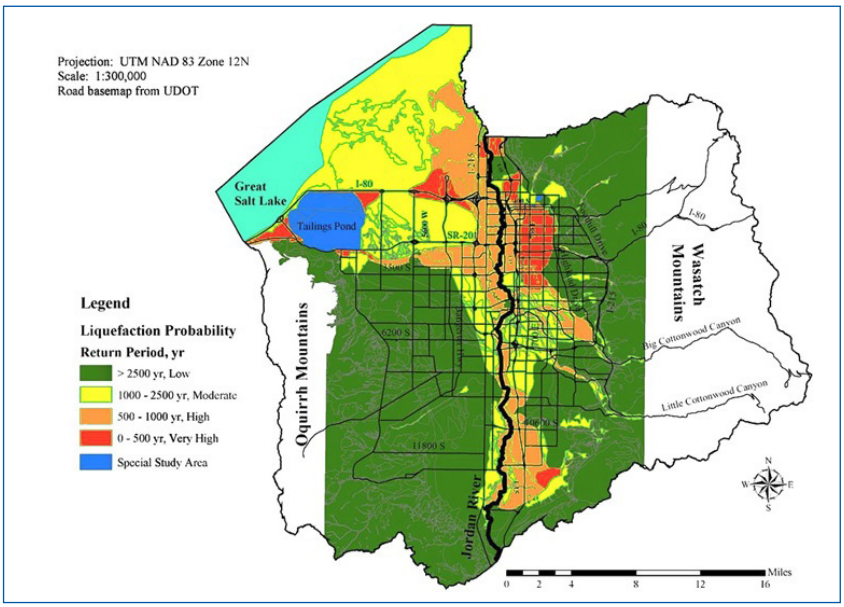
MOUNTAIN-PLAINS CONSORTIUM

MPC 22-477 | S.F. Bartlett and M. Hosseinali

DEVELOPMENT OF NEXT GENERATION LIQUEFACTION (NGL) DATABASE FOR LIQUEFACTION-INDUCED LATERAL SPREAD



MOUNTAIN-PLAINS CONSORTIUM REGION 8

A University Transportation Center sponsored by the U.S. Department of Transportation serving the Mountain-Plains Region. Consortium members:

Colorado State University
North Dakota State University
South Dakota State University
University of Colorado Denver
University of Denver
University of Utah
Utah State University
University of Wyoming

Development of Next Generation Liquefaction (NGL) Database for Liquefaction-induced Lateral Spread

Steve F. Bartlett

Massoud Hosseinali

Dept. of Civil and Environmental Engineering
University of Utah
Salt Lake City, Utah

August 2022

Acknowledgements

The Transportation Pooled Fund Program TPF-5(350) funded this research titled: “Development of Next Generation Liquefaction (NGL) Database for Liquefaction-induced Lateral Spread.” The Mountain Plains Consortium also sponsored this research under project MPC-524 titled “Development of Next Generation Liquefaction (NGL) Database for Liquefaction-induced Lateral Spread.” Additional funding was also given by Dominion Energy of Salt Lake City.

The researchers collected data in collaboration,using protocols given by the Pacific Earthquake Engineering Research (PEER) Center and the NGL program. These data-gathering activities were shared between the University of Utah (Steven. Bartlett) and Brigham Young University (Kevin Franke). Also, Kristin Ulmer of the Southwest Research Institute (SWRI) participated in gathering data for the Kobe, Japan Earthquake. In addition, the Utah Department of Transportation was the lead agency for the pooled fund program. The authors are especially grateful to David Stevens of UDOT for managing this research and the UDOT geotechnical group for their technical review. In addition, several reviewers from the pooled fund study participants contributed to guiding and improving this work. Finally, we are thankful for their efforts.

We also acknowledge the University of Utah dissertation review committee members, Professors Evert C. Lawton, Kevin W. Franke, Jeffrey R. Moore, and Chris P. Pantelides, for their valuable guidance and suggestions.

Disclaimer

ABSTRACT

This report presents several advancements in the empirical modeling of liquefaction-induced lateral spread. It starts with a newly collected dataset of 5,560 historical lateral spread displacement vectors, a sample size over 10 times larger than the existing databases and subsurface data comprising over 633 standard penetration test boreholes. This work presents a comprehensive comparison of state-of-the-art empirical models for lateral spreads through Monte Carlo simulations and sensitivity analyses and proposes new evaluation metrics to measure performance. It also quantifies the uncertainty of model weights of the Multiple Linear Regression (MLR) model using Bayesian Statistics. A new functional form is proposed for the MLR model using the least absolute shrinkage and selection operator method. Importantly, the conventional probabilistic framework for predicting lateral spread is expanded to account for the probability of lateral spread triggering given the triggering of liquefaction. This expansion allows us to model zero-displacement lateral spreads despite having liquefaction susceptibility. A convolutional neural network classifier is developed to model the probability of lateral spread triggering with an out-of-fold model accuracy of 90.5%. A new mathematical representation of soil types is presented and trained in the context of liquefaction and lateral spread and boosted model performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

This research was conducted in conjunction with the Pacific Earthquake Engineering Research (PEER) Center and various state DOTs via a pool-fund study managed by the Utah Department of Transportation (UDOT). The paper addresses the need to improve empirical, semi-empirical, analytical and numerical methods to estimate the amount of permanent ground displacement associated with liquefaction-induced lateral spread resulting from major earthquakes. The project is being executed in two phases: (1) database development and collection and (2) predictive model development.

This research is relevant to the MPC strategic goals in two areas: (1) safety and (2) sustainability. Potential earthquake damage to transportation systems will affect the operation and safety of such systems. This research will lead to methods to make such systems more resilient to such damage. Also, it will contribute to developing methods to proactively identify, quantify, visualize, prioritize, and mitigate risk resulting from earthquake hazards.

Liquefaction-induced lateral spread is a type of permanent ground deformation from the horizontal movement of surficial soil resulting from liquefaction that has occurred at depth. It generally is the most pervasive and damaging type of liquefaction-induced ground failure occurring during major earthquakes. Lateral spread displacement has caused significant damage to transportation infrastructure and other facilities during major earthquakes. Examples of such damage can be found in the engineering literature from the following earthquakes: 1964 Alaska; 1964 Niigata, Japan; 1983 Nihonkai-Chu, Japan; 1989 Loma Prieta, California; 1999 Kocaeli, Turkey; 1999 Chi-Chi, Taiwan; 2004 Northridge, California; 2005 Kobe, Japan; 2010 Chile; 2011 Tohoku, Japan; 2011 Christchurch, New Zealand. During these and other earthquakes, lateral spread horizontal ground displacement ranging from a few tenths of a meter to several meters was common in liquefaction-prone areas. These displacements resulted in hundreds of millions of dollars in damage to transportation facilities, such as bridges, embankments, culverts, and pavements.

The Next Generation Liquefaction (NGL) Project is advancing state-of-the-art liquefaction research and working toward providing end-users with a consensus approach to assess liquefaction potential within a probabilistic and risk-informed framework. Specifically, NGL aims to collect and organize liquefaction information in a standard and comprehensive database to provide all researchers with a substantially more extensive, consistent, and reliable source of liquefaction data than previously existed. Based on this database, we will create probabilistic models that provide hazard- and risk-consistent bases for assessing liquefaction susceptibility, the potential for liquefaction to be triggered in susceptible soils, and the likely consequences. NGL is committed to an open and objective evaluation and integration of data, models, and methods, as recommended in a National Academies report (2016). Following these principles will ensure that the resulting liquefaction susceptibility, triggering, and consequence models are reliable, robust, and vetted by the scientific community, providing a solid foundation for designing, constructing, and overseeing critical infrastructure projects.

This study contributes a newly collected dataset of 5,560 historical lateral spread displacement vectors, a sample size over 10 times larger than the existing databases, and subsurface data comprising over 633 standard penetration test boreholes (Chapter 3).

In addition, we propose two modifications to the functional form of the revised Multiple Linear Regression (MLR) model of Youd et al. (2002): 1) replacing R^* with R, replacing log(F15) with F15, and replacing log(T15) with T15 based on Least Absolute Shrinkage and Selection Operator (LASSO) regression technique, and 2) capping predicted horizontal displacements by using logit function instead of log function to transform target variable. These modifications result in predictions that are more consistent with our current understanding of the mechanics behind the lateral spread phenomenon and the trends observed in the lateral spread case histories database. The revised model was re-regressed to generate a new set of MLR equations, resulting in 12.3% and 9.1% higher coefficient of determination R^2 values in free face and sloping ground modes, respectively.

Most importantly, the authors developed a probabilistic framework for liquefaction-induced lateral spread screening criteria. We expanded the conventional probability chain for predicting lateral spread displacements exceeding a certain threshold to account for the probability of lateral spread triggering given liquefaction. The second outcome was to build a classifier using Convolutional Neural Networks (CNNs) to compute the newly added probability of lateral spread given liquefaction. Over 620 standard penetration testing (SPT) boreholes were collected from 11 earthquakes to train this model. The proposed CNN-based screening criteria used pairs of SPTs spaced less than 150 meters in the analysis. The proposed model can be used as a binary classifier to predict the probability of lateral spread triggering given liquefaction and classify the boreholes pair type as either within, on the boundary (i.e., margin), or outside the potential lateral spread zone. The out-of-fold model accuracy for triggering and multiclass classifiers are 81.4% and 90.5%, respectively. The uncertainty associated with the stratified k-fold cross-validation strategy was also studied. The reported accuracy for classification has a normal distribution with a mean of 81.4% and a standard deviation of 1.6%. Finally, as part of this study, a new mathematical representation of soil types was presented. These soil type latent vectors are trained in the context of liquefaction and lateral spread and resulted in a 2% improvement in the model's accuracy. These latent vectors could be used in conjunction with or as a substitute for soil index in developing predictive models for liquefaction or its consequences

1. INTRODUCTION

The effects of soil liquefaction on the built environment can be catastrophic and cause many infrastructure failures, loss of life, and significant financial loss. For instance, the 1964 Niigata earthquake in Japan caused over $1 billion of damage, and most of it was related to soil liquefaction. Another catastrophic example of soil liquefaction is the Aberfan, Wales, slide that killed 144 —116 of whom were children inundated in a school. Soil liquefaction was also one of the factors in the abandonment of the Nerlerk artificial island in Canada after more than $100 million had been spent on its construction (Jefferies and Been 2015). In addition, this phenomenon played a major role in the destruction in San Francisco's Marina District during the 1989 Loma Prieta earthquake and in Port of Kobe during the 1995 Great Hanshin earthquake. More recently, extensive liquefaction-induced damage to residential properties in Christchurch, New Zealand, was observed during the 2010 Canterbury earthquake and more extensively again Christchurch earthquakes that followed in early and mid-2011 and resulted in 185 deaths (Elliot et al. 2012, Poter et al. 2015).

Soil liquefaction that usually occurs at depths may cause damage only if it induces ground deformations, ground displacements, or ground failure that distresses human-made structures (refer to Figure 1.1. for examples). The most pervasive type of liquefaction-induced ground failure is lateral spread. As shown in Figure 1.2, during lateral spread, blocks of mostly intact, surficial soil layer migrate on liquefied soil either down gentle slopes or toward a free face (Youd 2018). Lateral spreads generally have translational movements, although some rotations and subsidence occur (Andrus & Youd 1987), and they can affect performance of structures (Gowda et al. 2022) even if not strong enough to completely destroy them. The magnitude of lateral spread typically ranges from a few centimeters to several meters (Bartlett & Youd 1995), and it may exhibit significant spatial variations within a specific site (Guan and Wang 2022).

Because of its critical impact and significant destructive nature, soil liquefaction and its consequences, particularly liquefaction-induced lateral spread, have been among the most critical topics of geotechnical engineering research over the past three decades. Researchers investigated lateral spread phenomenon using different approaches, including:

(1) Simplified analytical methods, such as Newmark's sliding block analysis proposed by Yegian et al. (1991) and Baziar et al. (1992), or the analysis based on minimum potential energy proposed by Towhata et al. (1992) and Tokida et al. (1993).
(2) Numerical methods, including finite element method, finite difference method (Finn 1991, Finn et al. 1994, Soroush & Koohi 2004, Little & Rathje 2021, and Yang & Kavazanjian 2021), and cubic interpolated pseudoparticle (Hadush et al. 2001); Vargas et al. (2022) presented a validation of numerical predictions on clean sands.
(3) Experimental testing, including laboratory (Seed et al. 1985, Kuwano & Ishihara 1988, Hadush et al. 2000, Dobry et al. 2010, and El-shamy et al. 2010) and centrifuge studies (Taboada-Urtuzuastegui & Dobry 1998, Sharp et al. 2003, and Kutter et al. 2004, Liu et al. 2018).
(4) Empirical methods, including statistical methods and machine learning algorithms, which will be subsequently discussed in detail.

Figure 1.1 Examples of soil liquefaction consequences on human-made structures (Idriss & Boulanger 2008)

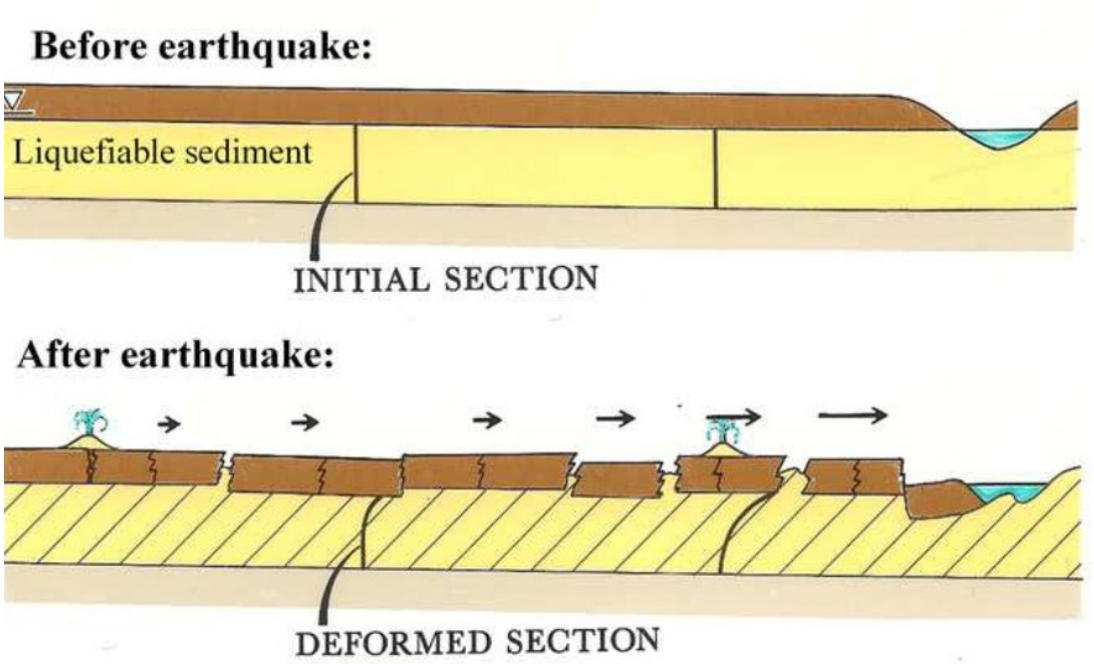


Figure 1.2 Diagram of the general character of lateral spread (Youd 2018)

Each of these approaches has advantages and disadvantages, and there has been continuous research in most of them. For example, regarding simplified analytical methods, Olsen and Johnson (2008) proposed a method for estimation of lateral spread displacements based on sliding block method analysis that ties the sliding resistance to the residual strength of the liquefied soil. However, Makdisi and Kramer (2019) investigated the capability of sliding block analyses for predicting lateral spread displacement and concluded that the sources of uncertainty and bias in the sliding block framework strongly impact their performance. These sources produce extremely low precision in both the predicted displacement and back-calculated residual shear strength.

Although analytical and numerical methods and experimental testing might help us better understand the mechanics behind lateral spread, they are less commonly used by engineers than empirical methods. Empirical methods are often used because of their simplicity and easy-to-interpret results. Hence, the empirical models continue to have a significant impact on engineering practice. Further, empirical procedures are based on compiled data; thus, they are constrained by case history observations. As new data become available or corrections are needed to the empirical models, the latest data is added to the model to improve their predictability. This dissertation focuses on the collection and addition of new lateral spread cases. These data will be subsequently used to evaluate and enhance the performance of empirical approaches.

1.1 Literature Survey

Because this study focuses on the development of empirical models, this literature survey covers only empirical models used for predicting horizontal ground displacement due to liquefaction-induced lateral spread. This evaluation includes statistical methods and machine learning algorithms.

Before 1992, there existed only a few empirical models for lateral spread displacements, notably the ones proposed by Hamada et al. (1986) and Youd and Perkins (1978, 1987). The model proposed by Hamada et al. (1986) was based on the ground displacement vectors generated from pre-and post-earthquake aerial photographs of the 1964 Niigata and 1983 Nihonkai-Chubu earthquakes. Their proposed regression model had only two predictive variables: (1) the thickness of the liquefied layer and (2) the maximum of ground slope and slope of the bottom of the liquefied layer. Their model emphasized the liquefied layer but did not address the importance of earthquake factors. They proposed the following regression model:

$$D = 0.75H^{0.5}\theta^{0.33} \tag{1.1}$$

where D is horizontal ground displacement in meters, H is the thickness of the liquefied layer in meters, and θ is the larger value of the ground slope and the slope of the bottom of the liquefied layer in percent.

One of the fundamental differences between the Hamada et al. (1986) model and its succeeding models is in that Hamada et al. (1986) divided subsurface cross-sections into segments that appeared to have moved as a discrete unit and then averaged their variables on those blocks while later models treated each displacement vector on the ground as one sample. Their model is primarily based on the thickness of the liquefiable layer and slope, and it does not include important earthquake factors, such as earthquake magnitude and earthquake source distance. Moreover, since their model is solely based on case histories from Japan with earthquake moment magnitudes around 7.5 and highly liquefiable sediments that are about 20 to 30 km away from the energy source, it results in reasonable predictions for only such conditions and should not generalize to other cases.

Youd and Perkins (1987) proposed the Liquefaction Severity Index (LSI) as a scale to represent the maximum value of ground displacement. They postulated the LSI to be primarily a function of amplitude and duration of strong ground motion. Their study was limited to a specific geologic setting (saturated cohesionless Holocene fluvial or deltaic deposits with standard penetration resistances ranging from 2 to 10 blows per foot) and lateral spreads that occurred on gentle slopes or into rive channels having widths greater than 10 meters. They proposed the following regression model:

$$\log LSI = -3.49 - 1.86\log R + 0.98 M_w \quad (1.2)$$

where LSI is the maximum, permanent, horizontal displacement in inches, R is the horizontal distance from the energy source in kilometers, and M_w is the moment magnitude of the earthquake.

Because the Youd and Perkins (1987) model was primarily based on case histories in California and Alaska, it applies only to seismic regions with high ground motion attenuation (Bartlett and Youd 1992). The LSI regression model of Youd and Perkins (1987) is based on earthquake factors only and is intended to provide a conservative upper bound prediction of lateral ground displacements at sites with moderate to high liquefaction susceptibility.

Bartlett and Youd (1992) strived to develop a more comprehensive model based on the studies by Hamada et al. (1986), Youd and Perkins (1978, 1987), and Bartlett and Youd (1990) that includes (1) earthquake factors, such as earthquake magnitude and distance from the zone of seismic energy release, (2) topographical factors, such as ground slope or distance to and height of a free face, if present, (3) geological factors, such as the thickness of the low SPT N value layer, and (4) soil factors, such as mean grain size and fines content of the liquefied layer. They have developed two multiple linear regression (MLR) models for sloping ground and free face modes based on eight earthquakes and their corresponding case histories from the United States and Japan. Bartlett and Youd (1992, 1995) compiled a more comprehensive database for lateral spread case histories consisting of 448 horizontal displacement vectors. To select the features to be used in their MLR models, they have investigated different variables, a complete list is given in Table 1.1.

Among the earthquake (or seismic) variables, they chose moment magnitude (M) and horizontal distance from the seismic source (R) over Peak Ground Acceleration (PGA). These former variables were used instead of PGA and the duration of earthquake shaking (D) because strong ground motion records were not available for almost all data sources. Hence, the estimation of PGA and D from existing relations resulted in less reliable predictions when inserted into the MLR model. Nonetheless, Bartlett and Youd (1992) believed that PGA and D are more fundamental measures of the seismic energy than M and R. If high-quality measurements of those variables are available, Bartlett and Youd (1992) decided to develop two separate MLR models for free face and sloping ground modes because they observed different displacement patterns for these conditions. An example map created by Hamada et al. (1986) is shown in Figure 1.3, which illustrates the displacement vectors along the Shinano River near the Echigo Railway Bridge in Niigata, Japan. As can be seen, the displacement vectors near the river follow a distinctly different pattern from those occurring north of the railroad embankment. The areas near the river lacked lateral resistance to deformation because of the incised channel and were directed toward the free face. At the same time, other displacement vectors north of the railroad embankment were generally smaller and headed down the 0.2 percent ground slope to the northeast.

Table 1.1 Summary of independent variables evaluated by Bartlett and Youd (1992)

Variables	Description
Earthquake variables	
M	Earthquake moment magnitude, M_w
R	Nearest horizontal distance to seismic energy source or fault rapture, km
A	Peak ground acceleration, g
D	Duration of strong ground motion (>0.05 g), s
Topographical variables	
S	Ground slope, %
L	Distance to the free face from the point of displacement, m
H	Height of free face, m
W	Free face ratio, % (i.e. 100 H / L)
Geological variables	
T_S	Thickness of liquefied zone(s) based on simplified procedure, m
T_L	Thickness of liquefied zone(s) based on Liao's 50% probability curve, m
T_{10}	Thickness of saturated cohesionless soils with $(N_1)_{60} \leq 10$, m
T_{15}	Thickness of saturated cohesionless soils with $(N_1)_{60} \leq 15$, m
T_{20}	Thickness of saturated cohesionless soils with $(N_1)_{60} \leq 20$, m
I_S	Index of liquefaction potential based on simplified procedure
I_L	Index of liquefaction potential based on Liao's 50% probability curve
Z_{tls}	Depth to top of liquefied zone based on simplified procedure, m
Z_{tll}	Depth to top of liquefied zone based on Liao's 50% probability curve, m
Z_{bls}	Depth to bottom of liquefied zone based on simplified procedure, m
Z_{bll}	Depth to bottom of liquefied zone based on Liao's 50% probability curve, m
Z_S	Depth to the lowest factor of safety based on simplified procedure, m
Z_L	Depth to the lowest factor of safety based on Liao's 50% probability curve, m
Z_N	Depth to lowest SPT N value in saturated cohesionless soil, m
Z_{N160}	Depth to lowest SPT $(N_1)_{60}$ value in saturated cohesionless soil, m
N	Lowest SPT value in saturated cohesionless sediments
$N1_{60}$	Lowest SPT $(N_1)_{60}$ value in saturated cohesionless sediments
J_S	Lowest factor of safety based on simplified procedure below water table
J_L	Lowest factor of safety based Liao's 50% probability curve below water table
$N1_{60S}$	$(N_1)_{60}$ value corresponding to J_s
$N1_{60L}$	$(N_1)_{60}$ value corresponding to J_L
K_S	Average factor of safety in T_S
K_L	Average factor of safety in T_L
O_S	Average $(N_1)_{60}$ in T_S
O_L	Average $(N_1)_{60}$ in T_L

Note: Simplified procedure from Seed and Idriss (1971); Liao's probability curves from Liao (1986) considered. Among the other variables, Bartlett and Youd (1992) chose those yielding higher values of the coefficient of determination (R2) in a stepwise regression process.

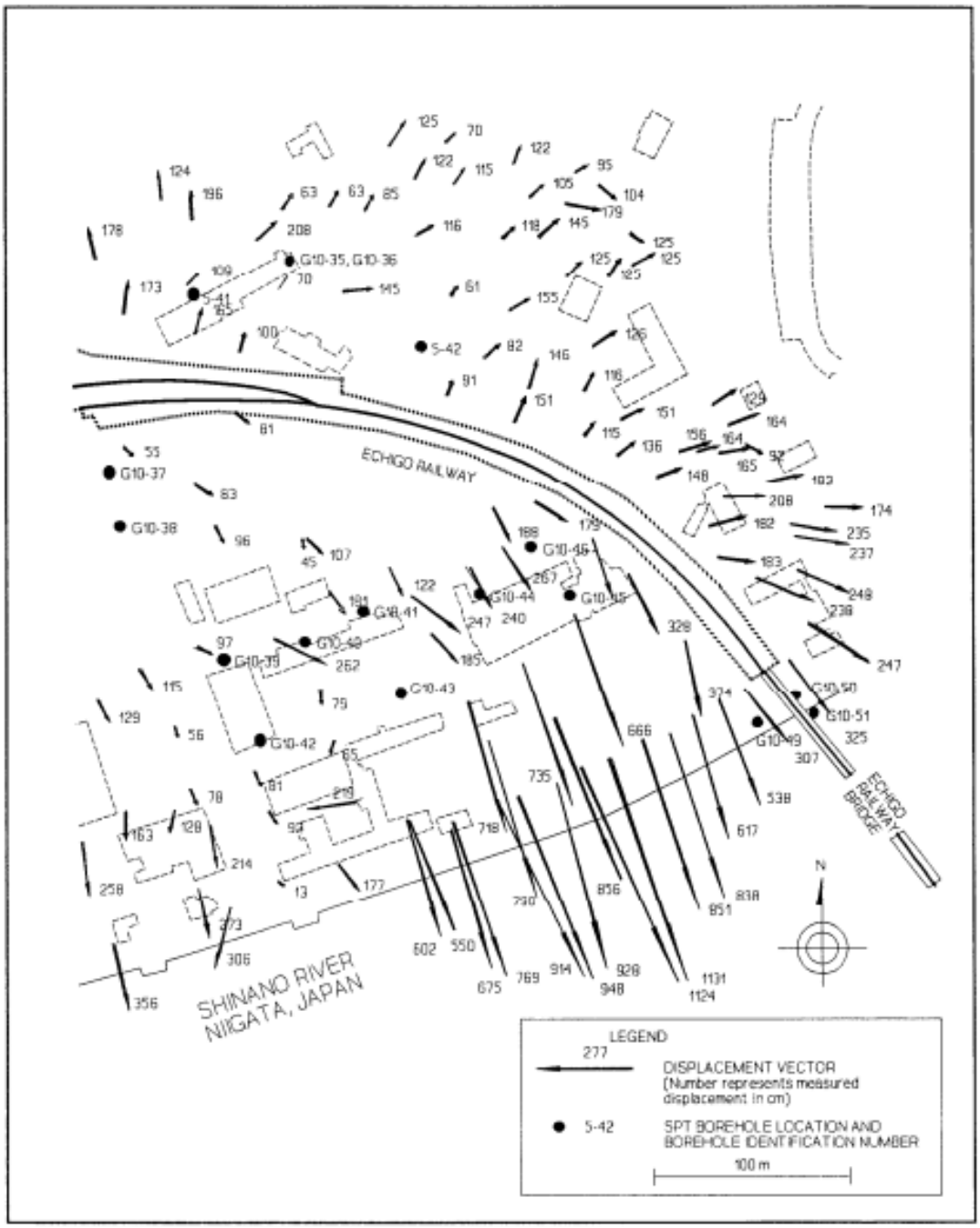


Figure 1.3 Displacement vectors near Shinano River in Niigata, Japan (source: unpublished ground failure maps courtesy of M. Hamada)

The final form of the MLR models developed by Bartlett and Youd (1992) for free face and sloping ground modes are given in Equations 1.3 and 1.4, respectively:

$$\log (D_H+0.01) = - 16.366 + 1.178\ M - 0.927\ \text{Log}\ R - 0.013\ R + 0.657\ \text{Log}\ W + 0.348\ \text{Log}\ T_{15} + 4.527\ \text{Log}\ (100-F_{15}) - 0.922\ D50_{15} \quad (1.3)$$

$$\log (D_H+0.01) = - 15.787 + 1.178\ M - 0.927\ \text{Log}\ R - 0.013\ R + 0.429\ \text{Log}\ S + 0.348\ \text{Log}\ T_{15} + 4.527\ \text{Log}\ (100-F_{15}) - 0.922\ D50_{15} \quad (1.4)$$

where D_H is the horizontal ground displacement, and all other variables are defined in Table A.1. Bartlett and Youd (1992) reported an overall coefficient of determination and adjusted coefficient of determination of 82.6% and 82.3%, respectively. The coefficient of determination is the proportion of the variance in the dependent variable that is predictable from the independent variables (Neter 1996). The adjusted coefficient of determination adjusts R^2 for the number of independent variables in a model relative to the number of data points because R^2 automatically and spuriously increases when extra explanatory variables are added (Theil 1961). The adjusted R^2 is always less than $R^{2,}$ and it increases only in cases when the increase in R^2 is more than one would expect to see by chance.

For non-western United States and Japanese sites and other soft soil conditions, Bartlett and Youd (1992) require corrections to their model. Moreover, because the bulk of their data came from $6 \leq M \leq 8$ earthquakes, they recommended not using their model for moment magnitudes out of this range. Such an attempt is an extrapolation, and unreliable results might be produced. They developed a flow chart to apply their models (Equations 1.3 and 1.4), shown in Figure 1.4. Finally, they concluded that the addition of more case histories would strengthen the MLR database and improve its reliability for extremely large earthquakes.

Bartlett and Youd (1995) added a few more case histories to the earlier version of their database and increased the number of displacement vectors to 467. Nonetheless, their independent variables remained the same as the ones in Equations 1.3 and 1.4, The final MLR models is defined as:

$$\log (D_H) = - 16.366 + 1.178\ M - 0.927\ \text{Log}\ R - 0.013\ R + 0.657\ \text{Log}\ W + 0.348\ \text{Log}\ T_{15} + 4.527\ \text{Log}\ (100-F_{15}) - 0.922\ D50_{15} \quad (1.5)$$

$$\log (D_H) = - 15.787 + 1.178\ M - 0.927\ \text{Log}\ R - 0.013\ R + 0.429\ \text{Log}\ S + 0.348\ \text{Log}\ T_{15} + 4.527\ \text{Log}\ (100-F_{15}) - 0.922\ D50_{15} \quad (1.6)$$

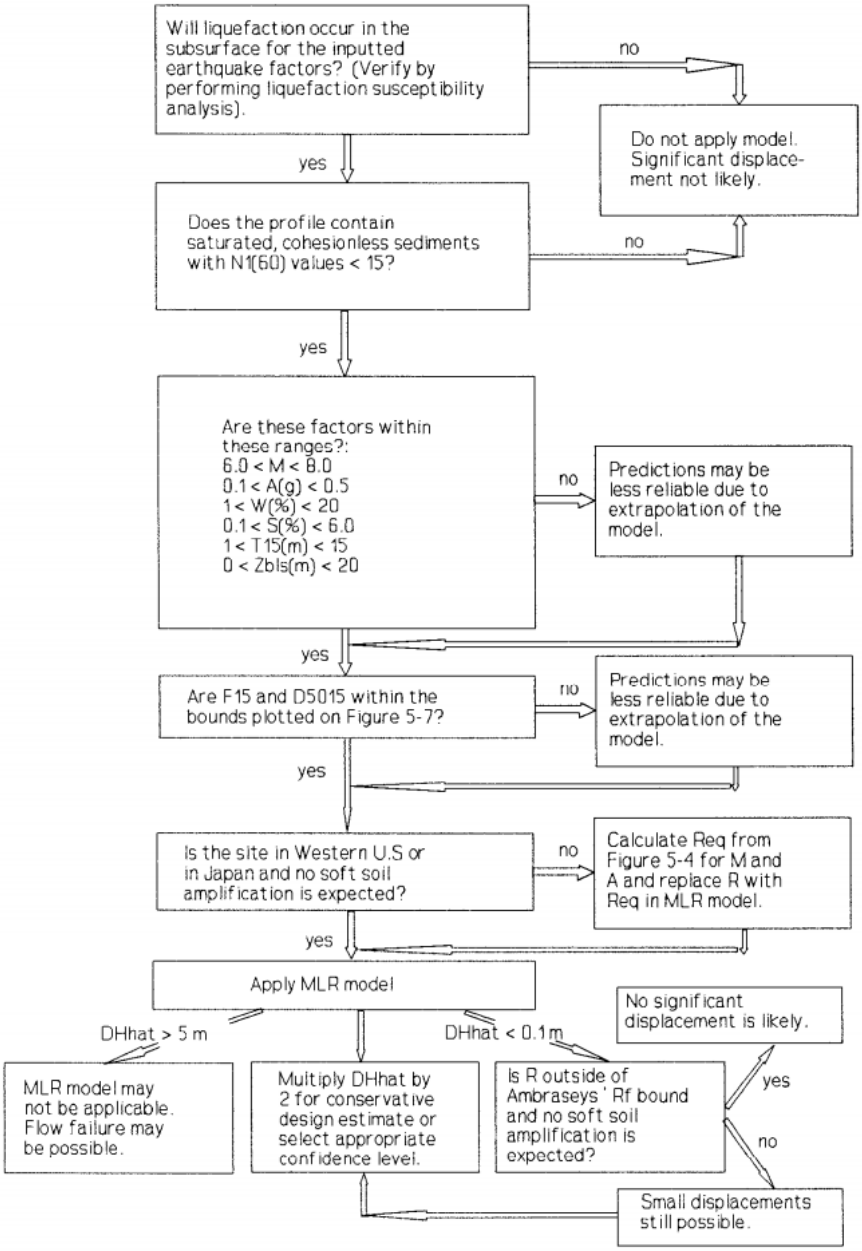


Figure 1.4 Flow chart for application of Bartlett and Youd (1992) model

After the compilation of the Bartlett and Youd (1992 and 1995) database, many researchers developed empirical models by applying different statistical methods and machine-learning algorithms to the database. These methods and algorithms include multiple linear regressions, artificial neural networks, tree-based algorithms, such as random forests, fuzzy algorithm, genetic programming, multivariate adaptive regression splines, etc. Machine-learning algorithms differ from statistical methods. These methods do not require a predefined functional form, and the algorithm will automatically detect the nonlinear pattern between predictive variables and target variable.

Wang & Rahman (1999) developed the first backpropagation artificial neural network for liquefaction-induced lateral spreads on top of the Bartlett and Youd (1992, 1995) database shortly after this new field of "soft computing" emerged for solving decision-making, modeling, and control problems. A detailed description of artificial neural networks could be found in Rashid (2016) and Geron (2017). Wang and Rahman (1999) started with using all the 28 variables in the database and then tried to eliminate a number of variables based on the methodology described by Garson (1991) that yields relative importance of variables using their weights in the model. They found that the most important variables among all the 28 variables were M, R, S, W, T_{15} and F_{15}. Nonetheless, the highest coefficient of determination was obtained when all the variables used by Bartlett and Youd (1992 and 1995) were used.

Despite that their reported coefficient of determination (0.816 and 0.654 for the training and testing stages of the combined model, respectively), was higher than the MLR model of Bartlett and Youd (1992, 1995), they did not provide the neural network they used — neither the model itself nor the details, such as neuron weights, activation functions, learning rate, hyperparameters, etc. Hence, it is not possible to reproduce their solution. They also emphasized the need for the collection of more data.

Rauch and Martin (2000) developed the Empirical Prediction of Liquefaction-induced Lateral Spreading (EPOLLS) model, a multiple linear regression based on 71 lateral spread case histories. They followed the approach of Hamada et al. (1986) in that they treated blocks of displacement vectors as one unit or case history. They defined an EPOLLS case study as a contiguous mass of soil that moved in one general direction and may enclose a multitude of measured displacements. An example of two of EPOLLS case studies they identified in Niigata is shown Figure 1.5. They also emphasized that their model and the preceding models were not intended to predict where a lateral spread might develop but only the potential deformations where a lateral spread is expected.

