial covariance functions with compact support.

From the perspective of Bayesian inference, the posterior probability of the function $f(x)$ given a set of data $D = \{x_i, y(x_i)\}$ can be derived using:

$$p(f|D) = \frac{p(f)p(D|f)}{p(D)} \tag{4.14}.$$

Therefore, with a set of prediction input variables $\boldsymbol{X}$, the joint distribution of the training output, $\mathbf{y}$, and the prediction function $\mathbf{f}_*$ is given as:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{m}(x) \\ \mathbf{m}_*(x) \end{bmatrix}, \begin{bmatrix} \boldsymbol{K}(X,X) & \boldsymbol{K}(X,X_*) \\ \boldsymbol{K}(X_*,X) & \boldsymbol{K}(X_*,X_*) \end{bmatrix} \right) \tag{4.15}$$

where $\mathbf{m}(\cdot)$ and $\boldsymbol{K}(\cdot)$ are mean and variance functions, respectively. The prediction $m_*(X)$ and $k_*(X,X')$ given by:

$$\begin{aligned} \mathrm{E}[\mathbf{f}_*|\mathbf{D}] &= m_*(X) = m(X_*) + \boldsymbol{K}(X_*,X)^{\mathrm{T}} \boldsymbol{K}(X,X)^{-1}(\boldsymbol{y} - m(X)) \\ \mathrm{Var}[\mathbf{f}_*|\mathbf{D}] &= k_*(X,X') = \boldsymbol{K}(X_*,X_*) - \boldsymbol{K}(X_*,X)^{\mathrm{T}} \boldsymbol{K}(X,X)^{-1} \boldsymbol{K}(X,X_*) \end{aligned} \tag{4.16}$$

The hyperparameters which defines the mean and covariance functions are not known in advance and they can be estimated using one of the statistic inference methods: variographic analysis (VA), maximum likelihood estimation (MLE) or Bayesian estimation (BE) (Dubourg *et al.* 2011). The MLE method is the mostly-used technique which obtains the optimal hyperparameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma_f^2, l\}$ by maximizing the likelihood function $\mathcal{L}(\boldsymbol{\theta})$. In practice, it is easier to minimize the negative log marginal likelihood $\mathcal{L}(\theta)$ which is given as follow:

$$-\log p(\mathbf{y}|x,\boldsymbol{\theta}) = \frac{1}{2}(\mathbf{y} - m(X))^{\mathrm{T}} \boldsymbol{K}_y^{-1}(\mathbf{y} - m(X)) + \frac{1}{2}\log|\boldsymbol{K}_y| + \frac{N}{2}\log(2\pi) \tag{4.17}$$

where $\boldsymbol{K}_y$ refers to $\boldsymbol{K}(X, X)$ in (4.16). The local optimal solution can be derived by taking the gradient of the marginal likelihood function with respect to the individual hyperparameter θ_j which can be written as:

$$\frac{\partial \mathcal{L}}{\partial \theta_j} = -\frac{1}{2}(\mathbf{y} - m(X))^{\mathrm{T}} \boldsymbol{K}_y^{-1} \frac{\partial \boldsymbol{K}_y}{\partial \theta_j} \boldsymbol{K}_y^{-1}(\mathbf{y} - m(X)) - \frac{1}{2}\mathrm{tr}(\boldsymbol{K}_y^{-1} \frac{\partial \boldsymbol{K}_y}{\partial \theta_j}) \qquad (4.18)$$

where tr( ) is trace of the squared matrix.

4.2.3 Remarks

Although both the MLS and GP belong to the response surface method, there are some differences between the two. First of all, in GP, all the training data are utilized and their covariance functions have to be calculated, whereas in MLS, only those within the influence radius will be selected. It indicates that the influence radius is a function of the training data with the proposed stochastic sample being the centre of the circle. This will not only result in a difference in computational time, but more importantly, the quality of the constructed regression function. To be more specific, a subset of the training data for GP has to be carefully selected. Researchers have studied on how to choose the subset of data for GP. The most straightforward way is to choose randomly which turned out to be effective but lack of accuracy (Snelson 2007). More advanced selecting algorithms were proposed, such as projected latent variables (Seeger *et al.* 2003), informative vector machine (Lawrence 2005), and sparse pseudo-input Gaussian processes (Snelson & Ghahramani 2006), to name just a few. The majority of the above-mentioned studies involve constructing various types of information criterion after adding or deleting data from the subset.

A unique feature of the GP model is that it is a probabilistic model which allows one to estimate not only the mean, but the uncertainty on predictions as well. However, this requires the formation of the covariance function and optimization of its hyperparameters. Although both the MLS and GP methods contain O (N^3) operations in their standard form, it is easier to fix the free parameters in MLS (Krishnamurthy & Romero 2002).

4.3 Proposed adaptive SS-MLS and SS-GP algorithms for reliability analysis[1]

In this section, a novel framework is proposed to estimate the reliability of a structure under earthquake-induced tsunami impact and at the same time alleviate the computational burden and yet accurately assess the structural reliability.

Since the metamodels are able to transform the original engineering simulation problem into a pure mathematical function, it can be utilized to achieve the system reliability assessment purpose without computing as many times as it requires in original MCS and Subset Simulation. Similar to the SS, the modified Metropolis-Hasting algorithm facilitates the implementation of reliability since the local random walk allows the stochastic samples to explore the failure domains rather than the entire boundary. Consequently, a novel integration of the Subset Simulation (SS) method (Au & Beck 2001) and two metamodels: (1) modified MLS approach (Taflanidis & Cheung 2012); and (2) GP regression model are proposed. The proposed algorithm can be summarized as follows:

1. Denote the number of MCS of the numerical process by *n*, the number of metamodel samples for each intermediate failure level by *M*, and the prescribed conditional probability by p_0.
2. Generate *n* samples $\boldsymbol{\theta}$ following $\boldsymbol{p}(\boldsymbol{\theta})$ using MCS and performing numerical simulations of the physical process to produce the response samples. The *n* samples are used later as supporting points for the metamodels.
3. Construct MLS/GP metamodels using the *n* supporting samples.
    a. For MLS, a subset of *n* samples which exceed a prescribed threshold are considered to follow the target distribution in order to estimate the entropy (Taflanidis & Cheung 2012).
4. Generate *M* stochastic samples $\boldsymbol{\theta}$ from $\boldsymbol{p}(\boldsymbol{\theta})$ using MCS with the limit state function computed using the metamodels.
5. Sort the *M* samples in a descending order according to the responses computed by the metamodels and select a certain number of samples (say *m*,

[1] The proposed methodology presented in this section is partially published in the 11th International Conference on Structural Safety & Reliability, New York, 2013. Cheung, S. H. and S. Zhe (2013).

m<M) that exceed a prescribed threshold to carry out the original numerical simulations.

6. Aggregate the n samples in step 2 and the newly generated m samples in step 5 as new set of supporting samples.
7. Construct MLS/GP metamodels using $n + m$ samples.
   a. For MLS, the samples with the responses exceeding a prescribed threshold are used to estimate the entropy. The influence radius D is updated.
   b. For GP, a subset of $n + m$ samples is chosen as the supporting samples.
8. Generate M stochastic samples $\boldsymbol{\theta}$ from $\boldsymbol{p}(\boldsymbol{\theta}|\boldsymbol{F_i})$ using the modified M-H algorithm with the first Mp_0 samples in Step 5 as the seeds. The limit state function is computed using the metamodels.
9. Repeat steps 5-8 until a desired failure probability is reached.

Compared to the SS framework presented in Section 4.1, the most distinguishable feature of the proposed procedure is Step 5 where the samples generated by considering metamodels $\tilde{f}(\boldsymbol{\theta})$ that are more likely to lie in the failure domain (i.e. response exceeding a prescribed threshold) are replaced by the actual original numerical simulation $f(\boldsymbol{\theta})$. The samples generated by the metamodel which are "away" from the limit state function are kept as it is. The reason behind it is that the straightforward use of the metamodels may lead to erroneous result, no matter how many samples are produced from the metamodels. The finite accuracy of the metamodel, to large extend, does not affect the accuracy of the failure probability estimates since the samples close to the limit state is evaluated using the original simulation. On the other hand, the total number of original simulation is much less than the MCS and SS.

Regarding the subset of supporting point mentioned in Step 7(b), the training point selecting mechanism is used because not all the samples obtained from the numerical simulation can be taken as the support points. It is due to the fact that GP takes into account all the training data feeding to it when estimating the approximation (both mean and covariance functions). This is, to a certain extent,

acting as a global regression which could result in an over or underestimation of the approximation, depending on whether the support points are "far away" from the interpolation points. According to Eqn.(4.2), a minimum number of supporting points, decided by the number of basis terms (*NB*), have to be employed to avoid singularity in the solution for the coefficients of GP. Furthermore, the thresholds derived from the SS algorithm naturally divide the supporting points into various classes. In fact, as the purpose of the metamodel is to approximate limit state function in the next intermediate failure level, the supporting points which have the responses exceeding the latest threshold (refer to as N_+) have priority to be selected as the support points for constructing the GP for next conditional failure level. Considering all the above-mentioned factors, the following mechanism is proposed. All the supporting points are sorted in descending order based on the maximum structural displacements. The support points with their responses higher than the current threshold value (N_+) are kept. If the number of N_+ is more than or equal to the number of basis terms ($NS_+ \geq NB$), the selection is completed: only the N_+ are included for constructing the GP model in the next conditional failure level. Otherwise, another portion of support points is randomly selected from the rest of the support points (N_-). The flowchart shown in Figure 29 summarizes the training points selecting framework.

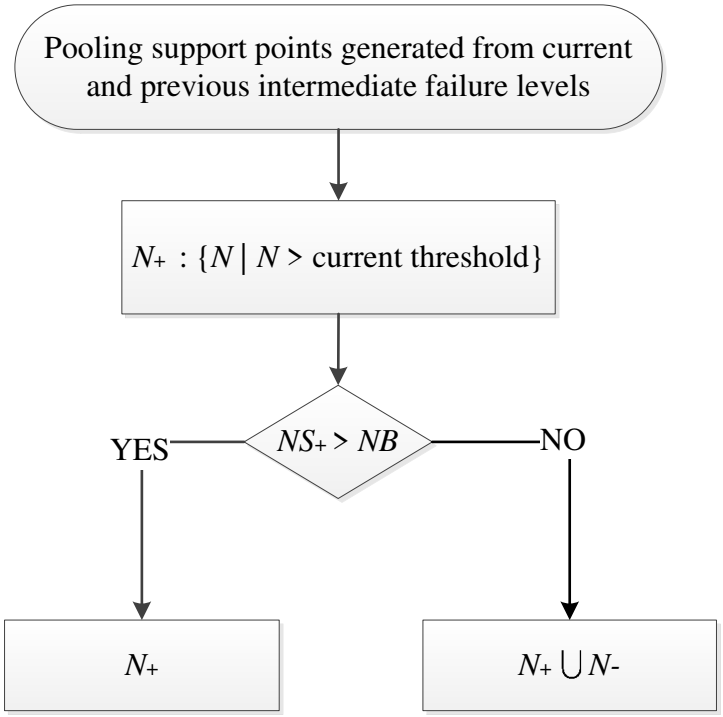


Figure 29 Training data selecting process in GP method

4.4 Illustrative example

4.4.1 SS-MLS and SS-GP

In this illustrative example, the reliability analysis of a structure under tsunami wave impact is presented using the proposed algorithms. For illustration, the limit state function $L(\boldsymbol{\theta})$ considered here is equal to t - $d(\boldsymbol{\theta})$ where t is the threshold corresponding to some pre-specified failure probability and $d(\boldsymbol{\theta})$ is the maximum displacement of the structure during the wave-structure interaction. Different ways of integrating the modified MLS and Gaussian Processes with Subset Simulation method are described below.

4.4.1.1 SS-MLS results and discussion

For comparison purpose, a non-adaptive way of using the metamodel in reliability assessment is firstly presented: a metamodel is constructed using the samples generated by the numerical simulations. The same metamodel is used throughout the reliability assessment process (e.g. Monte Carlo or Subset Simulation) to evaluate the limit state function. In fact, this is equivalent to steps 1-3 in the proposed algorithm. The detail of this counterexample is as follows. A total of 500

supporting points (NS = 500) are used to construct the full quadratic basis MLS. The relative information entropy for the MLS is formed using the samples which lead to the top 10% (p_0 = 0.1) of the maximum structural displacement, *i.e.,* $\boldsymbol{\theta}$ distributed according to $p(\boldsymbol{\theta}|F_1)$ and are considered as samples from $\pi(\boldsymbol{\theta})$ in MLS. For each intermediate level of SS, a number of M = 2,000 samples, including Mp_0samples from the previous intermediate failure level, are generated using the modified M-H algorithm. Therefore, the number of new samples produced at each conditional failure level is $M - Mp_0$ = 1,800. The constructed MLS are employed to replace the numerical simulation for the limit state function. The above version of the SS-MLS is named ori-SSMLS to distinguish from the proposed algorithm.

A total of 50 independent sets of simulations were carried out using the ori-SSMLS algorithm. The mean failure probability curve considering the 50 runs is plotted in Figure 30. The result is compared with the benchmark study. The fitness (error) is defined as the failure probability derived from the SS method divided by that from the SS-MLS approach at various threshold levels. A value of 1 indicates the performance is the same as the benchmark result. The fitness of the ori-SSMLS is 23.8 at the second conditional failure level and beyond comparison (Inf.) at the third level. As expected, deterioration develops with increasing of failure levels in the performance of the ori-SSMLS. It is mainly due to the fact that a lack of support points in the failure region corresponding to higher threshold levels and smaller failure probability leads to the underestimation of the maximum displacement in this problem.

To address this issue, a portion of the samples Mp_0 out of M samples distributed according to $p\left(\boldsymbol{\theta} \mid F_j\right)$ $\{j : j = 1, 2, 3\}$ which lead to the top 10% of the maximum displacement are selected and their maximum displacements are computed using the original physical numerical simulation, instead of using the metamodel. The basic idea is to gain more supporting samples lying in the failure domain in order to acquire a more accurate approximation function for the subsequent failure levels. As stated in the procedure of the proposed algorithm, the modified MLS has to be constructed at each intermediate failure level (denoted as MLS-lv$j$) since the

supporting points are expanding with the decreasing of the failure probability. Furthermore, the influence radius of the MLS is updated accordingly to reflect the change in supporting points. The integration of SS and MLS in adaptive manner described above is named adv-SSMLS.

Similarly, a total of 50 independent sets of simulations using adv-SSMLS algorithm were carried out. The mean failure probability curve of the adv-SSMLS over 50 runs is shown in Figure 30. As indicated from the figure, the estimation errors of adv-SSMLS are stable throughout different intermediate failure levels compared to the ori-SSMLS. A slightly lower fitness value of adv-SSMLS at the first threshold level is primarily due to the less supporting points used (NS = 200) compared to the ori-SSMLS (NS = 500). The fitness corresponding to the second and third threshold levels of adv-SSMLS are 1.86 and 1.40, respectively, which is a significant improvement compared to the result obtained from the ori-SSMLS algorithm. The above final version of the approach is taken as the proposed approach for the SS-MLS algorithm.

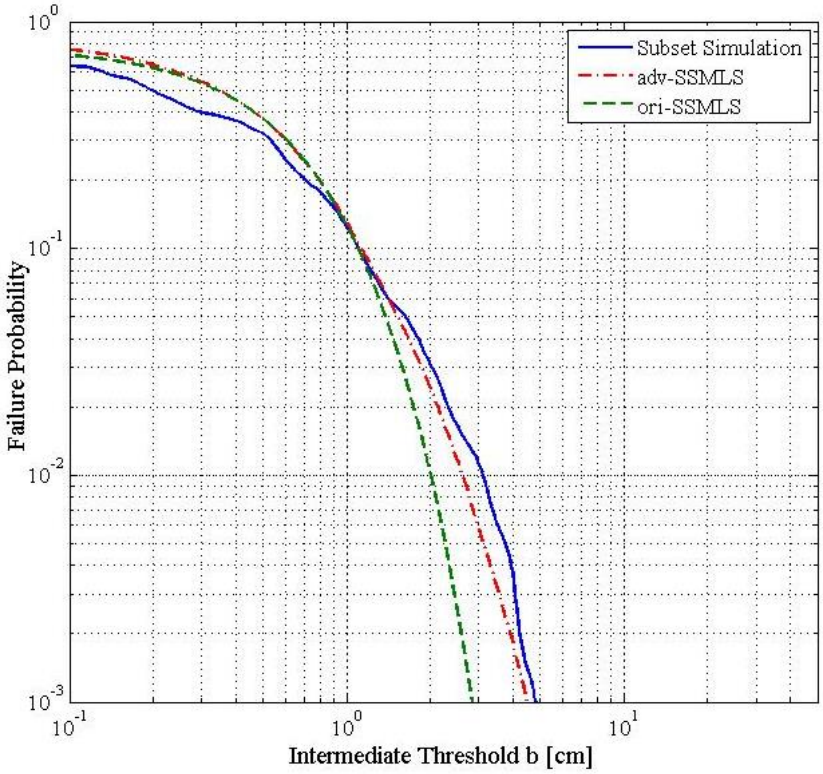


Figure 30 Failure probability estimates of SS, ori-SSMLS and adv-SSMLS

4.4.1.2 SS-GP results and discussion

The same procedure as described and implemented in section 4.4.3.1 for SSMLS is applied to the integration of SS with Gaussian processes. A total of 500 supporting points (NS = 500) are used to construct the GP. The same metamodel is used throughout the computation of the failure probability, *i.e.,* the non-adaptive SSGP or ori-SSGP. For each intermediate level, a number of M = 2,000 samples are simulated. The constructed GP are employed to replace the numerical simulation for the limit state function.

Concerning the adv-SSGP, *i.e.*, adaptive SSGP, a number of 200 supporting samples (NS = 200) are selected for the first level of the intermediate failure probability. Similar to the adv-SSMLS, a portion of the samples which lead to the top 10% of maximum displacement as computed by the Gaussian Processes are selected and their maximum displacements are computed using the original numerical simulation. After the proposed training point selecting mechanism presented earlier, a portion of the supporting points are activated for constructing the GP metamodel of the next intermediate failure conditional level. Repeating the process until the designed failure probability is computed.

After 50 independent running of the ori-SSGP and adv-SSGP algorithms, the mean failure probability over 50 runs is plotted in Figure 31 together with the benchmark result. Besides, Figure 31 also includes the 95-percentile and 5-percentile estimation which are derived from the covariance matrix in the GP method. As expected, the performance of adv-SSGP is consistent in every intermediate level, while the ori-SSGP underestimates the result from the second conditional level onwards. A relative better estimation of the ori-SSGP in the first level gives rise to the 500 support points used compared to the 200 in adv-SSGP. An important observation is that the benchmark result is always inside the upper and lower bounds of the adv-SSGP result which indicates a reliable performance of the GP model.

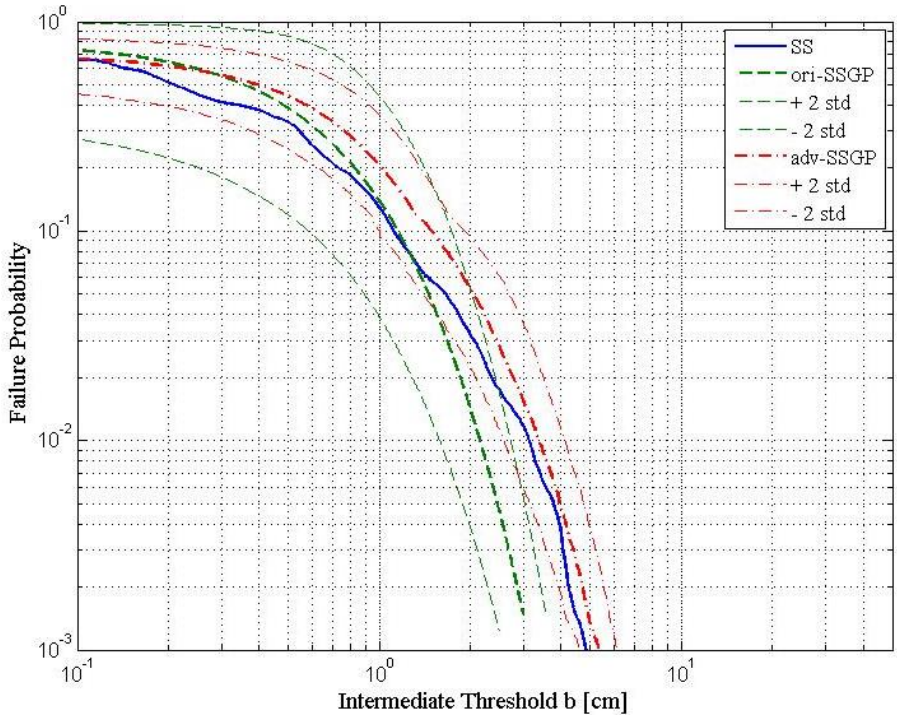


Figure 31 Failure probability estimates of SS, ori-SSGP and adv-SSGP with their associated uncertainty levels

4.4.1.3 Comparison between adv-SSMLS and adv-SSGP algorithms

In this section, the results from adv-SSMLS and adv-SSGP algorithms stand side by side for cross comparison. The effectiveness and computational efforts of the algorithms are evaluated. The results from non-adaptive version of the integration, *i.e.*, ori-SSMLS and ori-SSGP, are not included since the accuracy of these two algorithms is incomparable to the adaptive version. As mentioned in the previous sections, 30 independent runs were conducted for the benchmark and 50 for adv-SSMLS and adv-SSGP algorithms each. The c.o.v. for all the algorithms are presented in Figure 32. It can be seen that the variation of the adv-SSMLS algorithm is the largest among the three between the failure probabilities of 10^{0} – $10^{-2.5}$. The c.o.v. of the benchmark study rises after $10^{-2.5}$ and remains the highest up to $P_F = 10^{-3}$. On the other hand, the uncertainty level of adv-SSGP algorithm is consistently lower than the others with a mild uptrend all the way.