Rauch and Martin (2000) proposed three levels for their multiple linear regression model that should be used, depending on the availability of the data: (1) Regional-EPOLLS that require only data on the seismic source and local intensity of shaking, (2) Site-EPOLLS component that requires additional, site-specific data on the surface topography and dimensions of the anticipated area of sliding (but predictions are possible without detailed data on the subsurface soil conditions), and (3) Geotechnical-EPOLLS component that requires all of the inputs used in the other two components with additional data obtained from subsurface explorations. The final equations of these three models are given respectively in Equations 1.7 to 1.9:

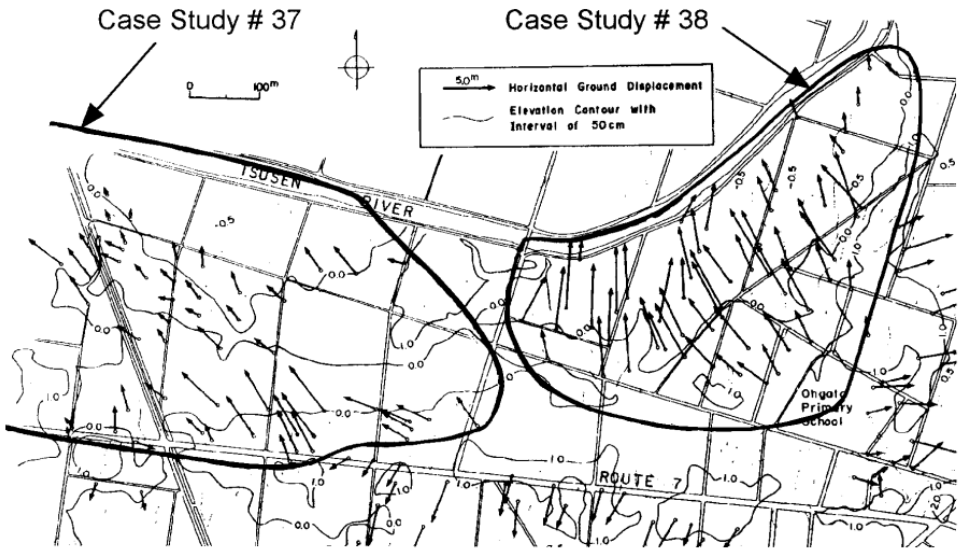


Figure 1.5 Two EPOLLS case studies in Niigata, Japan

$$\sqrt{Avg_{Horz}} = 0.613M_w - 0.0139R_f - 2.42A_{max} - 0.0114T_d - 2.21 \tag{1.7}$$

$$\sqrt{Avg_{Horz}} = 0.613M_w - 0.0139R_f - 2.42A_{max} - 0.0114T_d + 0.000523L_{slide} + 0.0423S_{top} + 0.0313H_{face} - 2.44 \tag{1.8}$$

$$\sqrt{Avg_{Horz}} = 0.613M_w - 0.0139R_f - 2.42A_{max} - 0.0114T_d + 0.000523L_{slide} + 0.0423S_{top} + 0.0313H_{face} + 0.0506Z_{FSmin} - 0.0861Z_{liq} - 2.49 \tag{1.9}$$

where Avg_Horz is the average horizontal ground displacement of the case study; M_w is the moment magnitude of the earthquake; R_f is the shortest horizontal distance from the site to surface projection of fault rupture or zone of seismic energy release; A_{max} is the peak horizontal acceleration at the ground surface of the site that would occur in absence of excess pore pressures or liquefaction generated by the earthquake; T_d is the duration of strong ground motions at the site, defined as the time between first and last occurrence of surface acceleration ≥ 0.05 g; L_{slide} is the maximum horizontal length from head to toe of lateral spread in the prevailing direction of movement; S_{top} is the average slope across the surface of lateral spread, measured as change in elevation over distance from head to toe; H_{face} is the height of the free face, measured vertically from toe to crest of the free face; Z_{FSmin} is average depth to the minimum factor of safety in potentially liquefiable soil, and Z_{liq} is average depth to the top of liquefied soil.

Youd et al. (2002) revised the empirical equations developed by Bartlett & Youd (1992, 1995) to correct and update the original analysis. Updates include: (1) corrections to errors due to erroneously overestimated displacement measurements for lateral spreads generated by the 1983 Nihonkai-Chubu, Japan Earthquake. Those displacements were miscalculated by a factor of 1.9, (2) deletion of eight displacement vectors where boundary shear impeded free lateral displacement to be consistent with a free lateral movement condition, (3) addition of three newer earthquakes – 1983 Borah Peak, Idaho, 1989 Loma Prieta, California, and 1995 Hyogo-Ken Nabu

(Kobe), Japan - to extend the predictive equation to coarser-grained materials, and (4) changes in the functional form to enhance model performance on coarser-grained sediments and consequently increase the accuracy of the model. Moreover, their changes to the functional form prevent unrealistic over predictions of displacements when R becomes small. Their revised and re-regressed models for sloping ground and free-face modes are respectively, as follows:

$$\begin{aligned}\log(Dh_{sg}) = {} & -16.213 + 1.532M - 1.406\log(R^*) - 0.012R + 0.338logS \\ & + 0.540logT_{15} + 3.413\log(100 - F_{15}) \\ & - 0.795\log(D50_{15} + 0.1)\end{aligned} \tag{1.10.a}$$

$$\begin{aligned}\log(Dh_{ff}) = {} & -16.713 + 1.532M - 1.406\log(R^*) - 0.012R + 0.592logW \\ & + 0.540logT_{15} + 3.413\log(100 - F_{15}) \\ & - 0.795\log(D50_{15} + 0.1)\end{aligned} \tag{1.10.b}$$

where M is moment magnitude; R is the horizontal or mapped distance from the site in question to the nearest bound of the seismic energy source; $R^*=10^{(0.89M-5.64)}+R$; S is the slope in percent; W is the free face ratio in percent; T_{15} is the thickness of saturated cohesionless soil with $(N_1)_{60} < 15$, in meters; F_{15} is average fines content in T_{15}, in percent; and $D50_{15}$ is average D_{50} in T_{15} in millimeters.

Considering the fundamental assumption of linear regression models that errors are distributed normally, one can rewrite Eqs. 1.10.a and 1.10.b in a probabilistic form:

$$P(Dh_{sg} > y) = 1 - \Phi\left(\frac{\log(Dh_{sg}) - y}{\sigma_{\log(Dh_{sg})}}\right) \tag{1.11.a}$$

$$P(Dh_{ff} > y) = 1 - \Phi\left(\frac{\log(Dh_{ff}) - y}{\sigma_{\log(Dh_{ff})}}\right) \tag{1.11.b}$$

where Φ() is the cumulative density function of normal distribution and σ is the standard deviation of the error term in the Youd et al. (2002) model, which is equal to 0.197 as given by Franke and Kramer (2013).

One of the limitations of the Youd et al. (2002) model is that because all the gravels included in their database had sufficient fines content, the use of their equations for coarse-grained sites is only valid for soil with impeded drainage. Since this limitation is rooted in the database, all subsequent models developed using this database would have the same limitation. Youd et al. (2002) also recommended not using their model for highly plastic fines because they are generally not liquefiable and therefore not susceptible to lateral spread. Another limitation mentioned by Youd et al. (2002) is the uncertainty of model for lateral displacement predictions larger than six meters. Similar to Bartlett and Youd (1992 & 1995), application of the revised MLR model has guidelines recommended by Youd et al. (2002) as follows:

1. The liquefaction susceptibility of the site should be verified through subsurface exploration before appling the MLR model. Nonliqufiable sites and liquefiable layers having all SPT $(N_1)_{60}$ greater than 15 are too dense and dilative for lateral spread to occur.

2. Use of the MLR model for earthquakes with magnitudes not between 6 to 8 has greater uncertainty. Nonetheless, the MLR model can be applied to earthquakes with magnitudes less than 6 because lateral spread displacements are generally small.
3. In case of free-face ratios (W) of less than 1%, the sloping ground mode controls predicted displacements and the equation of sloping ground mode should be used. When free-face ratio is greater than 5%, free-face conditions control the displacement behavior and the equation of free-face mode should be used. Where there is a question of which equation should be used, the maximum of the predictions from both modes should be used.
4. The applicable ranges of grain sizes for use with the MLR model is shown in Figure 1.6, which was later revised in Youd (2018).
5. The MLR models developed by Youd et al. (2002) are based on the stiff soil sites in the western United States and Japan with relatively high attenuation of strong ground motion with distance. For other seismic regions, such as the eastern United States or liquefiable sites underlain by soft soils that may amplify weak ground motions, an equivalent distance term (R_{eq}) is required.

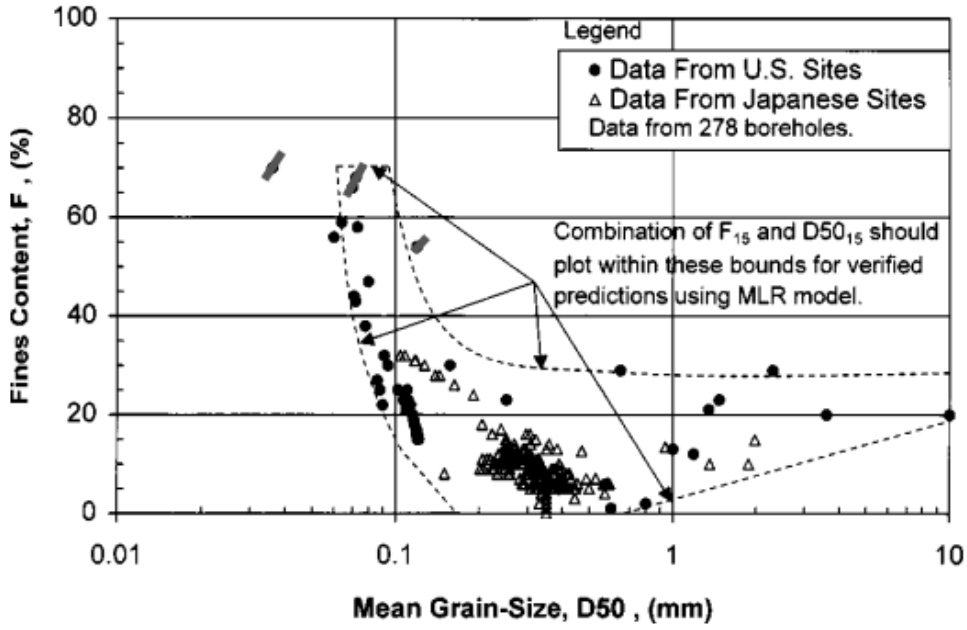


Figure 1.6 Applicable grain size ranges of F_{15} and $D50_{15}$

Bardet et al. (2002) dropped the two soil parameters in Bartlett and Youd (1992, 1995) and proposed a four-parameter MLR model to be used at a regional scale. They believed those two soil parameters are somewhat challenging to obtain from borehole data and to determine for large areas. Moreover, the uncertainty of those two variables is the largest in the Bartlett and Youd (1992) model. Also, Bardet et al. (2002) discarded larger displacements because it may be difficult to have engineering countermeasures against them. They divided data into two datasets (A) complete dataset consisted of 467 data points, and (B) data limited to displacement amplitudes smaller than 2 meters consisted of 283 data points. They then regressed two sets of MLR models on datasets A and B.

Bardet et al. (2002) also presented a probabilistic framework for assessing confidence intervals for predicting ground deformation and the probability of exceeding certain ground deformation levels, such as the ones given in Eqs. 1.11.a and 1.11.b. Their models are useful for risk

assessment to spatially distributed lifeline networks. They further concluded that there is a need to improve and extend the Bartlett and Youd (1992, 1995) database on which their model was based.

In 2004, Zhang et al. (2004) proposed a semiempirical approach to estimate liquefaction-induced lateral displacements using the standard penetration test (SPT) or cone penetration test (CPT) data to evaluate the potential maximum cyclic shear strains. They went one step further than pure empirical models and combined available SPT and CPT-based methods at the time with the laboratory test results for clean sands to estimate the potential maximum cyclic shear strains (γ_{max}) from which the lateral displacement index (LDI) could be obtained:

$$LDI = \int_0^{Z_{max}} \gamma_{max} dz \tag{1.12}$$

where Z_{max} is the maximum depth below all the potential liquefiable layers and should be less than 20 meters.

The LDI, which only accounts for the soil profile, soil properties, and earthquake characteristics, will then be used with the geometric parameters from case histories characterizing ground geometric to predict the magnitude of lateral spreads, LD. The final predictive equations for the sloping ground and free face modes are given respectively in Eqs. 1.13.a and 1.13.b:

$$LD = (S + 0.2)\ LDI \tag{1.13.a}$$

$$LD = 6\ (L/H)^{-0.8}\ LDI \tag{1.13.b}$$

where S is the ground slope, L is the horizontal distance from the toe of a free face to the site, and H is the elevation difference between the level ground surface and the toe of a free face.

Zhang et al. (2004) excluded several cases that they believed lateral spreads were impeded by shear forces along the margins of the failure zone or human-made structures. They also excluded case histories with the possibility of local slump failure or flow failure. Overall, they had 291 displacement vectors from 13 earthquakes. Their approach produced similar R^2 values when compared with the Youd et al. (2002) model.

Five years after Wang and Rahman (1999) developed the first artificial neural network, Baziar and Ghorbani (2005) presented another neural network for the evaluation of lateral spread. They based their model on the revised database of Youd et al. (2002), while the neural network model by Wang and Rahman (1999) was based on the Bartlett and Youd (1992, 1995) database. Another primary difference was the number of neurons used in hidden layers. Baziar and Ghorbani (2005) used 11 hidden neurons, while Wang and Rahman (1999) used five. Finally, the most important difference was that Baziar and Ghorbani (2005) presented the details of their neural network, such as the weights and activation functions, so readers could reproduce their results.

With the enormous expansion of genetic programming in the late 1990s and early 2000s, Javadi et al. (2006) applied this powerful technique to model liquefaction-induced lateral displacements. Their final predictive equations for the sloping ground and free-face modes are given respectively in Eqs. 1.14.a and 1.14.b:

$$D_h = -0.8\frac{F_{15}}{M} + 0.0014F_{15}^2 + 0.16T_{15} + 0.112S + 0.04\frac{ST_{15}}{D50_{15}} - 0.026RD50_{15} + 1.14 \quad (1.14.a)$$

$$D_h = -163.1\frac{1}{M^2} + 57\frac{1}{RF_{15}} - 0.0035\frac{T_{15}^2}{WD50_{15}^2} + 0.02\frac{T_{15}^2}{F_{15}D50_{15}^2} - 0.26\frac{T_{15}^2}{F_{15}^2} + 0.006T_{15}^2 - 0.0013W^2 + 0.0002M^2WT_{15} + 3.7 \quad (1.14.b)$$

Javadi et al. (2006) further proposed another set of equations for moderate lateral displacements where measured displacements are less than 1.5 meters. They proposed to use Eqs. 1.14.a and 1.14.b to predict lateral spreads and if the predicted displacements were in the moderate range (less than 1.5m) and to use Eqs. 1.15.a and 1.15.b for the sloping ground and free face modes, respectively:

$$D_h = -0.027\frac{T_{15}^2F_{15}}{M^2} + 0.05\frac{RT_{15}}{M^2D50_{15}} + 0.44\frac{1}{MR^2ST_{15}} - 0.03R - 0.02\frac{M}{ST_{15}} - 5\times10^{-5}\frac{MR}{D50_{15}^2} + 0.075M^2 - 2.4 \quad (1.15.a)$$

$$D_h = -234.1\frac{1}{M^2RW} - 156\frac{1}{M^2} - 0.008\frac{F_{15}}{R^2T_{15}} + 0.01\frac{WT_{15}}{R} - 2.9\frac{1}{F_{15}} - 0.036\frac{MT_{15}^2D50_{15}^2}{R^2W} + 9.4\frac{M}{RF_{15}} - 4\times10^{-6}\frac{MR^2}{D50_{15}} + 3.84 \quad (1.15.b)$$

As per their conclusions, genetic programming successfully modeled lateral spreads and improved the accuracy of predictions. Unfortunately, Javadi et al. (2006) used the equation for the coefficient of determination given in Eq. 1.16 based on the false assumption that the predictions had a mean of zero. Therefore, the coefficient of determination reported therein is invalid and could not be used to compare their model with other models:

$$R^2 = \frac{\sum_N(X_m)^2 - \sum_N\left(X_m - X_p\right)^2}{\sum_N(X_m)^2} \quad (1.16)$$

where X_m is the measured and X_p is the predicted displacement. The general formula for the coefficient of determination that does not assume zero mean is (Neter et al. 1996):

$$R^2 = 1 - \frac{\sum_N\left(X_m - X_p\right)^2}{\sum_N(X_m - \bar{X})^2} \quad (1.17)$$

where $\bar{X}$ is the mean of measured displacements.

Figure 1.7 clearly shows the effect of their false assumption. The reported coefficient of determination is 95.5%, which is very high, but almost all of the predictions are similar to each other (within the range of 1.1 to 1.3 meters), while the measured displacements are in the range of 0.5 to 1.5 meters. For the same data shown in Figure 1.7, the correct value of the coefficient of determination, calculated by the author using Eq. 1.17, is 64.9%.

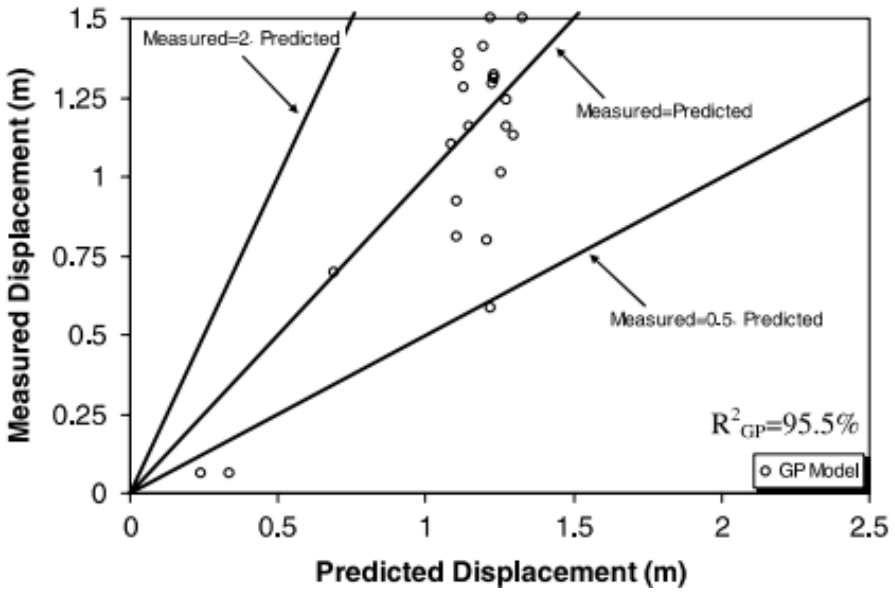


Figure 1.7 Measured versus predicted displacements of Javadi et al. (2006) – moderate displacements model for gently sloping mode

Neurofuzzy systems, which are another widespread statistical method in the 2000s, were not left unnoticed by the researchers working on empirical models of lateral spread. Garcia et al. (2008) proposed a model called NEFLAS (NEuroFuzzy) estimation of liquefaction-induced LAteral Spread) to capture the nonlinear relations between predictive variables and the target variable. Unlike the models using a predefined functional form, such as the MLR models, their model generates a functional form automatically within the model by recognizing the patterns in the data space. There are advantages and disadvantages in either using or not using a predefined functional form. In short, if the process is mechanics-based, then an appropriate functional form can be suggested based on the physics of the phenomenon. However, in phenomena affected by multi variables and complex interactions, a proper functional form may not be apparent or easily defined. Also, to eliminate or smooth the scattering caused by vague and contradictory data (i.e., noisy datasets), Garcia et al. (2008) preprocessed the data using a fuzzy clustering technique that transforms groups of points into single typicals.

Oommen and Baise (2010) proposed a Support Vector machine for Regression (SVR) model for only the free-face mode built upon the 219 free-face data points of the database compiled by Youd et al. (2002). Moreover, they presented an excellent discussion on validation techniques for machine-learning algorithms. Their method does not use a predefined functional form. They concluded that the performance measures used by Wang & Rahman (1999), Baziar & Ghorbani (2005), and Javadi et al. (2006) were biased. They also concluded that SVR performs better than the MLR model.

Oommen and Baise (2010) used the number of support vectors to indicate the uncertainty within the data space. Figure 1.8 shows the percentage of support vectors in the range of predictor variables where a low percentage of support vectors indicates that of all the instances in that region only a few are support vectors, meaning that there is less uncertainty and higher error in this region of the predictor variable is expected. Based on this chart, they concluded that the range of $D50_{15}$ values between 0.2 and 0.26 are of particular interest for future data collection to improve lateral spread displacement models.

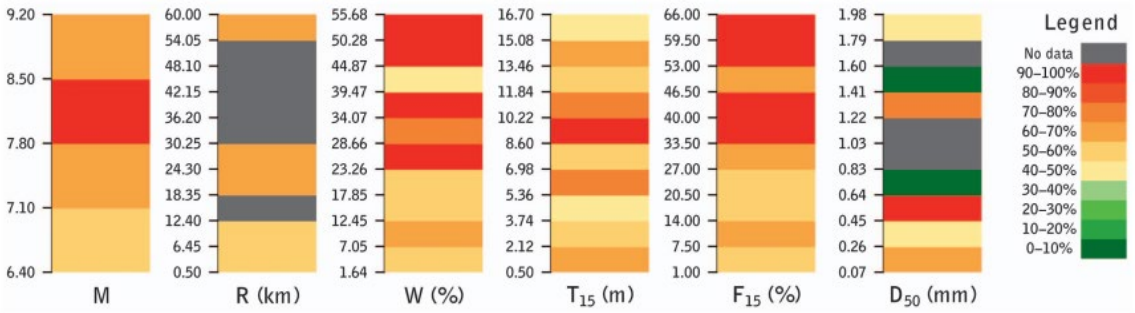


Figure 1.8 Percentage of support vectors in the range of predictor variables (Oommen & Baise 2010)

They further concluded that a single split of data might misrepresent the predictive power of the algorithms since they observed 25% variation in the coefficient of efficiency over the five splits of the free-face mode.

Application of newer statistical methods on liquefaction-induced lateral spread has continued to be the focus of researchers in this domain. Rezania et al. (2011) developed an evolutionary polynomial regression (EPR) model for the assessment of liquefaction potential and lateral spreading. EPR is a technique that combines the genetic algorithm to find feasible structures and the least square method to find the appropriate constants for those structures. Their final predictive equations for the sloping ground and free face modes are given respectively in Eqs. 1.18.a and 1.18.b:

$$D_h = -\frac{1.6941 T_{15}^{0.5} F_{15}^{0.5}}{M^2 D50_{15}^{0.5}} - \frac{0.78905 R^{0.5} S^{0.5} T_{15} F_{15}^{0.5}}{M^2} - \frac{2.2542 \times 10^{-12} M^{0.5} T_{15}^2 D50_{15}^2}{R^{0.5} S^{0.5} F_{15}^2} + \frac{0.036036 M S^{0.5} T_{15}}{D50_{15}^{0.5}} + 0.85441 \quad (1.18.a)$$

$$D_h = -\frac{2.1414 R^{0.5} W^{0.5}}{M^2 D50_{15}^{0.5}} - \frac{0.061863 T_{15} F_{15}}{M^{0.5} W^{0.5}} - \frac{11.1201 M^2}{R W^{0.5} F_{15}} + \frac{0.0017573 M^2 W^{0.5} T_{15}}{F_{15}^{0.5} D50_{15}} + 1.9671 \quad (1.18.b)$$

Similar to Javadi et al. (2006), Rezania et al. (2011) also proposed a different set of equations for the lateral displacements with a magnitude less than 1.5 meters for the sloping ground and free face modes, respectively as:

$$D_h = -\frac{1.8017 R D50_{15}^{0.5}}{M^2} - \frac{0.12148 F_{15}}{M} - \frac{31.5315}{M^{0.5}} + \frac{2.9385 M^{0.5}}{S^{0.5} T_{15}} + \frac{1.9056 \times 10^{-4} M^2 S^{0.5} T_{15} F_{15}^{0.5} D50_{15}^{0.5}}{R^2} + 13.3224 \quad (1.19.a)$$

$$D_h = -\frac{12.7493T_{15}D50_{15}^{0.5}}{M^2R^{0.5}W^{0.5}} - \frac{3.4311F_{15}}{M^2R^{0.5}} - \frac{24.0261}{MR^{0.5}T_{15}^{0.5}F_{15}D50_{15}} - \frac{355.8433}{MW^{0.5}} + \frac{4.0048\times10^{-6}R^{0.5}F_{15}^2}{MW^{0.5}D50_{15}^2} + \frac{128.522}{M^{0.5}W^{0.5}} - 5.3745\times10^{-4}M^2R^{0.5}W^{0.5} + 0.95605 \quad (1.19.b)$$

Unfortunately, the formula they used to calculate the coefficient of determination is similar to Javadi et al. (2006), Eq. 1.16, which is invalid. Figure 1.9 clearly shows the effect of the invalid assumption of Eq. 1.16 on the calculated R^2. The reported value of R^2 for the EPR model is 89.8%, while the correct value of R^2, calculated by the author using Eq. 1.17 is only 50.3%.

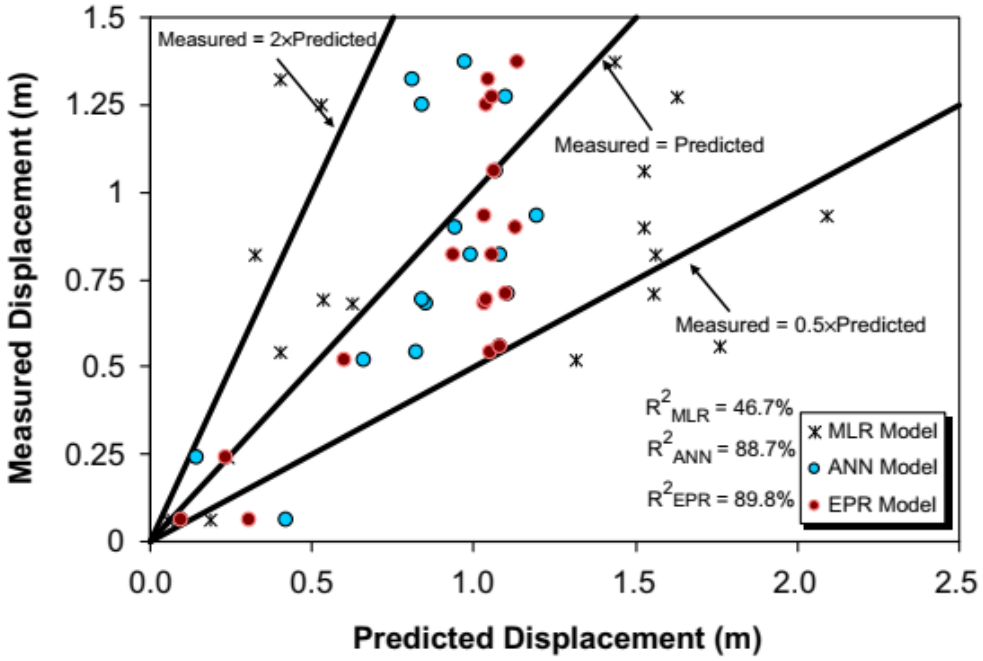


Figure 1.9 Measured versus predicted displacements of Rezania et al. (2006) - moderate displacements model for free face mode

Baziar and Azizkandi (2013) proposed an updated neural network model optimized using genetic programming. They also added 41 new case histories to the original database from Youd et al. (2002) and based their model on 525 data points. In their model, they used one hidden layer with 12 neurons. They presented details of their neural network, such as the weights and activation functions, so the reader can reproduce their results. Unlike other researchers, they did not divide the dataset into two modes (i.e., free face and sloping ground) and presented only one equation for all cases:

$$
\begin{aligned}
D_h = {} & \frac{54.36T_{15}}{D_{50}+0.6532} - \frac{55.34T_{15}}{D_{50}+0.6689} + \frac{196.9T_{15}}{W+0.9212} - \frac{199.8T_{15}}{W+0.9434} \\
& + \frac{0.0446(W-S)}{R} - \frac{1.718}{S+0.8956} - 0.02452T_{15}F_{15} \\
& - 0.00625F_{15}S + 0.001474R(W-T_{15}) \\
& - 0.06875T_{15}(W-S) \\
& + M(0.1058T_{15} + 0.009652T_{15}W - 0.1225) \\
& + 0.00024T_{15}F_{15}^2 - 0.00255RWS + 2.6
\end{aligned}
\tag{1.20}
$$

An updated MLR model was proposed by Gillins and Bartlett (2013), which was based on soil type and instead of the fines content and mean grain size. Their model is based on the same concept as the Bardet et al. (2002) model that these soil parameters are hard to obtain in mapping regional projects because laboratory soil data is lacking to estimate $D50_{15}$ and F_{15}. The key differences between their model and that of Bardet et al. (2002) is that (1) Gillins and Bartlett (2013) did not entirely discard those two soil parameters, but rather they assigned a soil index to each T_{15} layer in lieu of $D50_{15}$ and F_{15}, and (2) they gave a methodology whereby CPT data can be used in lieu of SPT N values. They introduced the proposed independent variables to the regression model and re-regressed the Youd et al., 2002 lateral spread database. Their model can be used in liquefaction-induced lateral spread hazard mappings similar to the work on Liu et al. (2016).

The final form of their equation is:

$$
\begin{aligned}
log D_h = {} & -8.208 - 0.344\alpha + 1.318M - 1.073logR^* - 0.016R \\
& + 0.445logW + 0.337logS + 0.592logT_{15} - 0.683x_1 \\
& - 1.200x_2 + 0.252x_3 - 0.40x_4 - 0.535x_5
\end{aligned}
\tag{1.21}
$$

where α is 1 for free face mode and 0 for sloping ground mode; and x_i is the thickness of the sublayers in the soil profile that comprises T_{15} with soil index, *i,* divided by the total cumulative thickness of the T_{15} layer.

Franke and Kramer (2013) developed a performance-based empirical model based on the lateral spread model of Youd et al. (2002) MLR model by utilizing the same principles used in probabilistic seismic hazard analysis (PSHA). The primary component of their framework is usually described by the mean annual rate of exceeding a given engineering demand parameter (EDP) level. Franke and Kramer (2013) expressed a model conditional upon the site parameter by assigning all the uncertainty in the Youd et al. (2002) model to the conditional displacement calculation as:

$$
\lambda_{D_H|S}(d|S) = \sum_{i=1}^{N_L} P(D_H > d|S, L_i)\Delta\lambda_{L_i} \tag{1.22}
$$

where N_L i s the number of loading parameter increments required to span the range of possible L values; L represents quantities describing apparent loading; S represents quantities describing site condition; and $\Delta\lambda_{Li}$ is the increment of loading measure hazard. Because the loading parameter, which is a function of the commonly assumed, independent variables of distance and magnitude Eq. 1.22, can be rewritten as:

$$\lambda_{D_H|S}(d|S) = \sum_{i=1}^{N_S} \nu_i \sum_{j=1}^{N_M} \sum_{k=1}^{N_R} P\left(D_H > d \middle| S, M = m_j, R = r_k\right) \times P\left(M = m_j, R = r_k\right) \tag{1.23}$$

where ν is the mean annual rate of exceeding a minimum magnitude of interest for a given seismic source; M is the earthquake magnitude; and R is the distance to the seismic energy source.

Finally, a seismic hazard curve for conditional lateral spread displacements can be obtained from the Poisson probability distribution as:

$$P[(D_H)_T > d|S] = 1 - e^{-\lambda_{D_H|S} \cdot T} \tag{1.24}$$

where T is the time period in years.

Goh et al. (2014) utilized a nonparametric regression technique called multivariate adaptive regression splines (MARS) to model lateral spreads. They based their model on the database compiled by Youd et al. (2002).

In 2015, Khoshnevisan et al. (2015) modified Zhang et al. (2004) model based on the maximum likelihood analysis on 47 cases histories of liquefaction-induced lateral spread they derived from the 2010–2011 Canterbury earthquake in New Zealand to turn Zhang et al. (2004) model into a CPT-based probabilistic framework. They further developed a mathematical formulation for the maximum cyclic shear strain (γ_{max}) of Eq. 1.12.

1.2 Research Outline

Surveying the literature on liquefaction-induced lateral spread empirical models shows that despite this topic has gained considerable attention from researchers over the past three decades, some aspects need further work and development, including (1) the collection of more data and compilation of a more extensive, community-driven, PEER-reviewed, and vetted database, (2) the development of new predictive empirical models based on a more extensive, vetted database, new methodologies to handle uncertainties, uniform assessment of the performance of existing and new empirical models, and re-evaluation of the independent variable selection process, and (3) the development of a probabilistic framework to answer not only "what is the probability of displacements exceeding a certain threshold?" but also "what is the probability of having lateral spread in the first place given certain seismological, topographical, soil and site factors?" Therefore, the research objectives accomplished by the research are:

1. Collecting additional data points to fill the current gaps in the existing databases and provide more robustness to models.
2. Evaluating and comparing of existing statistical and machine learning models for lateral spread to assess the shortcomings and identify the potential aspects that could be improved.
3. Evaluating and quantifying the uncertainty associated with handling missing values in the revised MLR model, including evaluation of different methods of inputting missing values and their contribution to the uncertainty; and the assessment and quantification of the uncertainty associated with the measurement errors of displacement vectors.

4. Improving the state-of-the-art empirical models for liquefaction-induced lateral spread by:
    a. Making necessary modifications to the revised MLR model to achieve lower prediction error and more conformity to the experimental results.
    b. Using Bayesian linear regression, which provides posterior distribution of regression parameters and more explicit quantification of model uncertainty.
    c. Changing the evaluation function to focus on the percentual error rather than the absolute errors.
    d. Developing an entirely new artificial neural network (ANN) model with a different architecture and higher regularization compared to the existing ANN models and the cost and evaluation functions of task 5.c, given that the size of the database is growing and simple linear regression models might not have enough capacity to fit the data.
5. Formulizing a probabilistic framework by introducing a new probability chain to include the probability of producing lateral spread given liquefaction in addition to the probability of having displacements exceeding a certain threshold amount.

1.3 Research Program and Organization of the Report

Almost all of the models developed for liquefaction-induced lateral spread are based on either the Bartlett and Youd (1992, 1995) database, which has 467 samples, or a later revision by Youd et al. (2002), which has 484 data points. However, some additional data points (less than 50) were added to the original database by later researchers. Despite these efforts, the accuracy of empirical models is bounded by the quality and quantity of the available data points. Almost all researchers (Bartlett & Youd 1992, Bartlett & Youd 1995, Bardet et al. 2002, Youd et al. 2002, Javadi et al. 2006, Goh and Zhang 2014) have attested that there are gaps in the database and the collection of more data is necessary. Consequently, empirical models need to be further refined as new data becomes available. Therefore, the collection of a more extensive database was one of the primary objectives of this research.

The data space of the existing database compiled by (Bartlett & Youd 1992 and 1995) and later revised by Youd et al. (2002) for free-face and sloping-ground modes are shown in Figures 1.10 and 1.11, respectively. Those plots also illustrate some of the select points used in Chapter 3 of this report. The potential gaps in the data space are easily observed in these figures. The presence of such gaps results in insufficient statistical support for predictions having predictive variables in those regions. One of the primary contributions of this work is the collection and presentation of an increased number of case histories and the compilation of new data samples from a more diverse set of earthquakes. Chapter 2 includes the data collection and development of a relational database.

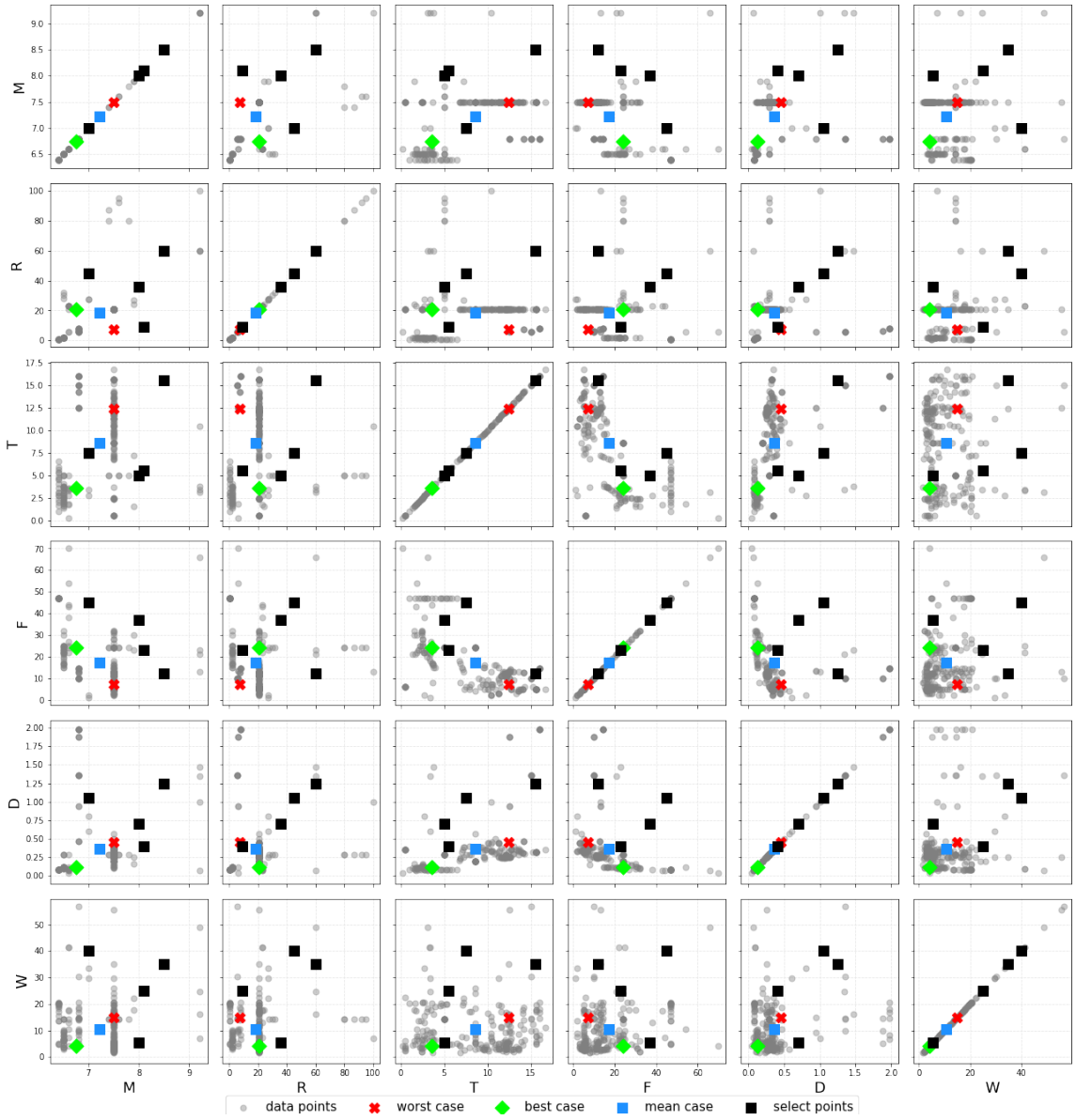


Figure 1.10 Data space of free face mode of existing lateral spread case histories database

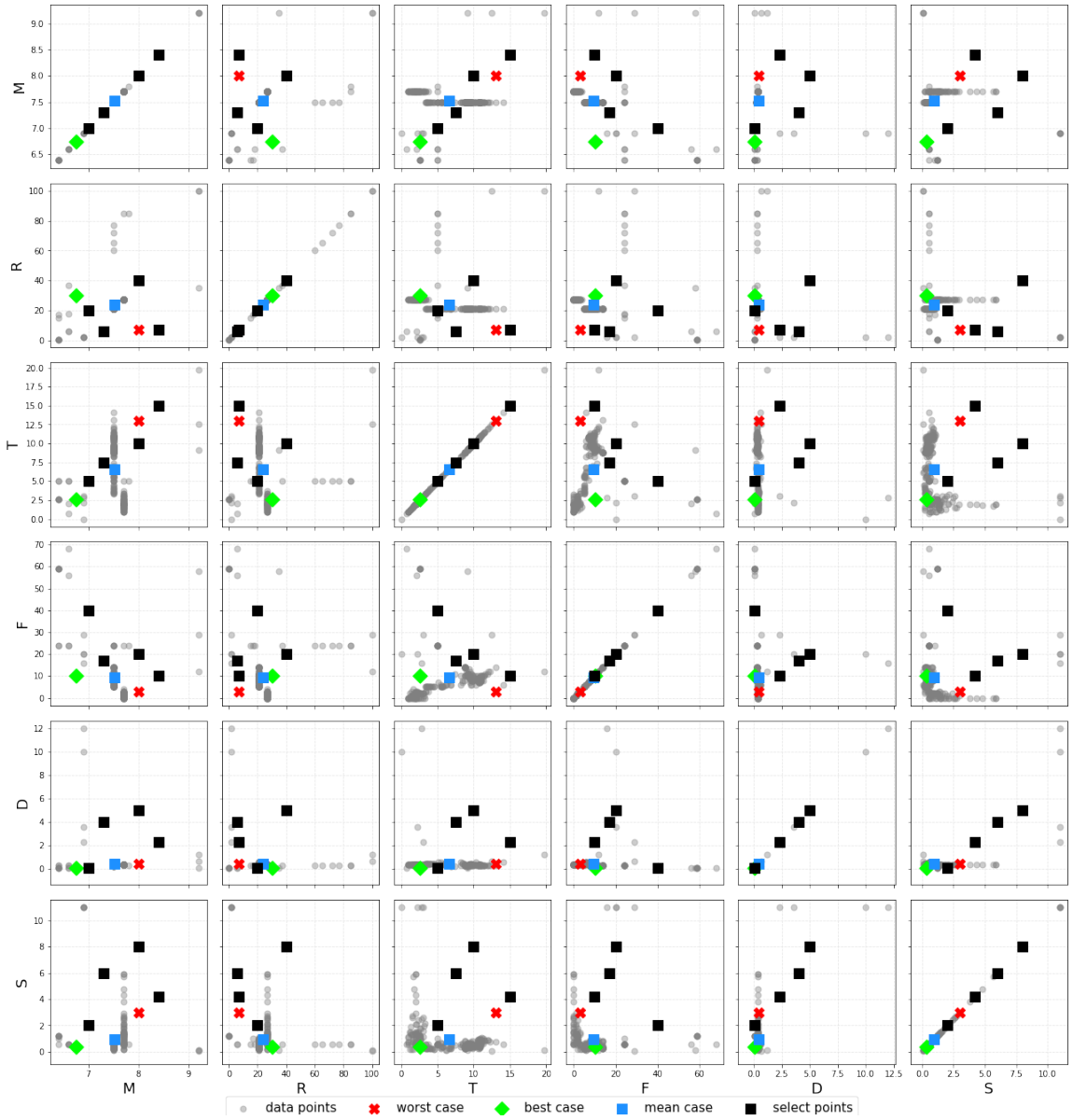


Figure 1.11 Data space of sloping ground mode of existing lateral spread case histories database

The abundance of similar empirical models confuses engineers/researchers regarding selecting the “best” model for engineering evaluations. Perhaps one comprehensive performance metric could be introduced to compare the empirical models to help resolve this issue. However, this metric does not currently exist. Instead, different evaluation metrics have been introduced, which are not comparable to each other in many cases. Chapter 3 contains a review and evaluation of most empirical lateral spread models' performance and assesses the advantages and disadvantages of each one. This evaluation will shed some light on the performance of existing empirical models.

Moreover, identifying the shortcomings of models can be used in future model developments. Zhang et al. (2012) published a paper comparing empirical models for predicting lateral spread. However, they only included in Youd et al. (2002) and Zhang & Zhao (2005) models. Chapter 3 presents a more extensive comparison and evaluation. The chapter also contains Bayesian uncertainty quantification of the MLR model, which is the most commonly used empirical model in practice.

Bartlett and Youd (1992, 1995) selected predictive variables of their MLR model based on feature importance in a multiple linear regression model, which does not necessarily result in the most critical features to be used in other algorithms, such as artificial neural networks or decision, tree-based ensembles, including random forest and gradient boosting decision methods. This limitation arises because ―dissimilar to newer machine learning algorithms ―linear models do not automatically consider interactions between variables and nonlinear patterns. Such interactions are only assessed in a rudimentary way via the use of cross-products in the model. Hence, the use of other features in more advanced artificial intelligence techniques has the potential to capture such interactions.

After compilation of a more comprehensive data set, as Youd (2018) discussed, the empirical models must be reformulated or re-regressed, or both. The addition of more data points may be used to re-regress existing models and give further insights to define/modify the functional forms or develop entirely new models. In addition to this, adding a new set of features (i.e., independent variables) might enable us to take advantage of a more advanced predictive algorithm and obtain more accurate models. The development of a new predictive model based on a more comprehensive database and a new set of features is presented in Chapter 4 of this report.

Finally, most developers of the empirical models did not describe guidelines regarding the application as was done by Wang and Rahman (1999), Bartlett and Youd (1992, 1995), Rauch and Martin (2000), and Youd et al. (2002). Also, practitioners or academics often neglect such guidance, as in the case of the master's thesis by Deterling (2015), discussed in Youd (2018). The reason (usually) is to facilitate computation and disregard that most of the models do not predict whether the site and subsurface conditions are favorable for generating lateral spread. In other words, many practitioners and researchers assume the probability of developing lateral spread given liquefaction is expected is 1.0. This assumption is not the case, as indicated in the case history dataset and further discussed and evaluated in this report. Unfortunately, this misuse of the MLR procedure or other empirical models can overestimate horizontal ground displacements, leading to unnecessary, costly mitigation strategies (Youd 2018). An intuitive way to prevent the issues discussed above is to formulate the flowcharts or "rules" into equations or logic trees so users can quickly implement them into engineering practice.

Also, a better, repeatable, formal mathematical approach to answering the question "under what conditions is lateral spread expected?" must be developed and implemented. Chapter 5 presents a new probabilistic framework for this by introducing a novel convolutional neural network to model the associated conditional probability. This approach includes trained latent vectors as mathematical representations for soil types. This approach can be generalized and used in any predictive model for liquefaction and lateral spread analysis.

2. DATA COLLECTION AND DATABASE DEVELOPMENT

An empirical model can provide reliable results when it is based on an adequate amount of data. An artificial intelligence algorithm learns from patterns within the data; hence the dataset must be of sufficient quantity and quality to reveal these patterns. Therefore, data quality and quantity play a significant role in developing statistical and machine-learning models.

Most empirical models of liquefaction-induced lateral spread are based on the 484 data samples of Youd et al. (2002) except for a few studies that added tens of more data points to the database. Almost all researchers who developed an empirical model for this phenomenon (Bartlett & Youd 1992, Bartlett & Youd 1995, Bardet et al. 2002, Youd et al. 2002, Javadi et al. 2006, Goh and Zhang 2014) attested that there are gaps in the database and collecting more data is necessary. Consequently, the models need further revisions and improvements as new data become available. To this end, as part of this report, we have collected over 10 times the largest existing lateral spread dataset, including 5,560 displacement vectors, 633 SPT boreholes, 11,586 subsurface data points, and 2,043 topology points from 12 earthquakes around the world. These case histories are mostly well-documented historical case histories that occurred between 1900 and 2000. This data collection was part of the Next Generation Liquefaction (NGL) project sponsored by the Pacific Earthquake Research Center https://apps.peer.berkeley.edu/lifelines/projects/ngl/. Funding was obtained from several state Departments of Transportation and the Mountain Plains Consortium. The compilation of lateral spread case histories was done jointly with researchers from Brigham Young University (BYU). BYU was responsible for case histories after 2000 that are not included in this report; displacement vectors in those case histories are obtained using newer techniques, such as optical image correlations (Little et al. 2021) for the 2011 Christchurch, New Zealand earthquake. The NGL is an Open Source Global Database and Model Development project for the Next-Generation of Liquefaction Assessment Procedures.

2.1 Next Generation Liquefaction

The Next Generation Liquefaction (NGL) Project is advancing state-of-the-art liquefaction research and working toward providing end-users with a consensus approach to assess liquefaction potential within a probabilistic and risk-informed framework. Specifically, NGL aims to collect and organize liquefaction information in a standard and comprehensive database to provide all researchers with a substantially more extensive, consistent, and reliable source of liquefaction data than previously existed. Based on this database, we will create probabilistic models that provide hazard- and risk-consistent bases for assessing liquefaction susceptibility, the potential for liquefaction to be triggered in susceptible soils, and likely consequences. NGL is committed to an open and objective evaluation and integration of data, models, and methods, as recommended in a National Academies report (2016). Following these principles will ensure that the resulting liquefaction susceptibility, triggering, and consequence models are reliable, robust, and vetted by the scientific community, providing a solid foundation for designing, constructing, and overseeing critical infrastructure projects.

Data published to the NGL project is publicly available for download at https://nextgenerationliquefaction.org/ through an interactive Graphical User Interface (GUI).

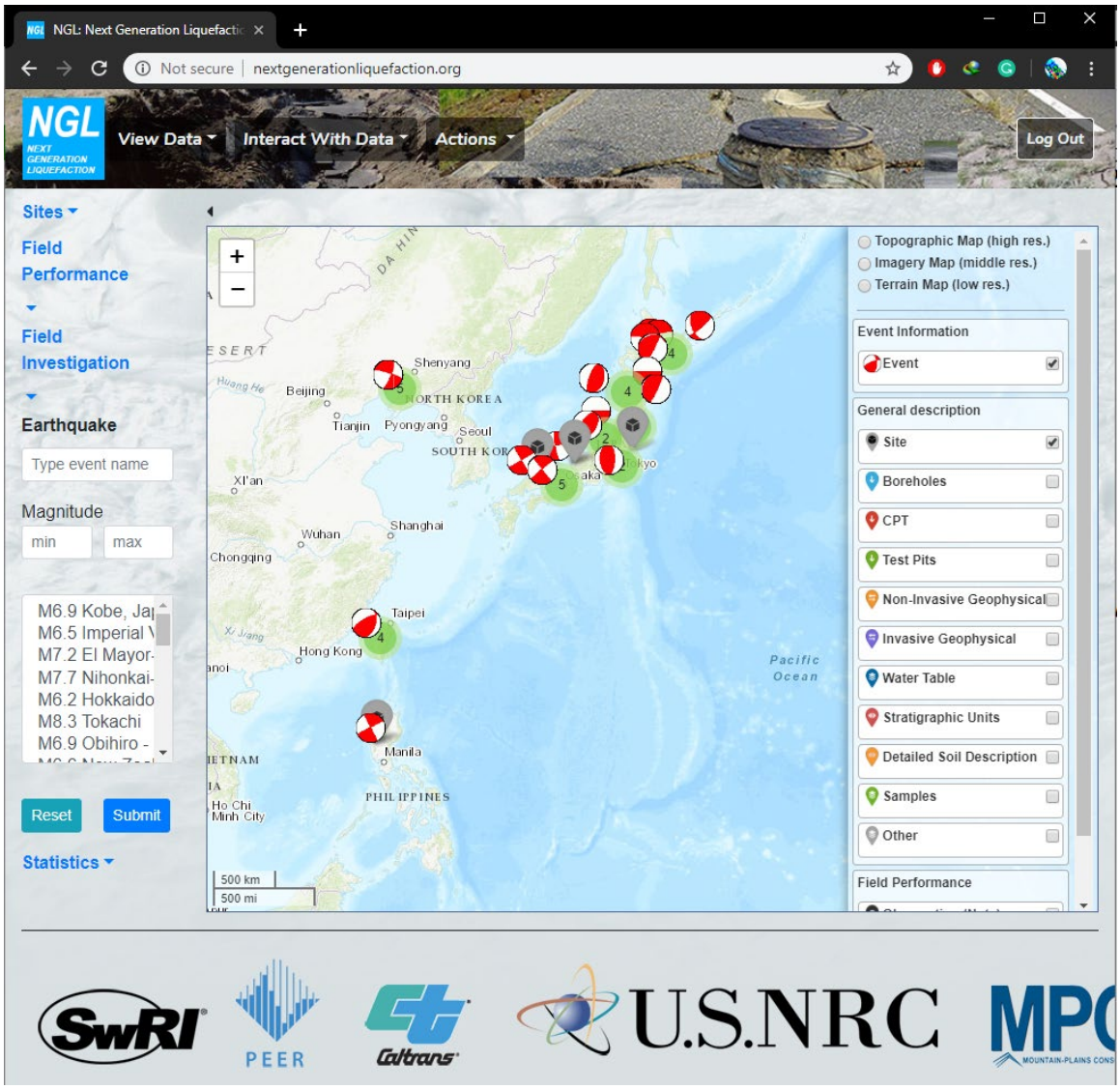


Figure 2.1 Graphical User Interface of the NGL project

Figure 2.1 shows the GUI of the NGL project, where one can interactively query and download the data.

2.2 Data Collection

The process of data collection consisted of researching and reviewing literature for any data related to liquefaction and lateral spread of case histories of interest. After identifying the documents, data had to be digitized and reformatted to fit the database schema. Data came in different languages, primarily English and Japanese and can be categorized into three main classes:

1. Displacement vectors; WGS 1984 coordinate the vector's starting point and its magnitude and azimuth.
2. Standard Penetration Test boreholes; WGS 1984 coordinate of the borehole; ground water level; testing method information, such as hammer energy ratio if available; and subsurface data, including blow counts, fines content, soil type, etc.

3. Topology; WGS 1984 coordinate of topology point andthe altitude.

2.3 Case Histories

2.3.1 1906 San Francisco

The 1906 San Francisco earthquake occurred on April 18, 1906, at 5:12 a.m. PST. This earthquake was perhaps the most important seismic event in U.S. history and among the most significant in the worldwide catalog of destructive earthquakes. From an engineering and planning perspective, this M_L 8.3 earthquake is of considerable interest. One of the distinctive features of this earthquake was its impact on lifeline systems. Data summarized in Table 2.1 was collected from O'Rourke et al. (1992), which is an expansion of the previous studies (Youd and Hoose, 1978; Hovland and Darragh, 1981; Roth and Kavazanjian, 1984; O'Rourke and Lane, 1989).

2.3.2 1964 Niigata, Japan

The 1964 Niigata earthquake occurred on June 16, 1964, at 04:01 UTC with a magnitude of 7.6. This catastrophic earthquake resulted in severe damage, including 3,534 houses destroyed and a further 11,000 people injured (Kawasumi 1968). Data summarized in Table 2.2, is collected from Hamada et al. (1986).

2.3.3 1964 Alaska, US

The March 27, 1964, Alaska earthquake was one of the most powerful earthquakes of the 20th century. The U.S. Coast Guard reported the mainshock and Geodetic Survey to be 8.3 to 8.4 on the M_L scale and 9.2 on the M_w scale (Hansen 1966; Kanamori, 1978). Despite this large event, most of the stricken region was sparsely populated, lessening the loss of life and property. However, the Alaska earthquake caused widespread crustal warping, and propagated long-lasting, intense ground shaking, which triggered numerous ground failures and landslides in south-central Alaska. Data summarized in Table 2.3 is collected from Bartlett and Youd (1992) and Ross et al. (1973).

Table 2.1 Summary of data collected from the 1906 San Francisco earthquake

Site	Displacement vectors	Boreholes	Subsurface data rows	Topology points
Mission creek zone	9	8	92	flatfile
South of market area	7	7	80	not available
Foot of market area	3	Quality data not available		not available
Total	19	15	172	

Table 2.2 Summary of data collected from the 1964 Niigata earthquake

Site	Displacement vectors	Boreholes	Subsurface data rows	Topology points
F10	179	24	359	429
G10	654	68	1574	256
H9	155	4	92	235
J9	442	45	192	297
K8	285	4	62	302
Total	1715	145	2279	1519

Table 2.3 Summary of data collected from the 1964 Alaska earthquake

Displacement vectors	Boreholes	Subsurface data rows	Topology points
14	20	411	flatfiles

2.3.4 1971 San Fernando, California

The 1971 San Fernando earthquake occurred on February 9, 1971, at 6:01 a.m. PST and was an event of considerable engineering and seismological significance. It registered as a M_L 6.4 event and affected an area of 220,000 km^2, including southern California, western Arizona, and southwestern Nevada (Coffman et al. 1982). From a geotechnical perspective, liquefaction-induced ground movements and slope failures cause substantial structural damage and left the lower San Fernando Dam precariously close to a catastrophic failure (Cortright, 1975). Data summarized in Table 2.4 is collected from Bennett (1989), O'Rourke et al. (1992), and Youd (1973).

2.3.5 1979 Imperial Valley, California

The Imperial Valley earthquake occurred on October 15, 1979, at 23:16 GMT. The Imperial Valley is a sparsely populated, primarily agricultural area located in Southern California near the Mexican border. The U.S. Geological Survey (USGS) array consisting of 13 stations recorded an M_L 6.6 earthquake. It was the largest earthquake occurring in California after the 1971 San Fernando event. The Imperial Valley earthquake mostly damaged irrigation and drainage systems. Data summarized in Table 2.5 is collected from Youd and Bennett (1983) and Bennett et al. (1984).

2.3.6 1983 Noshiro, Japan

The 1983 Noshiro earthquake (also known as the Sea of Japan earthquake or Nihonkai-Chubu earthquake) occurred on May 26, 1983, at 02:59 UTC. It was an M_W 7.8 event that occurred in the Sea of Japan, about 100km west of the coast of Noshiro in Akita Prefecture, Japan. This devastating earthquake resulted in 104 fatalities. Data summarized in Table 2.6 is collected from Hamada et al. (1986).

Table 2.4 Summary of data collected from the 1971 San Fernando earthquake

Site	Displacement vectors	Boreholes	Subsurface data rows	Topology points
Jensen water plant	69	33	494	flatfile
Juvenile hall	79	6	121	flatfile
Total	148	39	615	

Table 2.5 Summary of data collected from the 1979 Imperial Valley earthquake

Site	Displacement vectors	Boreholes	Subsurface data rows	Topology points
Heber road	29	7	135	flatfile
River park site	not available	4	62	not available
Total	29	11	197	

Table 2.6 Summary of data collected from the 1983 Noshiro earthquake

Site	Displacement vectors	Boreholes	Subsurface data rows	Topology points
South	266	128	462	176
North	147	59	848	348
Total	413	187	1310	524

2.3.7 1983 Borah Peak, Idaho

The 1983 Borah Peak earthquake occurred on October 28, 1983, at 8:06 MT in the western United States in the Lost River Range at Borah Peak in Central Idaho. The main shock was an M_W 6.9 event with a maximum Mercalli intensity of IX (Violent). Data summarized in Table 2.7 is collected from Youd et al. (1985), Andrus (1991), and Andrus and Youd (1987).

Table 2.7 Summary of data collected from the 1983 Borah Peak earthquake

Site	Displacement vectors	Boreholes	Subsurface data rows	Topology points
Whiskey springs	3	3	54	flatfile
Pence Ranch	3	6	69	flatfile
Total	6	9	54	

2.3.8 1987 Superstition Hills, California

The 1987 Superstition Hills earthquake occurred on November 24, 1987, at 01:54 UTC. This M_W 6.2 earthquake caused damage to places in California and Mexico. Data summarized in Table 2.8 is collected from Holzier et al. (1989).

Table 2.8 Summary of data collected from the 1987 Superstition Hills earthquake

Displacement vectors	Boreholes	Subsurface data rows	Topology points
7	2	53	flatfiles

2.3.9 1989 Loma Prieta, California

The 1989 Loma Prieta earthquake was one of the most catastrophic earthquakes in the United States. The event resulted in the death of 63 people, injuries to 3,757 people, and more than 12,000 left homeless (O'Rourke and Hamada 1992). This M_W 6.9 earthquake occurred on October 17, 1989, at 5:04 p.m. PST. Data summarized in Table 2.9 is collected from Robertson et al. (1999).

Table 2.9 Summary of data collected from the 1989 Loma Prieta earthquake

Displacement vectors	Boreholes	Subsurface data rows	Topology points
3	15	236	flatfiles

2.3.10 1990 Luzon, Philippines

The 1990 Luzon earthquake occurred on July 16, 1990, at 4:26 p.m. PDT with an estimated M_W 7.8 event and a maximum Mercalli intensity of IX (Violent). The earthquake epicenter was near the town of Rizal, Nueva Ecija, northeast of Cabanatuan City. Loss of life was estimated at 1,621. Data summarized in Table 2.10 is collected from Tokimatsu et al. (1994) and Ishihara et al. (1993).

Table 2.10 Summary of data collected from the 1990 Luzon earthquake

Displacement vectors	Boreholes	Subsurface data rows	Topology points
11	13	233	not available

2.3.11 1999 Chi-Chi, Taiwan

The Chi-Chi, Taiwan earthquake (also known as the Jiji earthquake) occurred on September 21, 1999, at 17:47 UTC with a moment magnitude scale of 7.7. As a result of this event, over 2,415 people perished and 11,305 were injured. Data summarized in Table 2.11 is collected from Chu et al. (2004).

Table 2.11 Summary of data collected from the 1999 Chi Chi earthquake

Site	Displacement vectors	Boreholes	Subsurface data rows	Topology points
Site C	20	15		flatfiles
Site C1	4	2		
Site B	4	6		
Site M	4	2		
Site N	4	3		
Total	34	23	2806	

2.3.12 1995 Kobe, Japan

The M_W 6.9 1995 Kobe earthquake (also known as Hyogoken-Nanbu earthquake) occurred on January 17, 1995. The event that struck Kobe, Japan, and the surrounding area was the most severe earthquake to affect the region in the 20th century. The earthquake resulted in more than 6,000 deaths and 30,000 injuries (Chung 1996). Data summarized in Table 2.12 is collected from Chu et al. (2004).

Table 2.12 Summary of data collected from the 1995 Kobe earthquake

Displacement vectors	Boreholes	Subsurface data rows	Topology points
3168	156	3273	not available

3. EMPIRICAL MODELS PERFORMANCE AND BAYESIAN UNCERTAINTY QUANTIFICATION OF THE MLR MODEL

3.1 Introduction

Soil liquefaction generally occurs in young, saturated, loosely deposited, uncemented, or poorly cemented sandy soils (Youd and Hoose, 1978, Youd and Perkins, 1978, Youd and Perkins, 1987, Bartlett and Youd, 1992, Gillins, 2012). Soil liquefaction could cause loss of bearing strength, sand boils, flow failure, floatation, settlement, ground oscillation, and lateral spreads, which is its most pervasive consequence. Lateral spread, as defined by Youd (2018), is lateral displacement of a soil layer atop a liquified layer of soil either toward a free face or a gentle slope. Displacements range from a few tenths of a meter to several meters (Bartlett and Youd 1992) and are due to a combination of gravitational and earthquake-generated inertial forces.

Because of the potential for severe and costly damage of lateral spreads to the infrastructure, many researchers have tried to quantify its effects using simplified analytical methods (Yegian et al. 1991, and Olson & Johnson 2008), empirical predictive models, numerical models (Finn et al. 1994), and laboratory devices (Sharp et al. 2003, and Kutter et al. 2004). Each of these approaches has advantages and disadvantages and domain of application. The focus of this report, however, is mainly on the empirical predictive models.

One of the first attempts in this field was made by Hamada et al. (1986); they proposed a regression model based on the slope and thickness of the liquefied layer. Youd and Perkins (1987) introduced the Liquefaction Severity Index (LSI) as the estimated maximum amount of lateral spread displacement normalized to the site conditions. Subsequently, Bartlett and Youd (1992, 1995) collected many data points from various lateral spread case histories of Japan and the United Statesand developed an empirical model to estimate lateral spread displacement. Youd et al. (2002) further updated the Multiple Linear Regression (MLR) model and included additional data points from more recent case histories. Bardet et al. (2002) suggested removing the geotechnical variables from the Bartlett and Youd empirical model and proposed a new set of MLR equations. Since then, several researchers have investigated different variations of the MLR model. For example, Gillins and Bartlett (2012) replaced the soil variables with soil type and showed how CPT data could be incorporated into the model. Goh and Zhang (2014) used multivariate adaptive regression (MARS) procedure to propose a predictive model.

Besides the above-mentioned regression models, several other methods, including machine learning algorithms, have been applied to this database. Wang and Rahman (1999) evaluated lateral spread displacements using an artificial neural network (ANN) and the database from Bartlett and Youd (1992). Later, Baziar and Ghorbani (2005) developed a neural network-based on the revised database of Youd et al. (2002). Javadi et al. (2006) utilized the genetic programming approach. Garcia et al. (2008) proposed a neuro-fuzzy system that profits from fuzzy and neural paradigms. Ommen and Baise (2010) used a machine-learning technique known as Support Vector Regression (SVR). Rezania et al. (2011) developed a model based on evolutionary polynomial regression (EPR). Finally, Baziar and Azizkandi (2013) proposed a newer neural network model.

The abundance of empirical predictive models could raise the question of which model should be used to make predictions? The empirical models are often compared to each other in terms of simple statistics, such as coefficient of determination (R-squared), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE). While these statistics could be one way to assess how the model predicts observed outcomes, they fail to describe the performance of the entire model.

This study aims to conduct a comprehensive comparison, beyond simple summary statistics, on the empirical predictive models of liquefaction-induced lateral spreads and assess their performances. For this purpose, two sets of empirical predictive models are chosen from the literature: (1) the regression models including Youd et al. (2002), Bardet et al. (2002), and Goh et al. (2014), and (2) the machine learning-based models without a predefined functional form including Javadi et al. (2006), Rezania et al. (2011), and Baziar & Azizkani (2013). These models are compared in terms of their reported statistics (e.g., R^2, RSME, and MSE and their variations and ability to extrapolate or predict points with poor statistical support. Practitioners often extrapolate the models in their evaluations; however, more robust models should allow some extrapolation. Also, the additional uncertainty from the extrapolation should be quantified.

3.2 Database Overview

Most empirical models are based on the revised database of Youd et al. (2002) or the addition of more data points to this database. Thus, knowledge of the formulation of this database is required. Unfortunately, various researchers have made different — sometimes incorrect — inferences from the database contents. The independent variables used in the database of Youd et al. (2002) are briefly described in Table 3.1. We emphasize that the T_{15} variable is not the thickness of the liquefiable layerbut represents the cumulative thickness of saturated granular layers with corrected blow counts less than 15. Bartlett and Youd (1992) chose this variable over the thickness of the liquefiable layer because of its high correlation with the observed horizontal ground displacements in their compiled database. Also, this factor can be independently calculated without the need for liquefaction analyses. The thickness of the deformable layer, T15, is a fundamental geometrical property if not a fundamental soil property. Theoretically, ground displacement is equal to the average shear strain multiplied by the thickness of the deformable layer; T_{15} is an estimate of that thickness. T_{15} has high statistical significance and hence, is an independent variable in the MLR model. Bartlett and Youd (1992) tested several possible independent variables as measures of the blow-count distribution, including the lowest $(N_1)_{60}$ in T_{15}, average $(N_1)_{60}$ in T_{15}, lowest calculated factor of safety against liquefaction in T_{15}, and so on. None of those potential indices of the blow-count distribution is statistically significant, and thus, they were not included in the 1992, 1995, and 2002 MLR models. The reason for that insignificance is not because the blow-count distribution is unimportant, but instead, because there appears to be only a small variation of the blow-count distribution within naturally occurring liquefiable layers compiled in the MLR database. It appears that nature has not deposited liquefiable layers with uniform blow counts of, say, 3 or 14 (Youd, 2018). For the sake of simplicity, some of the subscripts in variable names are eliminated hereafter.

Table 3.1 Description of parameters used in Youd et al. (2002)

	Unit	Description
M		Moment magnitude of earthquake
R	km	Shortest horizontal distance measured from the surface projection of the seismic energy source or fault rupture to the site of interest
T15	m	Cumulative thickness of saturated granular layers with corrected blow counts, (N1)60, less than 15
F15	%	Average fines content (fraction of sediment sample passing a No. 200 sieve) for granular materials included within T15
D50	mm	Average mean grain size for granular materials within T15
W	%	Free-face ratio defined as the height (H) of the free face divided by the distance (L) from the base of the free face to the point in question
S	%	Ground slope
Dh	m	Horizontal (lateral) ground displacement

Table 3.2 Summarized descriptive statistics of free-face datapoints of Youd et al. (2002)

	M	**R** (km)	$\mathbf{T}_{15}$ (m)	$\mathbf{F}_{15}$ (%)	$\mathbf{D}_{50}$ (mm)	**W** (%)	$\mathbf{D}_{h}$ (m)
mean	7.22	18.39	8.57	17.12	0.36	10.66	2.52
std	0.526	15.292	4.769	13.233	0.402	8.995	2.29
min	6.40	0.50	0.20	1.00	0.04	1.64	0.01
25%	6.75	7.25	3.60	7.00	0.12	4.24	0.93
50%	7.50	21.00	9.60	13.00	0.28	7.41	1.80
75%	7.50	21.00	12.40	24.00	0.35	15.06	3.29
max	9.20	100.00	16.70	70.00	1.98	56.80	10.16

Table 3.3 Summarized descriptive statistics of sloping-ground data points of Youd et al. (2002)

	M	**R** (km)	$\mathbf{T}_{15}$ (m)	$\mathbf{F}_{15}$ (%)	$\mathbf{D}_{50}$ (mm)	**S** (%)	$\mathbf{D}_{h}$ (m)
mean	7.52	23.79	6.56	9.36	0.43	0.95	1.83
std	0.311	11.721	3.726	10.763	0.978	1.520	0.983
min	6.40	0.20	0.01	0.00	0.06	0.05	0.01
25%	7.50	21.00	2.60	3.00	0.28	0.36	1.20
50%	7.50	21.00	8.20	8.00	0.35	0.55	1.61
75%	7.70	27.00	9.52	10.25	0.35	0.91	2.46
max	9.20	100.00	19.70	68.00	12.00	11.00	5.36

Tables 3.2 and 3.3 represent the summarized descriptive statistics, including mean, standard deviation, and quartiles for free-face and sloping-ground data points. However, the data space must be understood more deeply by users. The application domain of empirical models is usually reported as a set of variables ranges, such as the minimum and maximum values reported in Tables 3.2 and 3.3. When considering only three variables, this would result in a cubic data space while the actual data space might not be cubic, i.e., data points could be clustered at scattered volumes within the cube. As a result, despite some points are within the application domain of a model, there might not be enough statistical support around it. In such instances, the model being used is extrapolating, which is common in models constructed upon the revised version of the Youd et al. (2002) database when applied to new data points. For example, Figures 3.1 and 3.2 show the data space using matrix plots of variables for the free-face and sloping-ground modes to make this point clearer, respectively. In these figures, the data points of the database compiled by

Youd et al. (2002) are shown semi-transparently to visualize clustering (density of statistical support) of data within the data space. It is noteworthy in the case of free-face data points — the points with F > 20% and D > 0.5mm have almost no statistical support in the data space even though they are within the minimum and maximum ranges of the data space. Thus, caution is warranted when using empirical models to predict behavior with these inputted values.