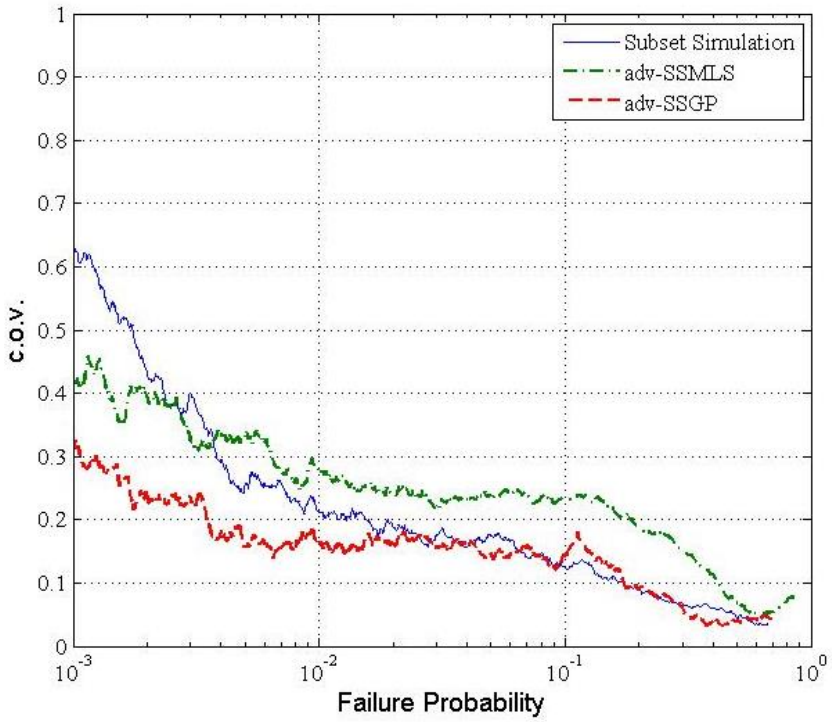


Figure 32 The c.o.v. of failure probability estimator for SS, SSMLS and SSGP

The exact values of the c.o.v. for the SS, adv-SSMLS, and adv-SSGP algorithms are listed in Table 9. A comparison of the computation efficiency of the three algorithms is also shown in Table 9. The "Nb. Smln" represents the total number of numerical evaluations used for each algorithm. Due to the nature of the MCMC algorithm, repetitive samples are unavoidable from the second conditional failure level onwards. Therefore, the total number of simulations for each run of the SSMLS and SSGP algorithms is subject to change and the number shown in the table is the averaged value from the 50 independent runs. The c.o.v. value δ at the target failure probability (P_F = 10^{-3}) is given. The unit c.o.v., denoted by $\Delta = \delta\sqrt{N_T}$ (Au & Beck 2003), is also listed in the table. This factor reflects the characteristic of the algorithm – the smaller the value, the more efficient the algorithm is. The adv-SSGP has the highest efficiency score among the three, on account of a lower number of the simulation and the lowest c.o.v. The adv-SSMLS is the second in terms of the efficiency owing to the least numerical simulation used, even though the c.o.v. is higher than SS.

Table 9 Comparison of the computation efficiency

$P_F = 10^{-3}$	Nb. Smln	δ	Δ
Subset Simulation	1500	0.987	38.226
adv-SSMLS	581	1.066	25.695
adv-SSGP	666	0.780	20.129

4.5 Closing remarks

In this chapter, an innovative algorithm which integrates the Subset Simulation with the second-order metamodels, *i.e.* MLS and GP, are proposed. The adaptive features of the algorithms enable the metamodels to update the support points and influence radius so as to better estimate the samples lie within failure domain corresponding to rarer events. The results obtained from the adaptive integration show great enhancement in terms of prediction accuracy compared to the naïve non-adaptive integration. They also reduce the total number of numerical simulations required to nearly one third of that for the benchmark. The efficiency of the algorithms is estimated based on the unit c.o.v. derived from 50 independent runs of the proposed algorithms. Some observations as well as a few constraints made in this chapter are summarized below.

First of all, the choice of 200 support points to start with for the adv- version of the SS-MLS and SS-GP methods is an ad hoc selection. The bottom line is to choose a number higher than the number of basis so as to avoid singularity issue in estimating the coefficients. Although additional 200 support points, *i.e.* 10% of the 2000 samples generated by metamodels, are supposed to be evaluated for each of the intermediate failure level, the 200 samples contain repetitive ones due to nature of the MCMC algorithms. More repetitive samples are observed with increasing intermediate failure probability level due to a higher rejection rate. Thus less numerical simulations are required. In fact, in each intermediate failure level of adv-SSMLS and adv-SSGP, increasing the sample size to a huge number ($M >> 2{,}000$) but restricting the number of additional support points can be used to bring down the c.o.v. of the failure probability estimator, since the computational time of the metamodel is negligible compared to the physical model simulation.

Secondly, it can be seen that the proposed SS-GP algorithm tends to overestimate the structural performance given the same failure probability. This indicates that the overestimation is related to the quality of the support points and thus, the nature of the training point selection mechanism proposed at the end of section 4.3. As mentioned in the section, Gaussian processes (GP) is different from moving least square (MLS) where the former takes all the support points into account when making a prediction. The proposed selection mechanism consists of two types of training points: those within the failure domain and those outside. In the first a few intermediate failure levels, a relatively large portion of the training points are those outside the failure domain due to the fact that only the top 10% samples are failure samples. This is gradually altered with the increasing number of samples. Although accurate failure estimation is undoubtedly the best result, overestimation which leads to a conservative measure is not unacceptable in reliability analysis when disaster management is the main concern.

According to the simulation results, both the proposed algorithms perform well compared to the benchmark study. The computational time of SS-MLS is faster than that of SS-GP as the latter needs to construct subset training data selection. However, the covariance matrix of GP enables the uncertainty to be quantified in failure probability estimate even with a single simulation.

Chapter 5 Higher Dimension Metamodel in Reliability Analysis

In Chapter 4, the adaptive SSMLS and SSGP algorithms are proposed and illustrated with an illustrative example of 16 random variables, *i.e.* earthquake subfault slip direction, is illustrated. Although response surface models offer comparative reliability estimation with higher efficiency compared to the original Subset Simulation method, they suffer from the *curse of dimensionality*. Especially in complex systems where the first and second-order functions are not adequate and higher order is necessary.

Furthermore, in real-time stochastic simulation, the earthquake parameters besides slip direction, such as earthquake magnitude, focal depth, and rake angle etc. are all subject to vary. Furthermore, the structural parameters are essential and their uncertainty should be considered in the stochastic simulation. Thus, the number of the random variables can easily exceed a hundred, which requires over 5,000 samples in order to formulate, for example, a second order approximation in MLS or the second order mean function in GP. Not to mention using a higher order polynomials for the basis function.

5.1 Sparse Bayesian modeling

In general, the training phase in the regression process requires $O(N^3)$ computational efforts where N denotes the number of data points. Due to the characteristic of the physical model, N can be very large. However, it turns out that the influence of an individual basis in a polynomial is heterogeneous and can be quantified using various methods. Introducing sparsity in the basis function can reduce the model complexity as well as avoid over-fitting problem. Furthermore, it minimizes the computational resource required. In recent years, attempts have been made for the sparse approximation, such as sparse pseudo-input Gaussian processes (SPGPs) (Snelson & Ghahramani 2006), automatic relevance determination (ARD) (Neal 1995), and relevance vector machine (RVM) (Tipping 2001) etc. In these studies, each coefficient in the basis function was assigned with a weight and only the basis with non-zero weight is kept in the function.

In this section, the sparse Bayesian modeling approach is integrated with the SSMLS and SSGP in order to include more random variables which render the simulation of the physical phenomenon more practical. Apart from this, higher order basis functions will be tested using the same number of variables as in Chapter 4 and the results will be discussed at the end of this chapter.

5.1.1 Sparse Bayesian Inference

Consider a standard linear regression problem for $i = 1, \ldots, N$ with prediction error ε, which gives:

$$y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \tag{5.1}$$

where $\boldsymbol{\beta}$ is the parameter vector and the $\boldsymbol{\varepsilon}_i$ is assumed to be normally distributed random variables: $\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2), i = 1, \ldots, N$. This error term indicates a probability model which corresponds to the following multivariate Gaussian likelihood function:

$$p(\mathbf{y}\,|\,\mathbf{X}, \boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2}{2\sigma^2} \right) \tag{5.2}$$

The prior probability $p(\boldsymbol{\beta})$ is chosen as a Gaussian distribution to reflect the uncertainty of the parameter vector. It is given by:

$$p(\boldsymbol{\beta}\,|\,\alpha) = (2\pi)^{-M/2} |\mathbf{A}|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{A}\boldsymbol{\beta} \right) \tag{5.3}$$

where diagonal matrix $\mathbf{A} = \mathrm{diag}(\alpha_1, \alpha_1, \ldots, \alpha_M)$ with each hyperparameter α_M controls the weight of the corresponding prior. The introduction of $\boldsymbol{\alpha} = [\alpha_1 \ldots \alpha_M]^{\mathrm{T}}$ enables the function to ultimately prune out the irrelevant basis terms and gains the sparsity property (Tipping 2001). Chang *et al.* (2008) categorized the Gaussian prior into three cases, in which the ARD prior corresponds to Tipping's approach. The conditional posterior distribution of the parameter $\boldsymbol{\beta}$ given data can be derived by combining the likelihood function and prior probability distribution of the parameter according to Bayes theorem, which is given by:

$$p(\boldsymbol{\beta}\,|\,\mathbf{y}, \boldsymbol{\alpha}, \sigma^2, \mathbf{X}) = \frac{p(\mathbf{y}\,|\,\boldsymbol{\beta}, \sigma^2, \mathbf{X})\, p(\boldsymbol{\beta}\,|\,\boldsymbol{\alpha})}{p(\mathbf{y}\,|\,\boldsymbol{\alpha}, \sigma^2)} \tag{5.4}$$

It is proven to follow $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (\mathbf{A} + \sigma^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ and $\boldsymbol{\mu} = \sigma^{-2}\Sigma\mathbf{X}^{\mathrm{T}}\mathbf{y}$.

Instead of extending the hyperparameters into Bayesian inference, Bayesian model class selection is adopted to estimate the most probable ones. To estimate the most probable model class $\boldsymbol{M}(\boldsymbol{\alpha}, \boldsymbol{\sigma}^2)$ condition on data $\mathbf{y}$ is equivalent to finding the hyperparameters which maximize the corresponding posterior probability. This can be derived using the Bayes theorem with the evidence $\boldsymbol{p}(\mathbf{y}|\boldsymbol{M})$ and the prior $\boldsymbol{p}(\boldsymbol{M})$. Under the assumption that the prior probability of selecting each model class is equal, the problem is reformulated to maximise the evidence or its equivalent, logarithm $L(\boldsymbol{\alpha}, \sigma^2)$:

$$\begin{aligned} L(\boldsymbol{\alpha}, \sigma^2) &= \log p(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2) = \log \int p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}|\boldsymbol{\alpha}) d\boldsymbol{\beta} \\ &= -\frac{1}{2}[N \log 2\pi + \log|C| + \mathbf{y}^{\mathrm{T}} C^{-1} \mathbf{y}] \end{aligned} \tag{5.5}$$

where $C = \sigma^2\mathbf{I} + \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^{-\mathrm{T}}$. By proper mathematical and matrix manipulation, the $L(\boldsymbol{\alpha},\sigma^2)$ becomes

$$\begin{aligned} L(\boldsymbol{\alpha}, \sigma^2) = -\frac{1}{2}[&N \log 2\pi + \log|C_{-i}| + \mathbf{y}^{\mathrm{T}} C^{-1} \mathbf{y} \\ &- \log \alpha_i + \log(\alpha_i + \mathbf{X}_i^{\mathrm{T}} C_{-i}^{-1} \mathbf{X}_i) - \frac{(\mathbf{X}_i^{\mathrm{T}} C_{-i}^{-1} \mathbf{y})^2}{\alpha_i + \mathbf{X}_i^{\mathrm{T}} C_{-i}^{-1} \mathbf{X}_i}] \\ = \ell(\boldsymbol{\alpha}_i, \sigma^2) &+ L(\boldsymbol{\alpha}_{-i}, \sigma^2) \end{aligned} \tag{5.6}$$

where $\boldsymbol{\alpha}_i = [\alpha_1 \ldots \alpha_{i-1}, \alpha_{i+1}, \ldots \alpha_M]^{\mathrm{T}}$. By setting this way, the maximum log likelihood is decomposed into two parts: the contribution from the i^{th} basis vectors $\ell(\alpha_i, \sigma^2)$ and the rest of the basis terms. The details in the derivation of Eqn. (5.6) can be found in both Tipping and Faul (2003) and Chang (2008) studies.

5.2 Proposed SB-SSMLS for higher order metamodels and higher dimensional random variables

Many reliability problems involving complex systems have adopted metamodels in order to reduce the computational time (Bailer-Jones, C. A., 2002, Nobile *et al.* 2008, Tang *et al.* 2010). However, as illustrated in Chapter 4, the accuracy of the metamodels deteriorated due to a lack of support points in low probability region. The proposed algorithms in the previous chapter try to solve the problem by

providing additional support points adaptively. The results have shown a great amount of improvement on the accuracy of the approximation. Note that all the previously used metamodels are second order functions due to a limited number of support points that can be afforded computationally. Since the response of a complex system with many random variables tends to be highly nonlinear, a lower order (first or second) metamodel may not be sufficient to handle such problem. Hence, an alternative is to increase the nonlinearity of the basis functions in the metamodels to allow them to apply to much complicated responses.

In this section, a sparse Bayesian-based metamodel is proposed. First, it copes with higher order (third and fourth) metamodels. The same benchmark study used in Chapter 4 is employed to demonstrate the efficiency of the proposed algorithm. Secondly, the proposed algorithm is adopted to include more random variables, including both structural and earthquake ones to make the reliability analysis more robust.

The algorithm starts with constructing the metamodel functions as described in Chapter 4. However, before estimating the coefficients for the basis function, the hyperparameters α_i for each basis term is calculated and updated sequentially by equalling the partial derivative of Eqn. (5.6) with respect to the corresponding α_i to zero. This is given as:

$$\frac{\partial L(\boldsymbol{\alpha},\sigma^2)}{\partial \alpha_i} = \frac{1}{\alpha_i} - \frac{1}{\alpha_i + s_i} - \frac{q_i^2}{\alpha_i + s_i} = 0 \tag{5.7}$$

which yields the unique solution:

$$\begin{cases} \alpha_i = \dfrac{s_i^2}{q_i^2 - s_i}, & if \quad q_i^2 > s_i \\ \alpha_i = \infty, & if \quad q_i^2 \le s_i \end{cases} \tag{5.8}$$

where $s_i = \mathbf{X}_i^T C_{-i}^{-1} \mathbf{X}_i$ and $q_i = \mathbf{X}_i^T C_{-i}^{-1} \mathbf{y}$. Therefore, only the terms with $\theta_i = q_i^2 - s_i > 0$ and a finite α_i value should be included in the basis function. The coefficients for the selected terms are then derived. This procedure is iteratively done until every basis term is evaluated. The framework is summarized in Figure 33.

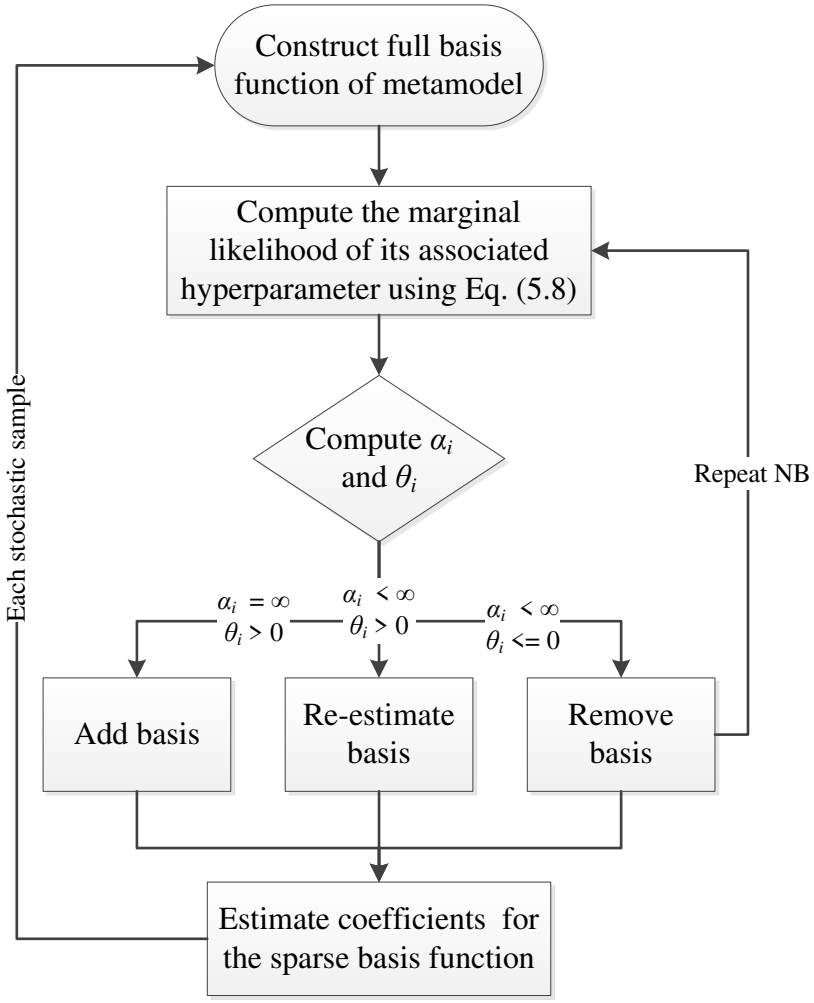


Figure 33 Sparse Bayesian metamodel framework

An extension of Tipping's work (2001) focuses on the modification of the variance term σ in C. In the MLS context, the errors of the support points are not i.i.d. samples. Therefore, the C is no longer expressed as $C = \sigma^2\mathbf{I} + \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^{-\mathrm{T}}$ in the previous section. Instead, the variance for the input vectors is represented by the inverse of the variable weight function $\boldsymbol{W}(c,k) = w(d(\boldsymbol{\theta};\ \boldsymbol{\theta}_i))$ as in Eqn.(4.6) which mainly depends on the distance between the interpolation point and the input vectors. For the simplicity of the expression, here it is defined that $\boldsymbol{\omega} = \boldsymbol{W}^{-1}(c,k)$. Thus, the matrix is transformed into $\boldsymbol{C} = \boldsymbol{\omega} + \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^{-\mathrm{T}}$. For reasons that will become clear later, the evidence maximization with respect to z which represents both c and k is written in the form:

$$\frac{\partial L}{\partial z} = -\frac{1}{2}\frac{\partial}{\partial z}\left[\log|\boldsymbol{\omega}| + \mathbf{y}^{\mathrm{T}}\boldsymbol{\omega}^{-1}\mathbf{y}\right] \tag{5.9}$$

In the equation, the derivative of the log determinant of a positive definite matrix $\boldsymbol{\omega}$ is given by $\frac{\partial \log|\boldsymbol{\omega}|}{\partial z} = \boldsymbol{\omega}^{-1}$ and the derivative of the inverse of the matrix $\frac{\partial \boldsymbol{\omega}^{-1}}{\partial x} = -\boldsymbol{\omega}^{-1}\frac{\partial \boldsymbol{\omega}}{\partial x}\boldsymbol{\omega}^{-1}$. Therefore, the gradient of the evidence is transformed into:

$$\frac{\partial L}{\partial x} = \frac{1}{2}\left(\boldsymbol{\omega}^{-1} - \boldsymbol{\omega}^{-1}t^{\mathrm{T}}\boldsymbol{\omega}^{-1}t\right)\frac{\partial \boldsymbol{\omega}}{\partial x} \tag{5.10}$$

By defining $S = c^{-2k}$ and $T = (\frac{cD}{d})^{-2k}$, the partial derivatives of the weight function with respect to *c* and *k* are given by:

$$\begin{cases} \dfrac{\partial \boldsymbol{w}}{\partial c} = -\dfrac{2kc^{-2k-1}e^{S-T}\left(-e^{S}\left(\dfrac{d}{D}\right)^{2k} + e^{T} + \left(\dfrac{d}{D}\right)^{2k} - 1\right)}{\left(1-e^{S}\right)^{2}} \\ \dfrac{\partial \boldsymbol{w}}{\partial k} = -\dfrac{2e^{-S}S\log(c)\left(e^{-T} - e^{-S}\right)}{\left(1-e^{-S}\right)^{2}} - \\ \qquad \dfrac{2e^{-S}c^{-2k}\log(c) + 2Te^{T}(\log(c) - \log(d) + \log(D))}{1-e^{-S}} \end{cases} \tag{5.11}$$

The above equations are incorporated into the sparse Bayesian scheme for the high dimensional variable studies presented in Section 5.3.2.

5.3 Illustrative example

In this section, integrating the sparse Bayesian method with the SSMLS and SSGP is illustrated. In Section 5.3.1, third- and fourth-order polynomial basis functions are utilized for the SSMLS and SSGP algorithms without computing additional numerical simulations at each intermediate failure level, *i.e.* ori-SSMLS and ori-SSGP. In Section 5.3.2, the sparse Bayesian method is integrated with the adaptive SSMLS algorithm proposed in Chapter 4 for evaluating the reliability of a system with a total of 53 random variables.

5.3.1 Higher-order SB-SSMLS and SB-SSMGP models

The sparse Bayesian inference greatly reduces the number of terms in the basis function. Thus it makes constructing higher order basis function possible for the metamodels. This feature enables the metamodel to create a more flexible and robust response surface for a complex system without using a large amount of training data which are often not affordable in practice.

5.3.1.1 SB-SSMLS

Before constructing a higher-order SSMLS using sparse Bayesian framework, it is worthwhile to compare the performance of the second-order SSMLS with and without sparse Bayesian framework. In order to compare the results obtained from different versions of the algorithms, 500 support points are used to construct the second order MLS. 50 independent runs for the sparse Bayesian SSMLS (SB-SSMLS) are carried out. The mean failure probability estimates over 50 runs by the SS, second order ori-SSMLS and second order SB-SSMLS are plotted in Figure 34. It can be seen from the figure that the SB-SSMLS estimate performs slightly better than the ori-SSMLS in each conditional failure level. This indicates that the metamodel with fewer basis terms can perform better than the one with more terms. In fact, this is the whole spirit behind the proposed algorithm that it is important to have appropriate combination of "terms" given some data points.

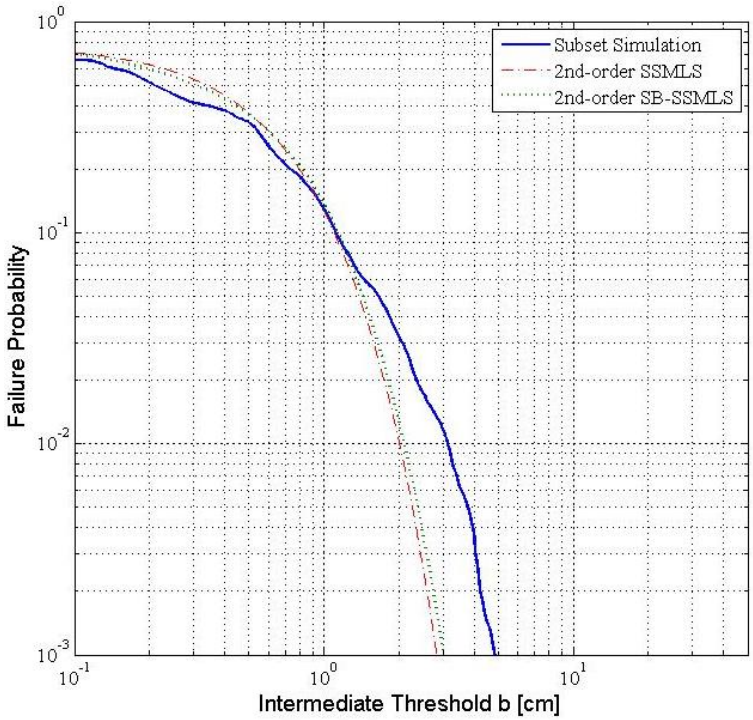


Figure 34 Failure probability estimates of SS, ori-SSMLS and SB-SSMLS

In terms of the appropriate combination of basis terms, Figure 35 summarizes the frequency of the basis term used for the 50 independent runs. The terms with frequency less than 3 are not shown. Overall, the sparsity of the basis function is significant. For the MLS method, the influence radius is prescribed to include 100 support points. Since each stochastic sample requires an iterative sparse Bayesian inference, it is not possible to present all the records. The figure only shows the last chosen terms of the sample out of the 2000 stochastic samples. It can be seen that some of the terms are never selected such as 7, 37, 79, etc., meaning these terms have little influence on the approximation. On the other hand, terms such as 38 and 53 have been chosen for more than half of the 50 simulations. It is also found that the first three frequently chosen terms 38, 53 and 103 correspond to the combination of variables x_2x_6, x_3x_7, and x_7x_{11}. Individually speaking, the random variables 2, 3, 6, 7, and 11 have relatively higher information entropies according to Table 8. It has to be emphasized that the 16 variables are equally important in the earthquake source modeling. It is perhaps only due to the mathematical manipulation which leads to the above-mentioned observation. The exact mechanism of selecting these combinations needs further investigation. The average

number of basis terms used for each sample is 11 with a minimum and maximum of 4 and 32, respectively.