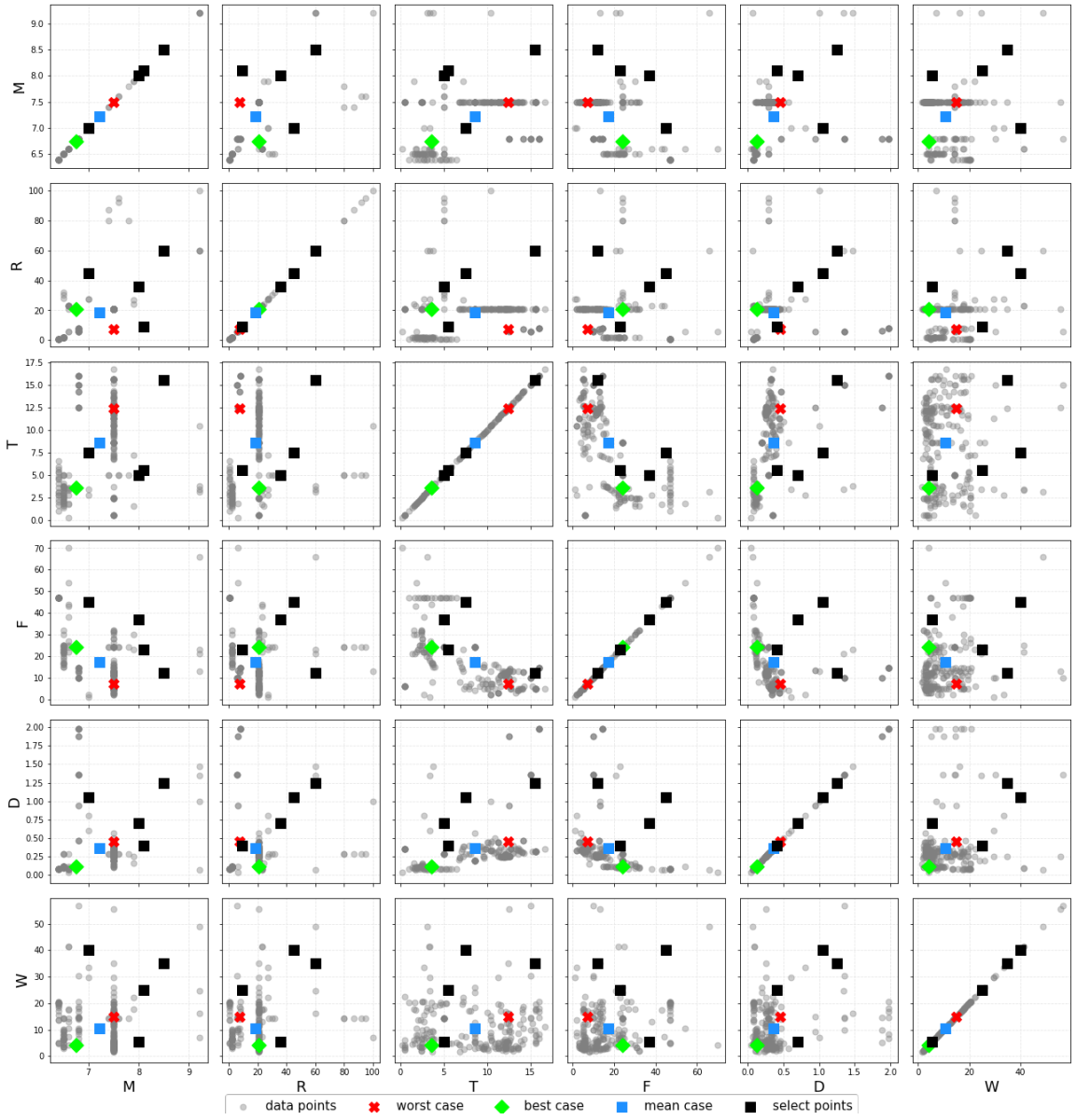


Figure 3.1 Data space of free-face mode

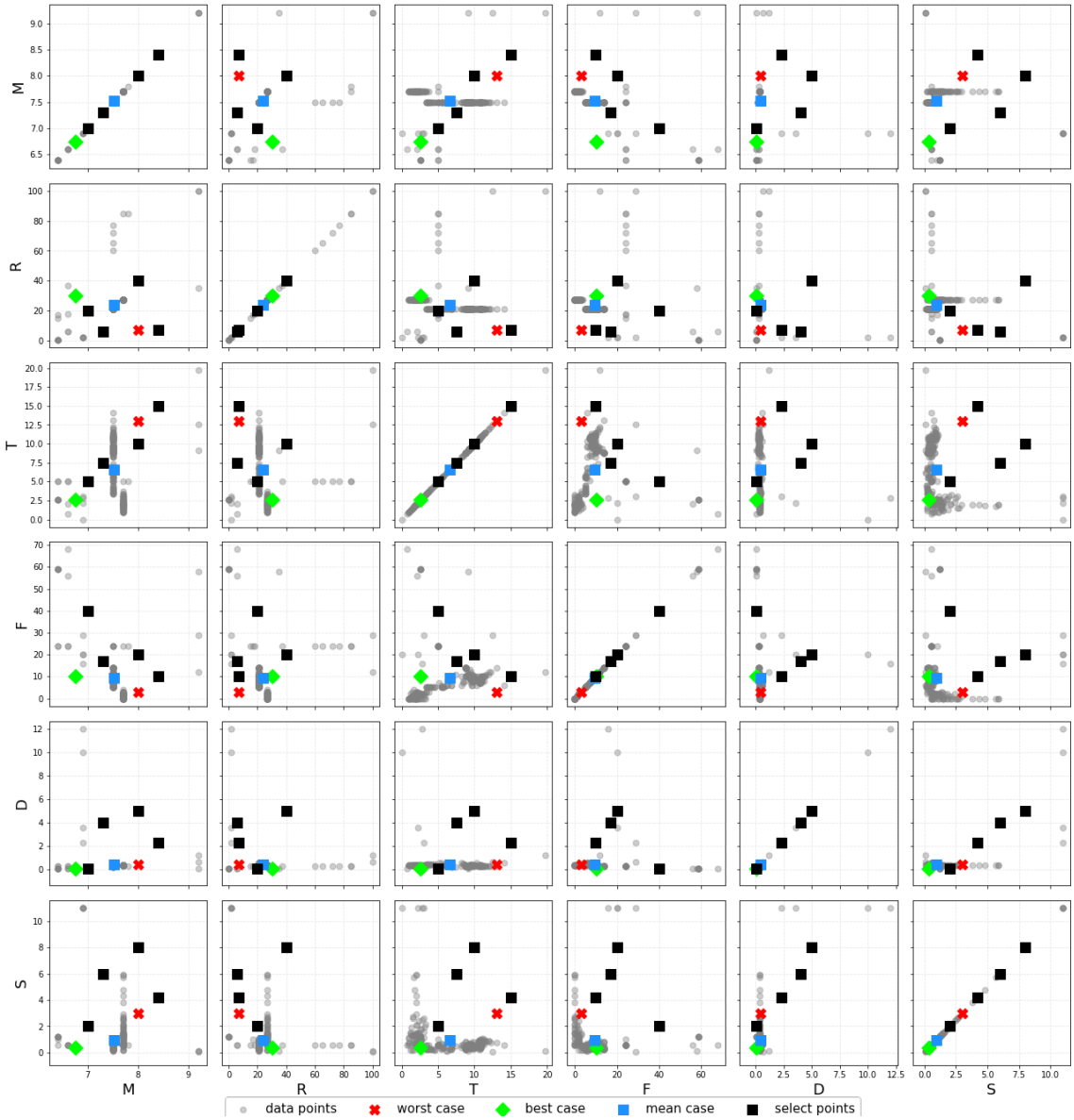


Figure 3.2 Data space of sloping-ground mode

3.3 Models

Six empirical models developed for liquefaction-induced lateral spreads are considered

- Regression models:
  - Multiple Linear Regression (MLR) (Youd et al. 2002)
  - Multiple Linear Regression (MLR) (Bardet et al. 2002)
  - Multivariate Adaptive Regression Splines (MARS) (Goh et al. 2014)
- Machine learning-based models
  - Genetic Programming (Javadi et al. 2006)
  - Evolutionary-based approach (Rezania et al. 2011)
  - Artificial Neural Network & Genetic Algorithm (Baziar & Azizkandi 2013)

It was found that the artificial neural network and the genetic algorithm used by Baziar & Azizkandi (2013) was incorrectly trained on the ground slope of 1.0 % in all free-face data points and the free-face ratio of 1.0 in all sloping ground data points, so the final equation is not valid. On the other hand, the authors could not replicate the results of the MARS model and thus, did not consider it further. Moreover, some of the other machine-earning-based models, such as Wang and Rahman (1999) and Garcia et al. (2008), were not considered because they had not presented their models in terms of predictive equations set for the end-user. Equations proposed by the other four models are briefed in Table 3.4.

Table 3.4 Empirical equations

Model	Mode	Equation
Youd et al. (2002)	Free face	$\log(D_h) = -16.713 + 1.532\ M - 1.406\ \log(R^*) - 0.012\ R + 0.592\ \log(W) + 0.540\ \log(T) + 3.413\ \log(100\text{-}F) - 0.795\ \log(D + 0.1)$
	Sloping ground	$\log(D_h) = -16.213 + 1.532\ M - 1.406\ \log(R^*) - 0.012\ R + 0.338\ \log(S) + 0.540\ \log(T) + 3.413\ \log(100\text{-}F) - 0.795\ \log(D + 0.1)$
Bardet et al. (2002)	Free face	$\log(D_h) = -6.815 - 0.465 + 1.017\ M - 0.278\ \log(R) - 0.026\ R + 0.497\ \log(W) + 0.558\ \log(T)$
	Sloping ground	$\log(D_h) = -6.815 + 1.017\ M - 0.278\ \log(R) - 0.026\ R + 0.454\ \log(S) + 0.558\ \log(T)$
Javadi et al. (2006)	Free face	$D_h = -163.1/M^2 + 57/(R\ F) - 0.0035\ T^2/(W\ D^2) + 0.02\ \ T^2/(F\ D^2) - 0.26\ (T/F)^2 + 0.006\ T^2 - 0.0013\ W^2 + 0.0002\ M^2\ W\ T + 3.7$
	Sloping ground	$D_h = -0.8\ F/M + 0.0014\ F^2 + 0.16\ T + 0.112\ S + 0.04\ S\ T/D - 0.026\ R\ D + 1.14$
Javadi et al. (2006) - moderate	Free face	$D_h = -234.1/(M^2\ R\ W) - 156/M^2 - 0.008\ F/(R^2\ T) + 0.01\ W\ T/R - 2.9/F - 0.036\ M\ T^2\ D^2/(R^2\ W) + 9.4\ M(R\ F) - 4\ (10^{-6})\ M\ R^2/D + 3.84$
	Sloping ground	$D_h = -0.027\ T^2\ F/M^2 + 0.05\ R\ T/(M^2\ D) + 0.44/(M\ R^2\ S\ T) - 0.03\ R - 0.02\ M/(S\ T) - 5\ (10^{-5})\ M\ R/D^2 + 0.075\ M^2 - 2.4$
Rezania et al. (2011)	Free face	$D_h = -2.1414\ R^{0.5}\ W^{0.5}/(M^2\ D^{0.5}) - 0.061863\ T\ F/(M^{0.5}\ W^{0.5}) - 11.1201\ M^2/(R\ W^{0.5}\ F) + 0.0017573\ M^2\ W^{0.5}\ T/(F^{0.5}\ D) + 1.9671$
	Sloping ground	$D_h = -1.6941\ T^{0.5}\ F^{0.5}/(M^2\ D^{0.5}) - 0.78905\ R^{0.5}\ S^{0.5}\ T\ F^{0.5}/M^2 - 2.2542\ (10^{-12})\ M^{0.5}\ T^2\ D^2/(R^{0.5}\ S^{0.5}\ F^2) + 0.036036\ M\ S^{0.5}\ T/D^{0.5} + 0.85441$

All the variables in Table 3.4 are described in Table 3.1 except R*, which is defined by Youd et al. (2002) as

$$R^* = R + 10^{0.89M-5.64} \tag{3.1}$$

Among the models considered here, the regression models of Youd et al. (2002) and Bardet et al. (2002) had predefined functional forms, while the other two models let the models recognize the functional forms automatically. Since Youd et al. (2002) and Bardet et al. (2002) models are similar in many aspects, they are hereafter referred to as the regression models, while Javadi et al. (2006) and Rezania et al. (2011) models are referred to as machine-learning models.

3.4 Models Comparison

The empirical models of liquefaction-induced lateral spreads are compared to each other from several aspects, including (1) basic summary statistics, (2) general inferences, (3) Monte Carlo experiments, (4) variations of models, and (5) extrapolation. Each of these comparisons is described in detail subsequently.

3.4.1 Statistical Measurements

The empirical models are often compared to each other in terms of simple summary statistics, such as coefficient of determination (R-squared), Root Mean Square Error (RMSE), and Mean Squared Error (MSE). The coefficient of determination is the proportion of the variance in the dependent variable that is predictable from the independent variable(s), defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.2}$$

where n = total number of observations; y_i = observed values; f_i = predicted values; and $\bar{y}$ = mean of the observed values.

Another popular statistic that is frequently used to quantify the difference between the values predicted by a model and the ones observed is Root Mean Square Error (RMSE), defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - f_i)^2}{n}} \tag{3.3}$$

The values of R^2 and RMSE reported by developers of each model is reported in Table 3.5. Note that the R^2 values marked with an asterisk are calculated based on the invalid assumption of $\bar{y}$=0. As an example, Figure 9.b of Rezania et al. (2011) has a reported R^2 value of 95.6% for the EPR model, while the model prediction for most cases is about 1.2 m. The correct value of R^2, calculated by the authors, is 64.8%.

Another point is that the root mean square error, which is the standard deviation of the residuals (prediction error), could be a reasonable statistic to compare models when all data points are treated equally. However, this is not a good statistic in the case of horizontal ground displacement. For example, having a 0.8-m residual in a 1m displacement vector is equivalent to 80% error in prediction, while the same amount of residual in a 10-m displacement vector is equal to only 8% error. Thus, RMSE and similarly, the Mean Absolute Error (MAE), are not good statistics to compare these models.

Table 3.5 Reported values of statistical measurements

Model	Mode	Reported R^2 (%)		Reported RMSE (m)	
		Training	Testing	Training	Testing
Bartlett & Youd (1992)	Compound	74.90		NR	
Youd et al. (2002)	Compound	83.60		NR	
Bardet et al. (2002)	Compound	64.26[A]		NR	
Javadi et al. (2006)	Free face	95.0*	89.6*	0.22	0.27
	Sloping ground	96.7*	95.5*	0.17	0.24
Rezania et al. (2011)	Free face	95.3*	89.8*	0.21	0.27
	Sloping ground	96.7*	95.6*	0.18	0.23
Baziar and Azizkandi (2013)	Compound	88.60*		0.8	
Goh et al. (2014)	Free face	93.68	85.59	0.1449	0.1530
	Sloping ground	90.47	89.39	0.1285	0.1429

** values of R^2 are calculated based on the invalid assumption that values have a mean zero.*
[A] Average of data sets A and B.
NR: not reported by authors

This point leads to another concept called Heteroscedasticity (i.e., non-uniform variance), which refers to the circumstance in which the variability of a variable is unequal across the range of values of a second variable that predicts it. In other words, the residuals are a function of the predictors. In statistical models, one tries to avoid heteroscedasticity and have a normal distribution for residual errors, but as mentioned earlier, heteroscedasticity is not necessarily harmful in this domain.

3.4.2 Anscombe's Quartet

As discussed by the statistician Anscombe (1973), "Numerical calculations are exact, but graphs are rough." He introduced four sets of data with an entirely different observable trend but had the same correlation R value of 0.816 between predictors (x_i) and predictions (y_i). He also showed that the same line could be regressed for all four datasets with a coefficient of determination of 0.67. Figure 3.3 shows the results. All four sets are identical when examined using simple summary statistics but vary considerably when data points are graphed.

While the coefficient of determination could be used as a metric to describe the performance evaluation of empirical models, it fails to explain potential curvature and clustering of data when used as the sole metric. A quick look at Table 3.5 suggests (misleadingly) that because the newer models developed more recently have higher R^2 and lower RMSE values, they are more accurate.

Also, considering the significant amount of noise in the MLR dataset, a high value of R^2 might be caused by a phenomenon known as "overfitting." In empirical predictive models, there is always a risk to fit random noise in the data by memorizing various peculiarities of the training data rather than finding a general predictive rule (Dietterich, 1995).

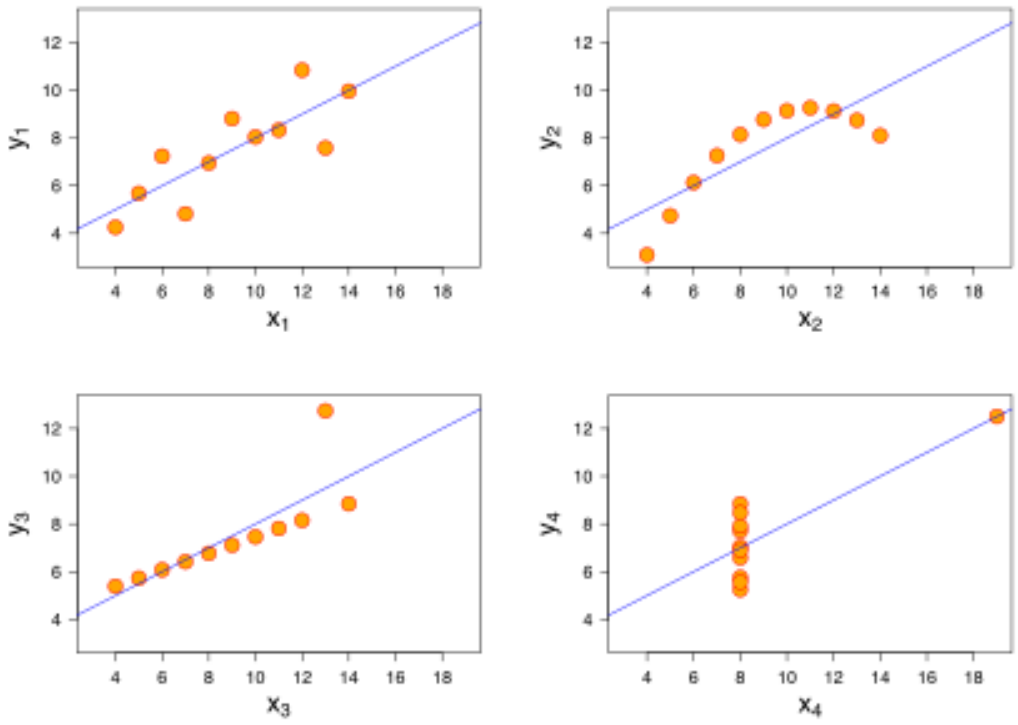


Figure 3.3 Anscombe's quartet

This overfitting could be continued until an R^2 of 1.0 is obtained on the training set. For example, Zhang et al. (2016) showed that 100% accuracy on the training set is obtainable in artificial neural networks even after randomizing the labels entirely. However, the accuracy of the test set drops significantly. As a general rule of thumb, a generalization gap or lower values of R^2 on the test set compared to that of the training set (as seen in the free face mode of Goh et al. (2014) in Table 5) indicates overfitting. Overfitted models result in divergence and unrealistic variability.

The regression models considered in this report do not have two separate training and test sets; hence, the metrics of model performance reported by the developers of the machine learning models are invalid. Therefore, these models could not be compared to each other in terms of the statistical metrics. Even if the performance metrics of the machine-learning models were valid, simply comparing models based on these statistics is not sufficient, as discussed by Anscombe (1973). To better compare these models, they will be assessed further by using other aspects, as described in the subsequent sections.

3.4.3 Functional Forms

One of the factors considered in this study is the general inferences that might be made on the final functional form of the liquefaction-induced lateral spread predictive equations.

Among the models considered herein, Youd et al. (2002) and Bardet et al. (2002) have predefined functional forms, while Javadi et al. (2006) and Rezania et al. (2011) let the models recognize their functional forms automatically. The latter is particularly useful in cases where the mechanics behind the phenomenon is primarily unknown. However, the mechanics behind liquefaction-induced lateral spreads has been the focus of many researchers for decades and is generally understood. On the other hand, automatic pattern recognition is also highly sensitive to the noise

in the dataset. Youd et al. (2002) developed their database mainly from historic case histories where some measurements had lower accuracy. Thus, there is some uncertainty (i.e., noise) in the dataset. For these reasons, having a functional form is vital in developing such models to avoid potential overfitting issues.
For Youd et al. (2002) and Bardet et al. (2002), the final predictive equation has a multiplicative form by transforming their formulation into the equivalent exponential equations. For example, in the free-face mode of Youd et al. (2002), the formulation will be

$$D_h = (10^{-16.713})(10^{1.532M})(R^*)^{-1.406}(10^{-0.012R})(W^{0.592})(T^{0.54})(100\text{-}F)^{3.413}(D+0.1)^{0.795} \quad (3.4)$$

The multiplicative form of this equation means that each and every variable in the equation could make a significant difference. For example, $\lim_{T\to 0} D_h = 0$ i.e. as the thickness of the liquefiable layer goes to zero, the predicted value of horizontal displacement also goes to zero. On the other hand, additive equations do not have this property, and each term of the equation could change the predictions by a limited magnitude. For example, in the sloping ground case of Javadi et al. (2006) and the free-face case of Rezania et al. (2014), given in Table 4, $\lim_{T\to 0} D_h \neq 0$ i.e. the equation could result in non-zero displacement predictions even without having a liquefiable layer whatsoever. Another issue associated with having an additive equation could be observed again in the sloping-ground case of Javadi et al. (2006) given in Table 3.4, whereas $F \rightarrow 0$ (clean sand) D_h will no longer be a function of moment magnitude of earthquake. In other words, in case of clean sands, the predictions are independent of the moment magnitude of earthquake.

Some terms of the predictive equations developed by Javadi et al. (2006) and Rezania et al. (2011) have independent variables in the denominator. These terms could result in a divergence in extreme cases. For example, both models have fines content (F) in some terms in the denominators, thus as F goes to zero, the predicted horizontal displacements go to infinity. This functional form could be problematic because prior research has shown that soil liquefaction is more likely to occur in clean sands with very low fines content (Idriss & Boulanger 2008). Thus, there is a high probability that this model will overpredict lateral spread displacement for clean sands.

Another point is that Rezania et al. (2011) and Javadi et al. (2006) proposed a separate set of equations for moderate lateral displacements ($D_h < 1.5$ m) to increase the prediction accuracy for moderate displacements. However, they did not describe how to determine whether or not the displacements will be less than 1.5 m; hence, it is unclear which equations should be applied to make the prediction.

The last general remark is the simplicity-accuracy tradeoff. Despite the R^2 reported by Javadi et al. (2006), Rezania et al. (2011), and Baziar and Azizkandi (2013) are misleading. Often, machine-learning algorithms yield results with higher R^2 than regression models for unnoisy datasets. However, higher R^2 values carry the cost of additional model complexity. Statisticians often strive to achieve parsimonious models, whereby the most influential variables are only used (i.e., those independent variables having high correlation with the dependent variable).

In contrast, more complex models with many variables are harder for researchers and engineers to implement and utilize and hard, if not impossible, to interpret. Regardless of their sophistication, the machine learning-based data models could be rendered powerless unless they can be interpreted by human experts (Vellido et al. 2012). In contrast, regression models are straightforward to interpret. For example, based on the Youd et al. (2002) model, the lateral spread magnitude is expected to be, on average, $10^{0.54} \approx 3.47$ times higher for each additional

increase in the cumulative thickness of saturated granular layers with $(N_1)_{60}$, values less than 15, controlling for other variables. Such inferences are hard to make, especially for deep neural networks with many hidden layers and/or random forests with deep decision trees.

3.4.4 Monte Carlo-based Sensitivity Analysis

Although some models could be divergent in extreme values of their predictors as described in previous sections, empirical models are usually meant to be used within the limits of the data space on which they are developed. In this section, a series of Monte Carlo experiments are conducted within those limits (minimum and maximum of the data space described in Tables 3.2 and 3.3). These experiments summarized in Table 3.6 give helpful insights into the expected outcome of these models. Every model was used to predict a million randomly generated data points. The values of variables are assumed to be equally likely within the data space.

As can be seen, different combinations of variables resulted in negative displacements in the machine-learning models, particularly in their sloping-ground modes where over 70% of all predictions were negative. The largest maximum and largest difference between mean and median predictions were observed in the Bardet et al. (2002) model, which could be interpreted as the most conservative model among the models evaluated in this study.

Table 3.6 Monte Carlo simulation results

Mode	Measure	Model			
		Youd et al. (2002)	Bardet et al. (2002)	Javadi et al. (2006)	Rezania et al. (2011)
free face	Median (m)	0.27	1.08	2.77	.19
	Mean (m)	3.06	25.4	4.12	0.40
	Minimum (m)	00.0	00.0	-177.7	-419.5
	Maximum (m)	477.6	4132.1	2119.5	245.0
	Standard deviation (m)	10.3	106.0	14.8	5.19
	Negative predictions	0	0	100,357	445,230
	Predictions between 10m and 20m	10,407	63,020	39,789	14,193
	Predictions greater than 20m	9,259	161,807	13,447	7,443
sloping ground	Median (m)	0.07	1.37	-3.60	-3.80
	Mean (m)	0.86	34.0	-4.91	-6.85
	Minimum (m)	00.0	00.0	-31.8	-88.7
	Maximum (m)	299.2	6105.2	133.6	63.2
	Standard deviation (m)	3.68	147.4	8.39	9.75
	Negative predictions	0	0	727,277	785,709
	Predictions between 10m and 20m	9,747	67,332	12,459	4,882
	Predictions greater than 20m	4,978	183,226	6,618	1,543

Besides the representative measures given in Table 3.6, matrix scatter plots of the predicted data points are generated. The results of the Monte Carlo experiments on free-face mode of Bardet et al. (2002) are shown in Figure 3.4. Similar plots of all other models are given in Appendix A. Note that in these figures, only 10,000 randomly selected data points — out of the one million predicted points — are shown for the sake of better visualization. These plots separate invalid predictions (negative values or values greater than 20 m in magnitude) from valid predictions. They can be used to show the areas of the dataspace where one could expect divergent results. They also show the general co-variability of predictions, particularly in models with no negative predictions.

To clarify this last point, in applying the free-face mode of Bardet et al. (2002), the highest values of horizontal displacements occur at a combination of high values of earthquake moment magnitude and areas found close to the energy source (Figure 3.4). Moreover, a somewhat fewer number of invalid prediction points are estimated at lower values of T_{15}. While these observations are consistent with our understanding of lateral spread behavior, and somewhat intuitive, the results show some models are not following these expected trends. For example, in the free-face mode of Javadi et al. (2006), predictions greater than 20m in the M-R subplot do not follow a general trend. They are distributed randomly over the M-R subplot (refer to Figures C.1 through C.7, Appendix C). This model also tends to predict negative values at lower values of T_{15}, which is nonsensical.

3.4.5 Partial Dependence

The Monte-Carlo analysis discussed in the previous section shows the simultaneous effect of variables on the predictions. In this section, a parametric study is discussed, which is conducted on each of the variables separately, controlling for other variables; three data points are used each time: (a) the mean case, where the average values of data space are used for all variables, (b) a less severe case, where lateral spreads are expected to be larger than the mean case, and (c) a more severe case, where lateral spreads are expected to be smaller than the mean case, according to our current understanding of lateral spreads.

Note that in this simulation, any combination of variables is possible. The more extreme cases could have been considered as "worst- and best-case" scenarios. However, these representative points, which are listed in Table 3.7 and shown in Figures C.1 through C.7 in Appendix C, are chosen to fall onto the areas within the data space in which there are many observed data points. The authors strived to choose these points so they also fall between the 25^{th} and 75^{th} percentiles of the variable limits. Thus, they have good statistical support, and the models are not being extrapolated at these points, so reasonable predictions are expected. The geotechnical variables in the worst-case scenario are similar to the soil condition with high to very high liquefaction susceptibility, described in Honegger et al. (2014): $D50_{15}$ = 0.3mm, F_{15} = 5%, $T15_{475\text{-yr}}$ = 13m.

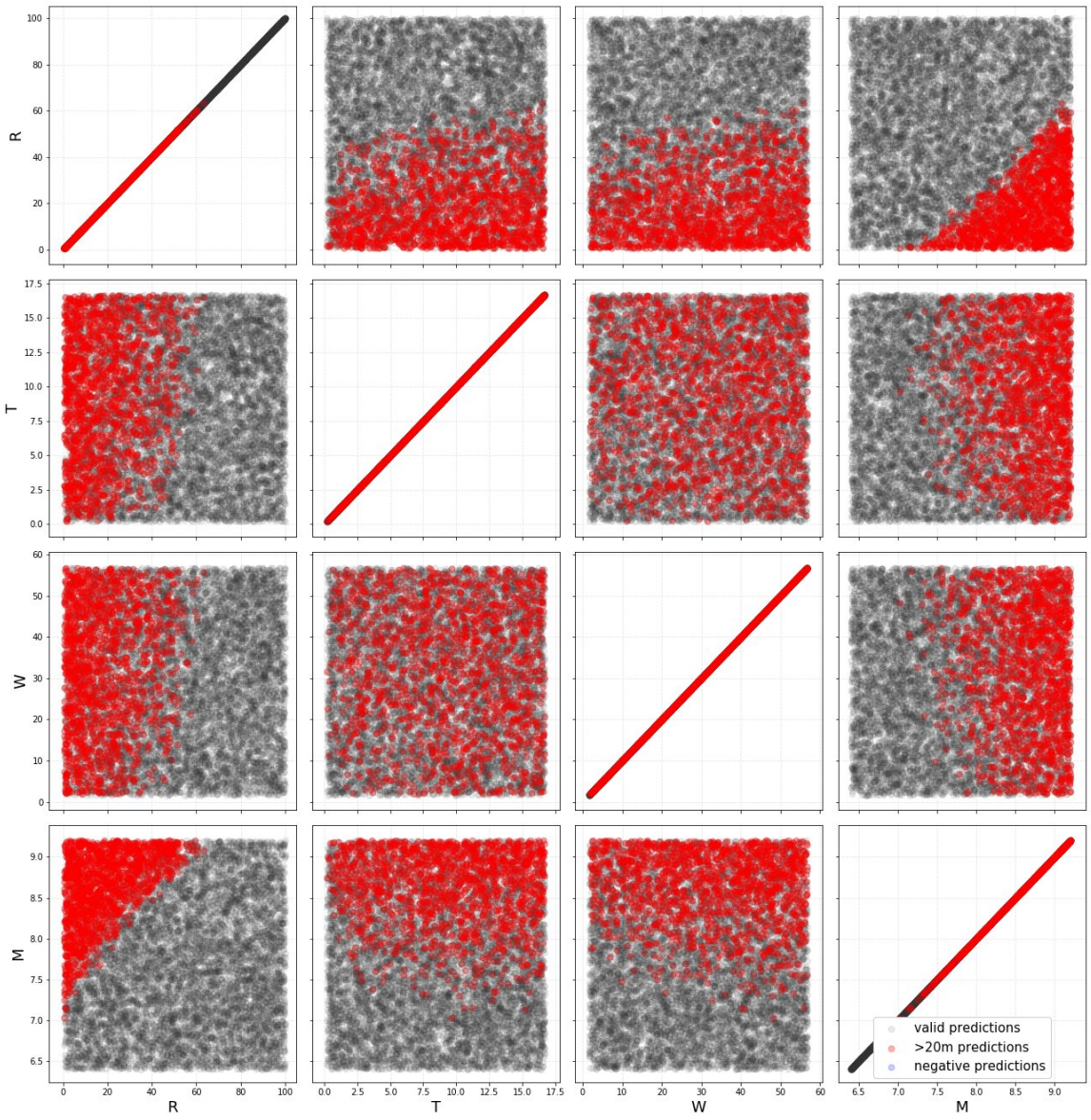


Figure 3.4 Results of Monte Carlo analysis on free-face case of Bardet et al. (2002)

Table 3.7 Data points used in parametric study

Mode	Case	M	R (km)	T (m)	F (%)	D (mm)	W (%)	S (%)
Free face	Mean	7.22	18.39	8.57	17.12	0.36	10.66	-
	More severe	7.50	7.25	12.39	7.00	0.45	15.06	-
	Less severe	6.75	21.0	3.59	24.0	0.12	4.23	-
Sloping ground	Mean	7.52	23.79	6.56	9.36	0.43	-	0.95
	More severe	8.00	7.25	13.0	3.0	0.45	-	3.00
	Less severe	6.75	30.0	2.60	10.25	0.06	-	0.36

The outcome of the parametric study is represented graphically in Figures 3.5 through 3.8 for all four models. These graphs give more useful insights for Javadi et al. (2006) and Rezania et al. (2011) models since their functional forms are generated automatically by the model and are not as interpretable as the other two models due to their cross-correlation between variables; although, the patterns found automatically by the machine-learning models are expected to be similar to those used in the predefined functional form of the regression models.

The variation of predicted horizontal ground displacement with the change in earthquake moment magnitude is exponential in the regression models, while in the machine-learning models, it is more or less linear, if not slightly logarithmic. Unlike others, predictions of the sloping-ground mode of Javadi et al. (2006) does not vary much with the change in the moment magnitude.

The changes in predictions for the distance from the seismic source do not show a similar trend for the free-face and sloping-ground modes in the machine-learning models. An inverse logarithmic form is observed in regression models and the free-face mode of Javadi et al. (2006) model. In contrast, the free-face model of Rezania et al. (2011) is positive logarithmic and predicts higher magnitudes of displacements for longer distances. Moreover, in the machine-learning sloping ground models, the predictions for the best-case scenario exceed that of the mean case at distances greater than about 80 km.

The positive logarithmic form of predictions with respect to the free-face ratio in machine-learning models is verifying the predefined functional form in the regression models. Yet for free-face ratios greater than 15, the Rezania et al. (2011) model is predicting the largest displacements for the best-case and the smallest displacements for the worst-case.

Changes in predictions with respect to the ground slope are positive logarithmic in all models except Javadi et al. (2006) model, where the relation is linear. The sloping-ground model of Javadi et al. (2006) illustrates how cross-correlation of variables might affect the predictions in unreasonable ways. Referring back to Table 3.4, the sloping-ground mode of Javadi et al. (2006) model is

$$D_h = -0.8\ F/M + 0.0014\ F^2 + 0.16\ T + 0.112\ S + 0.04\ S\ T/D - 0.026\ R\ D + 1.14 \tag{3.5}$$

which in our best-case scenario will change into

$$D_h = 0.441472 + 1.84533\ S \tag{3.6}$$

and in our worst-case scenario will be

$$D_h = 2.847775 + 1.26755\ S \tag{3.7}$$

Equations (6) and (7) show why predictions of our best-case scenario are initially smaller than the worst-case scenario due to the smaller constant term and then exceeds it because of its larger coefficient of “S.”

Unlike the regression models, which have a positive logarithmic form for the predictions with respect to the thickness of the liquefiable layer, the free-face model of Javadi et al. (2006) has an exponential form while the sloping-ground model, and both free-face and sloping-ground models of Rezania et al. (2011) have linear forms. Again, the potential for cross-correlation of variables, especially for the freeface model of Rezania et al. (2011), is problematic.

The changes in predictions with respect to the fines content are where the most discrepancy is observed between the models. The machine-learning models diverge as the fine content approaches zero. Moreover, in the sloping-ground model, the automatically recognized negative logarithmic form from Rezania et al. (2011) model conforms to the predefined functional form of the Youd et al. (2002) model. Javadi et al. (2006) model recognized a quadratic form whereby increasing the fines content, the predictions decrease down to a certain level and then increase afterward. On the other hand, in the free-face mode of the machine-learning models, except for the best case of Javadi et al. (2006) model, the ground displacements are predicted to increase initially with increasing fines content and then decrease linearly afterward. The best case of Javadi et al. (2006) is consistent with the negative logarithmic form in Youd et al. (2002). The Bardet et al. (2002) model does not use the fines content.

Despite the model divergence in the average mean grain sizes around zero in the machine-learning models, changes in predictions with respect to the average mean grain size in the machine-learning models, which is negative logarithmic, are is consistent with the functional form used in the regression model of Youd et al. (2002). The Bardet et al. (2002) model does not use the average mean grain size.

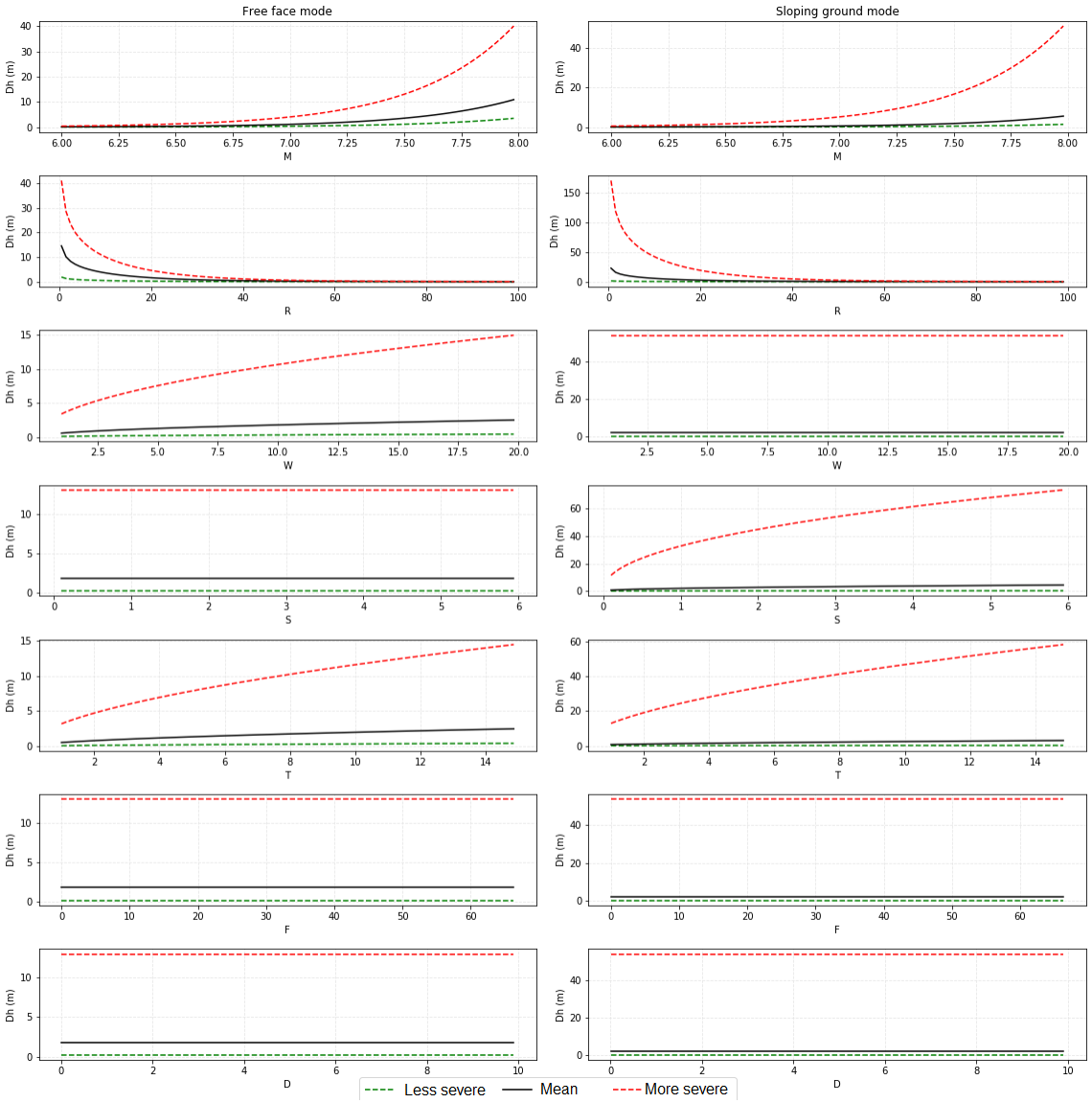


Figure 3.5 Partial Dependence Plot of Bardet et al. (2002) model

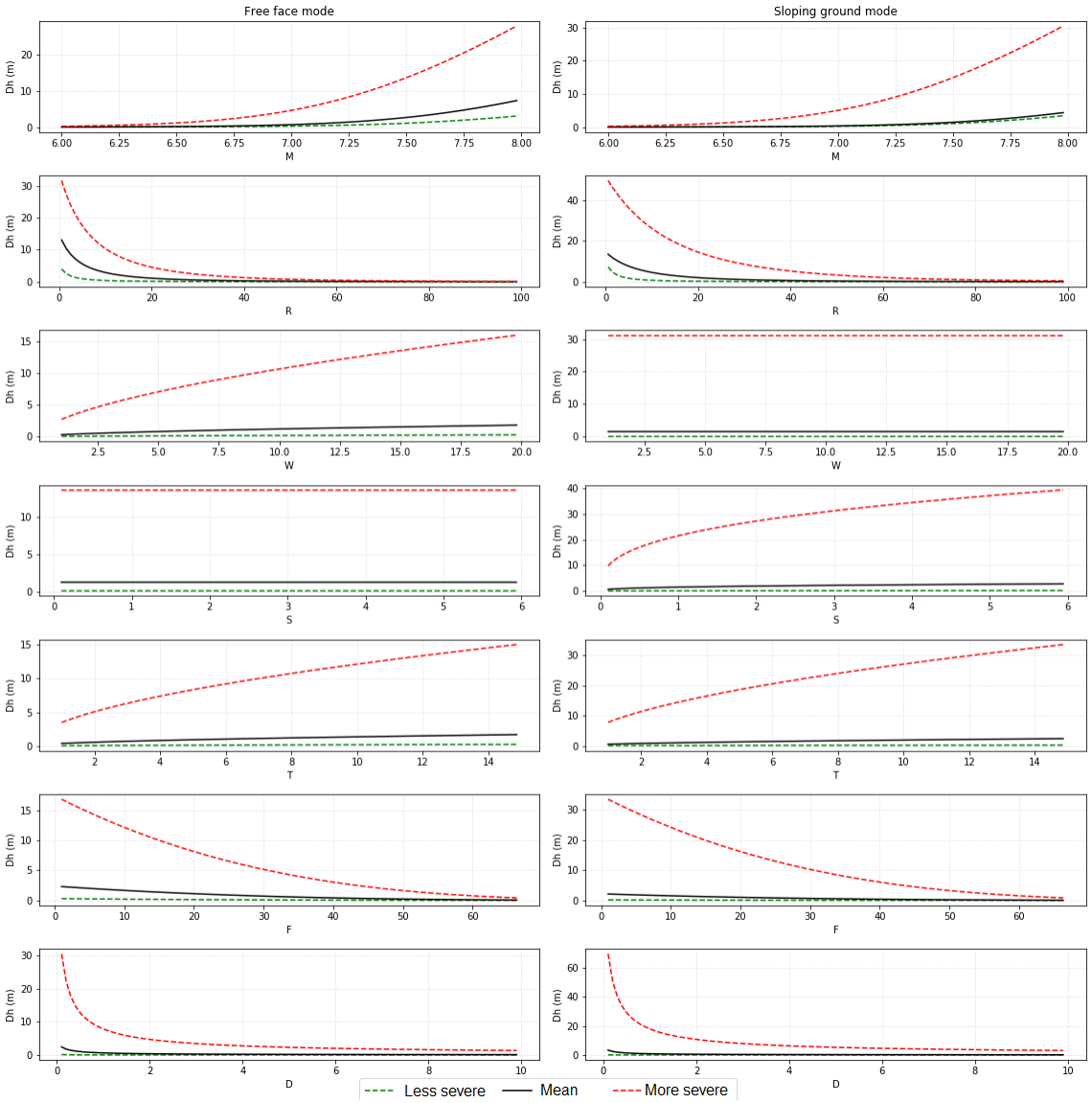


Figure 3.6 Partial Dependence Plot of Youd et al. (2002) model

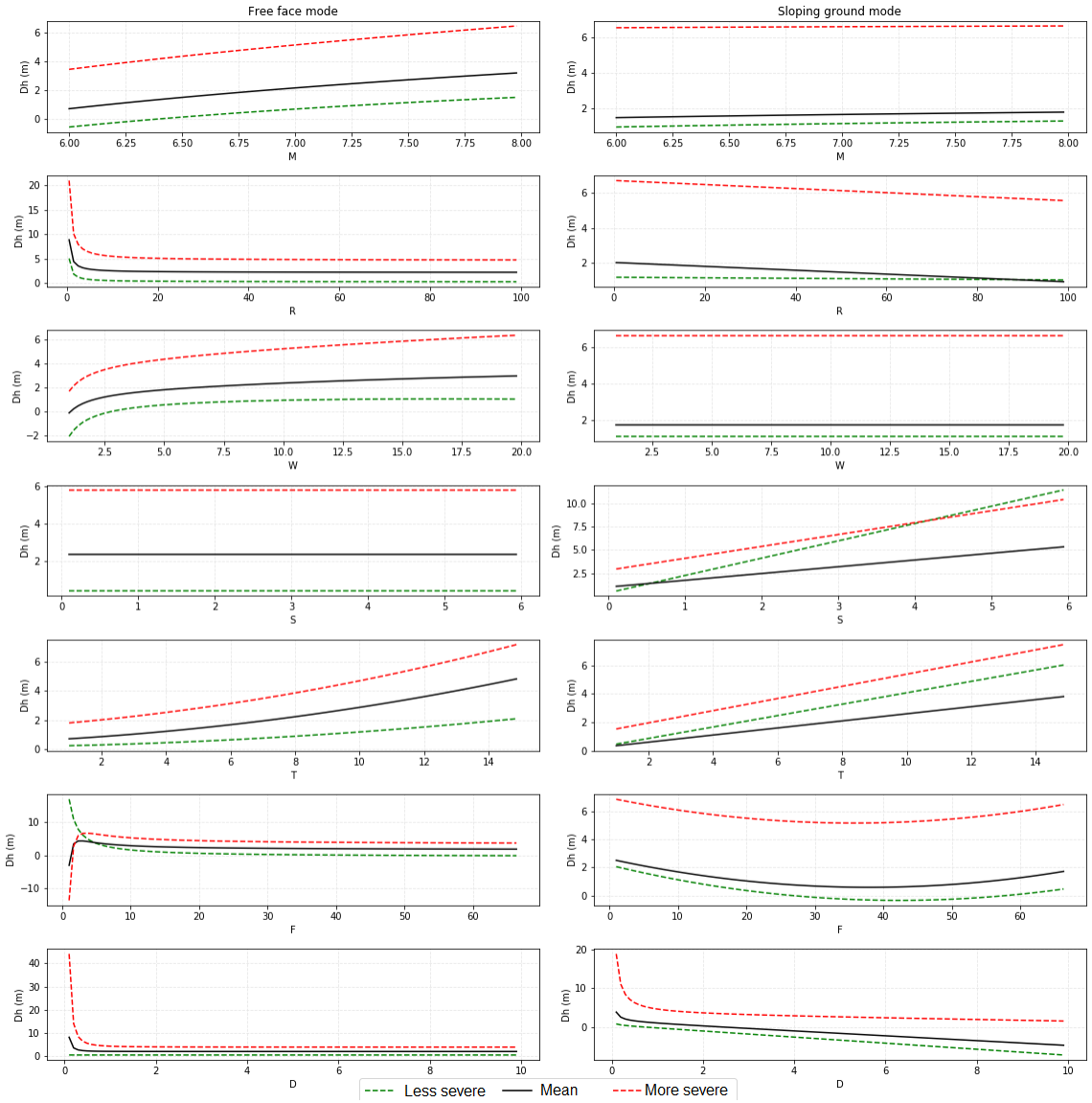


Figure 3.7 Partial Dependence Plot of Javadi et al. (2006) model

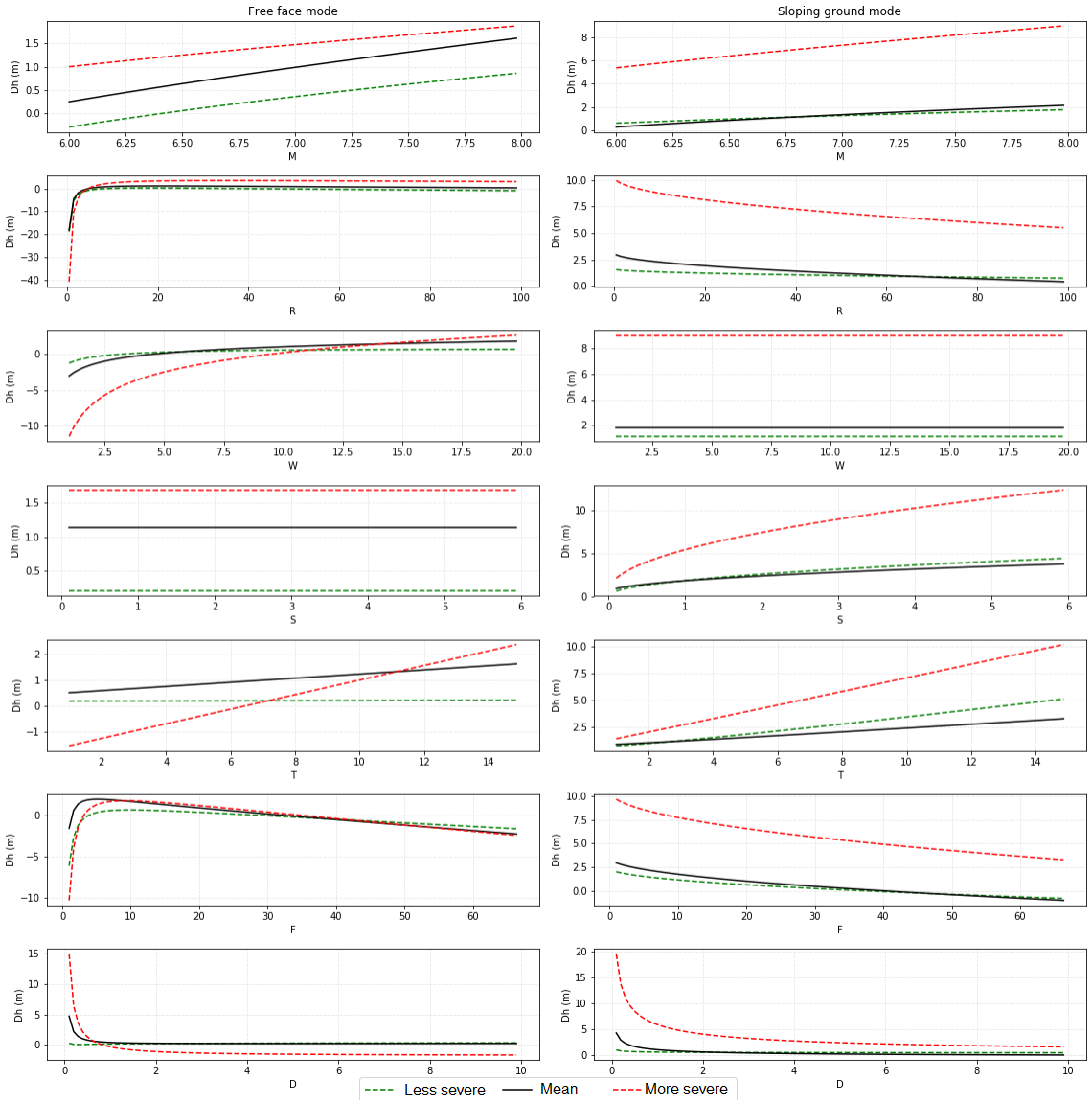


Figure 3.8 Partial Dependence Plot of Rezania et al. (2011) model

3.4.6 Extrapolatability

Unlike section 4.4, as part of the performance assessment, data points selected in this section are chosen such that they fall onto the areas within the data space for which there are not many observed data points. In other words, they lack sufficient statistical support, and the models are an extrapolation at those points. These select points are shown in Figures 3.1 and 3.2 for each of the free-face and sloping-ground modes, respectively. A summary of their exact values, and predictions made by the models under consideration, is given in Table 3.8 where the first four rows correspond to the sloping-ground mode and the other four rows correspond to the free-face mode.

Since these select points are chosen arbitrarily, a comparison between the predictions and the real displacements is not possible. However, the results of Youd et al. (2002) model seem more reasonable compared to other models because on one hand, Bardet et al. (2002) model has diverged at two points and predicted unrealistic values. On the other hand, the machine-learning models had negative predicted values. Other than these points, other values predicted by all four models seem to be in a reasonable range.

Another conclusion that could be drawn from these results is that the results coming from these four models are significantly inconsistent with each other so the average standard deviation of the predictions listed in Table 3.8 is 14.73m. Therefore, caution must be exercised in choosing which model to use for predicting lateral spreads, particularly when extrapolating.

Table 3.8 Select points for extrapolation

Variables							Predictions (m)			
M	R (km)	T (m)	F (%)	D (mm)	S (%)	W (%)	Youd et al. (2002)	Bardet et al. (2002)	Javadi et al. (2006)	Rezania et al. (2011)
8.4	7.3	15.0	10.0	2.3	4.2	-	13.78	173.52	3.86	3.87
7.3	6.0	7.5	17.0	4.0	6.0	-	1.32	11.98	1.38	0.34
8.0	40.0	10.0	20.0	5.0	8.0	-	0.56	6.36	-2.36	-5.53
7.0	20.0	5.0	40.0	0.05	2.0	-	0.34	0.89	7.81	3.42
8.1	9.3	5.5	23.0	0.4	-	25.0	14.07	36.02	2.78	1.60
8.5	60.0	15.5	12.0	1.25	-	35.0	4.40	5.52	9.02	2.59
7.0	45.0	7.5	45.0	1.05	-	40.0	0.03	0.32	1.61	-0.55
8.0	36.0	5.0	37.0	0.7	-	5.5	0.38	1.76	1.65	-0.24

3.4.7 Bayesian uncertainty quantification of MLR model parameters

As discussed earlier, the most commonly used empirical model for liquefaction-induced lateral spread is the MLR model of Youd et al. (2002). It is a multiple linear regression model whose parameters are obtained by a frequentist view of linear regression. The model assumes that the response variable (displacement) is a linear combination of weights multiplied by a set of predictor variables. The functional form also includes an error term to account for random sampling noise or the effect of variables not included in the model. Linear regression is a simple model, which makes it easy to interpret based on the weights (i.e., the regression coefficients). For example, if the weight of a predictive variable is 1.2, then for every unit increase in that variable, the response will increase by 1.2. In frequentist linear regression, the process of training model from training data is to find weights that best explain the data i.e., weights that minimize the residual sum of squares. We obtain from frequentist linear regression a single estimate for the model parameters based only on the training data, which is interpreted as the most likely estimate given the data. However, in case of liquefaction-induced lateral spread where data is not abundant, we might like to express our estimate as a distribution of possible values to quantify the uncertainty of the model input variables. Bayesian Linear regression does that.

3.4.8 Bayesian Linear Regression

In the Bayesian viewpoint, we formulate linear regression using probability distributions rather than point estimates. The response, y, is not estimated as a single value, but rather is assumed to be drawn from a probability distribution. The model for Bayesian Linear Regression with the response sampled from a normal distribution is

$$y \sim N\left(\beta^T X, \sigma^2 I\right) \tag{3.8}$$

where β^T = transposed vector of model parameters; X = vector of input data; I = the identity matrix; and σ=standard deviation of the model parameters.

The output, y, is generated from a normal (i.e., Gaussian) distribution characterized by a mean and variance. The mean for linear regression is the transpose of the weight matrix multiplied by the predictor matrix. The variance is the square of the standard deviation multiplied by the identity matrix because this is a multi-dimensional formulation of the model.

Bayesian Linear Regression aims not to find the single "best" value of the model parameters but rather to determine the posterior distribution for the model parameters. Not only is the response generated from an input probability distribution, but also, the model parameters are assumed to come from a distribution. The posterior probability of the model parameters is conditional upon the training inputs and outputs

$$P(\beta|y,x) = \frac{P(y|\beta,X)P(\beta|X)}{P(y|X)} \tag{3.9}$$

Here, the left side of equation is the posterior probability distribution of the model parameters given the inputs and outputs. Using Bayes Theorem, this is equal to the likelihood of the data, P(y | β, X), multiplied by the prior probability of the parameters, P(β | X) and divided by a normalization constant P(y | X).

In contrast to ordinary least square method, we have a posterior distribution for the model parameters proportional to the likelihood of the data multiplied by the prior probability of the parameters. There are two main benefits to using Bayesian Linear regression:

1. If we have domain knowledge or a guess for the model parameters, we can include them in our model via priors. This is unlike the frequentist approach, which assumes everything there is to know about the parameters comes from the data. For the case where we do not have any estimate ahead of time, we can use non-informative priors for the parameters, such as the normal distribution.
2. The result of performing Bayesian Linear Regression is a distribution of possible model parameters based on the data and the prior. This allows us to quantify our uncertainty about the model.

If we have fewer data points, the posterior distribution will be broader. As the amount of data points increases, the likelihood washes out the prior, and in the case of an infinite dataset, the outputs for the parameters converge to the value obtained from ordinary least square method.

3.4.9 Implementation

In practice, evaluating the posterior distribution for the model parameters is intractable for continuous variables, so we use sampling methods to draw samples from the posterior to approximate the posterior distribution. The technique of drawing random samples from a distribution as an approximation of that distribution is a common application of Monte Carlo methods. There are a number of algorithms for Monte Carlo sampling, the most common being variants of Markov Chain Monte Carlo.

The implementation of Bayesian Linear Regression follows the procedure:

1. Specify priors for the model parameters — in our case, normal distribution.
2. Create a model mapping the training inputs to the training outputs.
3. Use a Markov Chain Monte Carlo (MCMC) algorithm to draw samples from the posterior distribution for the model parameters.

In this study, the Bayesian Linear regression method is implemented in PyMC3. The main advantage of the PyMC3 Python package is that its user-facing features are written in pure Python. PyMC3 leverages Theano (Team et al. 2016) to transparently transcode models to C and compile them to machine code, thereby boosting performance. Theano is a library that allows expressions to be defined using generalized vector data structures called tensors, which are tightly integrated with the popular Numpy,nd array data structure.

3.4.10 Sampling Algorithm

Hamiltonian Monte Carlo (HMC) (Duane, 1987) is a Markov chain Monte Carlo (MCMC) algorithm that avoids the random walk behavior and sensitivity to correlated parameters, which plague many MCMC methods by taking a series of steps informed by first-order gradient information. These features allow it to converge to high-dimensional target distributions much more quickly than simpler methods (e.g., random walk Metropolis or Gibbs sampling). However, the performance of HMC's is highly sensitive to two user-specified parameters: the step size and the desired number of steps L. In particular, if L is too small, then the algorithm exhibits undesirable random walk behavior. While if L is too large, the algorithm wastes computation time. For these reasons, we used a more advanced MCMC sampling technique called No-U-Turn (NUTS; Hoffman, 2014), which is a self-tuning variant of the Hamiltonian Monte Carlo.

3.4.11 Results and Discussion

Figure 3.9 shows approximation of the posterior distributions of model parameters (i.e. coefficients of the multiple linear regression model). These result from the 5,000 steps of Markov Chain Monte Carlo (MCMC), meaning the algorithm drew 5,000 steps from the posterior distribution. Let's compare the mean values of the posterior distributions to those point estimates obtained from the ordinary least square method reported in Youd et al. (2002) (Table 3.9.). We see the results are similar. Since the numbers in Figure 3.9 are rounded to the nearest integer value, a summary of posterior distributions is given in Table 3.10. While we can use a single-point estimate, we also have a range of possible model parameters. As the number of data points increases, this range will shrink and converge to one single value representing greater confidence in the model parameters. This range of values is called the credible interval in Bayesian statistics. It is an interval in a posterior probability distribution domain or a predictive distribution (Edward et al. 1963). Credible intervals are analogous to confidence intervals in frequentist statistics (Lee 1989); however, they have a slightly different interpretation. Unlike frequentist linear regression, we do not get a single value from Bayesian linear regression — we get a distribution.

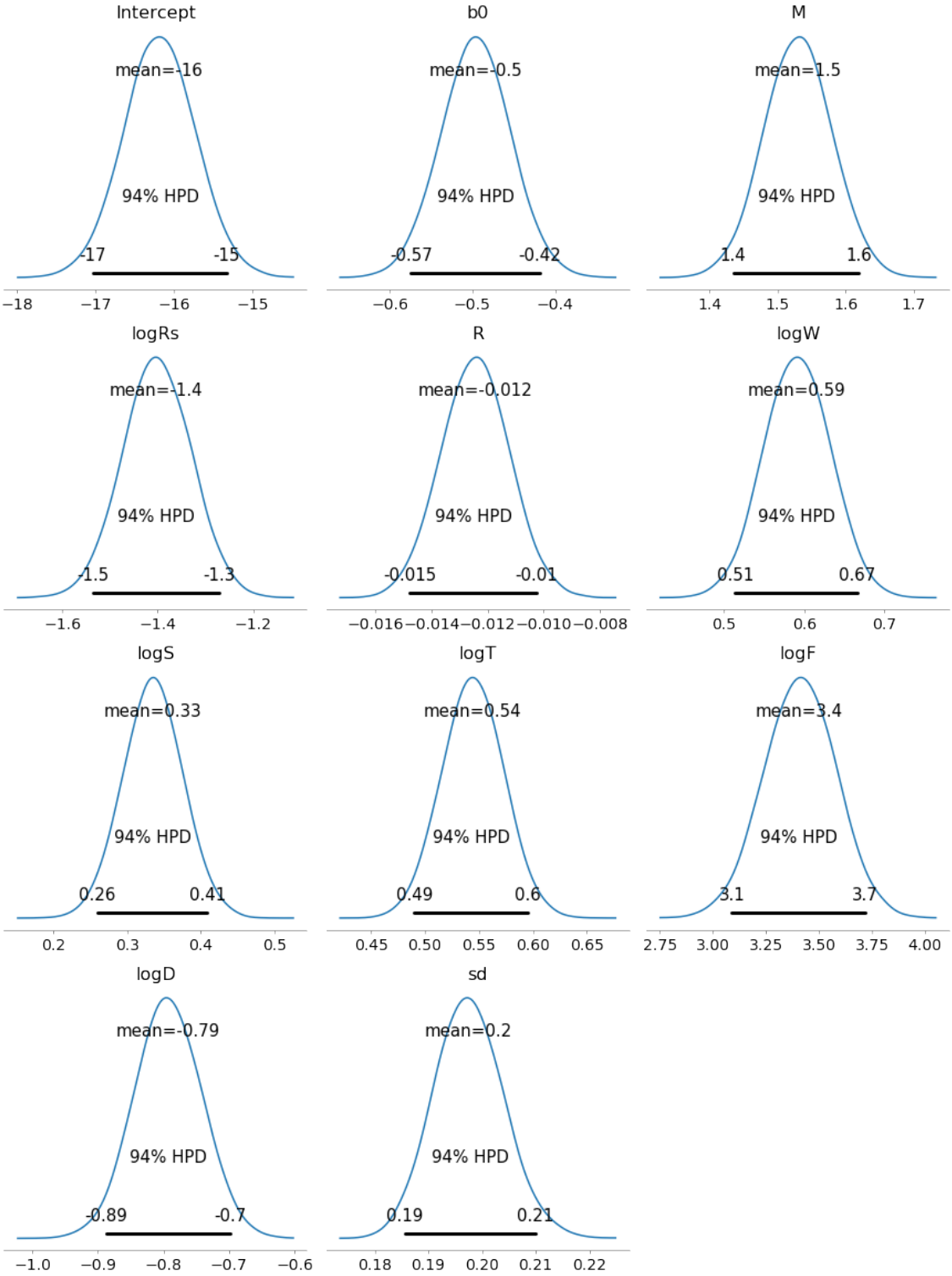


Figure 3.9 Posterior distribution of coefficients of the MLR model

Table 3.9 Point estimate of coefficients of the MLR model

b_0	b_{off}	M	$logR^*$	R	logW	logS	$logT_{15}$	$log(100-F_{15})$	$D50_{15}$
-16.213	-0.5	1.532	-1.406	-0.012	0.592	0.338	0.540	3.413	-0.795

As can be seen in Figure 3.9 and Table 3.10, the standard deviation of the model parameters varies from 0.45 (highest) for intercept to 0.001 (lowest) for R. The higher the standard deviation, the lower the confidence level and vice versa. It should also be noted that standard deviation is in units of the coefficient. For instance, for the intercept coefficient with a mean of -16.19, and standard deviation of 0.45, and given that the posterior distribution has normal distribution, we can expect the intercept coefficient to be roughly 68% of times in the range of (-16.19-0.45, -16.19+0.45), 95% of times in the range of (-16.19-0.9, -16.19+0.9) and 99.7% of times in the range of (-16.19-1.35, -16.19+1.35). These calculations hold for other coefficients, too. The empirical rule for normal distribution states that 99.7% of data observed following a normal distribution lies within three standard deviations of the mean. Similarly, 68% of the data falls within one standard deviation and 95% within two standard deviations from the mean.

Table 3.10 Summary of MCMC traces

	mean	sd	mc_error	hpd_2.5	hpd_97.5	n_eff	Rhat
Intercept	-16.188368	0.451536	0.005157	-17.065917	-15.289376	8994.522519	0.999976
b0	-0.496576	0.041378	0.000402	-0.578811	-0.416822	14841.50956	1.000253
M	1.529197	0.04915	0.000478	1.434469	1.627661	11365.51692	0.999982
logRs	-1.402801	0.070809	0.000738	-1.542889	-1.266819	9546.5402	1.000007
R	-0.012459	0.001195	0.00001	-0.014895	-0.01019	15732.37567	0.999927
logW	0.590545	0.040808	0.000388	0.50977	0.669086	15063.76892	1.000136
logS	0.334146	0.039645	0.000296	0.255991	0.411886	18478.04196	1.000022
logT	0.543828	0.027959	0.000206	0.490367	0.599447	20144.1537	0.999934
logF	3.407928	0.167858	0.001634	3.070137	3.729686	11758.33235	0.999936
logD	-0.793842	0.05048	0.000389	-0.889674	-0.692627	16848.3446	0.99999
sd	0.197484	0.006449	0.000046	0.184662	0.209974	20301.70348	1.000013

Valid inferences from sequences of MCMC samples assume that the samples are derived from the true posterior distribution of interest. Theory guarantees this condition as the number of iterations approaches infinity. Therefore, it is important to determine the minimum number of samples required to ensure a reasonable approximation to the target posterior density. From the trace plots shown in Figure 3.10, we see that the chains have converged. We can verify the convergence of the chains formally using the Gelman Rubin test. The Gelman–Rubin diagnostic evaluates MCMC convergence by analyzing the difference between multiple Markov chains. The convergence is assessed by comparing the estimated between-chains and within-chain variances for each model parameter. Significant differences between these variances indicate non-convergence. See Gelman and Rubin (1992) and Brooks and Gelman (1997) for the detailed description of the method. Values close to 1.0 mean convergence. The results of this test on the

chains of our model are given in Table 3.11. As can be seen, all values are close to 1. We can also test for correlation between samples in the chains.

Table 3.11 Gelman Rubin test results of MCMC chains

Coefficient	Intercept	0		ogRs		ogW	ogS	ogT	ogF	ogD
Gelman Rubin test	0.999	.00	.999	.000	.999	.000	.000	.999	.999	.999

3.5 Conclusion

In this study, the existing empirical models of horizontal ground displacements due to liquefaction-induced lateral spreads are compared to each other from several aspects. The following conclusions are drawn from this study:

1. The coefficients of determination (R^2) reported by Javadi et al. (2006) and Rezania et al. (2011) are invalid and could not be used to compare their models to others. Regardless, while the coefficient of determination could be used as a factor for performance evaluation of empirical models, it fails to describe the entire model and a more comprehensive comparison, such as the one conducted in this paper is required for future models.
2. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are not good measures for comparing predictive models of liquefaction-induced lateral spreads and should not be used.
3. Even though machine-learning models have proven to be promising in recent years, in the domain of liquefaction-induced lateral spread, the existing regression models of Youd et al. (2002) and Bardet et al. (2002) performed better than the existing machine learning models of Javadi et al. (2006) and Rezania et al. (2011). They are less divergent and more consistent with our current understanding of the mechanics behind liquefaction and lateral spreads.
4. The mechanistic-based functional forms of the regression models make them not only less divergent but also more interpretable than the existing machine-learning models; The use of more regularized algorithms is suggested, at least until a less noisy database of liquefaction-induced lateral spread is compiled to reduce the divergence of the machine-learning models developed in future.
5. The unreasonable variations observed in the machine-learning models of Javadi et al. (2006) and Rezania et al. (2011) are due to the cross-correlation between variables in their predictive equations.
6. The results of the existing empirical models are significantly inconsistent with each other in terms of their variations and predicted displacements, particularly when extrapolating.
7. The multiplicative form of the final predictive equations in the regression models of Youd et al. (2002) and Bardet et al. (2002) is proven to be more rational than those with the additive form in the machine-learning models of Javadi et al. (2006) and Rezania et al. (2011). Thus, additive forms of the final predictive equations are less rationale and perhaps, should be avoided in the future development of such models.
8. The log transformation used in the regression models of Youd et al. (2002) and Bardet et al. (2002) prevents negative predictions and is suggested to be used for future model developments. The sloping-ground mode of machine-learning models proposed by Javadi et al. (2006) and Rezania et al. (2011) predicted about 73% and 79% negative values, respectively.

9. Among the models assessed in this study, Bardet et al. (2002) resulted in the most conservative predictions.
10. The uncertainty of the MLR model parameters is essential and should be quantified.