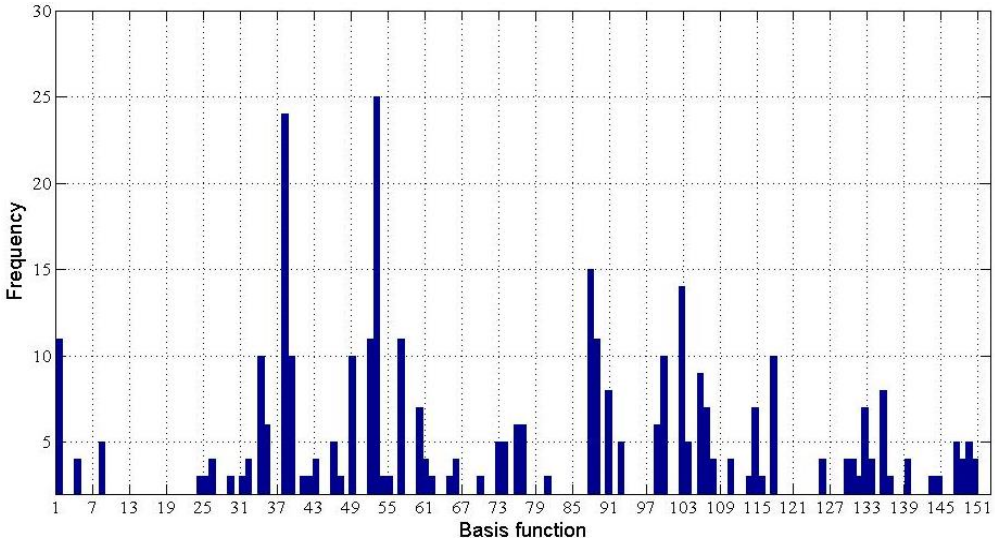


Figure 35 Frequency of the basis terms chosen by sparse Bayesian MLS

The third and fourth order SB-SSMLS use the same framework as in the second order SB-SSMLS except that higher order basis functions are applied. The number of basis terms for the full cubic and quartic polynomials are 969 and 4845, respectively. The mean failure probability estimates by the SS, third order and fourth order SB-SSMLS are plotted in Figure 36. The fitness of the two higher order SB-SSMLS algorithms are compared to the Subset Simulation. In general, the performances of the two are rather similar and both agree well with the benchmark reliability result. The fourth order SB- SSMLS is slightly better than the third order SB-SSMLS in the first and third conditional failure levels.

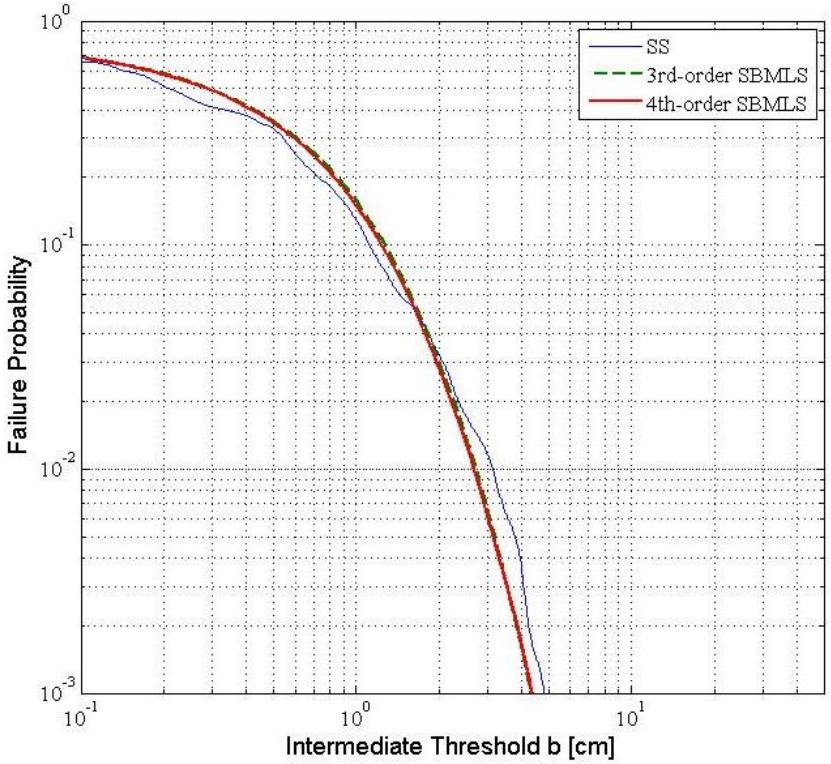
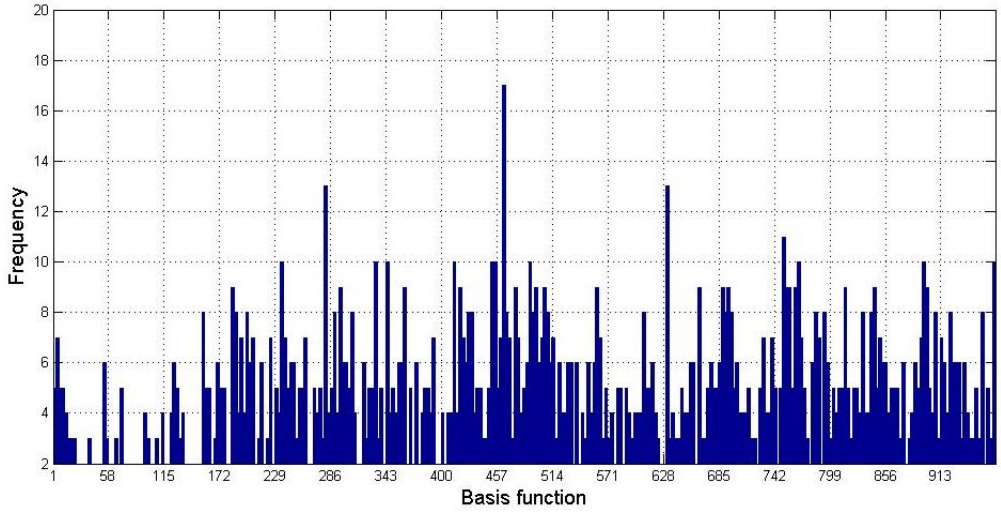
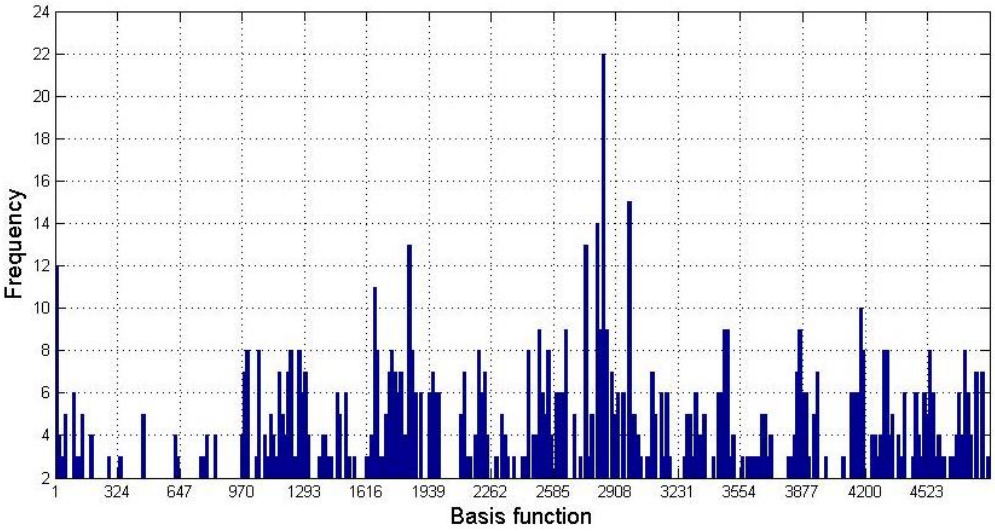


Figure 36 Failure probability estimates by SS, third order and fourth order SB-SSMLS

Again, the frequency of the basis term used for the 50 independent runs for both the third order and the fourth order SB-SSMLS are plotted in Figure 37. The terms with frequency less than 3 are not shown. In Figure 37(a), the first 153 terms represent the first and second order whereas the following 816 terms are third order. In Figure 37(b), the basis terms after 969 indicate the fourth order polynomials. It can be seen that the algorithm tends to use the higher order terms much more frequent than the lower order terms. The average total number of basis terms used for the third and fourth order SB-SSMLS is 31 and 24, respectively. The minimum and maximum number of basis terms for the third order is 5 and 95, respectively, compared to 5 and 85, respectively, for the fourth order case. According to the figure, the fourth order SB-SSMLS shows more sparsity compared to the third order one although the total number of basis terms for the quartic polynomial is five times the size of the cubic one. A possible explanation is that the fourth order terms are representative and containing more information than the third order ones. Further investigation is required to better understand this issue.



(a)



(b)

Figure 37 Frequency of the basis terms chosen by SB-SSMLS.
(a) third order and (b) fourth order

A comparison of the computation efficiency of the SS, third order and fourth order SB-SSMLS algorithms are presented in Table 10. Although the c.o.v. for the SB-SSMLS algorithms are a little higher compared to the benchmark, the total number of numerical simulations is only one third of the benchmark. This leads to a better efficiency in terms of Δ. The c.o.v. for all the algorithms are presented in Figure 38. Note that the total computational time for the two higher order SB-SSMLS are close since only 30 basis terms are used on average with the sparse Bayesian scheme.

Table 10 Comparison of the computation efficiency for SB-SSMLS algorithms

$P_F = 10^{-3}$	Nb. Smln	δ	Δ
Subset Simulation	1500	0.987	38.226
3rd order SB-SSMLS	500	1.018	22.763
4th order SB-SSMLS	500	1.017	22.741

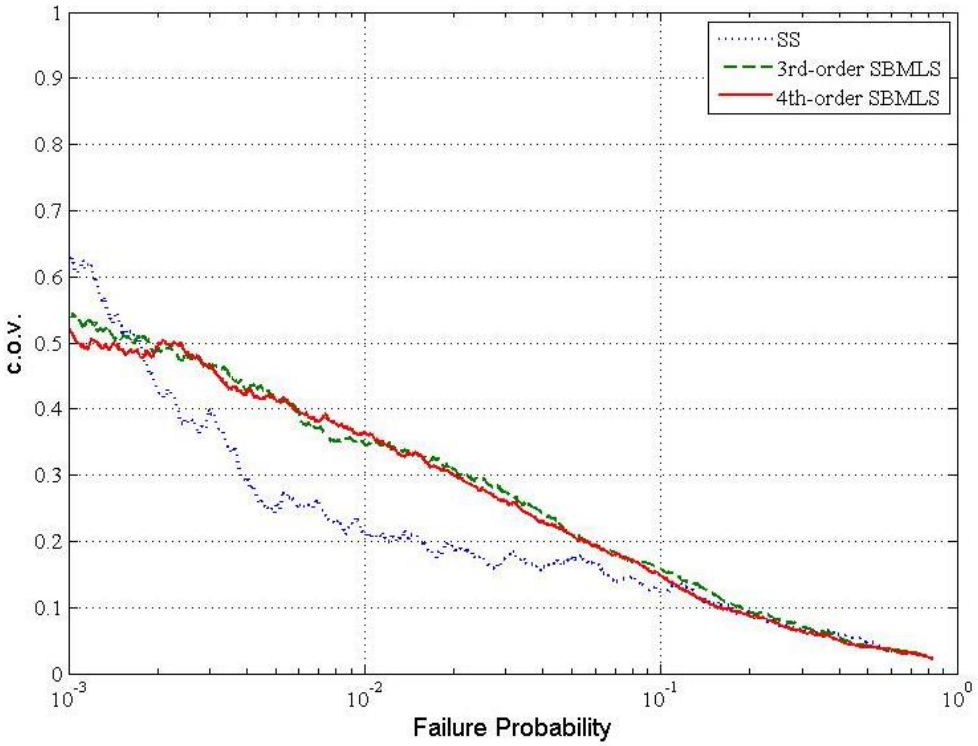


Figure 38 The c.o.v. of failure probability estimates for SS, third order and fourth order SB-SSMLS

5.3.1.2 SB-SSMGP

Initially, the same framework as used for the SB-SSMLS was applied to the SSGP. However, the performance of the higher order SB-SSGP does not overweigh the second order one. This is illustrated from the resultant reliability functions for the sparse Bayesian integrated GP metamodel with various orders of basis functions plotted in Figure 39. It is shown that the employment of a higher order basis function has limited influence on the GP approximation. The improvement obtained from the third and the fourth order SB-SSGP is hardly noticeable. In contrary to this, the accuracy of the SB-SSMLS algorithm is enhanced with increasing order of the

basis function. In fact, the SB-SSGP performs similarly to the ori-SSGP result as shown in Figure 31. On one hand, the improvement in accuracy of the adv-SSGP is achieved by adaptively introducing more numerical evaluations within or near the failure domain as well as properly choosing the subset support points. On the other hand, the SB-SSMLS in the previous section performs quite well without additional numerical simulations. This leads to the following proposed modification to the SB-SSGP algorithm.

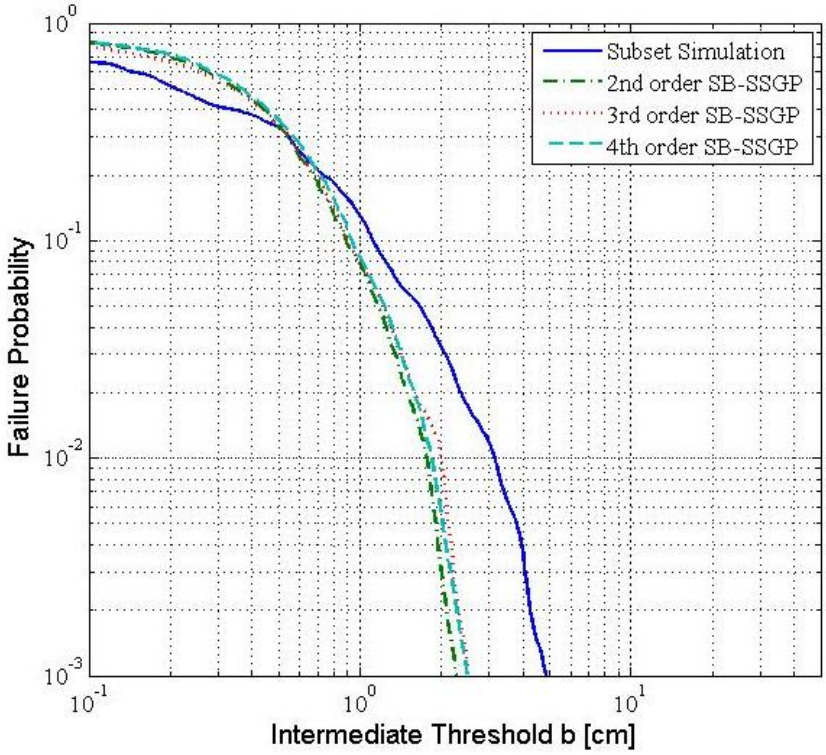


Figure 39 Failure probability estimates of SB-SSGP with various order basis function

As stated at the end of Section 4.2.2, an important characteristic of MLS is that only the training points within an influence radius will be selected. Although the subset selection of the training points for GP as shown in Section 4.3.3.2 helped in the SSGP approximation, the selecting mechanism is not applicable in this case since no extra training points are generated. Thus, a novel moving Gaussian processes (MGP) is proposed in this section to be integrated with the sparse Bayesian inference. The MGP works in a similar way as MLS except for the basis and covariance functions which are calculated by Gaussian processes. The distance

between the interpolated point and the supporting points is computed before carrying out the GP model. Only the support points within the influence radius are used in the GP regression. With this pre-processing step, the SB-SSMGP algorithm is implemented for the reliability assessment.

50 independent runs were conducted for both third order and fourth order SB-SSMGP algorithms. The resultant reliability estimates by SB-SSMGP with 2^{nd} – 4^{th} order basis functions are plotted in Figure 40. The fitness of the SB-SSMGP algorithm compared with the Subset Simulation is derived quantitatively and summarized in Table 11. The failure probability estimation results derived from the SB-SSMGP algorithms are greatly enhanced compared to the original version of SB-SSGP. This suggests that the ability of "moving" for each stochastically generated point allows the Gaussian processes to use the most relevant support points to construct the basis and covariance functions. Hence, a better approximation result can be achieved. According to Table 11, the third and fourth order SB-SSMGP algorithms have similar performance which is also observed in SB-SSMLS algorithms.

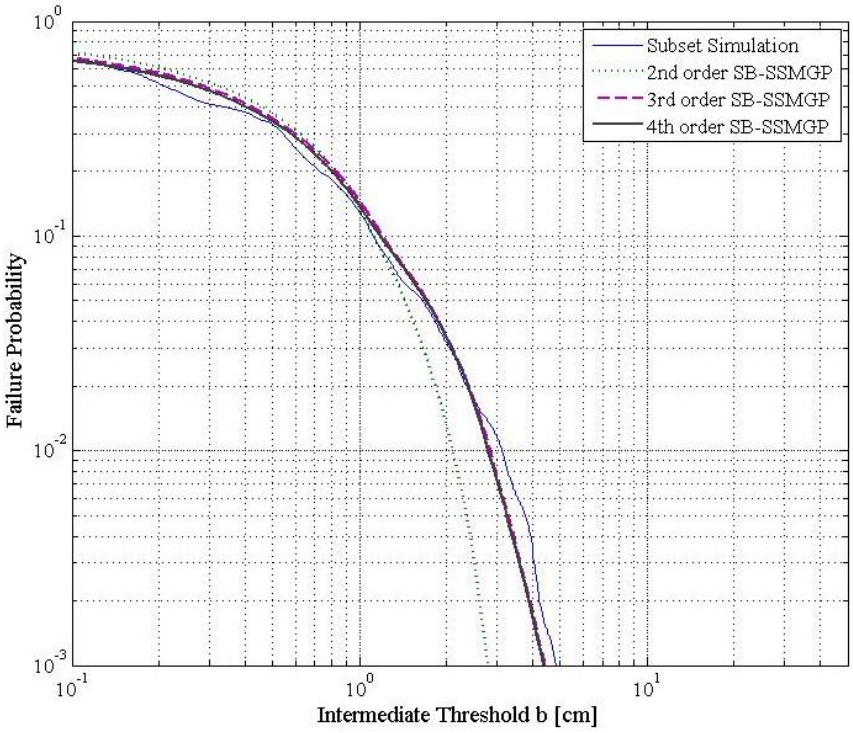


Figure 40 Failure probability estimates of SS and various orders SB-SSMGP

Table 11 Fitness of the higher order metamodel compared to benchmark

	Lv1	Lv2	Lv3
3rd order SB-SSMGP	0.86	1.61	1.64
4th order SB-SSMGP	0.91	1.67	1.87

The c.o.v. of the SB-SSMGP algorithms are presented against the benchmark study as shown in Figure 41. In general, the uptrend is slightly milder than the benchmark c.o.v. from the second conditional level onwards, due to a large number of samples used in each intermediate failure level. The computational efficiency of the SB-SSMGP algorithm is summarized in Table 12. Compared to Table 10, the SB-SSMGP outperforms the benchmark study not only in terms of the efficiency Δ but the c.o.v. as well. Similar to the case of adv-SSMLS and adv-SSGP, further reduction in c.o.v. of the failure probability estimator of 3rd-order and 4th-order SBMGP is expected with more samples without extra original numerical simulation.

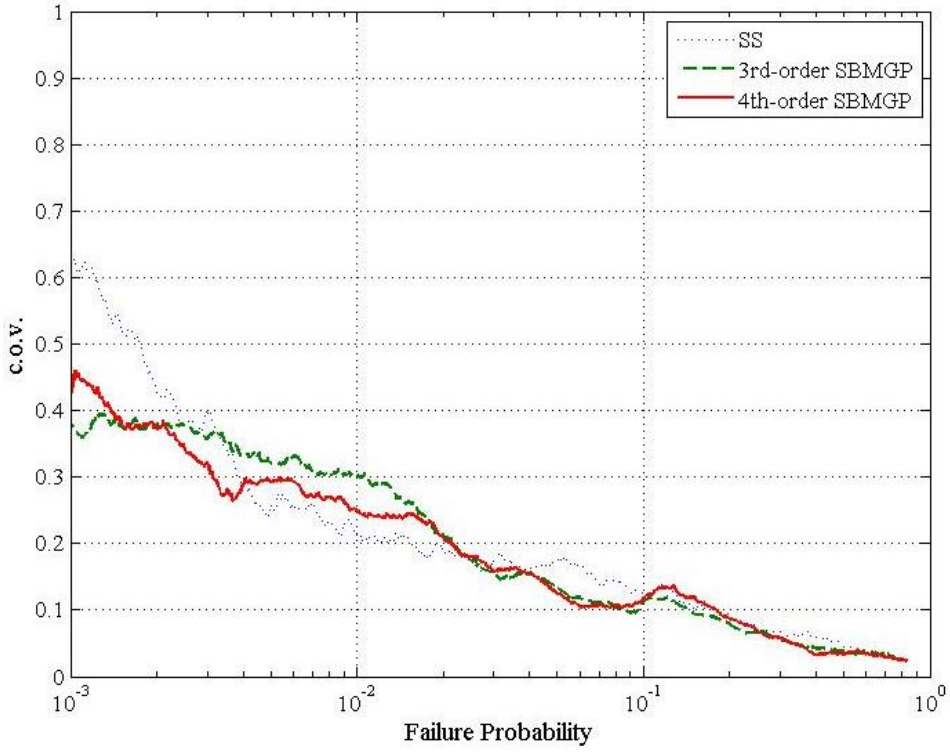


Figure 41 The c.o.v. of failure probability estimates for the third and fourth orders SB-SSMGP

Table 12 Comparison of the computation efficiency for SB-SSMGP algorithms

$P_F = 10^{-3}$	Nb. Smln.	δ	Δ
Subset Simulation	1500	0.987	38.226
3rd order SB-SSMLS	500	0.892	19.946
4th order SB-SSMLS	500	0.888	19.856

5.3.2 Adaptive SB-SSMLS for high dimensional random variables

In this section, a total of 53 random variables are considered, including both the earthquake and structural parameters, to better understand the tsunami-structure interaction mechanism. The number of subfault is increased to 8×8. The outskirts of the fault are assumed to be zero to reflect the actual earthquake source mechanism. Therefore, the 36 inner sub-faults are non-zero. The random variables consist of earthquake magnitude, focal depth, subfault strike angle, rake angle, dip angle, subfault slip magnitude, and structural variables, *i.e.* Young's modulus of steel reinforcement, yield strengths of steel reinforcement, tensile strength of concrete, as well as Poisson ratios for the two materials. The distributions used for the earthquake source parameters and material properties are consistent with the observation of historical earthquakes in Tohoku region (Yamamoto & Hori 2004) and study by Mirza, S. A., and MacGregor, J. G. (1982), respectively. The fault length and width are determined by the earthquake magnitude. The distributions of the 53 random variables and their basis are provided in Table 13. A sample COMCOT input for the case study is presented in Appendix A.2.

Since a full set of the second order metamodel has 1485 coefficients to be evaluated, applying the metamodel directly results in at least 1485 numerical simulations. More simulations are required for adaptive metamodel algorithms such as the one proposed in Chapter 4. Therefore, the integration of the sparse Bayesian method with the adaptive second order metamodel, *i.e.*, adaptive SB-SSMLS, is illustrated in this section. The purpose of such integration is to include more random variables and at the same time maintain the accuracy of the reliability estimate. As it will be shown later, fewer than 300 coefficients out of the full set of the second order coefficients are kept using the sparse Bayesian method. Thus, a total of 500 support

points are firstly assigned to construct the MLS function. For each intermediate level, a number of M = 2,500 samples are simulated. A portion of the samples which lead to the top 10% of maximum displacement as computed by the MLS is selected and their maximum displacements are computed using the original numerical simulation. The newly generated samples are included as support points to improve the accuracy of the approximation for samples from rare region.

Table 13 Distributions of the 53 random variables

Parameter	**PDFs**	**Nb. of variables**
Earthquake magnitude [M_w]	G-R relation [8.0, 9.0]	1
Focal depth [km]	U(20, 70)	1
Strike direction [deg]	N(204, 30)	1
Dip angle [deg]	N(25, 10)	1
Rake angle [deg]	N(90, 34)	36
Dislocation [m]	Hybrid K-squared	8
Young's modulus of steel	N(200GPa, 20GPa)	1
Yield strengths of steel	N(460MPa, 46MPa)	1
Poisson ratios of steel	N(0.3, 3e-3)	1
Tensile strength of concrete	N(3.2MPa, 0.57MPa)	1
Poisson ratios of concrete	N(0.2, 2e-3)	1
Total		**53**

It was shown in Chapter 4 that the influence radius for the MLS is updated when new support points are included. However, due to the higher dimensional random variables, the lower range of the influence radius is limited by the number of coefficients remained after the sparse Bayesian inference. This indicates that the optimization of the influence radius might be compromised. Therefore, the free parameters c and k which control the weight function in MLS are subject to change so as to retain the performance of MLS as in the case where the optimal radius is applied. The ranges for the c and k are 0.1 to 1.0, and 0.4 to 1.8, respectively. A sequential optimization scheme is adopted to obtain the c, k, and influence radius. The optimization objectives are the difference between the intermediate thresholds of the pre-samples and the support points and information entropy between the two.

The mean failure estimates by the proposed adaptive SB-SSMLS is shown in Figure 42, together with the Subset Simulation results involving the 53 random variables. It can be seen that the approximation by the proposed algorithm overestimates the failure probability before $P_F = 0.1$, and underestimates it afterwards. However, the gap is closer as the failure probability decreases. The greatest discrepancy between the two is at the end of the second intermediate failure level. The average cumulative number of support points for the three levels are 500, 750, and 951, respectively. The optimized value of influence radius, MLS shape factors c and k, are listed in Table 14. The c and k in the first intermediate level are at their original value (0.4, 0.1) as used in other literatures. As can be seen from the table, the radius is increasing with the number of conditional levels. The minimum radius is confined by the number of coefficients selected by the sparse Bayesian algorithm, which can be the reason for the overestimation in the first conditional failure level. For the subsequent levels, although the influence radius as well as the shape factors c and k vary, the performance of the MLS is barely passable due to a lack of support points in the rare region.

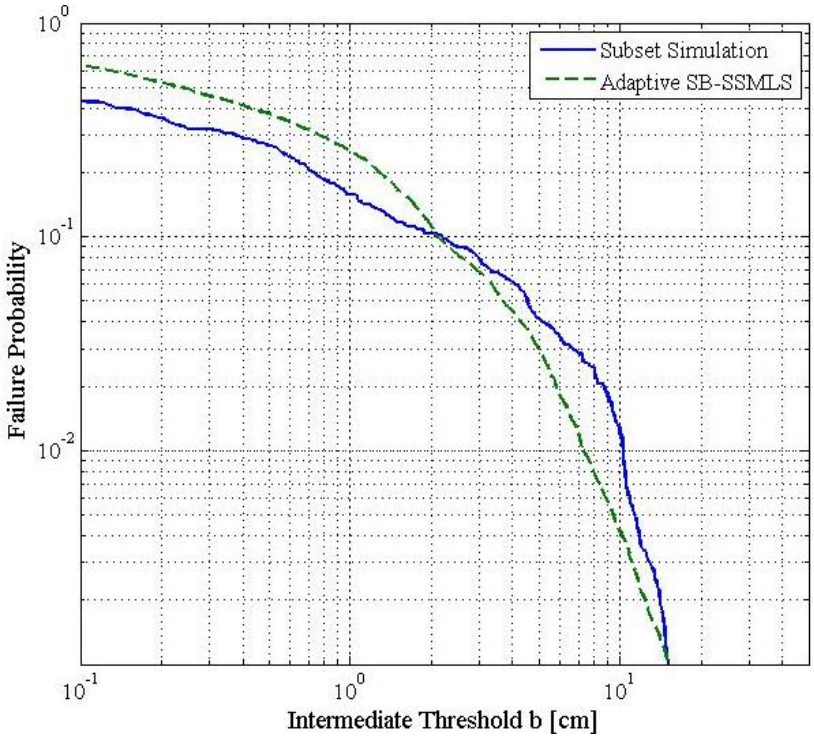


Figure 42 Failure probability estimates of SS and adaptive SB-SSMLS

Table 14 Sub-optimized MLS parameters in adaptive SB-SSMLS

	Radius D	c	k
1st level	190	0.40	1.00
2nd level	530	0.39	0.90
3rd level	710	0.48	1.11

Apart from the failure probability estimates, the interaction among different variables provides a better understanding of the physical process and help to comprehend the algorithm intuitively. The correlation matrix of the earthquake magnitude, earthquake depth, strike and dip angles at the third conditional failure level, is shown in Figure 43. All of the four parameters have high correlation coefficients, indicating a relatively strong linear relationship between the paired variables. The earthquake depth and dip angle shows the highest correlation coefficient in a negative value. In general, a shallower earthquake causes severer tsunami and the optimal dip angle for a maximum run-up is between 20° - 30° (Geist 1998). Therefore, the simulation result coincides with the observation in other studies.

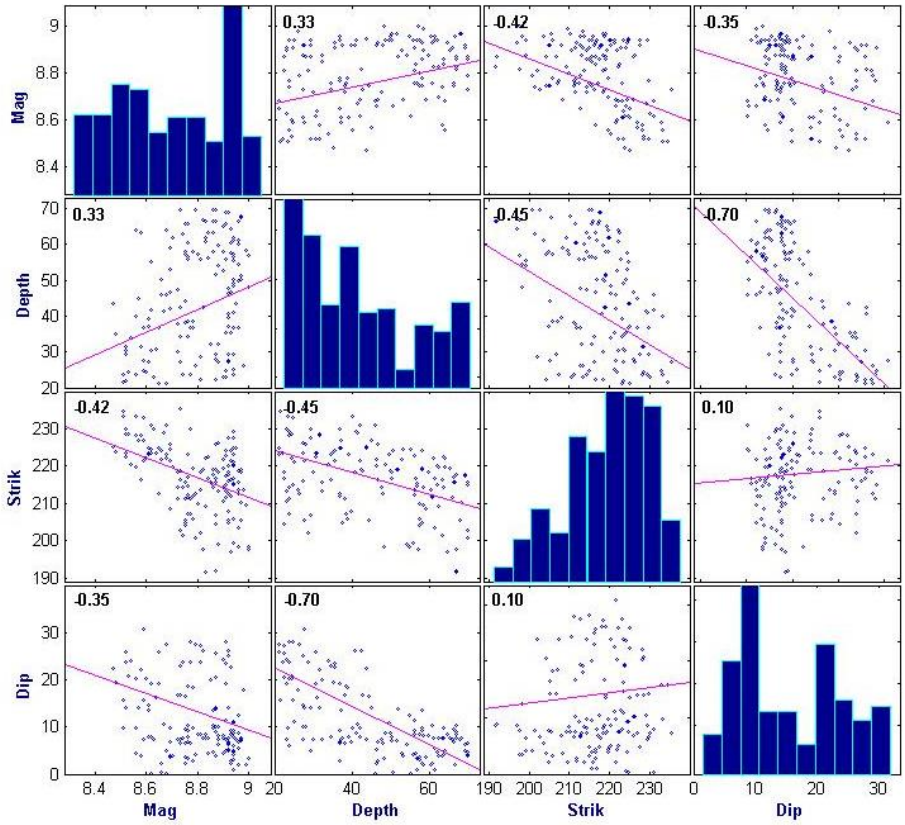


Figure 43 Correlation matrix of four earthquake parameters

The marginal probability distributions of the variables obtained from the simulation is also presented. For example, the marginal distributions of the earthquake magnitude of the three intermediate failure levels are shown in Figure 44. The kernel density estimates are derived based on the total samples generated at each level. As the conditional failure level increases, the peak of the magnitude density moves towards the upper boundary of this case study, *i.e.* M_w 9.0. The marginal PDF of the earthquake depth is plotted in Figure 45 where the depth becomes shallower end as the level increases.

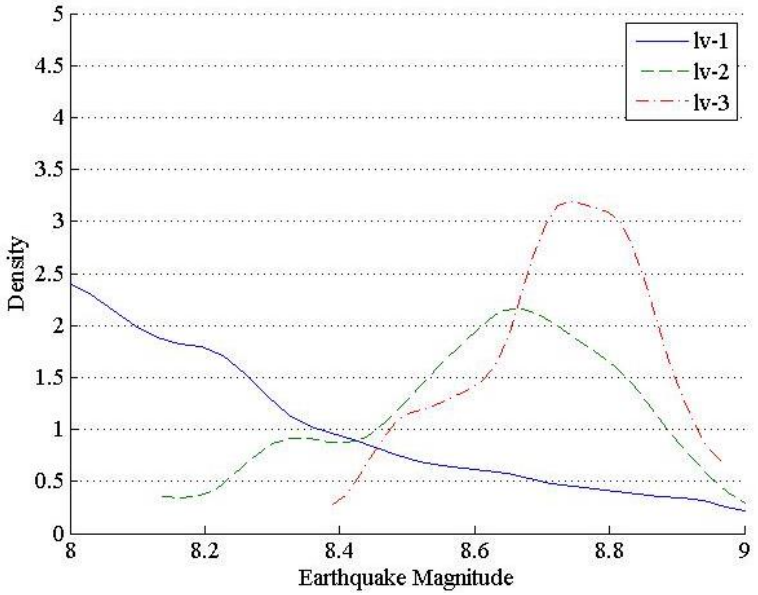


Figure 44 Kernel density estimates of earthquake magnitudes for the three intermediate failure levels

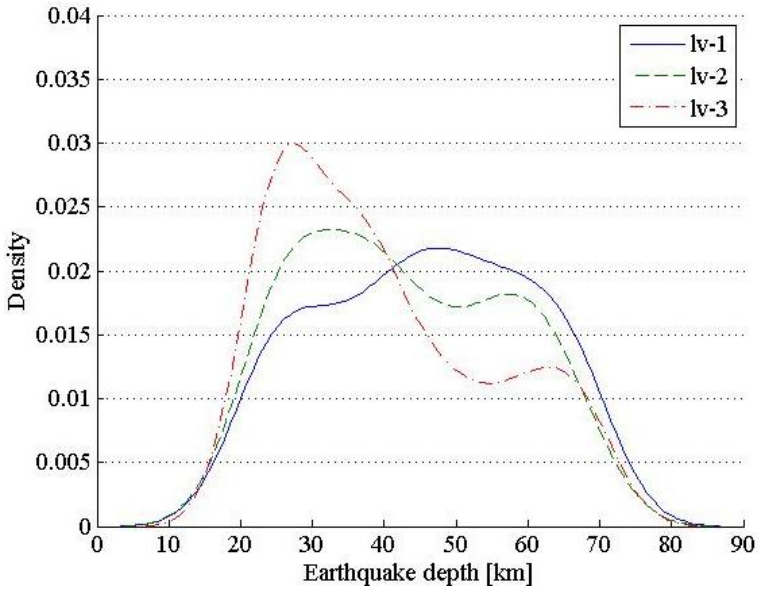


Figure 45 Kernel density estimates of earthquake depth for the three intermediate failure levels

For the 36 slip angles, an upper diagonal of the correlation matrix is showcased in Figure 46. Positive correlations are displayed in blue and negative correlations in red colour. The colour intensity and the size of the circle are proportional to the correlation coefficients. The legend colour shows the correlation coefficients and the corresponding colours. The paired variables of slip angle with higher correlation magnitude can be found between V1-V16, V3-V31, V4-V31, and V18-V21. The location of the 36 variables on an imaginary 6×6 fault plane is shown in Table 15. Note that the actual geological interaction on an earthquake fault is much more complex and requires further investigation.

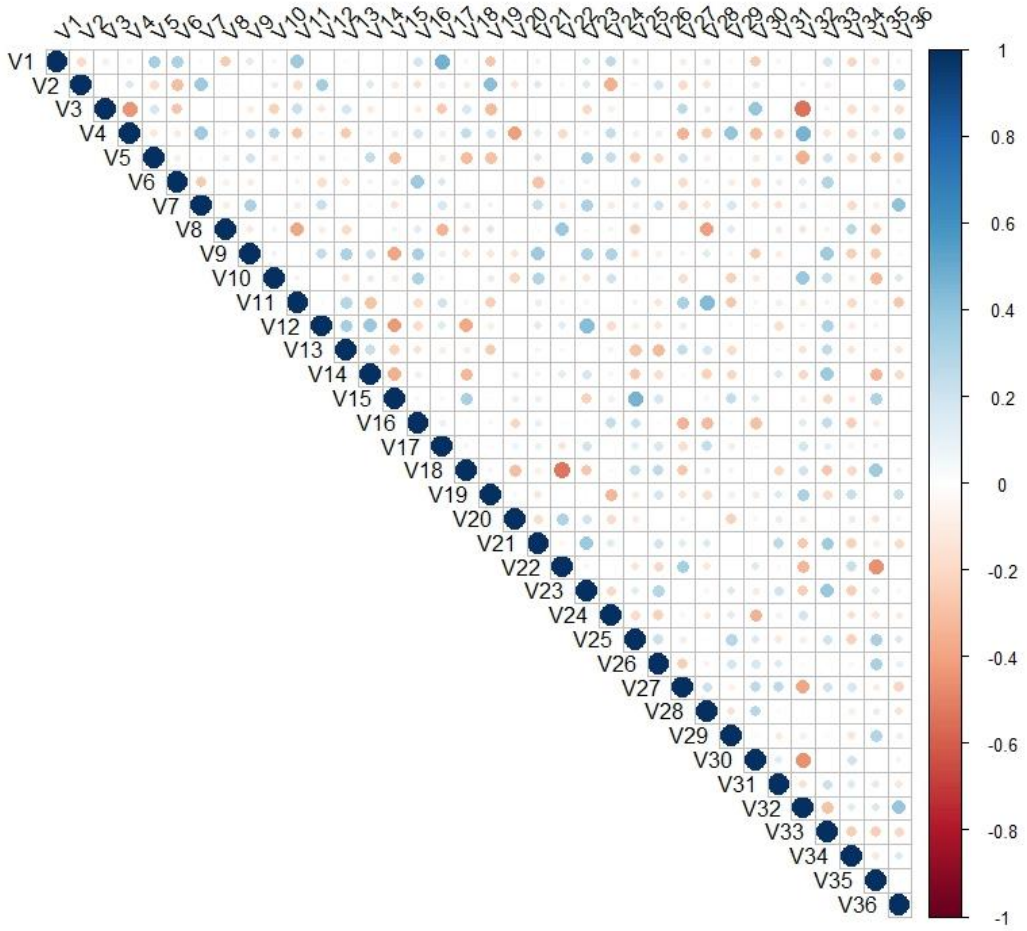


Figure 46 Upper diagonal correlation matrix of the slip angles

Table 15 Location of the 36 slip angles on a fault plane

V1	V7	V13	V19	V25	V31
V2	V8	V14	V20	V26	V32
V3	V9	V15	V21	V27	V33
V4	V10	V16	V22	V28	V34
V5	V11	V17	V23	V29	V35
V6	V12	V18	V24	V30	V36

Regarding the 8 slip amplitudes, the correlation matrix at the third conditional failure level is presented in Figure 47. The linear correlations among the pairs are hardly visible, except for V1-V7 which has a relatively higher correlation coefficient of -0.46. However, the correlation is not sufficient to imply dependence (causation). Again, the knowledge in earthquake physics is required for further interpretation of the simulation result.

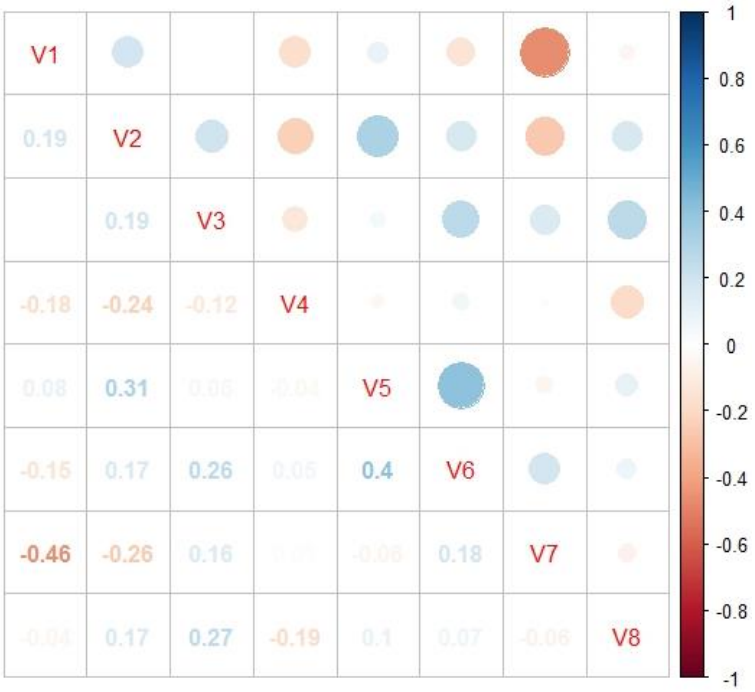


Figure 47 Upper diagonal correlation matrix of the slip amplitude

The marginal distribution of the structural parameters, *i.e.* yield strength, Young's modulus, and Poisson ratio of steel (labelled as Yd-S, E-S, and nu-S, respectively); as well as tensile strength and Poisson ratio of concrete (labelled as Ts-C and nu-C, respectively), are examined. The correlation matrix of the five variables at the third conditional failure level is presented in Figure 48. It can be seen from the figure that correlations are low in general, except for the Young's modulus and Poisson ratio of the steel which has a positive correlation of 0.58. This is reasonable since Young's modulus (E) is related to Poisson ratio (ν) by the following relationship: $E = 2G(1 + \nu)$, where G is the shear modulus.

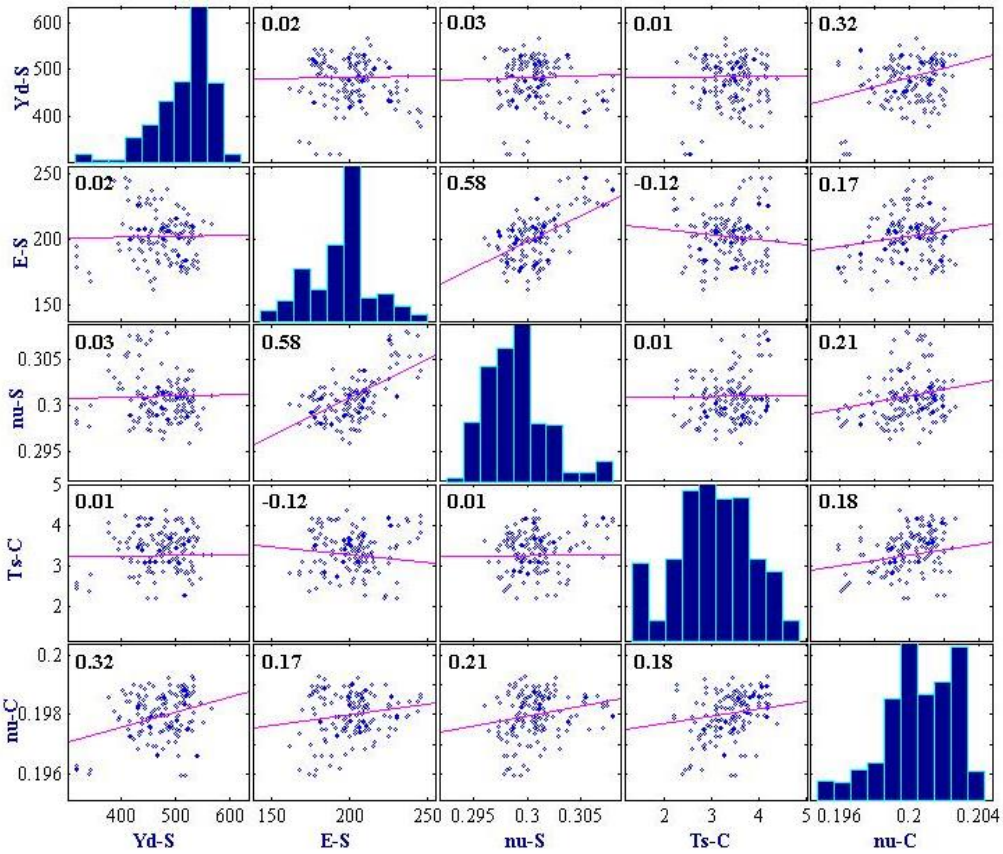


Figure 48 Correlation matrix of the five structural variables

5.4 Closing remarks

This chapter has shown the possibility of using metamodels with higher order basis functions as well as using metamodels to solve system with high dimensional random variables. One thing in common for these two problems is that they require thousands of support points derived from physical computations which are hardly affordable for such complex system. The elimination of the irrelevant terms in the basis function is the key to solving such challenge. In the case studies, only a few hundred basis terms are conserved after the sparse Bayesian inference. Therefore, only 500 simulations are required for the higher-order SB-SSMLS and 951 for the adaptive SB-SSMLS for high-dimensional problem, compared to 1,500 for the Subset Simulation benchmark study.

The sparse Bayesian inference is applied to both SSMLS and SSGP. The accuracy of the third and fourth order SB-SSMLS are greatly improved compared to the

second order case. However, the higher order SB-SSGP does not perform as expected due to the limitation in subset selection of the training points. It turns out that the integration of the sparse Bayesian with SSGP alone cannot fulfil the accuracy requirement. Therefore, a moving GP (MGP) conceptually similar to MLS is proposed, *i.e.*, the approximation only relies on the support points within the influence radius. As seen from Figure 40, the SB-SSMGP algorithm with higher order basis functions achieves a more accurate result compared to the second order case. The influence radius feature assists the GP to use a few yet important training points to construct the model. The MGP is a possible solution for a bunch of problems where uncertainty quantification is desired and subset data selection has difficulty in providing relevant training data.