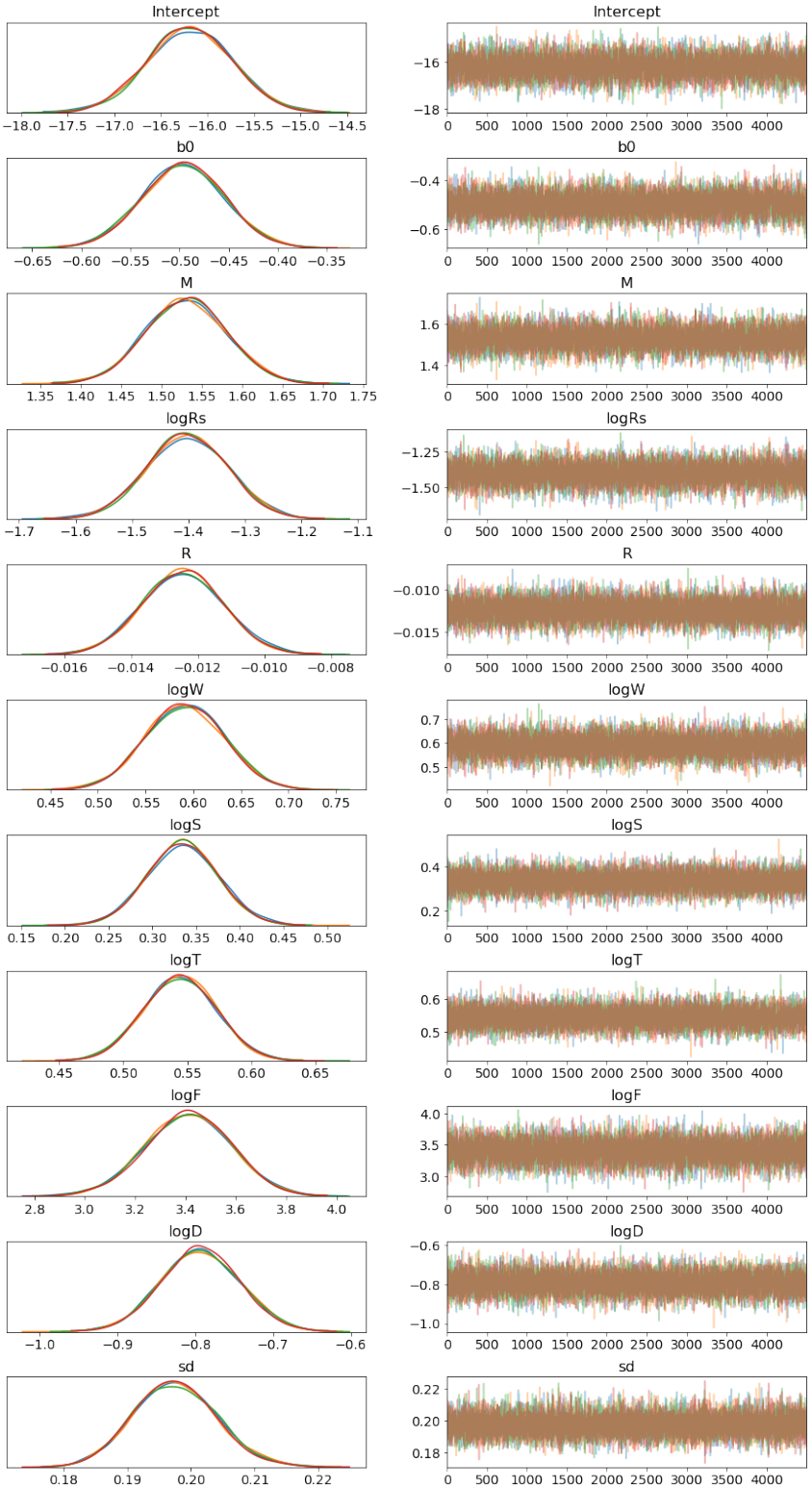


Figure 3.10 Traceplot of MCMC chains

4. LASSO-BASED MODIFICATIONS TO FUNCTIONAL FORM OF THE MLR MODEL

4.1 Introduction

The modeling and estimation of the severity of lateral spread as the most pervasive damage of liquefaction to infrastructures have gained renewed attention during the past few decades. For this purpose, there are various approaches researchers have taken, including 1) simplified analytical methods, such as Newmark's sliding block analysis (Olson and Johnson, 2008) and Minimum potential energy method (Towhata et al., 1992), 2) numerical methods (e.g., Finn et al., 1994), 3) empirical relations based on case histories, such as regression models (Youd et al., 2002; Bardet et al., 2002) and machine learning-based models (Rezania et al., 2011; Baziar and Saeedi Azizkandi, 2013), and 4) laboratory studies (e.g., Kuwano and Ishihara, 1988).

Each of these approaches has advantages and disadvantages. For example, while empirical models have gained extensive use in practice due to their simplicity, their accuracy is subject to the quality of the database on which they were derived, regardless of the modeling technique. A comprehensive comparison of empirical models for liquefaction-induced lateral spread is presented in Chapter 3. We found that considering the high volume of noise in the existing lateral spread case history database(s), having a predefined functional form plays an important role and is vital unless a less noisy database becomes available in the future. The choice of the functional form directly affects the performance of the model, particularly when extrapolating. The functional forms of the independent variables could come from either the mechanics governing the lateral spread phenomenon (i.e., physics-based models) or the trends observed in the database (empirical or machine-learning models), or a combination of both.

One of the most commonly used empirical relations is the Multiple Linear Regression (MLR) model initially proposed by Bartlett and Youd (1995, 1992), which was later modified by Youd et al. (2002) by adding data from three additional earthquakes and changing some terms of the functional form. It is clear that empirical equations must be revised as our understanding of liquefaction and lateral spread grows and as newer case histories become available (Youd 2018). The discrepancy between variations of different empirical models discussed in Chapter 3 indicates the need for re-evaluating the functional form used in the revised MLR model by Youd et al. (2002). This chapter proposes modifications to the functional form of the revised MLR model based on Least Absolute Shrinkage and Selection Operator (LASSO) regression technique, which is consistent with our understanding of the mechanics behind lateral spread the trends observed in the case histories database.

4.2 Least Absolute Shrinkage and Selection Operator

Least absolute shrinkage and selection operator (LASSO) is a regression analysis method in statistics and machine learning that performs both feature selection and regularization. LASSO regression was first introduced in geophysics by Santosa and Symes (1986). A few years later, this technique was independently rediscovered and popularized by Tibshirani (1996). Consider the usual linear regression of data (X^i , y_i), $i = 1, 2, 3...N$, where $X_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ and y_i are the regressors and response for the ith observation. The ordinary least squares (OLS) estimates α and $\beta = (\beta_1, \beta_2, ..., \beta p)^T$ are obtained by minimizing the residual squared error, i.e.

$$(\hat{\alpha}, \hat{\beta}) = argmin\{\sum_{i=1}^{N}\left(y_i - \alpha - \sum_j \beta_j x_{ij}\right)^2\} \quad (4.1)$$

There are mainly two reasons why the result of the OLS estimates are not satisfactory. First, the OLS estimates usually have a large variance despite their low bias. One way to improve prediction accuracy would be to set to 0 some coefficients to reduce the variance of the predicted values at the cost of introducing a little bias. The second reason is interpretation. With a large number of predictors, we often would like to determine a smaller subset that exhibits the strongest effects. The LASSO regression shrinks some coefficients and set others to zero and finds α and β by minimizing Eq. 4.1 subjected to a constraint on sum of absolute values of βs (Tibshirani, 1996), i.e.

$$(\hat{\alpha}, \hat{\beta}) = argmin\{\sum_{i=1}^{N}\left(y_i - \alpha - \sum_j \beta_j x_{ij}\right)^2\} \quad subject\ to \quad \sum_j |\beta_j| \le t \quad (4.2)$$

where $t \geq 0$ is a tuning parameter. Eq. 4.2 can be further decomposed into

$$(\hat{\alpha}, \hat{\beta}) = argmin\{\sum_{i=1}^{N}\frac{1}{2}\left(y_i - \alpha - \sum_j \beta_j x_{ij}\right)^2 + \lambda\sum_j |\beta_j|\} \quad (4.3)$$

where the parameter λ is predetermined and controls the amount of regularization. The higher the value of λ, the more elements of the estimated β vector are set to zero, and the more the nonzero entries are shrunken toward zero. Smaller λ implies less regularization and more nonzero β with larger (absolute) coefficients (Reid et al., 2016).

4.3 LASSO-Based Modifications to the Revised MLR Model

The revised MLR model by Youd et al. (2002) consists of two equations. One for the free-face mode

$$log(Dh_{ff}) = -16.713 + 1.532M - 1.406\log R^* - 0.012\,R + 0.592\,logW + 0.540\,logT_{15} + 3.413log100 - F_{15} - 0.795log(D50_{15} + 0.1) \quad (4.4)$$

and one for the sloping ground mode

$$log(Dh_{sg}) = -16.213 + 1.532M - 1.406logR^* - 0.012R + 0.338logS + 0.540log(T_{15}) + 3.413\log(100 - F_{15}) - 0.795\log(D50_{15} + 0.1) \quad (4.5)$$

where D_h is the magnitude of ground displacement in meters; M is moment magnitude; R is the horizontal or mapped distance from the site in question to the nearest bound of the seismic energy source; $R^* = 10^{0.89M-5.64} + R$; S is the slope in percent; W is the free face ratio in percent; T_{15} is the thickness of saturated cohesionless soil with $(N_1)_{60} < 15$, in meters; F_{15} is average fines content in T_{15}, in percent; and $D50_{15}$ is average D_{50} in T_{15} in millimeters. All of the logarithms are to base 10.

As can be seen in Eqs. 4.4 and 4.5, the set of predictive equations is made of a combination of seismic (M, R), geotechnical (T_{15}, F_{15}, and $D50_{15}$), and topographic variables (W and S) and their logarithms. The variable selection and decision of taking the logarithm of variables or using them directly in the functional form of the revised MLR model are mainly based on the observations of lateral spreads obtained from the research of Youd and Perkins (1978) and Hamada et al. (1986). Some choices, such as using T15 instead of the thickness of the liquefiable layer, were made based on the high statistical correlation between these predictive variables and the dependent variable.

In this study, we took advantage of the automatic variable selection of LASSO regression by regressing on 12 variables, including M, R, T_{15}, $D50_{15}$, F_{15}, and W (free face mode) or S (sloping ground mode), plus their logarithms to base ten including log(M), log(R), log(T_{15}), log($D50_{15}$ + 0.1), log(100 − F_{15}) and log(W) or log(S). Choice of the parameter λ in Eq. 4.3 controls the complexity of the model by determining the number of variables, which consequently affects the accuracy. Figure 4.1 shows the effect of λ on the free face model. As can be seen, as λ grows, the coefficient of more variables are set to zero, and the coefficient of determination, R^2, reduces. Based on these results, a λ = 0.00199 was chosen to get the highest coefficient of determination with the least number of variables.

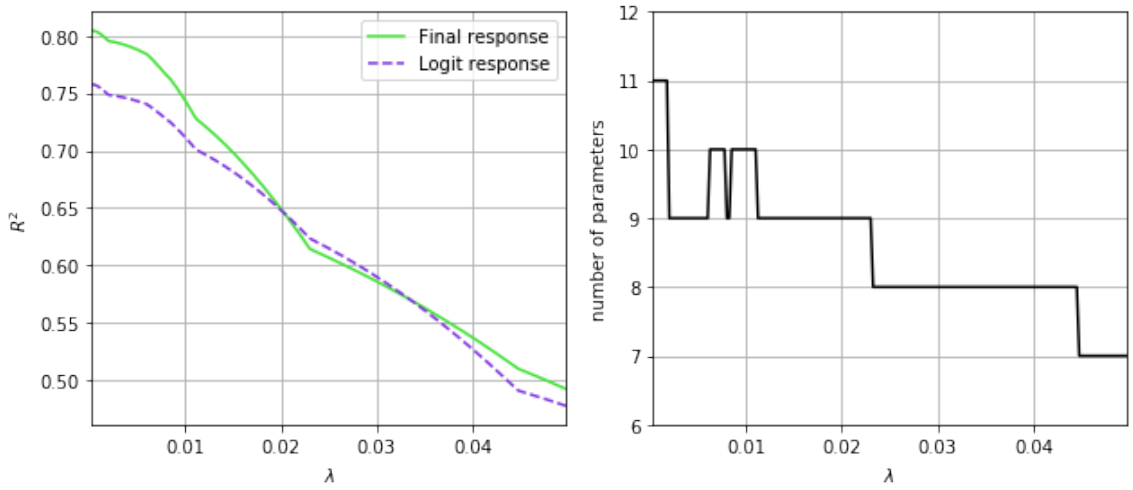


Figure 4.1 Effect of λ on the free face model

After this automatic variable selection, M, R, log(R), T_{15}, F_{15}, log($D50_{15}$ + 0.1), and log(W) or log(S) variables had non-zero coefficients i.e., they are the ones that LASSO regression has selected. Therefore, it is recommended that the functional form of the revised MLR be changed to

$$log(Dh) = f(M, R, W, S, T_{15}, D50_{15}, F_{15})$$

$$= \alpha + \beta_1 M + \beta_2 R + \beta_3 log(R) + \beta_4 T_{15} + \beta_5 F_{15} + \beta_6 log(D50_{15} + 0.1) + \beta_7 log(t) \quad (4.6)$$

where *t* is the topographic variable and equal to *W* in free face mode and *S* in sloping ground mode.

4.4 Capping Concept

One common problem of most predictive empirical equations for lateral spread found by Hosseinali and Bartlett (2019) is the unrealistically large displacements at extreme values of variables, such as very close distances to the energy source (when R approaches zero). Youd et al. (2002) tried to mitigate this problem of the original MLR model from Bardet et al. (2002) and Bartlett and Youd (1995) by replacing the term log(R) with log(R^*). Their efforts, however, only mitigates the problem associated with small values of R to some extent. As shown in Chapter 3, even their model, which has the lowest maximum predictions among all other empirical models of lateral spreads, could predict as large as 50 m of displacement in extreme cases. Chapter 3 concluded that the same issue with low values of R could also be found in high values of T15, W, S, and M and low values of D5015 and F15.

The diverging behavior of the predictive equations is due to the log transformation of horizontal displacements, D_H. Eqs. 4.4 and 4.5 could be re-written as Eq. 4.7 where the left-hand side would diverge as the right-hand side of the equation gets large.

$$log(D_H) = f(M, R, W, S, T_{15}, D50_{15}, F_{15}) \tag{4.7}$$

To resolve this issue, we propose the use of a logit function instead of the log function and regress on the following functional form

$$log\left(\frac{Dh}{\gamma - Dh}\right) = f(M, R, W, S, T_{15}, D50_{15}, F_{15}) \tag{4.8}$$

where γ is an arbitrary parameter, which defines the cap of regression model and should be set to 10, since magnitudes of lateral displacements larger than 10 meters are seldom seen in the case history database and might be considered flow failure may have a different mechanism (i.e., formulation of thin films of water with extremely low shear resistance). Note that the maximum horizontal displacement recorded in the database compiled by Youd et al. (2002) is 10.15 m from the 1964 Niigata, Japan Earthquake for free-face failure into the Shinano River (Bartlett and Youd, 1992). In their database, there are only four data points with displacements greater than 9m. The difference between the use of logit function (Eq. 4.8) and log function (Eq. 4.7) is shown graphically in Figure 4.2, which clearly demonstrates how using logit function could help mitigate the divergence problem at extreme values of predictive variables leading to large values of $f(M, R, W, S, T_{15}, D50_{15}, F_{15})$.

Based on the concept of capping described in this section, Eq. 4.6 could be re-written as Eq. 4.9, which is the final proposed functional form. The term $log(D50_{15})$ is changed to $log(D50_{15} + 0.1)$ because in case of $D50_{15} = 0$, $log(0)$ is undefined.

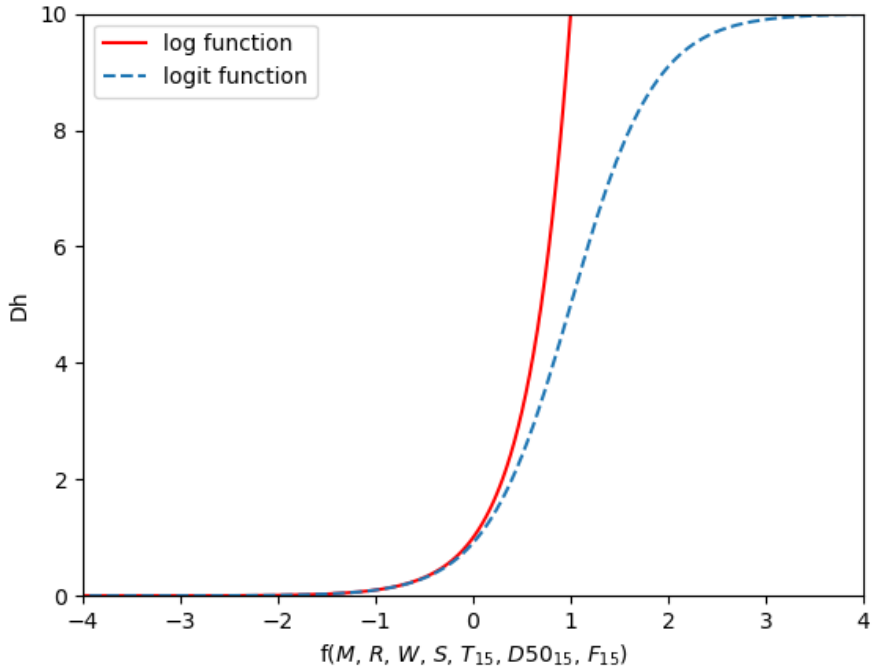


Figure 4.2 Variations of expected ground displacement with predictive variables

$$\begin{aligned} log\left(\frac{Dh}{10-Dh}\right) &= f(M,R,W,S,T_{15},D50_{15},F_{15}) \\ &= \alpha + \beta_1 M + \beta_2 R + \beta_3 log(R) + \beta_4 T_{15} + \beta_5 F_{15} \\ &\quad + \beta_6 log(D50_{15} + 0.1) + \beta_7 log(t) \end{aligned} \tag{4.9}$$

4.5 Results and Discussion

The MLR model was re-regressed on the database compiled by Youd et al. (2002) using the final functional form given in Eq. 4.9 and the resulting set of predictive equations are given in Eqs. 4.10 and 4.11 for free-face and sloping-ground modes, respectively.

$$\begin{aligned} log\left(\frac{Dh}{10-Dh}\right) &= -9.729 + 1.242M - 0.0122R - 1.069log(R) \\ &\quad + 0.0644T_{15} - 0.0387F_{15} \\ &\quad - 1.633log(D50_{15} + 0.1) + 0.985log(W) \end{aligned} \tag{4.10}$$

$$\begin{aligned} log\left(\frac{Dh}{10-Dh}\right) &= -6.262 + 0.882M - 0.0104R - 1.021log(R) \\ &\quad + 0.07049T_{15} + 0.0336F_{15} \\ &\quad - 1.044log(D50_{15} + 0.1) + 0.327log(S) \end{aligned} \tag{4.11}$$

Predicted displacements versus measured values of the MLR models of this study are plotted in Figure 4.3. Compared to the MLR model proposed by Youd et al. (2002), the results show an increase in regression coefficient of determination, R^2, of 12.3% (from 0.701 to 0.796) and 9.1% (from 0.536 to 0.585) for free-face and sloping ground modes, respectively. R_2 is calculated using

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{4.12}$$

where y_i are measured values; f_i are predicted values; and $\bar{y}$ is the mean of measured values.

The discussed modifications to the MLR model increased its predictability and is very consistent with our understanding behind the field behavior of lateral spread and modifications that other researchers have proposed. For example, the change of functional form from $D50_{15}$ to $log(D50_{15} + 0.1)$, which is one of the revisions made by Youd et al.

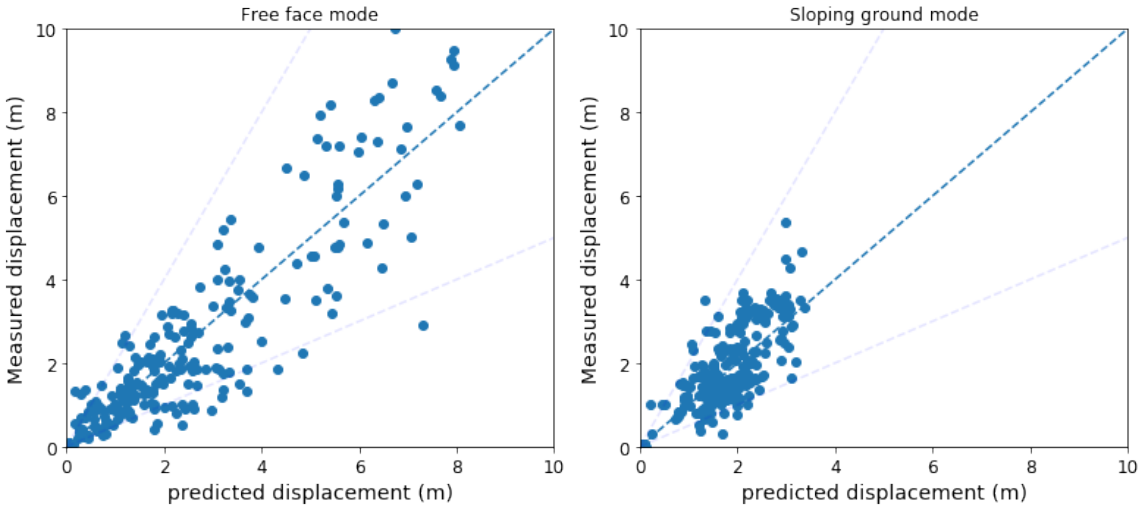


Figure 4.3 Measured versus predicted displacement using the proposed MLR model

(2002) to the original MLR model proposed by Bartlett and Youd (1992) and Bartlett and Youd (1995) to reduce models' sensitivity to mean-grain size, was detected in our automatic LASSO-based variable selection.

caveat of automatic variable selection is the potential for overfitting. Although the built-in regularization of the LASSO regression technique to some extent prevents overfitting, it is extremely important to investigate models' behavior from aspects other than simple summary statistics, such as R^2 (e.g., variations of target variable with respect to predictive variables) and check models compatibility with results obtained from independent researchers.

Figure 4.4 shows partial dependence plots (PDP) of the proposed model for the following three cases:

- The mean case, where the average values of data space are used for all variables (solid black lines).
- The best case, where large lateral spreads are highly likely to occur (dashed green lines).
- The worst case, where large lateral spreads are highly unlikely (dotted red lines).

The values of variables chosen for the best- and worst-case scenarios are listed in Table 4.1. A more comprehensive discussion on how these values are obtained can be found in Chapter 3.

Table 4.1 Values used in mean-, best- and worst-case scenarios

Mode	Case	M	R (km)	T (m)	F (%)	D (mm)	W (%)	S (%)
Free face	Mean	7.22	18.39	8.57	17.12	0.36	10.66	-
	Worst	7.50	7.25	12.39	7.00	0.45	15.06	-
	Best	6.75	21.00	3.59	24.00	0.12	4.23	-
Sloping ground	Mean	7.52	23.79	6.56	9.36	0.43	-	0.95
	Worst	8.00	7.25	13.00	3.00	0.45	-	3.00
	Best	6.75	30.00	2.60	10.25	0.06	-	0.36

4.6 Conclusion

A new set of multiple linear regressions for liquefaction-induced lateral spread is proposed in this study with some modifications to the functional form of Youd et al. (2002). This new functional form takes advantage of the logit transformation of the target variable, rather than the log function, to cap the target variable and therefore, prevents unrealistically large predictions. On the other hand, by using Least Absolute Shrinkage and Selection Operator regression technique, it was found that replacing R^* with R, replacing $\log(F_{15})$ with F_{15}, and replacing $\log(T_{15})$ with T_{15} result in an MLR model that has 12.3% and 9.1% higher coefficient of determination between predictions and measured displacements in the free face and sloping ground modes, respectively. Compared to the MLR model proposed by Youd et al. (2002), the variations of our proposed model are also more compatible with experimental centrifuge tests, numerical and empirical models from other researchers. Moreover, the regularization of the LASSO technique ensured there was not over fitting.

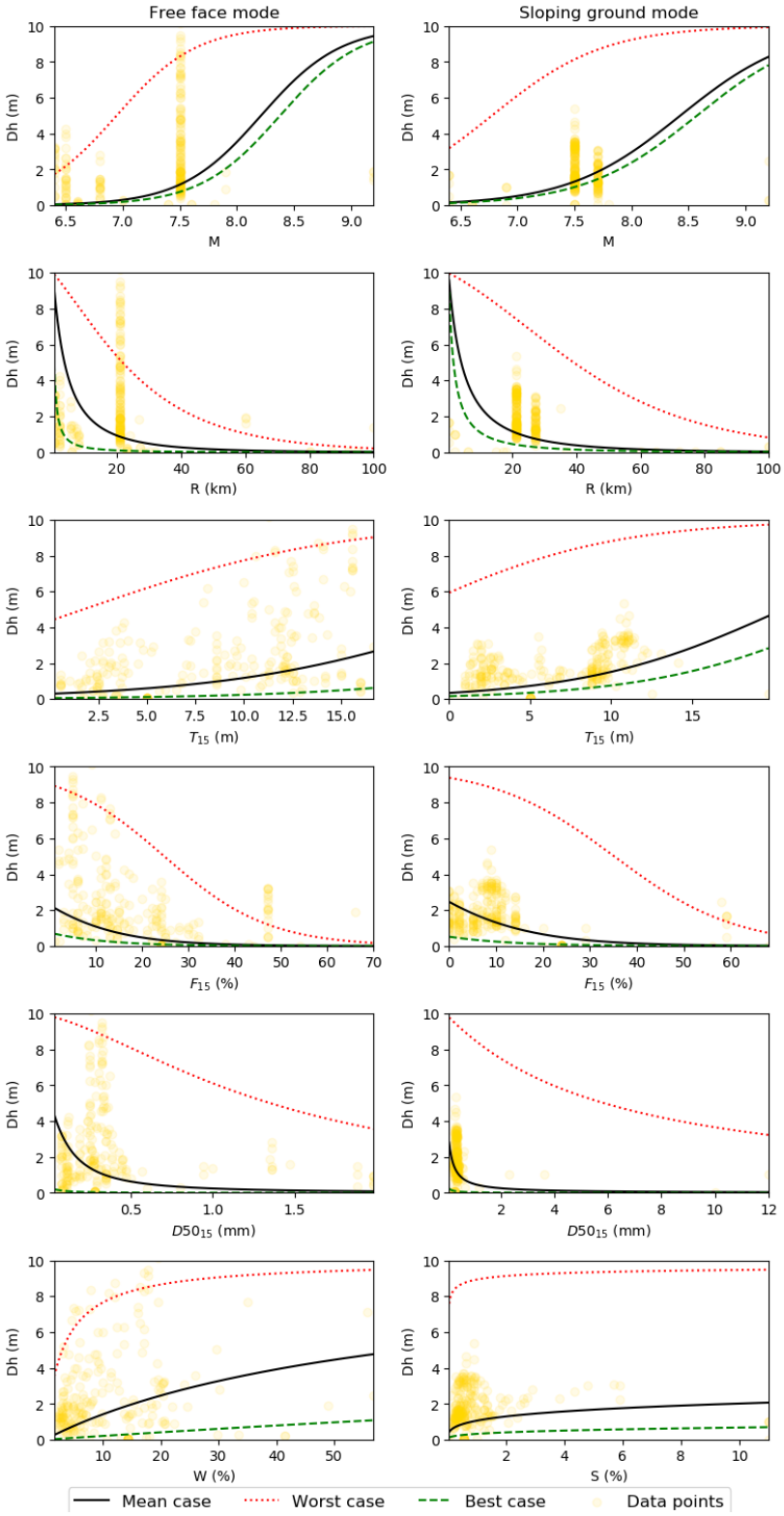


Figure 4.4 Partial dependence plots of proposed LASSO-based model

5. CONVOLUTIONAL NEURAL NETWORK-BASED SCREENING CRITERIA FOR LATERAL SPREAD

5.1 Introduction

One of the most common surface manifestations of liquefaction is lateral spread displacement. This type of failure occurs when blocks of relatively intact soil move downslope atop a liquefied soil layer. Bartlett and Youd (1992a) classified lateral spread failures as either free face (i.e., where the ground moves toward an incised river channel or other types of abrupt topographical depression) or a ground slope (ground movement down gentle sloping topography). Pervasive soil liquefaction is a requirement to generate lateral spread. However, the opposite is not necessarily true; having liquefaction does not necessarily translate into having lateral spread but rather, it means there is a chance for lateral spread. This often-neglected fact is a noticeable gap in the existing empirical equations for predicting liquefaction-induced lateral spread.

Liao (1988), Cetin et al. (2002), and later researchers have proposed empirical models that can be used to estimate the probability of triggering liquefaction. In addition, empirical equations for predicting the amount of lateral spread displacement have been developed primarily using Multiple Linear Regression (MLR) and other techniques (Bartlett and Youd, 1992a, Bartlett and Youd, 1995, Youd et al. 2002).

However, the triggering of liquefaction does not necessarily generate lateral spread displacement. Nonetheless, some engineering practitioners have mistakenly assumed that triggering liquefaction will also result in the development of lateral spread and have inappropriately applied the MLR lateral spread displacement predictive equations. This misapplication is equivalent to implying that the probability of generating lateral spread displacement given liquefaction is unity, which is false. Thus, we believe there is a notable gap in the existing hazard formulations regarding this issue.

In addition, many practitioners have extrapolated the lateral spread displacement models using sparse data or input data beyond the conditions or data bounds represented in the case history dataset (i.e., extrapolation of the model). Youd (2018) lists some typical misuses of the Youd et al. (2002) MLR model.

- Inferring critical layer continuity when it may not exist.
- Extrapolating model to relatively thin layers, i.e., layers less than one meter.
- Using the MLR equations in predominately non-plastic silts, etc.

As a possible solution to this problem, we propose a probabilistic “screening” or probabilistic lateral spread “susceptibility” model. This model will fill a missing gap in the conditional probably chain, as discussed later.

5.2 Theoretical Background

Many factors contribute to lateral spread susceptibility, including but not limited to

- Soil properties, such as fines contents, plasticity index, soil index, layer thickness, soil density, depth and lateral extent of the critical layer, and relative geologic susceptibility (i.e., geologic age and depositional environment).
- Seismic factors. including intensity and duration of shaking, seismic source distance, and earthquake moment magnitude.

- Topographical and geometrical factors including ground slope, proximity to a free-face, etc.
- Other possible seismic, site, soil, and topographic variables.

The factors listed above are candidates for developing a lateral spread susceptibility model; therefore, it is vital to understand how they influence lateral spread susceptibility.

Influence of Fines Content – The fines content is that percentage of the soil distribution with grain size finer than 0.075 mm. Field case histories indicate that fine-grained sediments, such as those beneath Adapazari, Turkey, although susceptible to liquefaction, did not generate lateral spread. Also, non-sensitive, clay-like soils appear to be immune to lateral spread. Also, the case history dataset suggests that lateral displacement decreases markedly with increasing fines content (Bartlett and Youd 1995, Youd et al., 2002, Youd 2018). This finding needs additional data support, definition, and research.

Influence of Plasticity Index (PI) – The plasticity index (ASTM D4943) is the liquid limit of the soil minus the plastic limit of the soil, as defined by Atterberg. Monotonic and cyclic undrained loading test data for silts and clays show that they transition, over a fairly narrow range of plasticity indices, from soils that behave more fundamentally like sands (sand-like behavior) to soils that behave more fundamentally like clays (clay-like behavior). Boulanger and Idriss (2006) propose that for practical purposes, clay-like behavior is expected for fine-grained soils with a plastic index equal to or greater than 7. Bray and Sancio (2006), using water in situ water content (wc), concluded that loose soils with PI < 12 and wc / LL> 0.85 were susceptible to liquefaction. However, loose soils with 12 < PI < 18, and wc / LL > 0.8 were systematically more resistant to liquefaction. Soils with PI > 18 tested at low effective confining stresses were not susceptible to liquefaction.

Influence of Soil Index (SI) – Gillins and Bartlett (2013) found that the soil classification obtained from borehole logs could supplant the use of fines content and mean grain size in the Youd et al. 2002 MLR model. They develop a soil-type factor based on soil classification called the soil index, SI. The SI was used instead of fines content and mean grain size because these laboratory measurements are often missing in many borehole logs. They show how SI might replace F_{15} and $D50_{15}$ for regional hazard-mapping studies. The SI is also used in the logit analyses herein.

Influence of Layer Thickness (T) – Bartlett and Youd (1992a, 1995) showed that the cumulative thickness of the loose, saturated, sandy deposits influences lateral spread and the resulting amount of horizontal displacement. They defined layer thickness factors (i.e., independent variables) in their MLR analysis that accounted for thickness. These factors are T_{10}, T_{15}, and T_{20}, where T is the cumulative thickness of saturated, granular deposits with $N_{1,60}$ values less than 10, 15, and 20, respectively. Bartlett and Youd (1992a, 1995) were careful not to infer that these independent variables represented the thickness of the "liquefied zone." Instead, they were introduced in their evaluation simply as soil factors correlated with lateral spread displacement, hence valuable without the need to perform liquefaction analysis procedures. The thinnest T_{15} layer in the Bartlett and Youd (1995) dataset associated with lateral spread occurrence was about 1.0 m. The thinnest layer observed in CPT data is about 0.6 m (Youd, 2018). Hence, it appears that layers with a thickness less than this are either not continuous across the site or do not have a sufficient thickness to generate sufficient water migration to induce lateral spread displacement (Bartlett and Youd, 1992a).

Influence of Soil Density (D) – From their MLR database, Bartlett and Youd (1992a) and Youd et al. (2002) concluded that sediments susceptible to lateral spread generally have SPT N160 values less than 10 and almost always have SPT $N_{1,60}$ less than 15. Nonetheless, the influence of soil density (i.e., SPT N values) on the probability of lateral spread occurrence will be explored using logistic analyses.

Influence of Depth (Z) – Bartlett and Youd (1992a) found that the depth to the critical zone, defined as the lowest $N_{1,60}$ value in saturated, granular deposits, was almost always in the upper 15m of the soil profile. Therefore, the influence of depth could also be significant.

Influence of Seismic Source Distance (R) – For similar magnitude earthquakes, liquefaction effects are known to attenuate with decreasing seismic energy associated with further distances from the seismic source (Youd and Perkins, 1978, 1987; Ambraseys, 1988; Bartlett and Youd, 1992a, 1995; Youd et al., 2002).

Influence of Earthquake Magnitude (M_w) – At susceptible sites and all other factors being equal, the occurrence of lateral spread and the magnitude of the associated displacement increases with earthquake magnitude (Bartlett and Youd, 1992a; 1995).

5.3 Probabilistic Framework

The conventional probability chain for lateral spread magnitude (D_H) exceeding a certain threshold (y) can be expressed as

$$P(D_H > y) = P(L)P(D_H > y|L) \tag{5.1}$$

where P(L) = probability of liquefaction; and $P(D_H > y \mid L)$ = probability of lateral spread magnitude exceeding threshold y given liquefaction. These probabilities could be obtained using the existing probabilistic empirical equations for liquefaction and lateral spread in the literature, such as Cetin et al. (2004) and Youd et al. (2002) if and only if the application requirements are met. This consideration is important for the second term on the right-hand side of Eq. (5.1). The regression equation proposed by Bartlett and Youd (1995), which was later revised by Youd et al. (2002), is a deterministic linear regression model and has the following form

$$log D_H = \sum_i b_i V_i + \varepsilon \tag{5.2}$$

where b_i is the coefficients, V_i are the input/predictive variables, and ε is the mean zero error term representing variations in log D_H not explained by the predictors. Eq. (5.2) can be turned into a probabilistic model given the fundamental assumption of linear regressions given in Eq. (5.3) that the error term (ε) is distributed normally around a mean of zero and have a constant standard deviation (σ) i.e., the variance does not change across different levels of predictors.

$$\varepsilon \sim N(0, \sigma^2) \tag{5.3}$$

In Eq. (5.2), the error term is a random variable (R.V.). Adding a constant term to a random variable with normal distribution does not change the form of distribution; it only shifts the mean by the amount added and keeps the variance constant. Hence, log D_H also has a normal distribution with a non-zero mean and standard deviation of σ. By using the abovementioned assumption of linear regressions, we can rewrite Eq. (5.2) to take a probabilistic form

$$p(\log D_H > y) = 1 - \phi\left(\frac{y - \sum_i b_i V_i}{\sigma}\right) \quad (5.4)$$

where ϕ is the cumulative density function of standard normal distribution. One of the model prerequisites presented by Youd et al. (2002) is the likelihood of liquefaction. Hence, Eq. (5.4) is conditioned on having liquefaction, which is needed in the second term on the right-hand side of Eq. (5.1). However, as Youd et al. (2002) and Youd (2018) presented, having liquefaction is not the only condition that must be met before applying the MLR model.