The performance of the adaptive SB-SSMLS in higher dimensional (53 dimensions) problems is not as good as the other cases which have 16 random variables. It seems that a relatively high portion of basis terms are relevant in the adaptive SB-SSMLS. The correlations among the source parameters indicate they are potentially connected in a certain way. The subfaults with high correlation can be proposed in a joint distribution manner in the stochastic sampling part, which will improve the performance of the proposed algorithm. Furthermore, the statistical analyses can shed some light on earthquake engineer or geologist who studies the fault mechanism aiming at better understanding the physical behaviour.

Nonetheless, there are a few limitations in this work which requires follow-up studies. First of all, the selection of the highest order of polynomial is not trivial. In this study, the third and fourth order polynomials are adopted. As it is seen from Figure 36 and Figure 40, minimum improvement in terms of accuracy are achieved by adopting the fourth order and that is one of the reasons to stop at fourth order. Another reason is the limit number of support points since full-basis polynomials requires C_{d+m}^{m} number of training points to avoid singularity where d is the number of variables and m is the highest order of the polynomials. The most straightforward way is to use statistical analyses for trial run of the metamodel in order to determine the order. On the other hand, Chebyshev polynomials can be adopted to select the

uncorrelated coefficients (Gavin and Yau 2008). Another approach involves Bayesian model class assessment in order to choose the appropriate highest order to be used (Cheung, S. H. and Beck, J. L. 2010).

Secondly, in the sparse Bayesian inference algorithm, the forward-backward feature selection algorithm adds or deletes one term at a time. It is confined in a way such that only a local maximum of marginal likelihood is achieved. In another word, global optimization of marginal likelihood requires adding or deleting more than one term at a time, which is not implemented in the algorithm. The main concern of operating with one term only comes from the computation time as the permutation of the terms is exponentially proportional to the number of terms. Although the sparse Bayesian algorithm tremendously reduces the number of terms in the basis function, the limit still exists when the influence radius is optimized. According to the correlation matrix among the random variables, some of them show strong linear relationship among them and therefore, a multi-dimensional proposal PDF reflecting the dependence among the random variables can be used in order to achieve a higher acceptance ratio in MCMC algorithm in the Subset Simulation scheme.

Chapter 6 Other Enhancements over Metamodel-Based Reliability Analysis

6.1 Consistent reliability evaluation using SS-MLS integrated with response conditioning method

The introducing of response surface models help to provide more insights of the system behavior by generating stochastic samples more efficiently. However, due to the approximation nature, its resultant reliability estimates are not consistent in the sense that they do not converge to that obtained by solely running the original simulations regardless of the number of the stochastic samples generated by metamodel-based approaches. Furthermore, more often than not, the metamodels are implemented solely without checking the accuracy of its reliability estimates. In the next section, the response conditioning method (RCM) proposed by Au (2007) is integrated with the previously-proposed metamodel-based algorithms and a case study is presented.

6.1.1 Response conditioning method

The governing principle of the RCM is the Theorem of Total Probability. The failure probability can be expressed as:

$$P(F) = \sum_{i=0}^{m} P(F \mid B_i) P(B_i) \tag{6.1}$$

where

$$\begin{aligned} B_0 &= \{\boldsymbol{\theta} \in \mathbb{R}^n : \hat{f}(\boldsymbol{\theta}) \le x_1\} \\ B_i &= \{\boldsymbol{\theta} \in \mathbb{R}^n : x_i < \hat{f}(\boldsymbol{\theta}) \le x_{i+1}\}, i = 1, 2..., m-1 \\ B_m &= \{\boldsymbol{\theta} \in \mathbb{R}^n : \hat{f}(\boldsymbol{\theta}) > x_m\} \end{aligned} \tag{6.2}$$

The metamodel function $\hat{f}(\theta)$ is assumed to be computed with negligible amount of effort compared to the original numerical physical simulation. From Eqn. (6.1), it is clear that the failure probability can be partitioned by $\{B_i : i = 0, 1, \ldots, m\}$. The x_i's in Eqn. (6.2) are prescribed thresholds. Therefore, instead of estimating the failure probability $P(F)$, the RCM assesses the bin probability $\{P(B_i) : i = 0, 1, ..., m\}$ and the conditional probability $\{P(F|B_i) : i = 0, 1, ..., m\}$.

The Subset Simulation (SS) method is naturally a good choice to be integrated with RCM. The bin probability $P(B_i)$ can be estimated based on a commonly specified value of p_0 in the SS algorithm. The bin probability is given by:

$$\begin{aligned} P(B_i) &\approx \tilde{P}(B_i) = p_0^i(1-p_0), \quad i = 0,...,m-1; \\ P(B_m) &\approx \tilde{P}(B_m) = p_0^m \end{aligned} \tag{6.3}$$

where $\tilde{P}$ denotes the sample estimate. A certain number of original numerical physical simulations which calculate the responses are required to evaluate the conditional probability $P(F|B_i) = P(L(\boldsymbol{\theta}) < 0|B_i)$ where $L(\boldsymbol{\theta})$ is the limit state function. To avoid the clustering effect for better estimating the lower probability regime in each bin, a *sub-bin* concept was adopted which partitions the bin to equally sized sub-bin. One sample will be taken from each sub-bin to run the actual simulation. A choice of 50 sub-bin was used in Au's case studies.

Let N denote the number of samples at each intermediate failure level, and N_1 is the number of samples from each bin to evaluate the conditional probability. Then the failure probability can be written as:

$$P(F) = \sum_{i=0}^{m}\sum_{j=1}^{N_1} P(F_i|B_{ij})P(B_{ij}) \approx \sum_{i=0}^{m}\sum_{j=1}^{N_1} I(F_i)\frac{P_i}{N_1} \tag{6.4}$$

where $I(F_i)$ is an indicator function and $I(F_i) = 1$ when $L(\boldsymbol{\theta}) < 0$ and zero otherwise; P_i approximates $P(B_i)$ as given in Eqn. (6.3). The framework of the RCM shares a similar concept of a method called Auxiliary Domain Method (ADM) in which it assumes an auxiliary failure domain in order to explore the dependency between the target and conditioning responses (Katafygiotis *et al.* 2007). ADM assumes the samples obtained by auxiliary response conditional on the auxiliary failure domain can be obtained by forming the failure domain as a union of linear half-space. It avoids the gradually propagation of the failure domain from frequent to rare region.

6.1.2 Adaptive RCM integrated with SS-MLS algorithm

The RCM has shown promising results in Au's case studies (Au 2007) where the "corrected" results are close to the ones obtained by MCS. However, as pointed out in his paper, the RCM may not perform well for those problems in which the approximation and the target response are not well correlated. This is proven by

using the RCM directly for the benchmark problem. The failure curves are plotted in Figure 49. The dashed lines indicate the eight independent failure probability curves derived using the RCM integrated with MLS without adaptation. From the result, one can see that the RCM does not provide reliable estimation even though the sub-bin size is doubled from 50 to 100 from the second conditional failure level onwards. For some of the cases when the failure probability increases, the RCM estimation returns zero which means no extra samples exceed a particular threshold. Nonetheless, it does give consistent value at the first intermediate failure level for most of the cases. Furthermore, the information provided by the RCM is generally correct, *i.e.* the RCM produces higher failure probability estimation when the metamodel approximation is lower than the actual result and vice versa. However, it tends to overcorrect the original results.

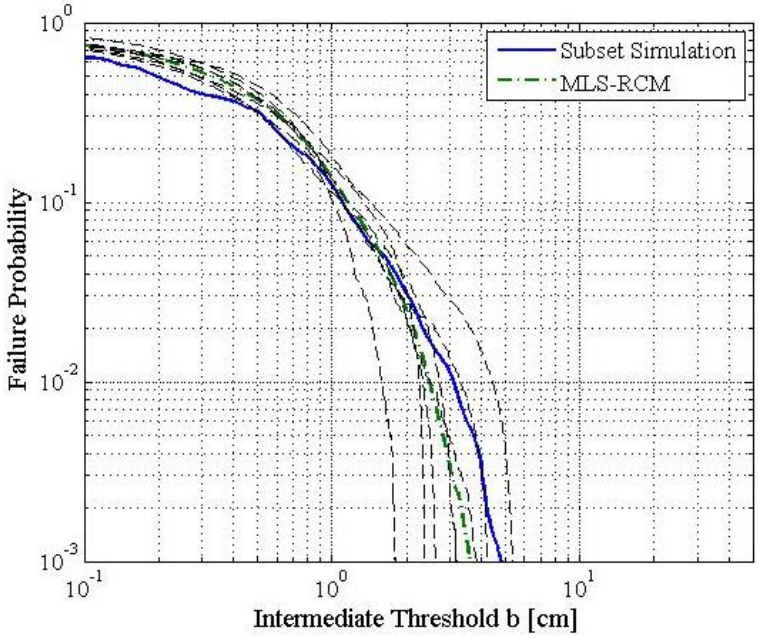


Figure 49 Results of a few typical runs using the RCM

Figure 50 shows the conditional samples derived from both numerical simulation (x-axis) and metamodel (y-axis) of the sub-bins. The umax is defined as the maximum structural displacement. The dashed line indicates the thresholds for the intermediate failure probability. Ideally, the sample points should fall in straight line with gradient 1, indicating that the approximation (umax-MLS) is the same as the numerical simulation (umax). However, this is impossible since approximation does not converge to the exact solution. Therefore the points are spread around. It

can be seen that the correlation between the two methods is low. The average correlation coefficients between the actual numerical simulation and approximation from metamodel for the four intermediate failure levels are 0.46, 0.13, 0.09, and 0.11, respectively. The nonlinearity of the system reduces the accuracy of the metamodel in the rest of the failure levels, which results in an unreliable estimate of the failure probability.

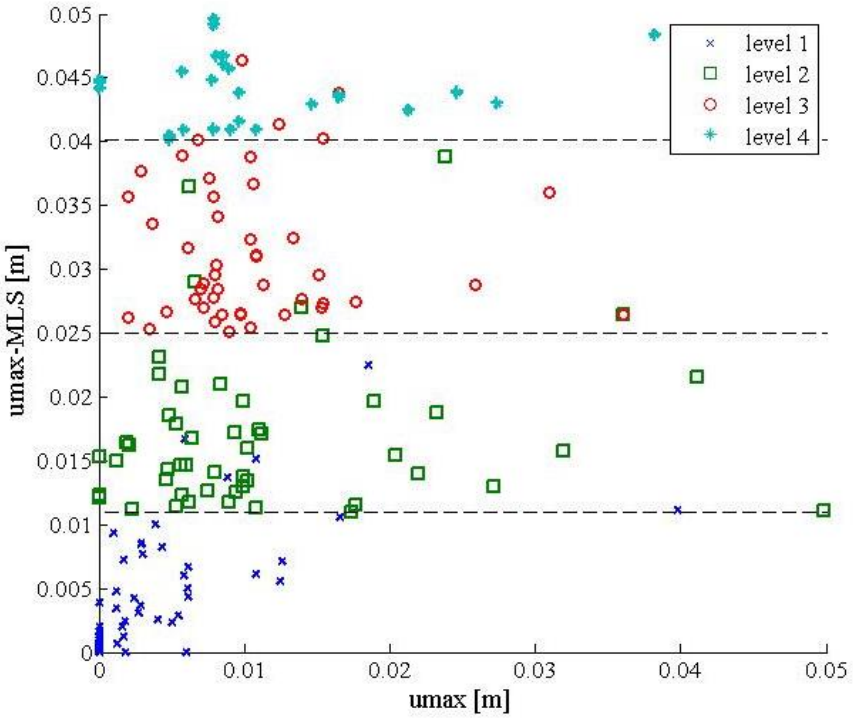


Figure 50 Scattergram of the samples selected from the sub-bins

Based on the observation from the direct implementation of RCM, there are two places where the integration of the RCM with the metamodel based algorithms can be improved in order for it to be used for problems involving highly nonlinear systems. First of all, a better approximation of the original model by metamodel beyond the first failure level can be achieved by adaptively introducing more support points in the part of the failure domain for each level as proposed in Chapter 4 thus gradually covering more and more parts of domain important for accurate failure probability estimation or through sparse Bayesian method as proposed in Chapter 5. Secondly, the results obtained from RCM contains information which can help construct a more robust metamodel by comparing the approximation by the metamodel and the RCM result at each conditional failure

level. Thus, the metamodel can be updated quantitatively according to the difference of the two.

The detailed procedure of the proposed algorithm using MLS as illustration is as follows.

1. Sampling using the MLS is firstly carried out. Similar to the adv-SSMLS algorithm, the influence radius is optimized according to the support points.
2. The RCM is implemented and the "correction" is obtained.
3. The influence radius D of the MLS is re-optimized in order to minimize the gap between the approximation and the correction.
4. Sampling of MLS using the newly derived radius for the failure level.
5. The sub-bin samples in RCM which have their corresponding responses computed by numerical simulations are included as support points for a better approximation of the original model by metamodel.
6. Iterate these five steps until a prescribed failure probability is reached.

The samples generated by MLS in Step 1 are not for estimating the failure probability but rather for RCM correction in the next step. Thus, the sampling in this stage is called pre-sampling to distinguish it from the samples for assessing the failure probability. The purpose of the pre-sampling in the first step is to acquire additional support points using the RCM and make use of the difference between the pre-sampling and RCM results to re-calibrate the influence radius of the MLS. The proposed algorithm has two advantages: (a) the sub-bin samples computed using the original simulation is taken into account as support points so as to enhance the performance of the metamodel; (b) the RCM is implemented before the sampling stage which allows a real-time adjustment of the metamodel. The adjustment of the influence radius D in step 3 is made based on the difference between the RCM result and the original MLS pre-samples. Among the entire set of candidate influence radii, the one which gives the absolute minimum error will be chosen to construct the MLS. An illustration of the influence radius adjustment is shown in Figure 51. As shown in the figure, the result from the RCM indicates that instead of a failure probability of 0.1 ($P_F = 10^{-1}$) obtained from the MLS pre-

sampling, the actual failure probability at this particular threshold is 0.18 ($P_{F\text{-}RCM}$ = 0.18). Thus, the influence radius D is updated to minimize the difference between the two results. The MLS sampling after adjustment is then carried out as shown in the figure.

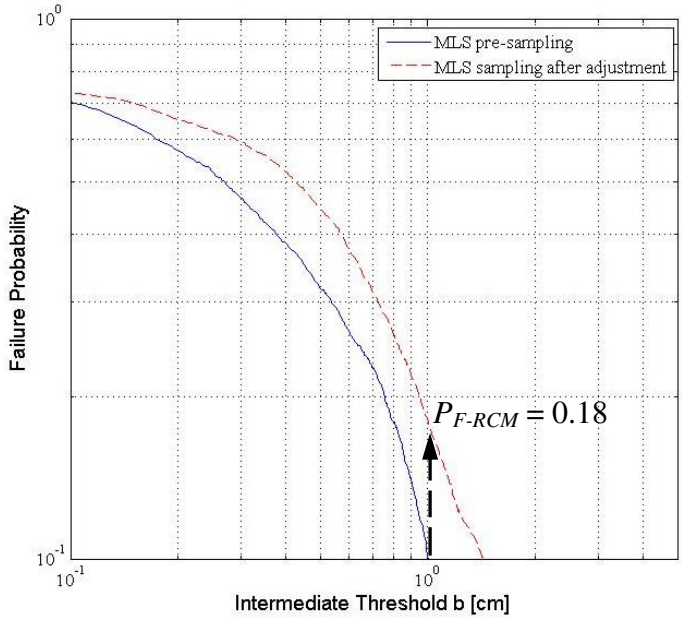


Figure 51 An illustration of the interaction between MLS and RCM

6.1.3 Illustrative example

The proposed adaptive SSMLS-RCM algorithm is tested using the 16-variable benchmark problem. Initially, 200 support points (NS = 200) is used to construct the MLS. A number of 1,000 pre-samples are generated using metamodel for each conditional failure level before the sampling stage. The pre-samples are further divided into N_1 = 50 and N_2 = 100 sub-bins for the first intermediate failure level and the subsequent intermediate failure levels, respectively. One sample from each sub-bin will be drawn randomly and proceed with the numerical simulation. The reason for a higher number of sub-bins for the subsequent intermediate failure levels is to improve the resolution of the RCM.

30 independent runs are carried out and the mean failure probability curve is plotted in Figure 52. The black dotted lines represent 10 out of the 30 independent runs. It can be observed that the mean failure probability curve agrees well with the

benchmark result at all the intermediate failure levels. More importantly, the spread of the independent runs are considerably reduced compared to the direct application of the RCM as shown in Figure 49.

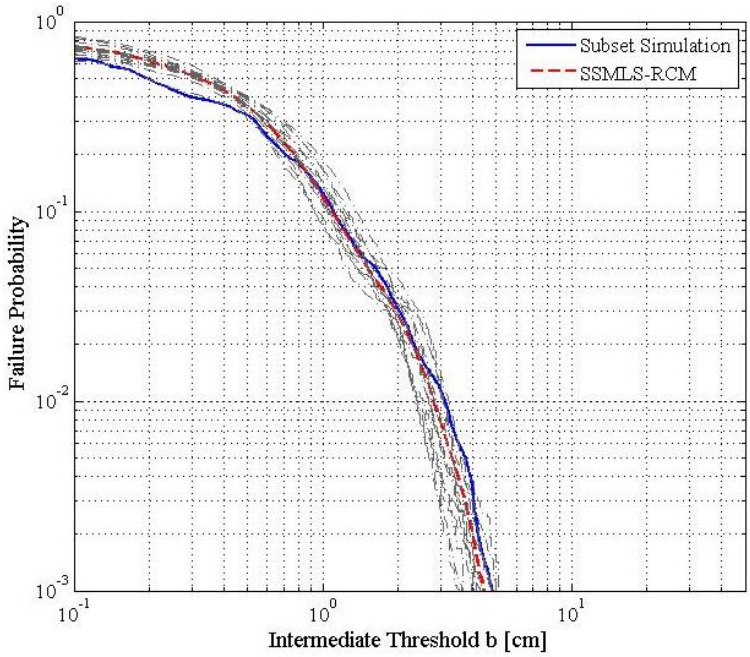


Figure 52 Failure probability estimates of SS and adaptive RCM estimates

The c.o.v. of the adaptive SSMLS-RCM algorithm is presented against the benchmark study as shown in Figure 53. An overall uptrend is observed similar to the Subset Simulation case. However, a drastic fluctuation is recorded at $P_F = 0.1$. The fluctuation may give rise to the varying influence radius used as well as lack of support points at the first failure level. As more support points are generated, the radius is steadier, so as to the approximation. This trend can be directly observed from Figure 52 where the spread of the independent runs is wider at both ends and narrows down at $P_F = 10^{-1.8}$. The computational efficiency for the adaptive SSMLS-RCM algorithms is summarized in Table 16. The number of simulations for adaptive SSMLS-RCM is supposed to be $NS+N_1+2\times N_2 = 450$. However, due to the nature of the MCMC algorithm, a few of the samples taken from the sub-bins are repetitive and therefore the original model simulation computed once for these samples. Although the overall c.o.v. of the adaptive SSMLS-RCM is slightly higher than the one of the Subset Simulation, the computational efficiency of the adaptive SSMLS-RCM is 40% higher than the SS. It should be noted that without increasing

the number of original model simulations, the c.o.v. of the proposed SSMLS-RCM can be reduced significantly by increasing the number of samples obtained by considering metamodel for each conditional failure level.

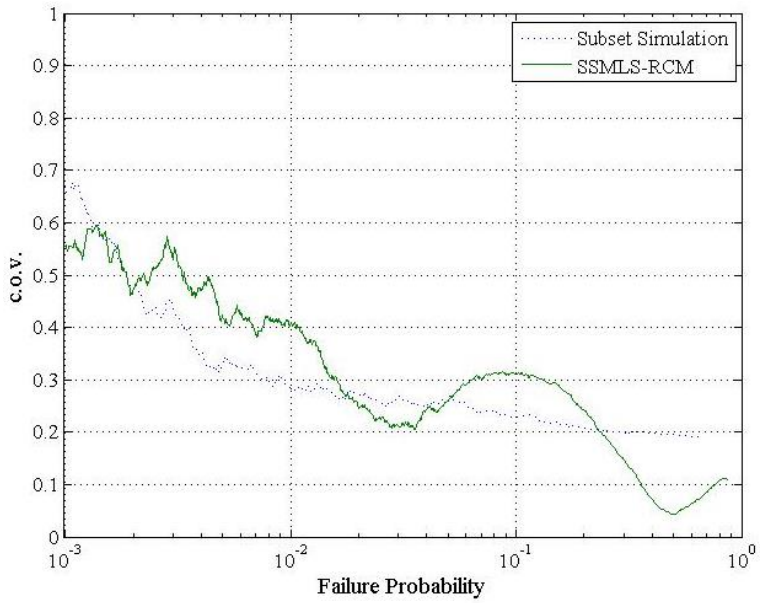


Figure 53 The c.o.v. of failure probability estimates for SS and the SSMLS-RCM

Table 16 Computation efficiency for SSMLS-RCM algorithms

$P_F = 10^{-3}$	Nb. Smln	δ	Δ
Subset Simulation	1500	0.987	38.226
SSMLS-RCM	415	1.129	22.999

Similar to Figure 50, the conditional samples of u_{max} derived from both numerical simulation (x-axis) and the metamodel (y-axis) of the sub-bins for the SSMLS-RCM are plotted in Figure 54. Two main differences can be observed from Figure 54 compared to Figure 50. First of all, the number of sub-bins is doubled from the second intermediate failure level onwards. Secondly, the linear correlations between the actual numerical simulation and approximation from metamodel for the third and fourth levels are higher which indicates a better performance of the metamodel. The correlation coefficients for the four conditional failure probabilities are presented in Figure 55. A triple and double increase of the linear correlations for the third and fourth levels, respectively, can be seen from the chart.

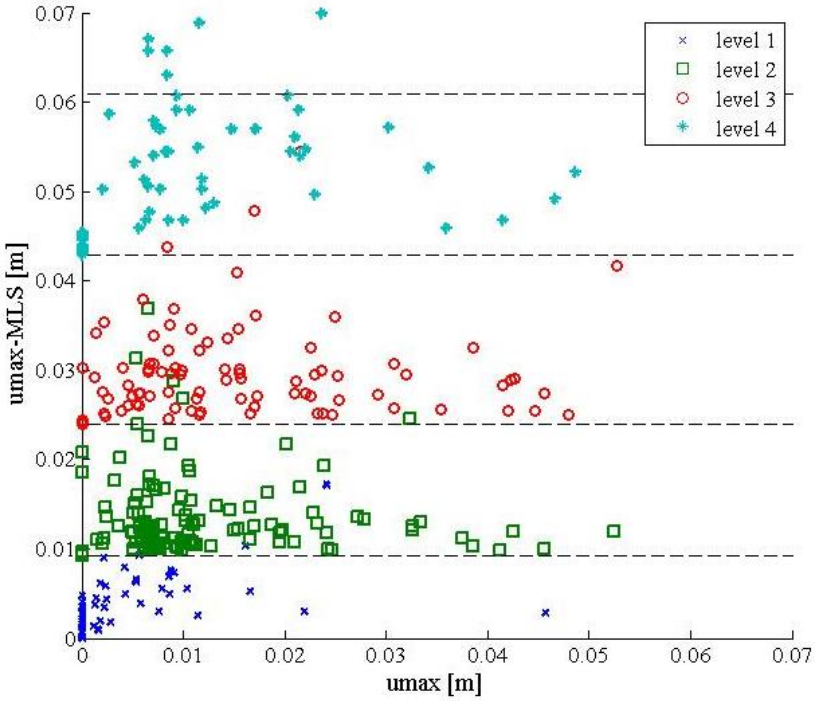
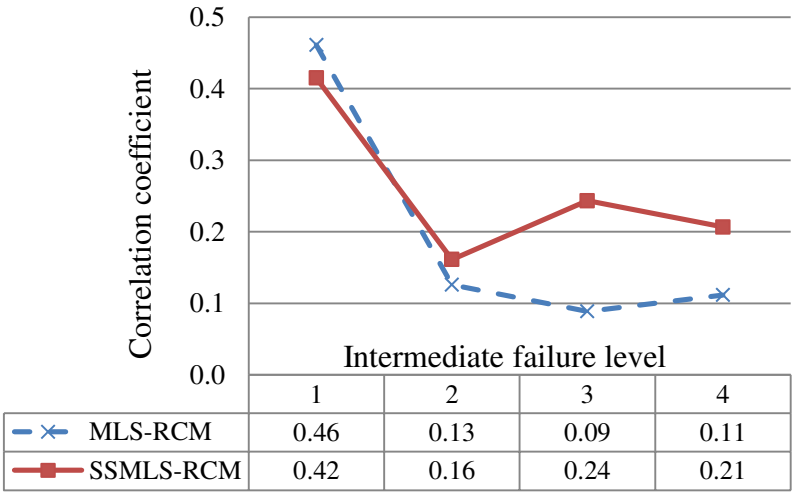


Figure 54 Scattergram of the samples selected from the sub-bins for SSMLS-RCM



	1	2	3	4
MLS-RCM	0.46	0.13	0.09	0.11
SSMLS-RCM	0.42	0.16	0.24	0.21

Figure 55 Comparison of the correlation coefficients for the direct RCM and SSMLS-RCM

6.2 Subset data selection with multiple criterion in GP regression

It was illustrated in Chapter 4 that the Gaussian processes regression is extremely sensitive to the input data. The subset data (SD) selection is a popular topic and many studies have contributed to this issue, such as (Seeger, Williams et al. 2003;

Snelson 2007). Most of the studies involve selecting an active data set based on some sort of criterion which essentially converts the problem into an optimization issue. Compared to single-objective optimization, multi-objective optimization (MOO) gives rise to a number of optimal solutions. It is a process of simultaneously optimizing more than one objective subject to certain constraints. Many algorithms were suggested in the past decades (Fonseca & Fleming 1993; Zitzler & Thiele 1999). In the following section, we present the MOO algorithm for the active data set selection in Gaussian processes regression and use the same benchmark study to illustrate the effectiveness of the proposed algorithm.

6.2.1 Subset data selection for Gaussian processes using multi-objective optimization

The non-dominated sorting genetic algorithm II (NSGA-II) proposed by Deb and Srinivas (1994; Deb *et al.* 2002) obtains a *Pareto front* based on non-domination sorting. The so-called non-domination has two-fold of meanings: (1) a solution which is better than all the other ones in at least one objective and (2) not dominated by all the other solutions. The resultant non-dominated solutions are theoretically equally optimal.

The proposed framework of the multi-objective optimizations integrated with the adaptive SSGP algorithm (MOO-SSGP) is as follows. First of all, the samples generated by numerical simulations are sorted in descending order based on the structural performance. The samples exceeding the current threshold are preceded with the MCMC algorithm without checking the failure condition (e.g. Metropolis-Hasting). This step serves the purpose of varying the elements of the random variables since they are part of the training data for the metamodel. The same set of input is given to the metamodel with different sets of possible combination of the training data. This is similar to the framework shown in Figure 29. The difference between the approximation and the exact result obtained by original model numerical simulations are evaluated. Instead of using single criterion for the subset data selection, three objectives are chosen for the study, namely, mean absolute error (MAE), two-sample Kolmogorov–Smirnov test (K-S test), and log-likelihood.

Upon obtaining the comparison results, non-dominated sorting is implemented which results in a few choices lying on the *Pareto front*. At last, one set of training data randomly chosen from the *Pareto front* is selected. Although the three particular objectives are chosen for the study, there are other measures which can quantitatively compare the differences between the two sets of data. The reason to use these three is because they are the most representative and well-established methods from various perspectives. The MAE focuses on the prediction error by checking the individual points. The two-sample K-S test is used to examine the difference between the prediction and simulations in the context of one-dimensional probability distributions. Finally, the log-likelihood comparison between the two data sets is made based on the best distribution that fits the data.

6.2.2 Illustrative example

An example using the same benchmark study is carried out to showcase the benefit of adopting MOO-SSGP over SSGP. In the MOO step, 20 points are used for the numerical simulation. 50 combinations of subset data were prepared for the three objectives, respectively. The general procedure of the MOO-SSGP follows the adaptive SSGP method.

30 independent runs were conducted for the MOO-SSGP algorithm. A sample result of non-dominated sorting is shown in Table 17. Objectives 1 to 3 refer to average absolute error, Kolmogorov–Smirnov test, and log-likelihood, respectively. Eight rankings are identified by the non-dominated sorting algorithm.

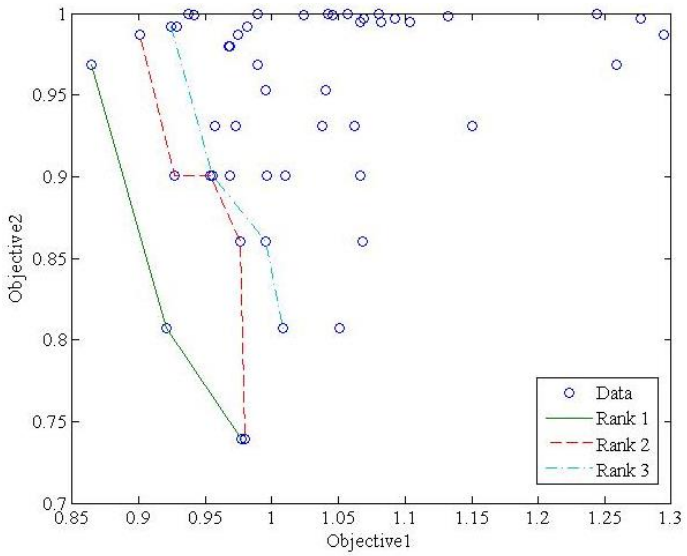
Table 17 Sample result of non-dominated sorting for MOO-SSGP

Nb.	Objective -1	Objective -2	Objective -3	Ranking
1	1.277	0.997	10.563	8
2	0.977	0.860	0.930	1
3	1.066	0.901	4.000	3
4	1.082	0.995	7.617	6
5	1.132	0.998	8.938	8
6	0.967	0.980	6.771	4

7	1.244	1.000	11.086	9
8	1.040	0.953	4.336	4
9	0.981	0.992	8.247	6
10	1.057	1.000	3.462	5
11	1.080	1.000	12.420	8
12	0.937	0.999	7.864	4
13	1.259	0.969	6.926	6
14	0.989	1.000	12.515	7
15	0.968	0.901	4.743	3
16	1.068	0.860	2.853	2
17	0.957	0.931	3.056	3
18	1.294	0.987	10.389	7
19	0.996	0.901	3.113	2
20	0.956	0.901	2.954	2
21	1.038	0.931	0.152	1
22	0.980	0.739	6.097	2
23	1.024	0.999	8.614	7
24	1.042	1.000	3.646	5
25	0.901	0.987	3.660	1
26	0.942	0.999	9.785	5
27	1.092	0.997	11.616	8
28	0.928	0.992	5.811	3
29	0.927	0.901	4.613	2
30	1.066	0.995	7.909	6
31	1.010	0.901	2.566	2
32	0.975	0.987	4.664	5
33	0.989	0.969	2.741	2
34	0.995	0.953	6.368	5
35	0.954	0.901	1.694	1
36	0.942	0.999	9.538	4
37	1.069	0.997	8.602	7
38	0.972	0.931	4.986	4

39	1.062	0.931	6.949	5
40	1.051	0.807	5.649	2
41	0.977	0.739	3.885	1
42	1.151	0.931	5.899	4
43	0.920	0.807	3.426	1
44	1.104	0.995	6.139	6
45	1.045	0.999	8.168	6
46	0.968	0.980	3.443	4
47	0.924	0.992	5.571	2
48	1.008	0.807	0.936	1
49	0.995	0.860	5.517	2
50	0.865	0.969	5.973	1

The paired results of the three objectives are plotted in Figure 56. The approximations by the metamodel are plotted in circles. The first three rankings are labeled in different line types.



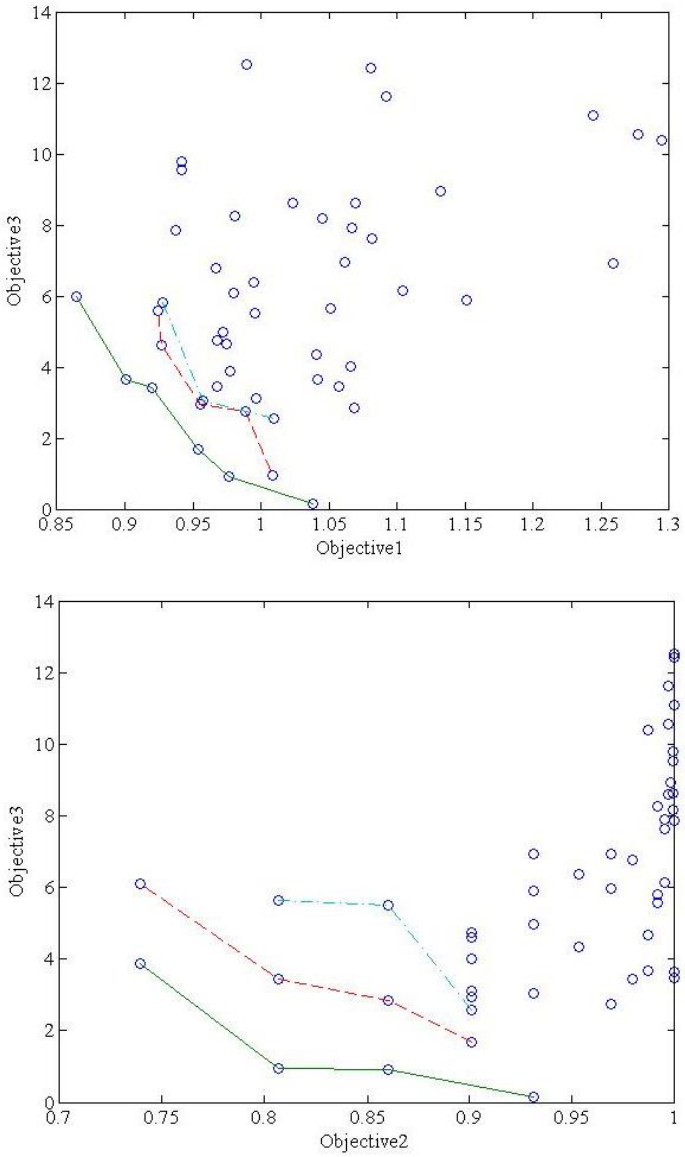


Figure 56 Ranking based on the non-domination sorting

The failure probability estimates by the SS, adv-SSGP and MOO-SSGP are plotted in Figure 57. The fitness of the SB-SSMGP algorithm compared with the Subset Simulation is derived quantitatively and summarized in Table 18. According to the figure and table, the difference between the two versions of the SSGP methods is relatively small. The adv-SSGP method has better performance up to $P_F = 0.1$ and the MOO-SSGP overtakes it beyond $P_F = 0.1$.

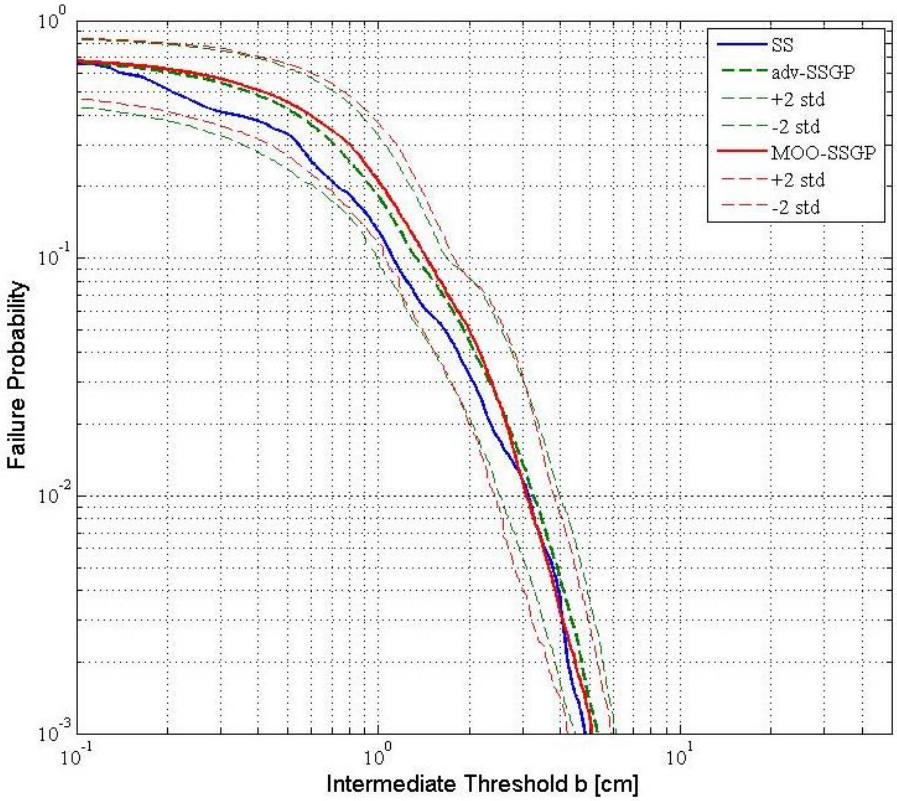


Figure 57 Failure probability estimates of SS, adv-SSGP, and MOO-SSGP with their associated uncertainty levels

Table 18 Fitness of the adv-SSGP and MOO-SSGP compared to SS

	Lv1	Lv2	Lv3
adv-SSGP	0.68	0.82	0.61
MOO-SSGP	0.59	1.04	0.72

The c.o.v. of the MOO-SSGP is plotted in Figure 58, together with the c.o.v. of the benchmark and adv-SSGP. The MOO-SSGP has a lower c.o.v. throughout the conditional failure levels compared to the other two. Both versions of the SSGP use half of the numerical simulations of the SS. The lowest simulation number and c.o.v. make the MOO-SSGP the most efficient algorithms among the three though the adv-SSGP is close. The c.o.v. of the MOO-SSGP can be reduced significantly by increasing the samples generated by metamodel at each intermediate failure level without additional original model numerical simulations.

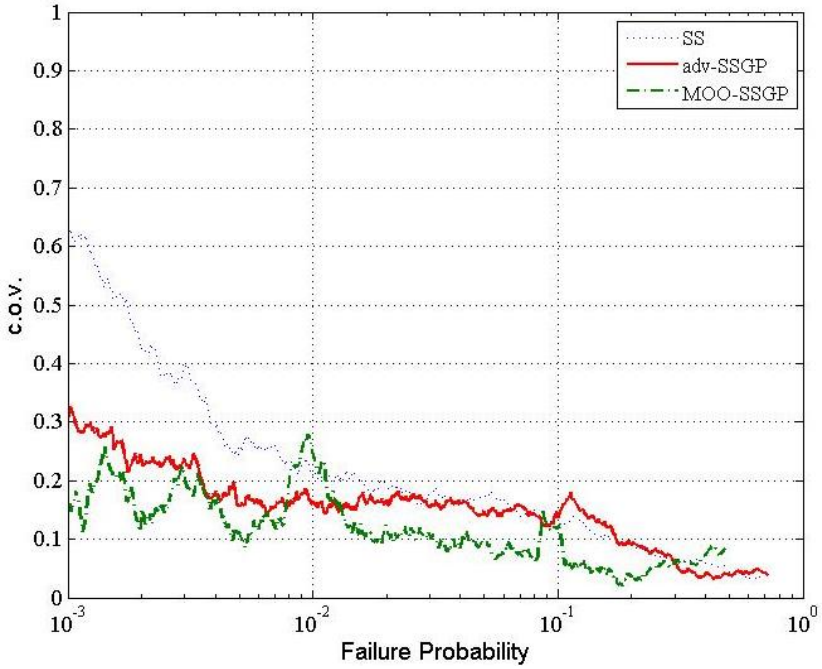


Figure 58 The c.o.v. of failure probability estimates for SS, adv-SSGP and MOO-SSGP

Table 19 Comparison of the computation efficiency

$P_F = 10^{-3}$	Nb. Smln	δ	Δ
Subset Simulation	1500	0.987	38.226
adv-SSGP	666	0.780	20.129
MOO-SSGP	630	0.738	18.524

6.3 Closing remarks

This chapter presents two algorithms, SSMLS-RCM and MOO-SSGP, aiming at further improving the performance of the metamodels based algorithms proposed in the work presented in the previous chapters. As stated at the beginning of this chapter, results obtained by metamodels cannot converge to the results obtained by the original model numerical simulations, due to its approximation nature. Using the original RCM algorithm proposed by Au (2007) can only solve problems where the approximations are more or less linearly correlated with the target response. Therefore, an interactive SSMLS-RCM algorithm is developed to release the constraint to a certain degree. The highlights of the SSMLS-RCM algorithms are: (1) incorporating the RCM-generated numerical simulations into the support points for

the metamodel; and (2) adjusting the influence radius in MLS based on the information obtained from the RCM. The purpose of the algorithm is to increase the linear correlation between the metamodel and the conditioning samples. The resultant exceedance probability curve performs well compared to the benchmark problem while the c.o.v. is significantly reduced compared to the case without the interaction between the RCM and the SSMLS.

The multi-objective optimization allows conflicting objectives to work interactively and provides a set of compromised solutions. The proposed MOO-SSGP algorithm uses the non-dominated sorting algorithm to select the best active data set for the GP regression. Three criteria, *i.e.*, mean absolute error, two-sample K–S test, and log-likelihood, are chosen for the study. Although the failure probability estimates do not significantly improve over the single objective case presented in 4.4.3.2, the c.o.v. is lower with less numerical simulations involved. Therefore a higher computational efficiency is achieved. It is clear that a stable performance in choosing the active data set for the GP reduces the variation of the independent runs.

Apart from the work mentioned above, a few observations which require future investigation are discussed. In the SSMLS-RCM algorithm, a total of 460 simulations were carried out. Increasing the number of sub-bins can further enhance the performance of the integrated metamodel and SS based algorithms which lead to a reduction of the overall c.o.v. To adopt a higher-order basis function as presented in Chapter 5 to improve the quality of the approximation is tempting and worth developing in future. The choice of the objectives in the MOO-SSGP algorithm is subjective. However, more objectives lead to more points in the first *Pareto front* increasing the uncertainty of choosing the optimal solution.

Chapter 7 Conclusion and Future Works

7.1 Conclusion

An increasing frequency of a tsunami event has inevitably raised the potential risk for the tsunami-prone area in terms of both human life and building infrastructure. The current study focuses on the latter in the hope of alleviating the potential structural damage by evaluating the structural reliability. Thus, an efficient yet rigorous approach of evaluating the reliability of critical structures is essential to mitigate the corresponding risk. In order to accomplish the goal, various types of novel reliability analysis algorithms are proposed. A brief review for each chapter is summarized below.

In Chapter 3, an innovative integration of the numerical simulations of the tsunami wave-structure interaction is realized. A stochastic tsunami generation scheme is proposed. The random variables cover all aspects of the earthquake mechanism, including the rake, dip, slip angle, dislocation, focal depth, fault length and width. The number of random variables largely depends on the number of subfaults since each subfault contains different value of the rake, slip angle, and dislocation. The COMCOT is adopted to simulate the tsunami process. The resultant inundation map of a particular area is validated using the 2011 Tohoku event. The output obtained from the tsunami modelling, *i.e.* wave velocity and height, is passed to the wave-structure model to calculate the structural response. The structural model is constructed using LS-DYNA which is able to simulate the wave-structure interaction through Arbitrary Lagrangian Eulerian (ALE) formulation. The reliability analysis of the dynamic system under the risk of earthquake-induced tsunami requires thousands of such numerical simulations. A benchmark problem is set up using the Subset Simulation method at the end of Chapter 3. It is used to compare the results obtained from the proposed algorithms in the following chapters. The main issue for the benchmark problem lies in the repetitive evaluation of the structural response which requires considerable computation efforts.

In Chapter 4, adaptive SS-MLS and adaptive SS-GP algorithms are introduced, focusing on efficiently as well as accurately assessing the structural failure

probability. Both algorithms are constructed based on the novel integration of the Subset Simulation framework and different metamodels. The former is used to estimate the probability for rare events, *i.e.*, the failure probability is evaluated by a sequence of intermediate failure events, and the latter alleviates the computational burden of the numerical simulation. Two metamodels for stochastic sampling are examined in this study: (1) a recently proposed entropy-based moving least squares response surface (MLS) and (2) Gaussian processes regression (GP). The case studies using the proposed algorithms showcase the potential benefits brought by the integration of Subset Simulation (SS) and adaptively constructed metamodels. The adaptive feature enables the metamodels to approximate the responses corresponding to rare probability region more accurately. It can be seen that a comparable failure probability estimate is obtained using the proposed algorithms with much fewer computational efforts compared with the Subset Simulation without integrating with the metamodels. However, the Gaussian processes regression tends to overestimate failure probability in the first few conditional levels, compared to the MLS. This may arise when the choice of the subset of the training data for GP is not optimal and the support points are not sufficient.

More often than not, engineers deal with a complex system containing variables more than a conventional metamodel can handle in practice. Besides, a lower order (first and second-order) basis function in metamodels is not adequate in the case of predicting responses of highly nonlinear systems. In both cases, the metamodel faces the *curse of dimensionality*, *i.e.*, the number of support points required for the metamodel can easily exceed a thousand or even tens of thousands for naïve choice of basis functions. Since the number of support points is determined by the number of coefficients in the basis function, feature selection using Bayesian inference is adopted in Chapter 5. The feature selection reduces the redundant terms in a considerable number which enables the metamodel to be used in higher dimensional problem. According to the results obtained from the two examples, the higher orders MLS with fewer random variables (16 parameters) performs better than the second order MLS with higher random variables (53 parameters). Furthermore, the latter has the number of remaining basis two or three times the former after sparse

Bayesian, even though the total number of basis terms is one quarter of the former. This indicates that more basis terms in the higher dimensional case are relevant compared to the higher orders MLS with lower dimension.