As a possible solution to this issue, we extended the probability chain in Eq. (5.1) by adding a term to predict lateral spread susceptibility

$$P(D_H > y) = P(L)P(LS|L)P(D_H > y|LS) \quad (5.5)$$

where P(LS|L) is the probability of lateral spread given liquefaction, and $P(D_H>y|LS)$ is the probability of lateral spread magnitude exceeding threshold y given lateral spread. The probability of lateral spread exceeding a threshold from Eq. (5.4) can be directly used in Eq. (5.5) because the added term considers lateral spread susceptibility.

5.4 Methodology

5.4.1 Introduction

Convolutional Neural Networks (CNN) are well-known, deep learning architectures inspired by the living creatures' natural visual perception mechanism. The modern framework of CNN was first introduced by LeCun et al. (1989) then later improved by LeCun et al. (1998) to classify handwritten digits via a model called LeNet-5. Since then, due to the excellent performance of CNNs, numerous models have been developed to solve tasks from various domains, including computer vision, natural language processing, etc. In recent years, deep-learning models and CNNs have had a broad range of applications in Civil Engineering, such as Cha et al. (2017), Nazaraki et al. (2018), Woldesellasse and Tesfamariam (2022). An excellent summary of recent advances in CNNs can be found in (Gu et al. 2015). Durante and Rathje (2021) present an exploration of the use cases of Machine Learning to predict lateral spreading.

5.4.2 Convolution layer

A convolution layer contains input and filter arrays. The filter array — sometimes referred to as kernel — is smaller than the input array in shape and consists of trainable weights and a bias initially generated at random; this array is defined by a width and height. A convolutional layer convolves the input and passes its results to the next layer. This process is similar to the response of a neuron in the visual cortex to a specific stimulus. A convolution ($\otimes$) is the dot-product (element-wise multiplication) between the input and filter arrays. The resulting array is summed, and a bias term is added, resulting in a single value. The filter is smaller than the input to allow the same filter (set of weights) to be multiplied by the input array multiple times at different points on the input. Because the filter is multiplied multiple times the result is a two-dimensional array of output values called a feature map. An example of a convolution layer with an input array of shape 3 by 2, filter size of 2 by 2, a stride of 1 is shown in Figure 5.1.

5.4.3 Pooling Layer

A key aspect of CNNs is a pooling layer, which helps reduce the spatial size of an input array. This reduction process is often called down sampling, which is specifically important when the training data is rare because the more trainable parameters a model has, the more training samples are needed. There are two commonly used pooling options. Max pooling takes the max values from an input subarray, whereas mean pooling takes the mean values. Similar to convolutional layers, a pooling layer is also defined by width, height, and stride.

5.4.4 Embedding Layer

Categorical data refers to input features that represent discrete items from a finite set of choices. For example, soil indices introduced by Gillins and Bartlett (2013) has six possible types and is a categorical feature. Suppose the raw soil indices are fed to an algorithm as a numerical feature. In that case, it is implicitly introduced to the algorithm that the distance between soil index 1 and soil index 2 is equal to the distance between soil index 5 and soil index 6, which is not necessarily true for categorical data. It is desirable to represent each category as a vector of numbers to represent categorical features in machine learning. Thus doing, semantically similar items have similar distances in the vector space. The idea was first introduced by Mikolov et al. (2013) to estimate word representations in vector space.

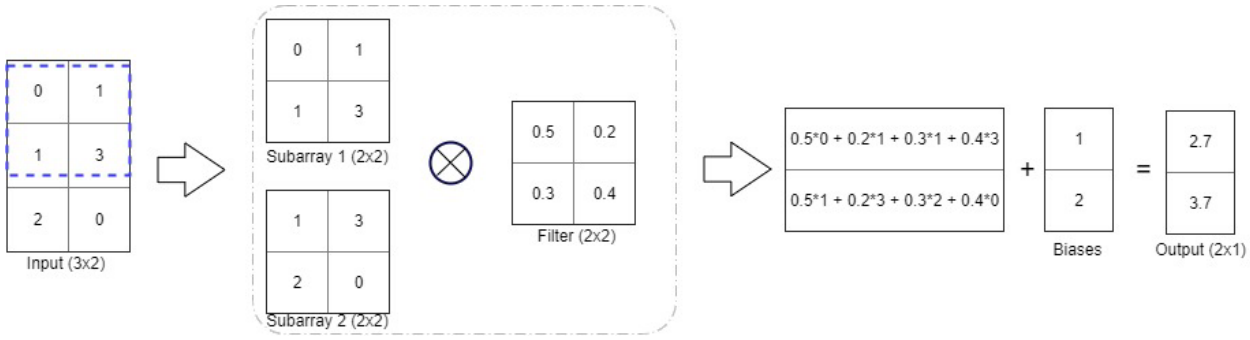


Figure 5.1 An example of the convolution layer

5.4.5 One-hot-encoding

An alternative approach to using embedding layers for categorical features is called one-hot-encoding technique. This technique removes the integer encoded variable, and a new binary variable is added for each unique integer value. For example, in the case of the soil index, where only six outcomes are possible, one-hot-encoding would result in six separate binary indicators and convert each category into a vector of zeros and ones of length 6. A soil index 1 will be converted to $v_1 = [1, 0, 0, 0, 0, 0]$ while soil index 5 will be converted to $v_5 = [0, 0, 0, 0, 1, 0]$. This transformation, as opposed to integer encoding, eliminates any implicit relationship between categories.

5.4.6 Activation Layer

Activation layers are the most typical way to introduce nonlinearity in artificial neural networks. Nonlinear activation functions allow neural networks to compute nontrivial problems using only a small number of nodes. Three most commonly used nonlinear activation functions are sigmoid, hyperbolic tangent, and rectified linear units (ReLu) introduced by Nair and Hinton (2010). Given

the gradients in ReLu's are all zeros and ones, the computational cost of using these activation functions is much less than the other two alternatives mentioned earlier.

5.4.7 Dropout Layer

One of the primary issues associated with machine learning — specifically training datasets with a few examples — is overfitting. This phenomenon occurs when the performance on the training set is satisfactory, but the model fails to generalize well on unseen validation and test data. There are several ways to avoid overfitting, including reducing model complexity, using regularizations, stopping training early, etc. Dropout, introduced by Srivastava et al. (2014), is a regularization method approximating training many neural networks with different architectures in parallel. During training, some number of layer outputs are randomly ignored or "dropped out."

5.4.8 Softmax Layer

A softmax layer, which is usually the last activation function of CNN architectures, is a generalization of the logistic function to multiple dimensions. The softmax function takes as input a vector of *n* real numbers. It normalizes it into a probability distribution consisting of *n* probabilities proportional to the exponentials of the input numbers. That is, before applying softmax, some vector components could be negative or greater than one; and might not sum to 1; but after applying softmax, each component will be in the interval (0, 1), and the components will add up to 1 so they can be interpreted as probabilities. Furthermore, the larger input components will correspond to higher probabilities.

$$P\left(y^{(i)}=n\middle|x^{(i)};W\right)=\begin{bmatrix}p\left(y^{(i)}=1\middle|x^{(i)};W\right)\\ y^{(i)}=2|x^{(i)};W\\ \vdots\\ y^{(i)}=n|x^{(i)};W\end{bmatrix}=\frac{1}{\sum_{j=1}^{n}e^{W_j^T x^{(i)}}}\begin{bmatrix}e^{W_1^T x^{(i)}}\\ e^{W_2^T x^{(i)}}\\ \vdots\\ e^{W_n^T x^{(i)}}\end{bmatrix} \tag{5.6}$$

where y = target classes; x = inputs to the activation function; and W = trainable weights.

5.4.9 Loss Function and Optimizer

The training process of artificial neural networks consists of optimizing (usually minimizing) a loss function that explains deviations of predictions from true labels. For accurate predictions, one needs to optimize and update randomly initialized weights and biases of a network in each epoch. There are several known methods, but Stochastic Gradient Descent (SGD) using backpropagation is considered the most efficient and most straightforward way to minimize the deviations (LeCun et al., 2012). The standard gradient descent algorithm performs updating weights and biases on the entire training data set, but the SGD algorithm performs it on single or several training samples. using the following equation

$$W_{new}=W_{old}(1-lr\nabla_W L) \tag{5.7}$$

where W = weights and biases; lr = learning rate; ∇_W= gradient of loss function with respect to weights and biases defined in Eq. (5.8.a); and L = softmax loss function defined in Eq. (5.8.b).

$$\nabla_W L\left(W;x^{(i)},y^{(i)}\right)=\frac{1}{m}\sum_{i=1}^{m}\left[x^{(i)}\left\{1\left\{y^{(i)}=j\right\}-p\left(y^{(i)}=j\middle|x^{(i)};W\right)\right\}\right]+\lambda W_j \tag{5.8.a}$$

$$L = \frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} 1\{y^{(i)} = j\}\, log\, \frac{e^{W_j^T x^{(i)}}}{\sum_{k=1}^{n} e^{W_k^T x^{(i)}}}\right] + \frac{\lambda}{2}\sum_{j=1}^{n} W_j^2 \qquad (5.8.b)$$

where 1{} = logical function that always returns either zero or one; and λ = regularization coefficient to penalize large weights, which is another expediency to prevent overfitting (Bengio, 2012; Bottou, 2012).

5.5 Development of Screening Criteria

5.5.1 Problem Formulation

Figure 5.2 shows the lateral spread in Heber Road during the 1979 Imperial Valley, California earthquake. This case history is used as an example to illustrate the problem formulation more clearly. This figure also includes the location of SPT boreholes. The fundamental building block of the proposed screening criteria is a pair of boreholes, which have three types:

- Pairs inside lateral spread, i.e., both boreholes are within the area where lateral spread has occurred. For example, the SPT 4 - SPT 6 pair.
- Pairs outside lateral spread, i.e., both boreholes are outside the area where lateral spread has occurred. For example, the SPT 1 – SPT 12 pair.
- Pairs on the fringe of lateral spread, i.e., one borehole is inside the lateral spread while the one borehole is outside the area where lateral spread has occurred. For example, the SPT 11 – SPT 4 pair.

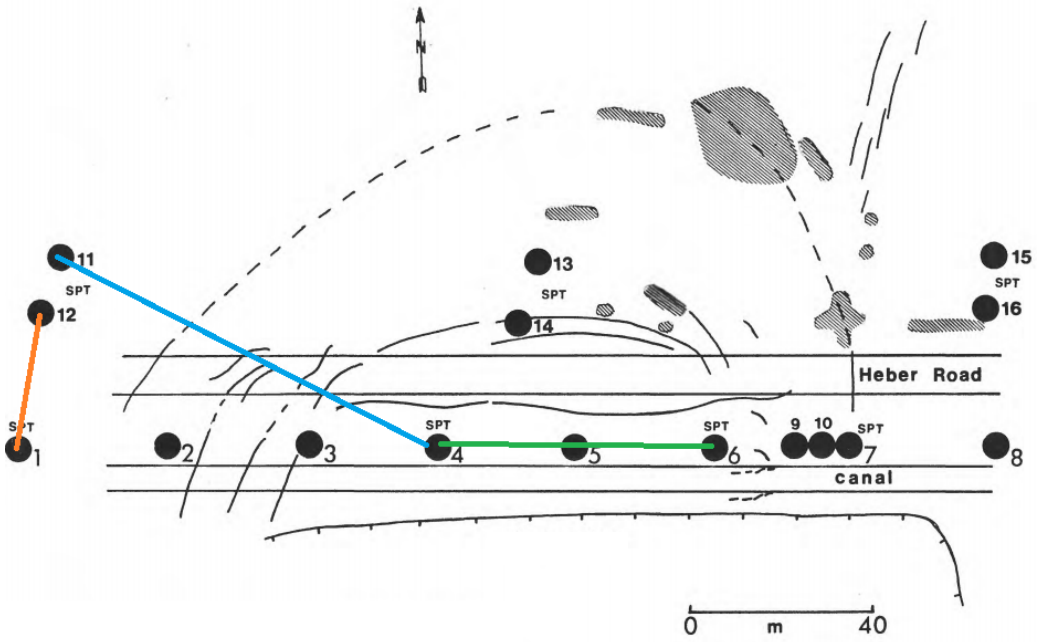


Figure 5.2 SPT boreholes and lateral spread in Heber road at 1979 Imperial Valley case history

For each borehole pair, interpolated soil density indicators ($N_{1,60}$), soil saturation indicators, and soil index at subsurface sampling intervals to a depth of 30 meters below the ground surface are used as input features (*S*). The distance between boreholes in a pair (d) is also used as another input feature andsubsurface information. Hence, the second term on the right side of Eq. (5.5) can be written as

$$P(LS|L) = f(S, d) \tag{5.9}$$

The function *f()* is defined by a convolutional neural network described in detail in subsequent sections. This function outputs probabilities for three cases (1) inside the lateral spread zone, (2) on the boundary of the lateral spread, and (3) outside the lateral spread zone. To implement these probabilities in Eq. (5.5), one can combine probabilities of inside and on the boundary of lateral spread and use that as the probability of having a lateral spread. This combination is equal to the probability of being within the lateral spread zone. Also, the probability of having two boreholes outside the lateral spread zone is the same as the probability of not having a lateral spread, which is one minus the probability of being within the lateral spread zone. These calculations can be done because all three outcomes are considered to have disjoint probabilities.

Note that because the probability obtained from Eq. (5.9) is to be used in combination with other probabilities in the conditional probability chain given in Eq. (5.5), other factors playing a role in lateral spread susceptibility, previously discussed in the Theoretical Background Section, are not included in Eq. (5.9). This omission is done to avoid the possible confounding of their effect on the prediction of lateral spread. Note that many of these factors are included in the other terms in Eq. (5.5). Also, the model developed below focuses on soil and site factors and does not include geologic factors, such as sediment age, depositional environment, etc. These factors are known to affect liquefaction and lateral spread susceptibility (Youd and Perkins, 1978), but it is difficult to quantify their influence in the probabilistic model. Therefore, this research will focus on a geotechnical layer and readily glean soil properties from the current case history dataset.

5.5.2 Model Architecture

The neural network architecture of the proposed screening criteria consists of an embedding layer to embed a categorical soil index input feature into a vector space, convolutional layers to find existing patterns between boreholes in a pair, and a dense feed-forward network at the end. Figure 5.3 shows the best-performing neural network architecture for the screening criteria. Several alternatives to this architecture were also considered, which will be discussed later. Note that the dropout layers and batch normalization layers are not included in the visualization. This model has 4,473 trainable weights and biases. A complete summary is given in Figure 5.4.

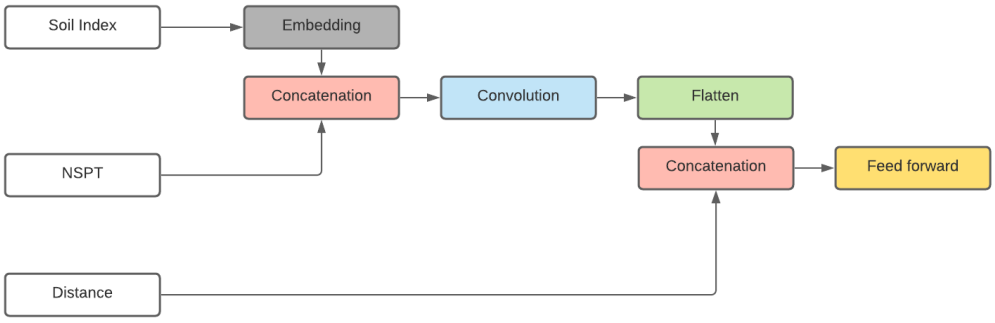


Figure 5.3 The neural network architecture of our best performing model

```
Model: "model_41"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
boreholes (InputLayer)          [(None, 60, 2, 3)]   0
__________________________________________________________________________________________________
tf_op_layer_strided_slice_91 (T [(None, 60, 2)]      0           boreholes[0][0]
__________________________________________________________________________________________________
tf_op_layer_strided_slice_90 (T [(None, 60, 2, 2)]   0           boreholes[0][0]
__________________________________________________________________________________________________
embedding_45 (Embedding)        (None, 60, 2, 3)     21          tf_op_layer_strided_slice_91[0][0
__________________________________________________________________________________________________
concatenate_86 (Concatenate)    (None, 60, 2, 5)     0           tf_op_layer_strided_slice_90[0][0
                                                                 embedding_45[0][0]
__________________________________________________________________________________________________
conv2d_42 (Conv2D)              (None, 52, 1, 18)    1638        concatenate_86[0][0]
__________________________________________________________________________________________________
tf_op_layer_Relu_41 (TensorFlow [(None, 52, 1, 18)]  0           conv2d_42[0][0]
__________________________________________________________________________________________________
dense_features (InputLayer)     [(None, 1)]          0
__________________________________________________________________________________________________
flatten_41 (Flatten)            (None, 936)          0           tf_op_layer_Relu_41[0][0]
__________________________________________________________________________________________________
concatenate_87 (Concatenate)    (None, 937)          0           dense_features[0][0]
                                                                 flatten_41[0][0]
__________________________________________________________________________________________________
dropout_13 (Dropout)            (None, 937)          0           concatenate_87[0][0]
__________________________________________________________________________________________________
dense_70 (Dense)                (None, 3)            2814        dropout_13[0][0]
==================================================================================================
Total params: 4,473
Trainable params: 4,473
Non-trainable params: 0
__________________________________________________________________________________________________
None
```

Figure 5.4 Model description and parameters

5.5.3 Data

Data Collection

Data from 11 earthquakes was collected from multiple sources, mostly from historical case histories with no digital version. The data came in different languages, primarily English and Japanese. For this study, 620 SPT boreholes were collected from which 439 pairs were formed for the analyses. Any low-quality boreholes, including shallow boreholes, were dropped before selecting the paired boreholes. Boreholes more than 150 meters apart were also not paired. Table 5.1 summarizes the evaluated dataset.

The digitization process is shown in Figure 5.5, which consisted of scanning aerial photographs and overlaying them on top of Google Maps™. Subsequently, the location of boreholes and displacement vectors were recorded. This spatial information was needed to compute the distance between boreholes, which is one of the input features and to decide on a borehole status in terms of whether it falls outside, inside, or on the boundary of a lateral spread.

Another critical aspect of the borehole data is the subsurface information (i.e., depth of sampling interval and groundwater, N_{SPT}, soil fines content, if available, and the soil type. This information was collected down to 30 meters below the ground surface. Figure 5.6 shows a typical borehole (BH 3-5) from the Niigata, Japan earthquake found on plat F-10 (unpublished but from Hamada).

Data wrangling

Data wrangling is the process of cleaning and unifying messy and complex data sets for easy access and analysis. This process is crucial in developing the proposed screening criteria because the consistency between input data points from the different sources must be ensured.

Table 5.1 Summary of paired types per earthquake

Earthquake	Boreholes count	Pair type	Pairs count	Reference(s)
1964 Alaska, US	20	boundary	3	Bartlett and Youd 1992b; Ross et al. 1973
		outside	3	
		within	5	
1964 Niigata, Japan	145	boundary	38	Hamada et al. 1986
		outside	33	
		within	169	
1971 San Fernando, California	39	boundary	12	Bennett 1989; O'Rourke et al. 1992; Youd 1973
		outside	4	
		within	44	
1979 Imperial valley, California	11	boundary	12	Youd and Bennett 1983; Bennett et al. 1984
		outside	2	
		within	6	
1983 Borah peak, Idaho	9	boundary	11	Youd et al. 1985; Andrus 1991; Andrus and Youd 1987
		outside	3	
		within	4	
1983 Noshiro, Japan	187	outside	17	Hamada et al. 1986
		within	8	
1987 Superstition hills, California	2	within	1	Holzier et al. 1989
1989 Loma Prieta, California	15	boundary	15	Robertson et al. 1999
		outside	3	
		within	7	
1990 Luzon, Philippines	13	outside	1	Tokimatsu et al. 1994; Ishihara et al. 1993
		within	1	
1995 Kobe, Japan	156	within	32	Chu et al. 2004
1999 ChiChi, Taiwan	23	boundary	3	
		within	2	

(a)

(b)

(c)

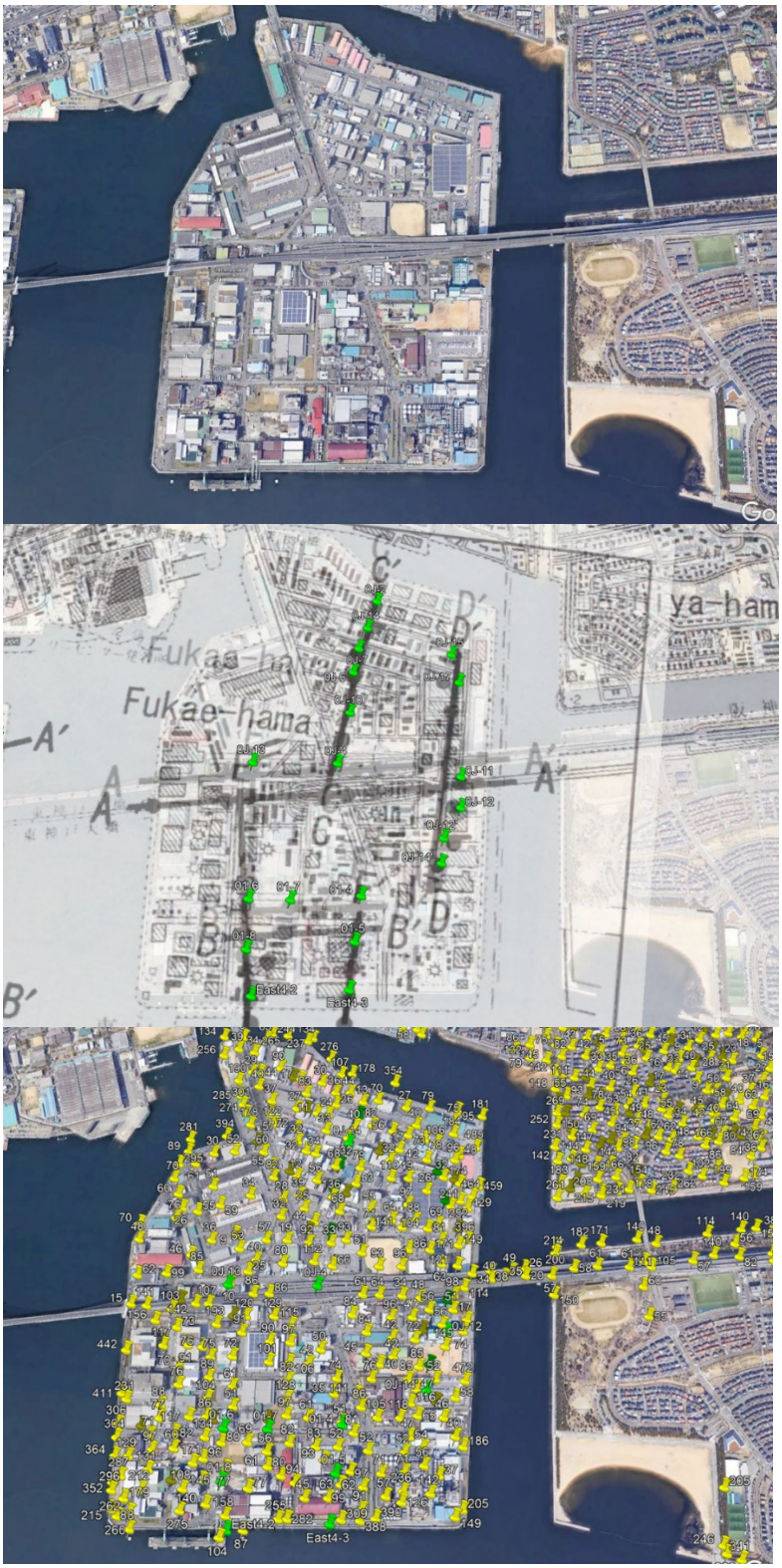
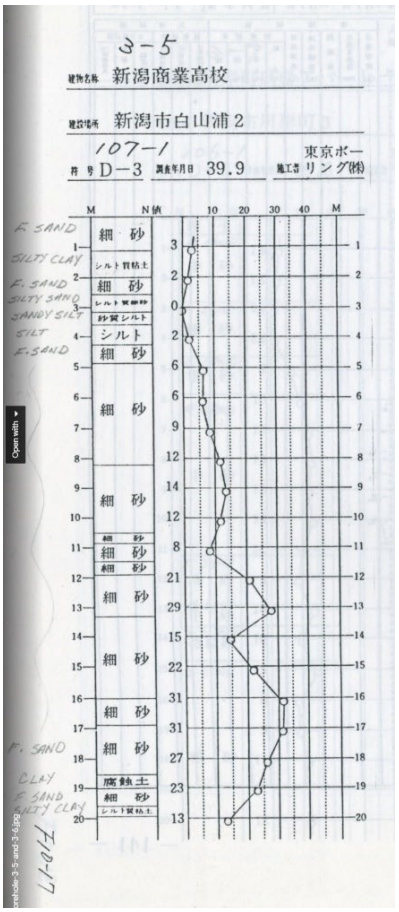


Figure 5.5 Process of borehole location digitization (a) Google maps view of a site in Kobe, Japan, (b) boreholes overlaid and digitized, and (c) displacements (yellow pins) digitized with boreholes (green pins)



(a)

Number	Japanese	English
(1)	標尺 (m)	Level rod
(2)	標高 (m)	Gland level
(3)	深度 (m)	Depth
(4)	柱状図	Boring log
	土質記号	Soil mark
	断面記号	Section mark
	地質図表	Geological map
	図表	Map
(5)	地質名	Geological name
	土質名	Soil name
(6)	標準貫入試験	Standard penetration test
(7)	N値 (回/30cm)	N-value
(8)	粒度試験	Mechanical analysis
(9)	粒度組成 (%)	Texture
	粒度配合	Grandation
(10)	礫	Gravel
(11)	砂	Sand
(12)	シルト	Silt
(13)	粘土	Clay
(14)	粒径 (mm)	Grain size
(15)	自然含水比 ω (%)	Natural water content
(16)	比重 Gs	Specific gravity
(17)	湿潤密度 P^t (g/cm^3)	Wet density
(18)	液性限界 W_L (%)	Liquid limit
(19)	塑性限界 W_P (%)	Plastic limit
(20)	三軸圧縮試験	Triaxial compression test
(21)	粘着力 C (Kgf/cm^2)	Cohesion
(22)	セン断抵抗角 ϕ (°)	Angle of shearing resistance
	内部摩擦角 ϕ (°)	Angle of infernal friction
(23)	一軸圧縮強さ q_u (Kgf/cm^2)	Unconfined compression strength
(24)	体積圧縮係数 M_v (Kgf/cm^2)	Coefficient of volume compressibility
(25)	圧密係数 C_v (cm/s)	Coefficient of consolidation
(26)	色調	Color tone
	岩色	Rock color
(27)	観察	Visual inspection
	記事	Comment
(28)	試料採取位置 (m)	Sampling depth
	採取深度	Borrow depth
(29)	物理試験	Physical test
(30)	間隙比	Void ratio
(31)	力学試験	Dynamic test Mechanical test
(32)	圧密試験	Consolidation test
(33)	圧密降伏応力 (Kgf/cm^2)	Consolidation yield stress
(34)	層厚 (m)	Layer thickness
(35)	相対密度及びｺﾝｼｽﾃﾝｼｰ Dr	Relative consistency
(36)	10cm毎の打撃回数 N	Times nailed
(37)	硬軟	Soil density
	硬質	Hard
	軟質	Soft
(38)	排水状況	Drainage condition
(39)	水位 (m)	Water level

(b)

Figure 5.6 Sample of flat files (a) subsurface data in Japanese, and (b) Japanese to English instructions of soil description

The primary input data are pairs of boreholes. In other words, each pair of adjacent boreholes form one sample in our dataset. However, every borehole has the potential to be included in more than one pair. Before pairing boreholes, each borehole was classified as either inside or outside lateral spread. For example, in Figure 5.2, SPT boreholes 1, 11, 12, and 15 are outside the lateral spread zone, while all other SPT boreholes are inside the lateral spread. This step is required for labeling our data which will be used as the target in our model training process.

There are four main components to the input predictive data:

1. Soil relative density indicator ($N_{1,60}$) at different depths for each borehole in a pair.,
2. Soil saturation indicator at each depth for each borehole in a pair.
3. Soil Index as defined in Gillins and Bartlett (2013) at different depths for each borehole in a pair.
4. Distance between pair of boreholes.

The effects of the operator, equipment, and test procedures on the Standard Penetration Test (SPT) blow counts were accounted for by using correction factors and the practices described by Salgado (2006). The N_{SPT} values were first corrected for the hammer type, rod length, sampler type, and borehole diameter. Then they are adjusted for effective overburden pressure of 100 kPa $(N_1)_{60}$. Finally, the SPT counts were interpolated and resampled at regular depth intervals within the borehole.

The corrected SPT blow count is the standard blow count denoted as $N_{60,}$ defined as

$$N_{60} = C_h C_r C_s C_d N_{SPT} \tag{5.10}$$

where C_h = hammer correction; C_r = rod length correction; C_s = sampler correction; and C_d = borehole diameter correction.

The hammer correction factor (C_h) is expressed as the ratio of energy ratio for the hammer used in the test (ER_{hammer}) to the standard 60% energy ratio for the safety hammer

$$C_h = \frac{ER_{hammer}}{ER_{safety}} = \frac{ER_{hammer}}{60\%} \tag{5.11}$$

In cases where the energy ratio for the SPT hammer Eq. (5.11) was recorded in the original case history documents, this was used. However, in most cases, such information was not measured or recorded. Thus, approximate values of the energy ratio and C_h (Table 5.2) were used.

Table 5.2 Approximate values of energy ratio and correction factor C_h for common hammer types

Hammer	Approximate energy ratio (%)	C_h
Safety	60	1
Donut	45	0.75
Automatic trip	80	1.33
Pin weight	72	1.2

The sampler correction factor (C_s) accounts for the occasional use of a liner sampler without the liner. According to Salgado (2006), the blow counts when the standard sampler is used are 10%–30% higher when the ASTM sampler is used without the liner. Hence, C_s was set to 1 for the standard ASTM sampler with a liner sample, while C_s was set to 1.2 samplings without a liner. The rod length correction factor (C_r) that was used is given in Table 5.3. The borehole diameter correction factor (C_d) equals 1 for borehole diameters with the standard range of 66 to 115 mm (2.5 to 4.5 in) or greater than 1 for larger borehole diameters. According to Salgado (2006), C_d is 1.05 and 1.15 for borehole diameters of 150 and 200 mm, respectively. Interpolation was used for other values within this range. Because the N_{SPT} blow counts depend on the vertical stress, it varies approximately with the square root of the effective stress (Eq. 5.12). Hence, to obtained N_{SPT} values normalized for the vertical effective stress we can write

$$N_1 = N_{SPT}\sqrt{\frac{p_A}{\sigma'_v}} \tag{5.12}$$

where p_A = reference stress of 100 kPA; and σ'_v= effective vertical stress.

Table 5.3 Rod length correction factor C_r

C_r	Rod length
0.75	Less than 4 meters
0.85	Between 4 and 6 meters
0.95	Between 6 and 10 meters
1	Greater than or equal to 10 meters

Once the N_{SPT} values are corrected for different factors and adjusted for overburden pressure, they were resampled at different intervals based on linear interpolation between the two closest blow counts recorded in each borehole. Figure 5.7 shows the effect of the selected sampling interval on the interpolated values. In most cases, a resampling interval of 0.5 meters has complete consistency with the original blow counts, while for larger values, resampling may cause loss of information. Having a 0.5 meters interval also helps to capture the effect of thin layers in the model.

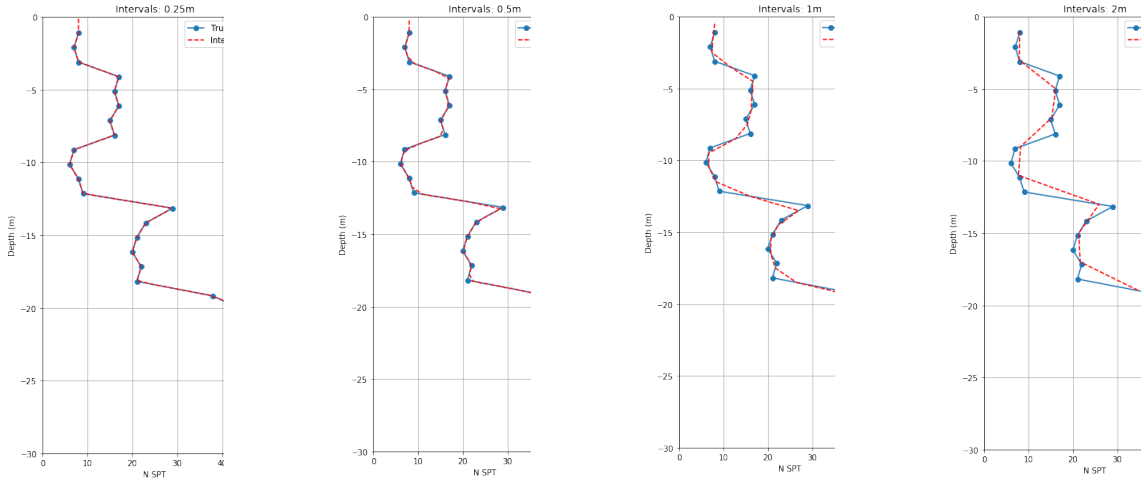


Figure 5.7 Blow counts resampling at various intervals

Gillins and Bartlett (2013) introduced Soil Index to replace soil properties, such as fines content and median grain size with a classification based on soil description. This index was introduced to resolve the issue caused by missing laboratory soil properties, which is common in many geotechnical soil logs. The soil index values presented by Gillins and Bartlett (2013) are given in Table 5.4. This table is used as a lookup table to encode all soil descriptions in our database. A list of encoded soils and their corresponding soil index is given in Table 5.5. This encoding allows us to use this categorical soil description factor in our models as a numerical input.

Table 5.4 Soil Index and their corresponding soil descriptions

Typical soil descriptions	General USCS symbol	Soil Index, SI
Silty gravel with sand, silty gravel, fine gravel	GM	1
Very coarse sand, sand, and gravel, gravelly sand	GM-SP	2
Coarse sand, sand with some gravel	SP	2
Sand, medium to fine sand, sand with some silt	SP-SM	3
Fine sand, sand with silt	SM	4
Very fine sand, silty sand, dirty sand, silty/clayey sand	SM-ML	4
Sandy silt, silt with sand	ML	5
Silty clay, lean clay	CL	6

Table 5.5 Soil Index of all soil descriptions in database

Soil Index (SI)	Soil descriptions
1	gravel mix. fine sand
2	coarse sand, coarse-medium sand, coarse to medium sand, organic, coarse sand, medium ~ coarse sand
3	medium sand, medium-fine sand, medium sand, sand, medium to coarse sand, fine to medium sand, medium to fine sand, medium sand, medium grain sand, fine-medium sand, fine ~ medium sand
4	alter silt and sand, fine sand, silty medium sand, very fine sand, clayey, fine sand, silty, fine sand, silty very fine sand, fine sand mixed with silty sand, fine sand, fine to very fine sand, very fine to fine sand, fine grain sand, fine sand, silty, fine sand, organic, fine sand, fine - very fine sand
5	silt, silty sand, sandy silt, clayey silt, coarse silt, silty sand mixed with silt, silty sand to sandy silt, silty coarse sand mixed with silt, sandy silt, medium sandy silt, silt
6	not defined, fill, silty clay, sandy clay, humus, surface soil, clay, sandy loam, reclaimed land, silty clay to sand, silty sandy clay, peat, organic soil/surface soil, topsoil, fill soil, cohesive soil, or hummus/peat, organic silt, all missing values

After collecting and converting subsurface information in our database, the largest category was soil index 6 (silty clay, lean clay) (Figure 5.8). To ensure there was no significant bias in the distance between boreholes, a histogram of distance is plotted in Figure 5.9. The summary statistics for this factor are given in Table 5.6.