Chapter 6 deals with two other important aspects of the metamodels in reliability assessment. As the result of the metamodel does not converge to the actual result regardless of the number of samples, it is essential to provide a measure for the consistency of the metamodel. The response conditioning method (RCM) and Auxiliary Domain Method (ADM) methods proposed by Au (2007) and Katafygiotis *et al*. (2007), respectively, have shed light on the consistent reliability estimation at the price of extra numerical simulations. However, the original RCM cannot provide consistent correction of the approximation due to the poorly correlated conditional samples and the samples based on considering metamodels. The proposed SSMLS-RCM algorithm becomes an effective tool even for problems involving highly nonlinear systems by modifications made to the RCM including (1) expanding the sub-bins for higher conditional failure levels, as well as (2) creating the communication between the MLS and the RCM. The resultant reliability estimate and c.o.v. are greatly improved when compared to using the RCM and MLS independently.

In conclusion, the objective of the thesis is to develop efficient algorithms to evaluate the performance and reliability of dynamic systems under earthquake-induced tsunami. The main issue is the computational burden introduced by the numerical simulation of the wave generation, propagation, and wave-structure interaction problem. Various algorithms proposed in Chapter 4 – Chapter 6 attempt to crack the problem from different aspects. The accuracy of the metamodel is the most important factor and therefore discussed first. Chapter 5 intends to apply the metamodel on often computationally prohibitive high dimensional problems. Last but not least, the efficiency and consistency of the algorithms are also of great importance. The flowchart of the core structure of the thesis is summarized in Figure 59.

.

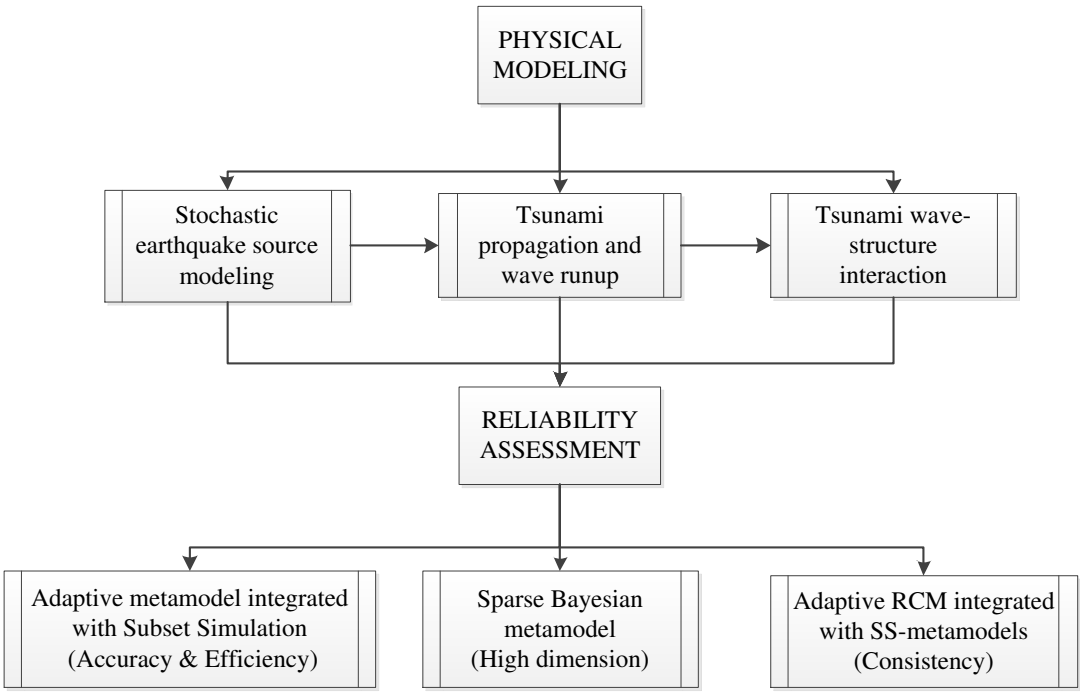


Figure 59 Schematic structure of the study of the structural reliability analysis

7.2 Future works

Due to time constraints, the limitations and potential releases described at the end of each chapter are for future works. They will be carried out to further improve the proposed algorithms.

From a global point of view, it should be emphasized that many other metamodels, such as artificial neural network (ANN), stochastic polynomial chaos (Ghanem & Spanos 1990; Xiu & Karniadakis 2002; Xiu & Karniadakis 2003) and stochastic collocation methods (Babuska, Nobile et al. 2007; Nobile, Tempone et al. 2008), to name just a few, are valid candidates of metamodels to partially replace the original physical model for the reliability analysis. There are other advanced stochastic simulation algorithms such as spherical subset simulation (S^3) (Katafygiotis *et al.* 2010) and domain decomposition method (Katafygiotis & Cheung 2006), etc., which are capable in generating failure samples. These algorithms have a similar function as the Subset Simulation used in the proposed algorithms but more efficient in generating failure samples. Incorporating these advanced algorithms to further develop the proposed methodology is one of our future works.

From a micro-scale perspective, there are two directions worth being explored in the near future. Firstly, metamodels which incorporate multiple outputs (Boyle 2003) is useful when more than one output (max. structural displacement in this case) is desired. Secondly, the adaptive RCM method proposed in Chapter 6 can be applied to the adaptive SB-SSMLS algorithm, which could potentially improve the performance of the metamodel further in higher dimensional system.

Furthermore, the scope will be expanded to cover other civil engineering systems including transportation system, etc. The quantification of economic loss due to earthquake-induced tsunamis is helpful in decision making of disaster mitigation and is a valuable supplement to the thesis.

Reference

Aki, K. (1967). "Scaling law of seismic spectrum." Journal of Geophysical Research **72**(4): 1217-1231.

Aki, K. and P. G. Richards (1980). Quantitative Seismology: Theory and Methods, WH Freeman and Company, San Francisco.

Andrews, D. J. (1980). "A Stochastic Fault Model 1. Static Case." J. Geophys. Res. **85**(B7): 3867-3877.

Andrews, D. J. (1981). "A stochastic fault model: 2. Time-dependent case." Journal of Geophysical Research: Solid Earth **86**(B11): 10821-10834.

Andrieu, C., N. De Freitas, A. Doucet and M. I. Jordan (2003). "An introduction to MCMC for machine learning." Machine learning **50**(1): 5-43.

Ang, A. H.-S. and W. H. Tang (1984). Probability concepts in engineering planning and design.

Ang, G. L., A. H.-S. Ang and W. H. Tang (1992). "Optimal importance-sampling density estimator." Journal of engineering mechanics **118**(6): 1146-1163.

Atsushi Nozu, Strong motion pulses observed during the 2011 Tohoku earthquake and their modeling. l0 CUEE Conference Proceedings, 10th International Conference on Urban Earthquake Engineering, March 1-2, 2013, Tokyo Institute of Technology, Tokyo, Japan.

Au, S.-K. and J. L. Beck (2001). "Estimation of small failure probabilities in high dimensions by subset simulation." Probabilistic Engineering Mechanics **16**(4): 263-277.

Au, S. (2007). "Augmenting approximate solutions for consistent reliability analysis." Probabilistic Engineering Mechanics **22**(1): 77-87.

Au, S. and J. Beck (2003). "Important sampling in high dimensions." Structural safety **25**(2): 139-163.

Babuska, I., F. Nobile and R. Tempone (2007). "A stochastic collocation method for elliptic partial differential equations with random input data." SIAM Journal on Numerical Analysis **45**(3): 1005-1034.

Bailer-Jones, C. A. (2002). "A summary of gaussian processes." University of Cambridge.< http://www. mpia-hd. mpg. de/homes/calj/gps. html>. Acesso em: junho.

Ben-Menahem, A. (1961). "Radiation of seismic surface-waves from finite moving sources." Bulletin of the Seismological Society of America **51**(3): 401-435.

Beroza, G. C., and T. Mikumo (1996). Short slip duration in dynamic rupture in the presence of heterogeneous fault properties, J. Geophys. Res. 101, 22,449–22,460.

Bernard, P., A. Herrero and C. Berge (1996). "Modeling directivity of heterogeneous earthquake ruptures." Bulletin of the Seismological Society of America **86**(4): 1149-1160.

Bilek, S. and T. Lay (2002). "Tsunami earthquakes possibly widespread manifestations of frictional conditional stability." GEOPHYSICAL RESEARCH LETTERS **29**(14): 1673.

Bilek, S. L. and T. Lay (1999). "Rigidity variations with depth along interplate megathrust faults in subduction zones." Nature **400**(6743): 443-446.

Blaser, L., F. Krüger, M. Ohrnberger and F. Scherbaum (2010). "Scaling relations of earthquake source parameter estimates with special focus on subduction environment." Bulletin of the Seismological Society of America **100**(6): 2914-2926.

Bolt, B., Ed. (2005). Earthquakes: 2006 Centennial Update.

Boyle, P., & Frean, M. (2003). Coastal Construction Manuals 3rd edition, FEMA 55. F. E. M. Agency. **3 Vols.**

Breanyn, T. M., A. R. Gusman, R. J. LeVeque and Y. Tanioka (2013). "Comparison of earthquake source models for the 2011 Tohoku-oki event using tsunami simulations and near field observations." Seismological Research Letters **84**(2): 399.

Bryant, E. and J. Nott (2001). "Geological indicators of large tsunami in Australia." Natural Hazards **24**(3): 231-249.

Bucher, C. G. (1988). "Adaptive sampling—an iterative fast Monte Carlo procedure." Structural safety **5**(2): 119-126.

Burjánek, J. and J. Zahradník (2007). "Dynamic stress field of a kinematic earthquake source model with k-squared slip distribution." Geophysical Journal International **171**(3): 1082-1097.

Carden, L., G. Chock, I. Robertson and G. Yu (2012). "Structural analysis case studies of buildings damaged during the Tohoku tsunami."

Carlson, J. M. and J. Langer (1989). "Properties of earthquakes generated by fault dynamics." Physical Review Letters **62**(22): 2632-2635.

Chang, K. O., J. L. Beck and M. Yamada (2008). "Bayesian learning using automatic relevance determination prior with an application to earthquake early warning." Journal of engineering mechanics **134**(12): 1013-1020.

Cheung, S. H., & Beck, J. L. (2010). Calculation of posterior probabilities for Bayesian model class assessment and averaging from posterior samples based on dynamic system data. *Computer-Aided Civil and Infrastructure Engineering*, *25*(5), 304-321.

Chinnery, M. A. (1961). "The deformation of the ground around surface faults." Bulletin of the Seismological Society of America **51**(3): 355-372.

Chock, G., I. Robertson, L. Carden and G. Yu (2012). "Tohoku tsunami-induced building damage analysis including the contribution of earthquake resistant design to tsunami resilience of multi-story buildings."

Choi, S.-K., R. Grandhi and R. A. Canfield (2006). Reliability-based structural design, Springer Science & Business Media.

Cornell, C. A. (1969). A probability-based structural code. ACI Journal Proceedings, ACI.

Dames and Moore (1980). Design and Construction Standards for Residential Construction in Tsunami-Prone Areas in Hawaii (Prepared for the FEMA).

Deb, K., A. Pratap, S. Agarwal and T. Meyarivan (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II." Evolutionary Computation, IEEE Transactions on **6**(2): 182-197.

Dubourg, V., B. Sudret and J.-M. Bourinet (2011). "Reliability-based design optimization using kriging surrogates and subset simulation." Structural and Multidisciplinary Optimization **44**(5): 673-690.

Engelund, S. and R. Rackwitz (1993). "A benchmark study on importance sampling techniques in structural reliability." Structural safety **12**(4): 255-276.

F. Imamura, N.Shuto and C.Goto (1988). Numerical simulations of the transoceanic propagation of tsunamis. Sixth Congress of the Asian and Pacific Regional Division, Int. Assoc. Hydraul. Res. Kyoto, Japan**:** 265-272.

Fonseca, C. M. and P. J. Fleming (1993). Genetic Algorithms for Multiobjective Optimization: FormulationDiscussion and Generalization. ICGA.

Fries, T.-P. and H. G. Matthies (2003). "Classification and overview of meshfree methods." Department of Mathematics and Computer Science, Technical University of Braunschweig.

Fujii, Y. and K. Satake (2007). "Tsunami source of the 2004 Sumatra–Andaman earthquake inferred from tide gauge and satellite data." Bulletin of the Seismological Society of America **97**(1A): S192-S207.

Fujii, Y., K. Satake, S. i. Sakai, M. Shinohara and T. Kanazawa (2011). "Tsunami source of the 2011 off the Pacific coast of Tohoku Earthquake." Earth, Planets and Space **63**(7): 815-820.

Gallovič, F. and J. Brokešová (2004). "On strong ground motion synthesis with k −2 slip distributions." Journal of Seismology **8**(2): 211-224.

Gavin, H. P., & Yau, S. C. (2008). High-order limit state functions in the response surface method for structural reliability analysis. *Structural Safety*,*30*(2), 162-179.

Geist, E. L. (1998). Local Tsunamis and Earthquake Source Parameters. Advances in Geophysics. D. Renata and S. Barry, Elsevier. **Volume 39:** 117-209.

Geist, E. L. (2005). "Rapid tsunami models and earthquake source parameters: far-field and local applications." ISET J. Earthq. Technol **42**: 127-136.

Geist, E. L. and T. Parsons (2011). "Assessing historical rate changes in global tsunami occurrence." Geophysical Journal International **187**(1): 497-509.

Gelfand, A. E. and A. F. Smith (1990). "Sampling-based approaches to calculating marginal densities." Journal of the American statistical Association **85**(410): 398-409.

Gelman, A. and K. Shirley (2011). "Inference from simulations and monitoring convergence." Handbook of Markov chain Monte Carlo: 163-174.

Geman, S. and D. Geman (1984). "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images." Pattern Analysis and Machine Intelligence, IEEE Transactions on(6): 721-741.

Geyer, C. J. and E. A. Thompson (1992). "Constrained Monte Carlo maximum likelihood for dependent data." Journal of the Royal Statistical Society. Series B (Methodological): 657-699.

Ghanem, R. and P. Spanos (1990). "Polynomial chaos in stochastic finite elements." Journal of Applied Mechanics **57**(1): 197-202.

Gilks, W. R. (2005). Markov chain monte carlo, Wiley Online Library.

Gilks, W. R. and G. O. Roberts (1996). "Strategies for improving MCMC." Markov chain Monte Carlo in practice **6**: 89-114.

Griffith, A. A. (1921). "The phenomena of rupture and flow in solids." Philosophical transactions of the royal society of london. Series A, containing papers of a mathematical or physical character **221**: 163-198.

Grilli, S. T., J. C. Harris, T. S. Tajalli Bakhsh, T. L. Masterlark, C. Kyriakopoulos, J. T. Kirby and F. Shi (2012). "Numerical simulation of the 2011 Tohoku tsunami based on a new transient FEM co-seismic source: Comparison to far-and near-field observations." Pure and Applied Geophysics: 1-27.

Guatteri, M., P. Martin Mai, G. C. Beroza, and J. Boatwright (2003). Strong ground-motion prediction from stochastic-dynamic source models, Bulletin of The Seismological Society of America 93, 301–313.

Guatteri, M., P. Martin Mai, and G. C. Beroza (2004). A pseudo-dynamic approximation to dynamic rupture models for strong ground motion prediction, Bulletin of The Seismological Society of America 94, 2051–2063.

Hanks, T. C. (1979). "b values and ω−γ seismic source models: Implications for tectonic stress variations along active crustal fault zones and the estimation of high-frequency strong ground motion." Journal of Geophysical Research: Solid Earth **84**(B5): 2235-2242.

Haritos, N., T. Ngo and P. Mendis (2005). "Evaluating Tsunami Wave Forces on Structures." Monash University.

Hartzell, S., Guatteri, M., Mai, P. M., Liu, P. C., & Fisk, M. (2005). Calculation of broadband time histories of ground motion, Part II: Kinematic and dynamic

modeling using theoretical Green's functions and comparison with the 1994 Northridge earthquake. Bulletin of the Seismological Society of America, 95(2), 614-645.

Haskell, N. (1964). "Total energy and energy spectral density of elastic wave radiation from propagating faults." Bulletin of The Seismological Society of America **54**(6A): 1811-1841.

Haskell, N. (1966). "Total energy and energy spectral density of elastic wave radiation from propagating faults. Part II. A statistical source model." Bulletin of The Seismological Society of America **56**(1): 125-140.

Hasofer, A. M. and N. C. Lind (1974). "Exact and invariant second-moment code format." Journal of the Engineering Mechanics division **100**(1): 111-121.

Hayes, G. P. (2011). "Rapid source characterization of the 2011 Mw 9.0 off the Pacific coast of Tohoku Earthquake." Earth, Planets and Space **63**(7): 529-534.

Herrero, A. and P. Bernard (1994). "A kinematic self-similar rupture process for earthquakes." Bulletin of The Seismological Society of America **84**(4): 1216-1228.

Hisada, Y. (2000). "A theoretical omega-square model considering the spatial variation in slip and rupture velocity." Bulletin of The Seismological Society of America **90**(2): 387-400.

http://www.globalcmt.org "Web Page." Global Centroid Moment Tensor Project http://www.globalcmt.org/ (accessed 30 March 2011).

Ihmlé, P. F. (1996). "Monte Carlo slip inversion in the frequency domain:: Application to the 1992 Nicaragua Slow Earthquake." GEOPHYSICAL RESEARCH LETTERS **23**(9): 913-916.

Iizuka, H. and H. Matsutomi (2000). Damage due to the Flooding Flow of Tsunami. Proceedings of Coastal Engineering, JSCE.

Imamura, F. (1995). "Tsunami numerical simulation with the staggered leap-frog scheme (numerical code of TUNAMI-N1 and N2)." Disaster Control Research Center, Tohoku University.

Kajiura, K. (1982). "Tsunami energy in relation to parameters of the earthquake fault model."

Kanamori, H. (1972). "Mechanism of tsunami earthquakes." Physics of the Earth and Planetary Interiors **6**: 346-359.

Kanamori, H. and E. E. Brodsky (2004). "The physics of earthquakes." Reports on Progress in Physics **67**(8): 1429.

Kanamori, H. and M. Kikuchi (1993). "The 1992 Nicaragua earthquake: a slow tsunami earthquake associated with subducted sediments." Nature **361**(6414): 714-716.

Katafygiotis, L. and S. H. Cheung (2006). "Domain decomposition method for calculating the failure probability of linear dynamic systems subjected to Gaussian stochastic loads." Journal of engineering mechanics **132**(5): 475-486.

Katafygiotis, L., S. H. Cheung and K.-V. Yuen (2010). "Spherical subset simulation (S^3) for solving non-linear dynamical reliability problems." International Journal of Reliability and Safety **4**(2): 122-138.

Katafygiotis, L., T. Moan and S. H. Cheung (2007). "Auxiliary domain method for solving multi-objective dynamic reliability problems for nonlinear structures." Structural Engineering and Mechanics **25**(3): 347-363.

Katafygiotis, L. S. and K. M. Zuev (2008). "Geometric insight into the challenges of solving high-dimensional reliability problems." Probabilistic Engineering Mechanics **23**(2): 208-218.

Kelleher, J., L. Sykes and J. Oliver (1973). "Possible criteria for predicting earthquake locations and their application to major plate boundaries of the Pacific and the Caribbean." Journal of Geophysical Research **78**(14): 2547-2585.

Kikuchi, M. and Y. Fukao (1987). "Inversion of long-period P-waves from great earthquakes along subduction zones." Tectonophysics **144**(1): 231-247.

Kikuchi, M. and H. Kanamori (1995). "Source characteristics of the 1992 Nicaragua tsunami earthquake inferred from teleseismic body waves." Pure and Applied Geophysics **144**(3): 441-453.

Kirkoz, M. (1983). Breaking and run-up of long waves, tsunamis: their science and engineering. Proceedings of the 10th IUGG International Tsunami Symposium, Sendai-shi/Miyagi-ken, Japan. Terra Scientific Publishing, Tokyo, Japan.

Koshimura, S., T. Oie, H. Yanagisawa and F. Imamura (2009). "Developing fragility functions for tsunami damage estimation using numerical model and post-tsunami data from Banda Aceh, Indonesia." Coastal Engineering Journal **51**(03): 243-273.

Kostrov, B. V. and S. Das (1989). Principles of earthquake source mechanics, Cambridge University Press.

Koyama, J. (1997). The complex faulting process of earthquakes, Kluwer.

Krishnamurthy, T. and V. Romero (2002). Construction of response surface with higher order continuity and its application to reliability engineering. Proceedings of the 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference.

Lancaster, P. and K. Salkauskas (1981). "Surfaces generated by moving least squares methods." Mathematics of computation **37**(155): 141-158.

Lau, T. L., T. Ohmachi, S. Inoue and P. Lukkunaprasit (2011). "Experimental and Numerical Modeling of Tsunami Force on Bridge Decks."

Lavallée, D., P. Liu and R. J. Archuleta (2006). "Stochastic model of heterogeneity in earthquake slip spatial distributions." Geophysical Journal International **165**(2): 622-640.

Lawrence, N. (2005). "Probabilistic non-linear principal component analysis with Gaussian process latent variable models." The Journal of Machine Learning Research **6**: 1783-1816.

Lay, T., C. J. Ammon, H. Kanamori, M. J. Kim and L. Xue (2011). "Outer trench-slope faulting and the 2011 Mw 9.0 off the Pacific coast of Tohoku Earthquake." Earth, Planets and Space **63**(7): 713-718.

Lay, T. and T. C. Wallace (1995). Modern global seismology, Academic press.

Levin, D. (1998). "The approximation power of moving least-squares." Mathematics of computation **67**(224): 1517-1532.

Liu, P.-L. and A. Der Kiureghian (1991). "Optimization algorithms for structural reliability." Structural safety **9**(3): 161-177.

Løvholt, F., G. Pedersen, S. Bazin, D. Kühn, R. Bredesen and C. Harbitz (2012). "Stochastic analysis of tsunami runup due to heterogeneous coseismic slip and dispersion." Journal of Geophysical Research **117**(C3): C03047.

Madariaga, R. and A. Cochard (1994). "Seismic source dynamics, heterogeneity and friction." Annals of Geophysics **37**(6).

Madariaga, R. and K. B. Olsen (2002). "Earthquake dynamics." International Geophysics **81**: 175-III.

Magallanes, J. M., Y. Wu, L. J. Malvar and J. E. Crawford (2010). Recent Improvements to Release III of the K&C Concrete Model. 11th International LS-DYNA Users Conference.

Mai, P. M. and G. C. Beroza (2002). "A spatial random field model to characterize complexity in earthquake slip." J. Geophys. Res. **107**(B11): 2308.

Malvar, L. J. and J. E. Crawford (1998). Dynamic increase factors for concrete, DTIC Document.

Matheron, G. (1973). "The intrinsic random functions and their applications." Advances in applied probability: 439-468.

Melchers, R. (1989). "Importance sampling in structural systems." Structural safety **6**(1): 3-10.

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

Mirza, S. A., & MacGregor, J. G. (1982). Probabilistic study of strength of reinforced concrete members. Canadian Journal of Civil Engineering, 9(3), 431-448.

Myers, R. H., D. C. Montgomery and C. M. Anderson-Cook (2009). Response surface methodology: process and product optimization using designed experiments, Wiley.

Nagano, O., F. Imamura and N. Shuto (1991). "A numerical model for far-field tsunamis and its application to predict damages done to aquaculture." Natural Hazards **4**(2): 235-255.

Neal, R. M. (1995). Bayesian learning for neural networks, University of Toronto.

Nimmala, S. B., S. C. Yim, K. F. Cheung and Y. Wei (2006). Tsunami Design Criteria for Coastal Infrastructure: A Case Study for Spencer Creek Bridge, Oregon.

Nistor Ioan, Dan Palermo, Younes Nouri, Tad Murty and M. Saatcioglu (2010). Tsunami-Induced Forces on Structures. Handbook of Coastal and Ocean Engineering. K. Young. Singapore, World Scientific**:** 261-285.

Nobile, F., R. Tempone and C. G. Webster (2008). "A sparse grid stochastic collocation method for partial differential equations with random input data." SIAM Journal on Numerical Analysis **46**(5): 2309-2345.

O'Hagan, A. and J. Kingman (1978). "Curve fitting and optimal design for prediction." Journal of the Royal Statistical Society. Series B (Methodological): 1-42.

Oh, C. K. (2008). Bayesian learning for earthquake engineering applications and structural health monitoring, California Institute of Technology.

Okada Tsuneo, Tadashi Sugano, Tadashi Ishikawa, Takero Ohgi, Shigemitsu Takai and C. Hamabe (2005). "Structural Design Method of Buildings for Tsunami Resistance, (SMBTR)." The Building Letter (In Japanese).

Okada, Y. (1985). "Surface deformation due to shear and tensile faults in a half-space." Bulletin of The Seismological Society of America **75**(4): 1135-1154.

Okal, E. A. (1988). "Seismic parameters controlling far-field tsunami amplitudes: a review." Natural Hazards **1**(1): 67-96.

Omira, R., M. Baptista, J. Miranda, E. Toto, C. Catita and J. Catalao (2010). "Tsunami vulnerability assessment of Casablanca-Morocco using numerical modelling and GIS tools." Natural Hazards **54**(1): 75-95.

Ozaki, T. (2011). "Outline of the 2011 off the Pacific coast of Tohoku Earthquake(M (w) 9. 0)-Tsunami warnings/advisories and observations." Earth, Planets and Space **63**(7): 827-830.

Ozawa, S., T. Nishimura, H. Suito, T. Kobayashi, M. Tobita and T. Imakiire (2011). "Coseismic and postseismic slip of the 2011 magnitude-9 Tohoku-Oki earthquake." Nature **475**(7356): 373-376.

Pacheco, J. F. and L. R. Sykes (1992). "Seismic moment catalog of large shallow earthquakes, 1900 to 1989." Bulletin of The Seismological Society of America **82**(3): 1306-1349.

Pelayo, A. M. and D. A. Wiens (1992). "Tsunami earthquakes: slow thrust-faulting events in the accretionary wedge." Journal of Geophysical Research **97**(B11): 15321-15315,15337.

Peyrat, S., and K. B. Olsen (2004). Nonlinear dynamic rupture inversion of the 2000 western Tottori, Japan, earthquake, Geophys. Res. Lett. 31, L05604.

Piatanesi, A., S. Tinti and I. Gavagni (1996). "The slip distribution of the 1992 Nicaragua earthquake." GEOPHYSICAL RESEARCH LETTERS **23**(1): 37-40.

Rasmussen, C. E. and C. Williams (2006). "Gaussian processes for machine learning. 2006." The MIT Press, Cambridge, MA, USA **38**: 715-719.

Rice, J. (1993). "Spatio-temporal complexity of slip on a fault." Journal of Geophysical Research **98**: 9885-9907.

Richter, T. (2010). "Numerical Methods for Fluid-Structure Interaction Problems." Course of lectures.

Ripley, B. D. (1991). Statistical inference for spatial processes, Cambridge university press.

Robertson, I., L. Carden, H. Riggs, S. Yim, Y. Young, K. Paczkowski and D. Witt (2010). Reconnaissance following the September 29, 2009 tsunami in Samoa, Research report UHM/CEE/10-01, Department of Civil and Environmental Engineering, University of Hawaii at Manoa, Honolulu, Hawaii, http://www.cee.hawaii.edu/reports/UHM-CEE-10-01.pdf.

Romanowicz, B. and J. B. Rundle (1993). "On scaling relations for large earthquakes." BULLETIN OF THE SEISMOLOGICAL SOCIETY OF AMERICA **83**(4): 1294-1297.

Rundle, J. B. (1989). "Derivation of the complete Gutenberg-Richter magnitude-frequency relation using the principle of scale invariance." Journal of Geophysical Research **94**(B9): 12337-12312,12342.

Saffer, D. M. and B. A. Bekins (1999). "Fluid budgets at convergent plate margins: Implications for the extent and duration of fault-zone dilation." Geology **27**(12): 1095-1098.

Satake, K. (1994). "Mechanism of the 1992 Nicaragua tsunami earthquake." GEOPHYSICAL RESEARCH LETTERS **21**(23): 2519-2522.

Satake, K. and H. Kanamori (1991). "Use of tsunami waveforms for earthquake source study." Natural Hazards **4**(2): 193-208.

Scholz, C. H. (1982). "Scaling laws for large earthquakes: consequences for physical models." BULLETIN OF THE SEISMOLOGICAL SOCIETY OF AMERICA **72**(1): 1-14.

Schwartz, D. P. and K. J. Coppersmith (1984). "Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas fault zones." Journal of Geophysical Research **89**(B7): 5681-5698.

Schwer, L. E. and L. J. Malvar (2005). "Simplified concrete modeling with* MAT_CONCRETE_DAMAGE_REL3." LS-DYNA Anwenderforum, Bamberg **2005**.

Seeger, M., C. Williams and N. Lawrence (2003). Fast forward selection to speed up sparse Gaussian process regression. Artificial Intelligence and Statistics 9.

Seno, T. (2002). "Tsunami earthquakes as transient phenomena." GEOPHYSICAL RESEARCH LETTERS **29**(10): 1419.

Shao, G., X. Li, C. Ji and T. Maeda (2011). "Focal mechanism and slip history of the 2011 Mw 9.1 off the Pacific coast of Tohoku Earthquake, constrained with teleseismic body and surface waves." Earth, Planets and Space **63**(7): 559-564.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. Proceedings of the 1968 23rd ACM national conference, ACM.

Shibata, H. (1998). "Design basis earthquakes for critical industrial facilities and their characteristics, and the Southern Hyogo prefecture earthquake, 17 January 1995." Reliability Engineering & System Safety **62**(3): 157-169.

Shimazaki, K. and T. Nakata (1980). "Time-predictable recurrence model for large earthquakes." GEOPHYSICAL RESEARCH LETTERS **7**(4): 279-282.

Shinozuka, M. (1983). "Basic analysis of structural safety." Journal of structural engineering.

Shuto, N. (1991). "Numerical simulation of tsunamis—Its present and near future." Natural Hazards **4**(2): 171-191.

Shuto, N. and K. Fujima (2009). "A short history of tsunami research and countermeasures in Japan." Proceedings of the Japan Academy. Series B, Physical and biological sciences **85**(8): 267.

Snelson, E. and Z. Ghahramani (2006). "Sparse Gaussian processes using pseudo-inputs."

Snelson, E. L. (2007). Flexible and efficient Gaussian process models for machine learning, Citeseer.

Somerville, P., K. Irikura, R. Graves, S. Sawada, D. Wald, N. Abrahamson, Y. Iwasaki, T. Kagawa, N. Smith and A. Kowada (1999). "Characterizing crustal earthquake slip models for the prediction of strong ground motion." Seismological Research Letters **70**(1): 59-80.

Srinivas, N. and K. Deb (1994). "Muiltiobjective optimization using nondominated sorting in genetic algorithms." Evolutionary computation **2**(3): 221-248.

Steketee, J. A. (1958). "ON VOLTERRA'S DISLOCATIONS IN A SEMI-INFINITE ELASTIC MEDIUM." Canadian Journal of Physics **36**(2): 192-205.

Taflanidis, A. (2009). Stochastic subset optimization with response surface methodologies for stochastic design. 1st International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering.

Taflanidis, A. A. and S.-H. Cheung (2012). "Stochastic sampling using moving least squares response surface approximations." Probabilistic Engineering Mechanics **28**: 216-224.

Tang, Q., Y. B. Lau, S. Hu, W. Yan, Y. Yang and T. Chen (2010). "Response surface methodology using Gaussian processes: towards optimizing the trans-stilbene epoxidation over Co 2+–NaX catalysts." Chemical Engineering Journal **156**(2): 423-431.

Tanner, M. A. and W. H. Wong (1987). "The calculation of posterior distributions by data augmentation." Journal of the American statistical Association **82**(398): 528-540.

Thatcher, W. (1990). "Order and diversity in the modes of circum-Pacific earthquake recurrence." Journal of Geophysical Research **95**(B3): 2609-2623.

Tipping, M. E. (2001). "Sparse Bayesian learning and the relevance vector machine." The Journal of Machine Learning Research **1**: 211-244.

Tipping, M. E. and A. C. Faul (2003). Fast marginal likelihood maximisation for sparse Bayesian models. Proceedings of the ninth international workshop on artificial intelligence and statistics.

Titov, V. and F. Gonzalez (1997). Implementation and testing of the method of splitting tsunami (MOST) model, US Department of Commerce, National Oceanic and Atmospheric Administration, Environmental Research Laboratories, Pacific Marine Environmental Laboratory.

Titov, V. V. (1997). Numerical modeling of long wave runup.

Titov, V. V., F. I. Gonzalez, E. Bernard, M. C. Eble, H. O. Mofjeld, J. C. Newman and A. J. Venturato (2005). "Real-time tsunami forecasting: Challenges and solutions." Natural Hazards **35**(1): 35-41.

Tukey, J. W. (1967). "An introduction to the calculations of numerical spectrum analysis." Spectral analysis of time series **25**(6).

Volterra, V. (1907). "Sur luilibre des corps elastiques multiplement connexes." Ann. Éc. Norm. **3**(24): 401-517.

Wang, X. M. (2009). "USER MANUAL FOR COMCOT VERSION 1.7 ".

Wei, Y., C. Chamberlin, V. V. Titov, L. Tang and E. N. Bernard (2012). "Modeling of the 2011 Japan Tsunami: Lessons for Near-Field Forecast." Pure and Applied Geophysics: 1-23.

Wells, D. L. and K. J. Coppersmith (1994). "New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement." BULLETIN OF THE SEISMOLOGICAL SOCIETY OF AMERICA **84**(4): 974-1002.

Wiener, N. (1949). Extrapolation, interpolation, and smoothing of stationary time series, MIT press Cambridge, MA.

Williams, C. K. and C. E. Rasmussen (1996). "Gaussian processes for regression."

Wyss, M. and J. N. Brune (1967). "The Alaska earthquake of 28 March 1964: A complex multiple rupture." BULLETIN OF THE SEISMOLOGICAL SOCIETY OF AMERICA **57**(5): 1017-1023.

Xiu, D. and G. E. Karniadakis (2002). "The Wiener--Askey Polynomial Chaos for Stochastic Differential Equations." SIAM Journal on Scientific Computing **24**(2): 619-644.

Xiu, D. and G. E. Karniadakis (2003). "Modeling uncertainty in flow simulations via generalized polynomial chaos." Journal of Computational Physics **187**(1): 137-167.

Yamamoto, S. and N. Hori (2004). "Average and variation of focal mechanism around Tohoku subduction zone."

Yamashita, T. and R. Sato (1974). "Generation of tsunami by a fault model." Journal of Physics of the Earth **22**(4): 415-440.

Yee, K. S. (1966). Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. IEEE Trans. Antennas Propag, *14*(3), 302-307.

Yeh, H. H. and I. Robertson (2005). Development of Design Guideline for Tsunami Shelters. First International Conference on Urban Disaster Reduction. Kobe, Japan, Washington State Department of Natural Resources, Division of Geology and Earth Resources.

Yim, S. C., K. F. Cheung, M. J. Olsen and Y. Yamazaki (2012). "Tohoku tsunami survey, modeling and probabilistic load estimation applications."

Yolsal, S. and T. Taymaz (2010). "Sensitivity analysis on relations between earthquake source rupture parameters and far-field tsunami waves: case studies in the eastern Mediterranean region." Turk. J. Earth Sci **19**: 313-349.

Zeng, Y., J. G. Anderson and G. Yu (1994). "A composite source model for computing realistic synthetic strong ground motions." GEOPHYSICAL RESEARCH LETTERS **21**(8): 725-728.

Zitzler, E. and L. Thiele (1999). "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach." Evolutionary Computation, IEEE Transactions on **3**(4): 257-271.

Zuev, K. M., J. L. Beck, S.-K. Au and L. S. Katafygiotis (2012). "Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions." Computers & structures **92**: 283-296.

Appendix A.1 Sample COMCOT Input for the Benchmark Study

No.	Time	Longitude	Latitude	Fault Length[m]	Fault Width [m]	Focal depth [m]	Strike [deg]	Dip [deg]	Rake [deg]	Slip [m]
1	0	142.2	38.4	45923.0	18156.4	24400.0	199	10	6.4	0.0
2	0	142.0	37.9	45923.0	18156.4	24400.0	199	10	133.4	0.0
3	0	141.9	37.4	45923.0	18156.4	24400.0	199	10	79.9	0.0
4	0	141.7	36.9	45923.0	18156.4	24400.0	199	10	119.7	0.0
5	0	141.5	36.5	45923.0	18156.4	24400.0	199	10	122.5	0.0
6	0	141.4	36.0	45923.0	18156.4	24400.0	199	10	103.8	0.0
7	0	142.4	38.4	45923.0	18156.4	27552.8	199	10	92.9	0.0
8	0	142.2	37.9	45923.0	18156.4	27552.8	199	10	132.9	15.9
9	0	142.1	37.4	45923.0	18156.4	27552.8	199	10	46.9	21.6
10	0	141.9	36.9	45923.0	18156.4	27552.8	199	10	70.3	25.0
11	0	141.7	36.4	45923.0	18156.4	27552.8	199	10	51.4	19.3
12	0	141.6	35.9	45923.0	18156.4	27552.8	199	10	75.9	0.0
13	0	142.6	38.3	45923.0	18156.4	30705.7	199	10	45.5	0.0
14	0	142.4	37.8	45923.0	18156.4	30705.7	199	10	129.7	18.2
15	0	142.3	37.3	45923.0	18156.4	30705.7	199	10	95.1	24.4
16	0	142.1	36.8	45923.0	18156.4	30705.7	199	10	106.1	27.8
17	0	141.9	36.3	45923.0	18156.4	30705.7	199	10	108.0	21.6

18	0	141.8	35.8	45923.0	18156.4	30705.7	199	10	140.6	0.0
19	0	142.8	38.2	45923.0	18156.4	33858.5	199	10	70.2	0.0
20	0	142.6	37.7	45923.0	18156.4	33858.5	199	10	48.3	15.9
21	0	142.5	37.2	45923.0	18156.4	33858.5	199	10	56.5	21.0
22	0	142.3	36.7	45923.0	18156.4	33858.5	199	10	106.5	23.3
23	0	142.1	36.3	45923.0	18156.4	33858.5	199	10	95.2	18.2
24	0	142.0	35.8	45923.0	18156.4	33858.5	199	10	98.4	0.0
25	0	143.0	38.2	45923.0	18156.4	37011.3	199	10	96.7	0.0
26	0	142.8	37.7	45923.0	18156.4	37011.3	199	10	81.1	11.3
27	0	142.7	37.2	45923.0	18156.4	37011.3	199	10	38.4	14.7
28	0	142.5	36.7	45923.0	18156.4	37011.3	199	10	96.0	15.9
29	0	142.3	36.2	45923.0	18156.4	37011.3	199	10	138.1	12.5
30	0	142.1	35.7	45923.0	18156.4	37011.3	199	10	54.5	0.0
31	0	143.2	38.1	45923.0	18156.4	40164.1	199	10	120.3	0.0
32	0	143.0	37.6	45923.0	18156.4	40164.1	199	10	131.3	0.0
33	0	142.9	37.1	45923.0	18156.4	40164.1	199	10	64.8	0.0
34	0	142.7	36.6	45923.0	18156.4	40164.1	199	10	89.2	0.0
35	0	142.5	36.1	45923.0	18156.4	40164.1	199	10	97.9	0.0
36	0	142.3	35.6	45923.0	18156.4	40164.1	199	10	123.0	0.0

Appendix A.2 Sample COMCOT Input for Higher Dimensional Case Study

No.	Time	Longitude	Latitude	Fault Length[m]	Fault Width [m]	Focal depth [m]	Strike [deg]	Dip [deg]	Rake [deg]	Slip [m]
1	0.0	142.7	39.5	49272.2	18009.5	33190.1	250.3	25.9	45.3	0.0
2	0.0	142.2	39.3	49272.2	18009.5	33190.1	250.3	25.9	67.7	0.0
3	0.0	141.7	39.1	49272.2	18009.5	33190.1	250.3	25.9	58.2	0.0
4	0.0	141.1	38.9	49272.2	18009.5	33190.1	250.3	25.9	160.5	0.0
5	0.0	140.6	38.7	49272.2	18009.5	33190.1	250.3	25.9	71.5	0.0
6	0.0	140.1	38.5	49272.2	18009.5	33190.1	250.3	25.9	112.4	0.0
7	0.0	139.5	38.3	49272.2	18009.5	33190.1	250.3	25.9	84.2	0.0
8	0.0	139.0	38.1	49272.2	18009.5	33190.1	250.3	25.9	116.7	0.0
9	0.0	142.8	39.3	49272.2	18009.5	41045.2	250.3	25.9	67.1	0.0
10	0.0	142.3	39.1	49272.2	18009.5	41045.2	250.3	25.9	47.9	13.3
11	0.0	141.7	38.9	49272.2	18009.5	41045.2	250.3	25.9	47.3	21.9
12	0.0	141.2	38.7	49272.2	18009.5	41045.2	250.3	25.9	104.6	21.4
13	0.0	140.7	38.5	49272.2	18009.5	41045.2	250.3	25.9	84.7	24.4
14	0.0	140.1	38.3	49272.2	18009.5	41045.2	250.3	25.9	84.1	17.2
15	0.0	139.6	38.1	49272.2	18009.5	41045.2	250.3	25.9	132.6	20.2
16	0.0	139.1	37.9	49272.2	18009.5	41045.2	250.3	25.9	98.7	0.0

17	0.0	142.9	39.1	49272.2	18009.5	48900.3	250.3	25.9	95.9	0.0
18	0.0	142.3	38.9	49272.2	18009.5	48900.3	250.3	25.9	137.6	22.5
19	0.0	141.8	38.7	49272.2	18009.5	48900.3	250.3	25.9	65.9	25.2
20	0.0	141.3	38.5	49272.2	18009.5	48900.3	250.3	25.9	110.9	28.3
21	0.0	140.7	38.3	49272.2	18009.5	48900.3	250.3	25.9	115.1	31.2
22	0.0	140.2	38.1	49272.2	18009.5	48900.3	250.3	25.9	82.7	24.0
23	0.0	139.7	37.9	49272.2	18009.5	48900.3	250.3	25.9	96.5	21.4
24	0.0	139.1	37.7	49272.2	18009.5	48900.3	250.3	25.9	55.0	0.0
25	0.0	142.9	38.9	49272.2	18009.5	56755.4	250.3	25.9	55.6	0.0
26	0.0	142.4	38.7	49272.2	18009.5	56755.4	250.3	25.9	93.1	23.5
27	0.0	141.9	38.5	49272.2	18009.5	56755.4	250.3	25.9	111.7	21.9
28	0.0	141.3	38.3	49272.2	18009.5	56755.4	250.3	25.9	167.6	33.4
29	0.0	140.8	38.1	49272.2	18009.5	56755.4	250.3	25.9	70.0	25.7
30	0.0	140.3	37.9	49272.2	18009.5	56755.4	250.3	25.9	95.6	28.0
31	0.0	139.7	37.7	49272.2	18009.5	56755.4	250.3	25.9	87.5	19.7
32	0.0	139.2	37.5	49272.2	18009.5	56755.4	250.3	25.9	32.0	0.0
33	0.0	143.0	38.7	49272.2	18009.5	64610.5	250.3	25.9	76.8	0.0
34	0.0	142.5	38.5	49272.2	18009.5	64610.5	250.3	25.9	36.2	20.6
35	0.0	141.9	38.3	49272.2	18009.5	64610.5	250.3	25.9	115.2	22.3
36	0.0	141.4	38.1	49272.2	18009.5	64610.5	250.3	25.9	63.4	25.2

37	0.0	140.9	37.9	49272.2	18009.5	64610.5	250.3	25.9	93.0	27.0
38	0.0	140.3	37.7	49272.2	18009.5	64610.5	250.3	25.9	73.7	22.1
39	0.0	139.8	37.5	49272.2	18009.5	64610.5	250.3	25.9	99.1	18.5
40	0.0	139.3	37.4	49272.2	18009.5	64610.5	250.3	25.9	72.0	0.0
41	0.0	143.1	38.5	49272.2	18009.5	72465.5	250.3	25.9	104.7	0.0
42	0.0	142.5	38.3	49272.2	18009.5	72465.5	250.3	25.9	112.2	12.7
43	0.0	142.0	38.1	49272.2	18009.5	72465.5	250.3	25.9	141.4	23.0
44	0.0	141.5	37.9	49272.2	18009.5	72465.5	250.3	25.9	84.2	20.2
45	0.0	140.9	37.7	49272.2	18009.5	72465.5	250.3	25.9	25.8	24.3
46	0.0	140.4	37.5	49272.2	18009.5	72465.5	250.3	25.9	64.8	17.3
47	0.0	139.9	37.4	49272.2	18009.5	72465.5	250.3	25.9	130.6	20.7
48	0.0	139.4	37.2	49272.2	18009.5	72465.5	250.3	25.9	57.8	0.0
49	0.0	143.1	38.3	49272.2	18009.5	80320.6	250.3	25.9	118.8	0.0
50	0.0	142.6	38.1	49272.2	18009.5	80320.6	250.3	25.9	93.7	14.4
51	0.0	142.1	37.9	49272.2	18009.5	80320.6	250.3	25.9	133.1	13.4
52	0.0	141.5	37.7	49272.2	18009.5	80320.6	250.3	25.9	31.2	15.9
53	0.0	141.0	37.5	49272.2	18009.5	80320.6	250.3	25.9	84.1	15.7
54	0.0	140.5	37.3	49272.2	18009.5	80320.6	250.3	25.9	53.8	15.2
55	0.0	139.9	37.2	49272.2	18009.5	80320.6	250.3	25.9	177.2	10.2
56	0.0	139.4	37.0	49272.2	18009.5	80320.6	250.3	25.9	114.8	0.0

57	0.0	143.2	38.1	49272.2	18009.5	88175.7	250.3	25.9	131.4	0.0
58	0.0	142.7	37.9	49272.2	18009.5	88175.7	250.3	25.9	58.3	0.0
59	0.0	142.1	37.7	49272.2	18009.5	88175.7	250.3	25.9	75.9	0.0
60	0.0	141.6	37.5	49272.2	18009.5	88175.7	250.3	25.9	81.8	0.0
61	0.0	141.1	37.3	49272.2	18009.5	88175.7	250.3	25.9	123.0	0.0
62	0.0	140.5	37.2	49272.2	18009.5	88175.7	250.3	25.9	81.7	0.0
63	0.0	140.0	37.0	49272.2	18009.5	88175.7	250.3	25.9	111.0	0.0
64	0.0	139.5	36.8	49272.2	18009.5	88175.7	250.3	25.9	28.4	0.0