Table 5.6 Summary statistics of the distance between boreholes in a pair per pair type

Pair type	**Distance (m)**			
	mean	median	min	max
boundary	90.0	97.7	3.0	149.4
outside	77.1	77.3	3.2	149.0
within	81.5	79.1	1.3	149.3

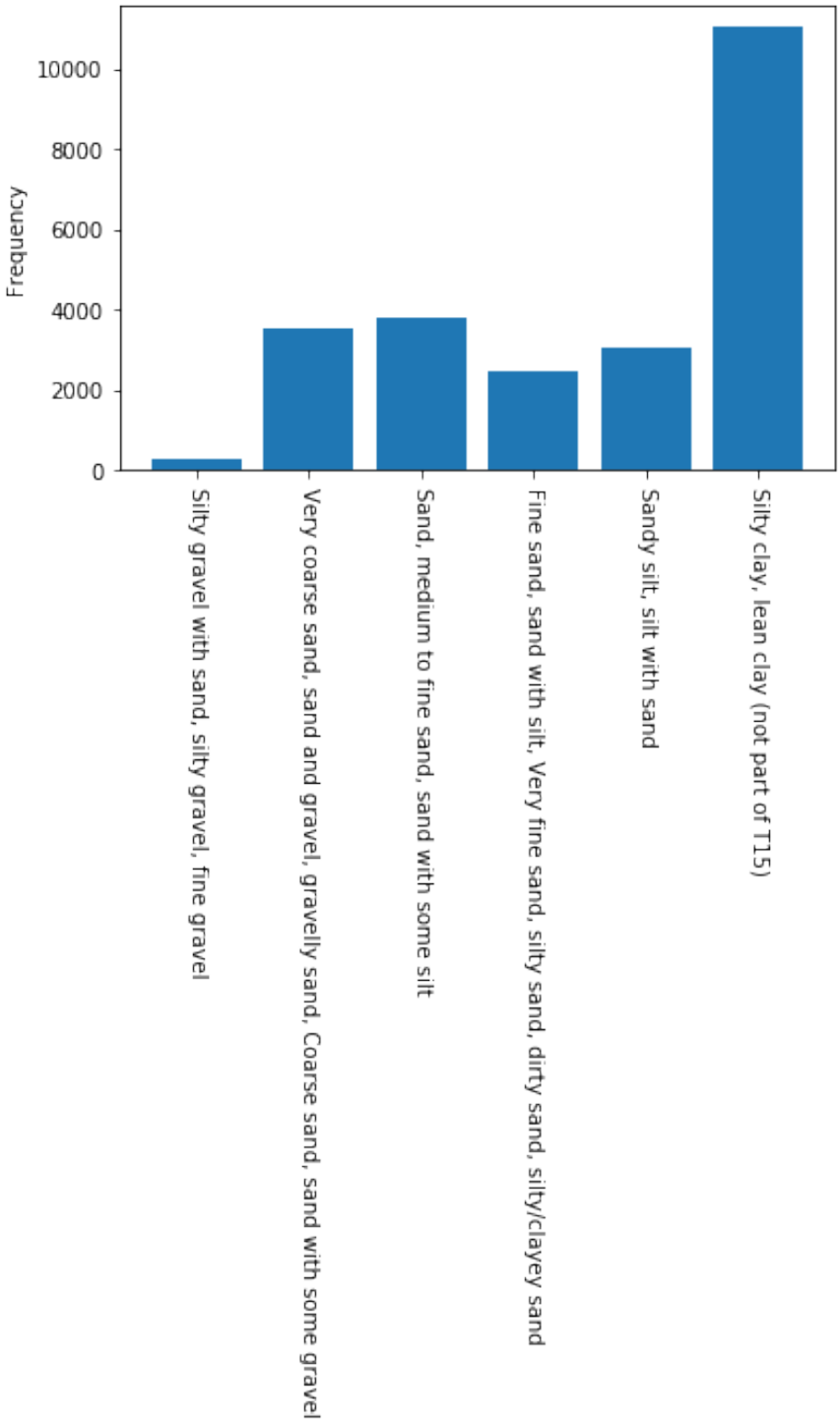


Figure 5.8 Distribution of soil indices in the dataset

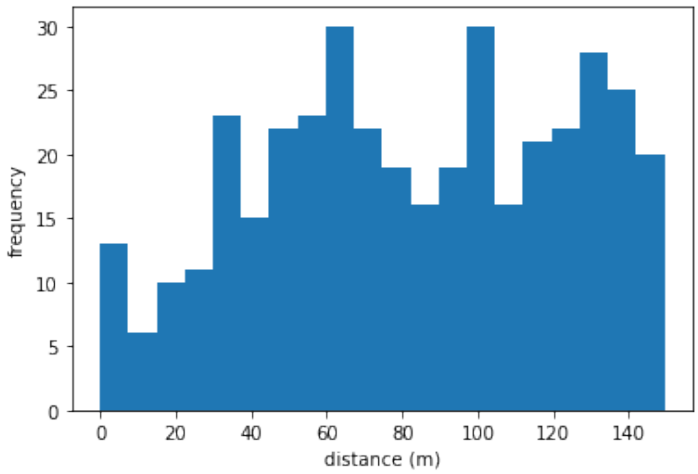


Figure 5.9 Distribution of distance between boreholes in pairs

5.5.4 Hyperparameters

Hyperparameters are parameters that are not directly learned within estimators. The choice of these hyperparameters can affect the model efficacy and should be optimized. For example, the type and number of layers, nodes, learning rate, number of filters, and kernel width in the convolutional layer are hyperparameters. The effect of different layers or network topology is subsequently described in the alternative network architectures section.

The most fundamental block of our CNN model is the convolution layer responsible for capturing interactions between boreholes in a pair. The two most important hyperparameters tuned in this study are (1) the number of filters and (2) the width of the convolution kernel.

A grid search algorithm was used to find the optimal number of filters and kernel width. This optimization is shown in Figure 5.10 as a heatmap; brighter colors correspond to higher accuracy and vice versa. The top four performing hyperparameters are given in Table 5.7. The best performing set of hyperparameter resulted from 16 filters and a kernel width 19.

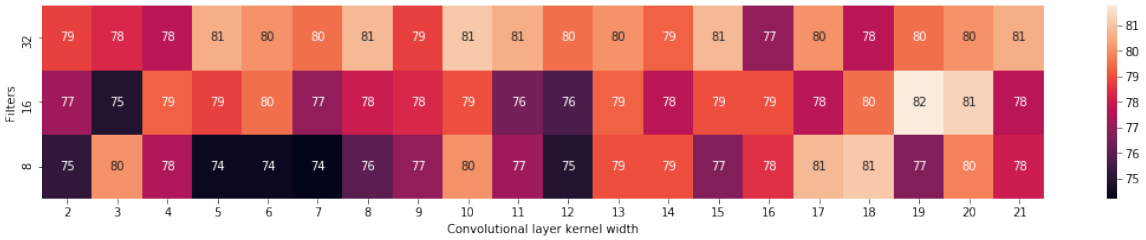


Figure 5.10 Hyperparameter tuning results

Table 5.7 Top 4 model accuracies in hyperparameter search

Convolutional layer kernel width	Number of convolutional layer filters	Model accuracy	Number of parameters
19	16	81.8	5099
20	16	81.3	5211
10	32	81.1	8155
18	8	81.1	2507

Another consideration is the number of trainable parameters in our network, which represents the number of weights and biases in all layers. In general, the higher the number of parameters used increases the model complexity and also increases the undesirable consequence of overfitting. Figure 5.11 shows the influence of both hyperparameters on the number of parameters. As expected, an increase in either the number of filters or kernel width increases as the number of parameters is increased. However, the rate of increase is less for the kernel width when compared to that of the number of filters.

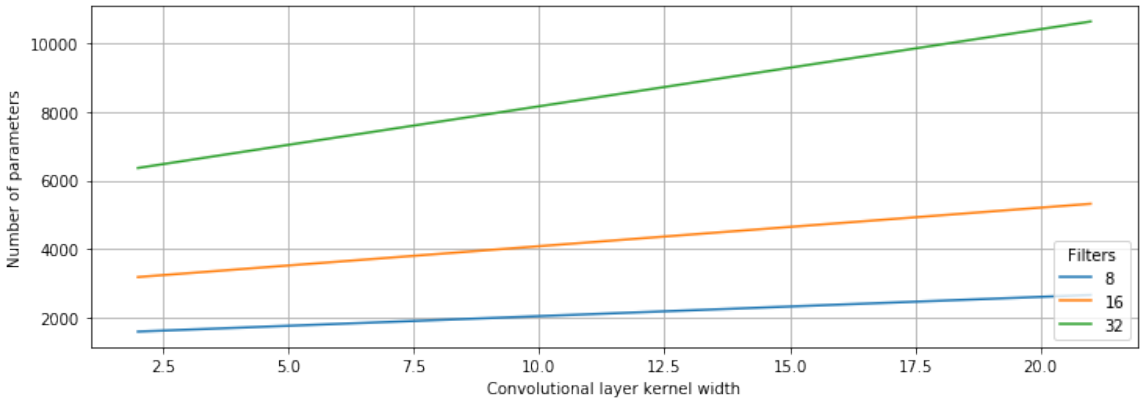


Figure 5.11 Effect of hyperparameters on number of parameters in the model

A minimal version of the hyperparameter optimization process was done for the alternative model architectures before choosing the final architecture. This process was done to ensure that hyperparameter tuning is not the controlling factor in selecting the "best" model architecture.

5.5.5 Training and Validation

In machine learning, the dataset is typically split into a training and a validation set. The training dataset is used to train the model, while the validation dataset is held in reserve to evaluate the model's performance. Comparing the model's performance between the training and validation sets indicates the model's ability to generalize. If the training performance is high and model performance on the unseen data is relatively poor, this is called overfitting (Figure 5.12). In this case, the model has memorized everything in the training set but cannot generalize well to unseen data. If the performance on both training and validation data is low, it is called "underfitting." In this case, the model has not learned the patterns in the training dataset and cannot generalize to the unseen validation data. The goal is to balance the model's performance between the training and validation sets, so the error rate of both groups goes down during the training process.

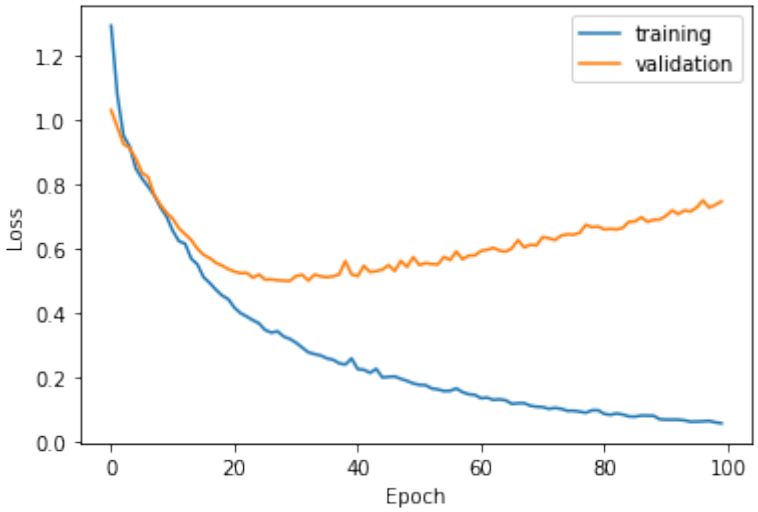


Figure 5.12 Training and validation losses illustrating overfitting

The idea of splitting data into two subsets and using one for training and the other for validation would prevent us from using validation data for training. Given our dataset size is relatively small, all information contained therein should be used for training. Therefore, initially, a five-fold, stratified cross-validation was selected to find the best hyperparameters of the model.

Lastly, the model was trained on the entire dataset with the optimized hyperparameters found in cross-validation. In k-fold cross-validation (see Figure 5.13), the data is split into k folds. Subsequently, each time k-1 folds are used for training and the remaining one-fold for validation. This process will be repeated k times, and each time a different fold will be used for the validation while the other folds are used for training. In this validation scheme, every sample in the dataset is used once for validation. In this way, k-fold cross validation ensures our validation set is not biased. The term "stratified" means that different target classes will be proportional in all folds.

Figure 5.13 K-fold cross-validation

Also, an "epoch" is a term used in machine learning, indicating the number of passes of the entire training dataset completed by the machine-learning algorithm. Our training process was continued for 100 epochs nearby until the validation loss no longer improved (Figure 5.14).

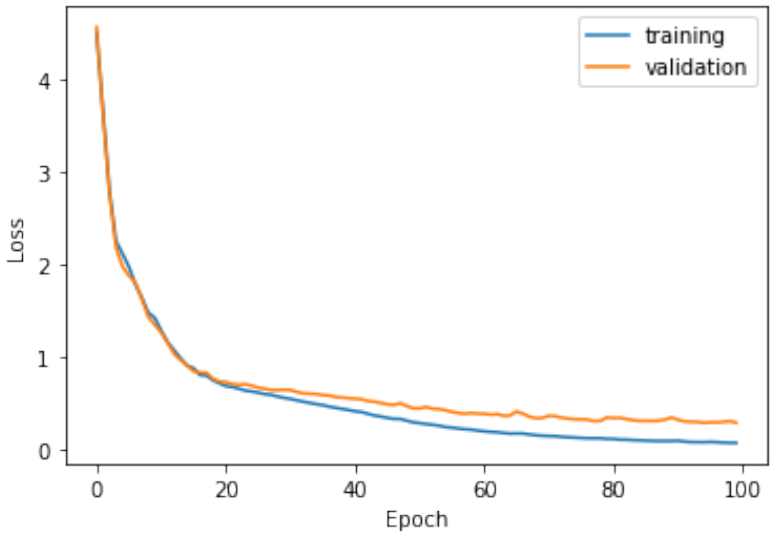


Figure 5.14 Training and validation loss over epochs of training

5.5.6 Alternative Network Architectures

Several variations of the primary network architecture were considered as alternatives before choosing the best-performing model. Table summarizes the top four architectures evaluated in this study in addition to the main model. Alternative model architectures are shown in Figures 5.15 through 5.18. All models were trained on the same folds with near-optimal hyperparameters. Their accuracy is shown in Table is an average of three runs for each model.

Table 5.8 Summary of alternative model architectures

Alternative model	Summary	Accuracy %	Parameters
1	Soil Index one hot encoded; separate convolutions on soil index and N_{SPT}	75.9	4,911
2	Similar to alternative 1 except embeddings instead of one-hot-encoding	78.7	7,299
3	Similar to alternative 2 without embeddings	76.7	6,630
4	Soil Index as is; one convolution over inputs after concatenation of different variables	74.9	3,804

Alternative Model Architecture No. 1

This model shown in Figure 5.15 is dissimilar to the main model from two aspects (1) the categorical feature Soil Index is being encoded using a one-hot-encoding technique as opposed to embedding in the primary model, and (2) convolutional layers are applied separately to soil index encodings and subsurface information before concatenation with each other and distance to be fed into the feed-forward section of the model.

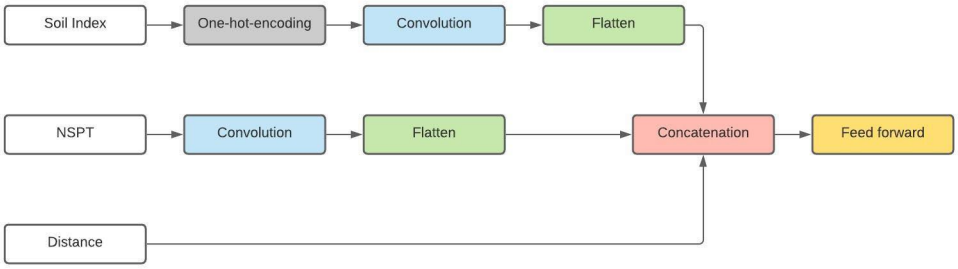


Figure 5.15 Alternative model architecture no. 1

Alternative Model Architecture No. 2

Among the alternatives assessed in this study, alternative no. 2, shown in Figure 5.16, is most similar to the main model. The only difference is that this alternative has separate convolutions on the Soil Index and subsurface information. In contrast, the main model has only one convolution layer after concatenating soil index and subsurface input data. Given this alternative had an accuracy 78.7% while that of alternative no. 1 was 75.9% shows the performance gain of using embedding layers instead of the one-hot encoding technique. A similar architecture to the main model that uses one-hot-encoding instead of embeddings was also considered. However, it was not among the top four most performing alternatives; hence it is not listed here.

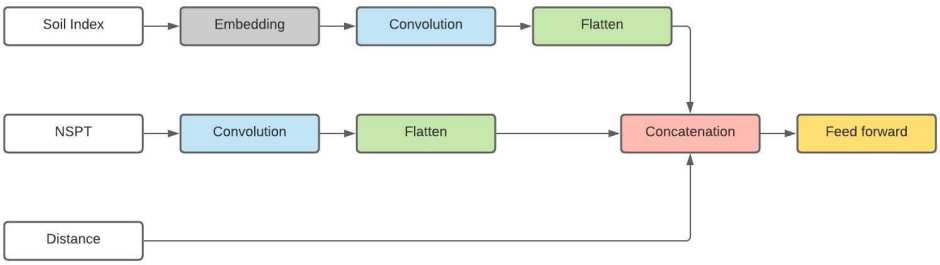


Figure 5.16 Alternative model architecture no. 2

Alternative Model Architecture No. 3

The alternative model shown in Figure 5.17 does not encode the categorical feature Soil Index before passing it into the convolutional layer. Because the soil index defined by Gillins & Bartlett (2014) has an ordering to it, with fine-grain soils being the largest index and coarse grain soils with the smallest, the soil index is an ordinal categorical feature itself, which could be used without encoding. The purpose of this model architecture is to evaluate performance if the Soil Index is not encoded.

Alternative Model Architecture No. 4

This alternative model shown in Figure 5.18 is similar to alternative 3 in that it does not use an encoding technique for the Soil Index. In regard to the convolutional layer, it is similar to the best-performing model.

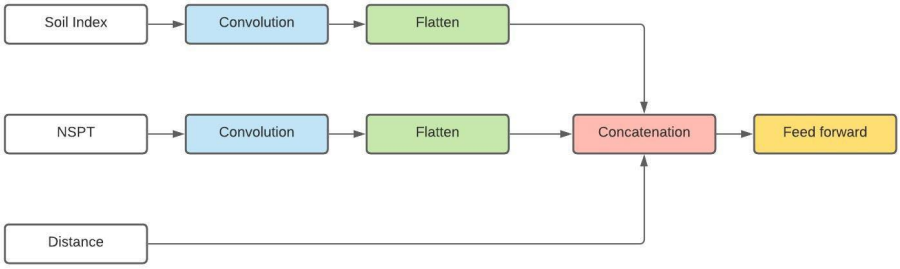


Figure 5.17 Alternative model architecture no. 3

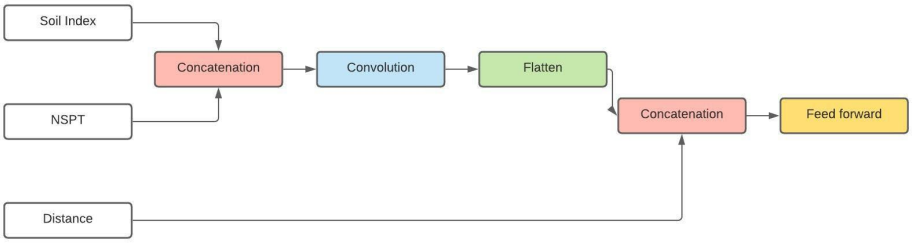


Figure 5.18 Alternative model architecture no. 4

5.6 Results and Discussion

5.6.1 Possible Model Outcomes: 2 or 3

The CNN model developed in this study is a multiclass classification model with three possible outcomes (i.e., within, on the boundary, and outside the lateral spread zone). To use this model in the probabilistic framework, one can combine the probabilities of being inside and on the boundary of lateral spread. This probability also represents the probability of lateral spread. One minus this probability is the probability of being outside the lateral spread zone or not having a lateral spread. This is because all three outcomes considered in this model have disjoint probabilities and hence, based on the addition rule of probability and that the joint probability of inside and on boundary of lateral spread zone is zero, to obtain probability of lateral spread those two probabilities could be simply added together. An analogy to this is outcomes of rolling a dice. There are six possible disjoint outcomes to that event. However, this event could be differently stated as observing an even number versus an odd number.

5.6.2 Performance

The performance of our proposed multiclass CNN model is given as an extended confusion matrix in Table 5.9. This matrix shows true predictions as on-diagonal elements and false predictions as off-diagonal elements. A brief explanation of the evaluation metrics used in this section is given in Appendix A. Classification based on the probabilities outputted from our proposed CNN model is based on the argmax function, i.e., the class with the highest probability is assigned to the pair as the final prediction. The overall accuracy of our proposed model from the confusion matrix is 81.1%, which is based on a trained model chosen to eliminate the effect of uncertainty associated with the folding strategy, which is discussed subsequently. As can be seen, the model accuracy is not uniform for all classes; the model has the highest accuracy in predicting the inside lateral spread zone class and the lowest accuracy in predicting the outside lateral spread zone class. This effect might be attributed to the class imbalance in the training data, as there are

almost four times more inside pairs in the training dataset when compared with outside pairs. Hence, this performance unbalance is somewhat expected because the model learns the majority class better than other classes found in the dataset. Also, this class imbalance should be considered in applying our proposed model. According to Table 5.9, the boundary class was misclassified 17 times as inside class, while only eight times as outside class. This comparison means our model might have an internal bias toward favoring inside class over outside class predictions. This bias may result in unintended over predictions of the inside class.

Table 5.9 Confusion matrix of out of fold predictions, multiclass classification

		Prediction			
		Outside	Boundary	Inside	Accuracy %
Ground truth	Outside	44	12	10	66.7
	Boundary	8	66	17	72.5
	Inside	7	20	207	88.5

As discussed earlier, the boundary and outside classes could be combined to convert the multiclass classification problem into a binary classification problem (i.e., lateral spread triggering versus not triggering). The performance of our proposed model, in this case, is given in Table 5.10. The overall accuracy of the binary classification problem goes up to 90.5% compared with the overall accuracy of 81.1% for the multiclass classification. The decrease in misclassification for the inside and boundary classes is trivial (i.e., the 17 boundary pairs previously misclassified as inside pairs are correctly predicted by the binary classification). Similarly, the 20 inside pairs previously misclassified as part of the boundary class are correctly predicted for the binary classification case.

Table 5.10 Confusion matrix of out of fold predictions, triggering

		Prediction		
		Positive	Negative	Accuracy %
Ground truth	Positive	310	15	95.4
	Negative	22	44	66.7

Precision and recall scores are valuable measures of predictive success when the classes are imbalanced. The precision score represents the proportion of positive identifications (i.e., the prediction was correct). The precision score of the binary classification referred to as a “lateral spread” triggering prediction, is 93.4% (Table 5.10.) The recall score, which represents what proportion of actual positives was identified correctly, is 95.4%. Either of those scores could be more important, depending on the relative importance of false positives versus false negatives. Also, a more comprehensive evaluation metric that could be used is the geometric mean of the precision and recall scores, called the F Score. The F score of the proposed model is 94.4%.

It is important to note that using an argmax function based on probabilities is equivalent to setting the output probabilities at a 0.5 threshold for the binary classification case. Thus, the values reported in Table 5.10 are based on using argmax function, which is threshold dependent. However, for future application of our model, a different threshold (i.e., other than 0.5 could) be used if the precision score is preferred over the recall score or vice versa.

5.6.3 Uncertainty

Although the accuracy of the one five-fold stratified cross-validation of our best performing model is about 82%, given our dataset is relatively small, the reported model performance could be affected by how data was split into different folds. Therefore, the model-training process was repeated 100 times. Each time, samples were distributed differently in folds, and accuracy was measured to quantify the uncertainty associated with our cross-validation strategy. It was found that the accuracies, shown in Figure 5.19, have an approximately bell-shaped normal distribution with a mean of 81.4% and a standard deviation of 1.6.

5.6.4 Mathematical Representation of Soil Types

An interesting observation from exploring alternative network architectures is that the same model not using any encoding technique performed better than the one using one hot encoding. This better performance might be explained by the fact that the soil index defined by Gillins and Bartlett (2014) is based on grain size, and therefore, this ordering is meaningful. One hot encoding, in a sense, removes the ordering between classes and introduces this assumption to the model that each class has the same Euclidean distance of 1 to all other classes. For example, soil index 1 has a Euclidean distance of one to soil index 2 and a distance of five to soil index 6, if not encoded, while in the case of one hot encoding, both these distances are equal to one. Therefore, the higher accuracy in case of no encoding than one hot encoding means the ordering introduced by Gillins and Bartlett (2014) contains useful information to the model that could help performance.

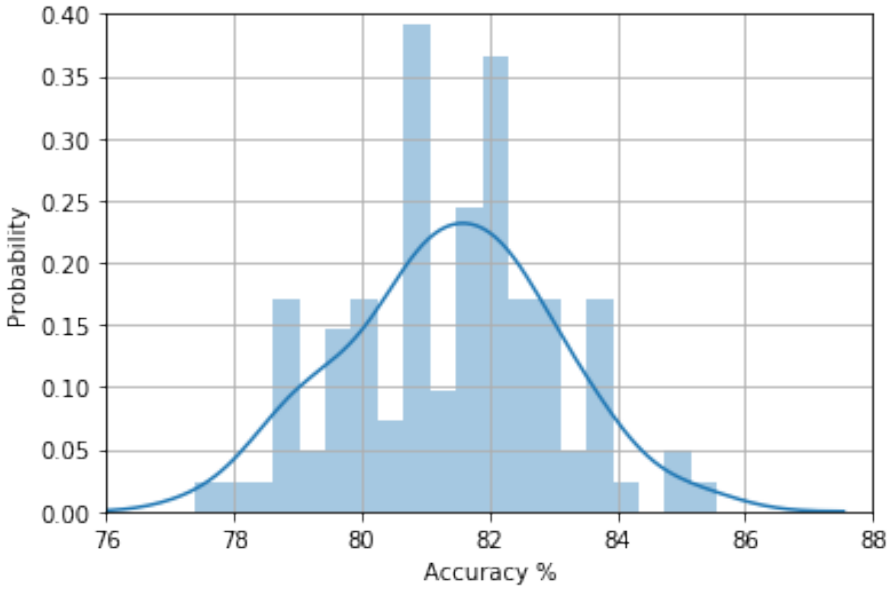


Figure 5.19 Uncertainty associated with cross-validation

While one hot encoding was not an effective technique, embeddings proved to be better than no encodings by giving a 2% accuracy boost. Word embedding such as GloVe (Pennington et al. 2014) is one of the most popular representations of document vocabulary. It can capture the context of a word in a document, semantic and syntactic similarity, relation with other words, etc. The embedding weights are trainable parameters of a model and are learned during training. By training the model, soil indices are placed into a three-dimensional Euclidean space where the distance between classes is optimized based on the data to represent the relationship between classes. For example, the distance between soil index 6 (clay) and soil index 5 (silt) might not be equal to that between soil index 1 (gravel) and soil index 2 (sand). Figure 5.20 shows the Euclidean distance between soil index 1 (GM) and all soil indices comparing the soil index introduced by Gillins and Bartlett (2014) and embeddings found in this study. As can be seen in the case of soil index, the distance between gravelly soil with other soil indices increases linearly with grain size. On the other hand, in the embeddings found in this study, the distance between gravelly soil and clay soil is smaller than the distance between gravelly soil and sandy soils. This finding surprisingly matches the fundamentals of liquefaction and lateral spread in the sense that gravelly and clay soils are both less likely to experience liquefaction than sandy soils.

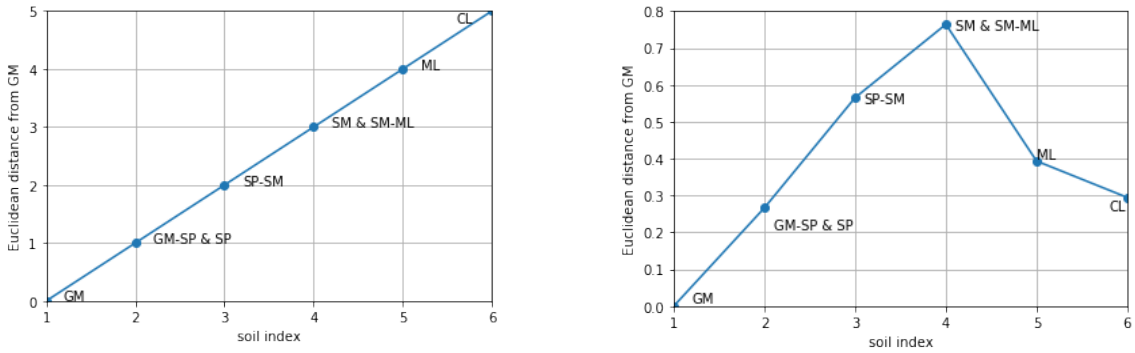


Figure 5.20 Euclidean distance between GM and other soil types in (a) Soil Index defined by Gillins and Bartlett (2014) and (b) our proposed model

The mathematical representation of soil type could help the predictive power by allowing it to learn nonlinear relationships. For instance, assuming liquefaction susceptibility as a linear function of only soil type

$$susceptibility = f(soil\ type) = \alpha S + \beta \tag{5.13}$$

where α and β = function coefficients; and S = soil type. Suppose the soil index introduced

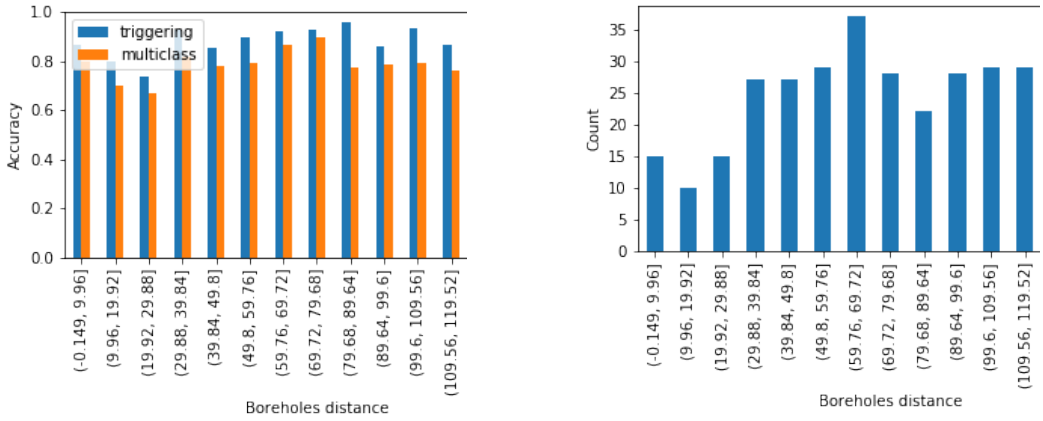
by Gillins and Bartlett (2014) is used for all positive α's, while sandy soils have higher susceptibility than gravelly soils. In that case, clay soils also have higher susceptibility than sandy soils. The opposite is true for all negative α's. On the other hand, using the latent vectors presented in Table 5.11, a linear function similar to Eq. (5.13) could be formed so that sandy soils have higher susceptibility than clay and gravelly soils. The latent vectors found in this study are trained in the context of liquefaction and lateral spread triggering and could be used in conjunction with or as an alternative to Gillins and Bartlett (2014) soil index in liquefaction and lateral spread analysis predictive models.

Table 5.11 Learned latent vectors of soil types

Typical soil descriptions	General USCS symbol	Latent vector
Silty gravel with sand, silty gravel, fine gravel	GM	(0.04199022, 0.0477515, 0.01456286)
Very coarse sand, sand and gravel, gravelly sand	GM-SP	(0.28888193, 0.14305559, 0.04663583)
Coarse sand, sand with some gravel	SP	(0.28888193, 0.14305559, 0.04663583)
Sand, medium to fine sand, sand with some silt	SP-SM	(-0.34709752, 0.39776, 0.22838174)
Fine sand, sand with silt	SM	(-0.29968923, -0.5980874, 0.23763911)
Very fine sand, silty sand, dirty sand, silty/clayey sand	SM-ML	(-0.29968923, -0.5980874, 0.23763911)
Sandy silt, silt with sand	ML	(0.17052442, -0.1486131, -0.3012689)
Silty clay, lean clay	CL	(0.3287804, -0.004597, -0.02813338)

5.6.5 Effect of Distance between Boreholes in a Pair

The dataset was split into 15 buckets based on boreholes distance, and each bucket had a range of approximately 10 meters to investigate the effect of boreholes distance in pairs. The accuracy of the predictions and the number of samples in each bucket were measured for both the multiclass classifier and the triggering model. The results are shown in Figure 5.21.



(a) (b)

Figure 5.21 Effect of boreholes distance in a pair

As can be seen, none of the buckets has significantly lower or higher accuracy. Neither is a monotonically increasing or decreasing pattern in accuracy with respect to borehole distance. Thus, we concluded that the proposed CNN model does not have an internal bias against closer or more distant boreholes.

5.7 Conclusion

This study expanded the conventional probabilistic framework for predicting the magnitude of liquefaction-induced lateral spread by adding a new conditional probability term to account for the probability of lateral spread given liquefaction. A multiclass convolutional neural network classifier was then designed and developed to model the added term to the probabilistic framework by classifying a borehole pair as (1) within, (2) on the boundary, or (3) outside of the lateral spread zone. We proposed that this scheme could be used as a binary triggering classifier to predict the probability of lateral spread triggering given liquefaction. The accuracy of our proposed multiclass classifier and triggering models are, respectively, 81.4% and 90.0%.

As part of this study, a new mathematical representation of soil types was presented. These latent vectors are trained in the context of liquefaction and lateral spread and resulted in a 2% boost in accuracy compared to the soil index developed earlier by Gillins and Bartlett (2014). Although this is a modest improvement, the approach may be helpful in future modeling endeavors. For example, these soil type latent vectors could be used in conjunction with or as a substitute for soil index in developing predictive models for liquefaction or its consequences, including lateral spread and settlement.

The proposed CNN model tends to slightly overpredict the probability of triggering lateral spread. This tendency should be kept in mind when applying the model. Finally, the uncertainty of the CNN model associated with the size of the dataset and folding strategy were investigated. The model's accuracy is expected to have a bell-shaped distribution with a mean of 81.4 percent and a standard deviation 1.6 percent.

REFERENCES

Ambraseys, N. N., & Menu, J. M. (1988). “Earthquake-induced ground displacements.” Earthquake engineering & structural dynamics, 16(7), 985-1006.

Andrus, R. D. (1991). “Liquefaction of gravelly soil at Pence Ranch during the 1983 Borah Peak, Idaho earthquake.” Soil Dynamics and Earthquake Engineering V. Comp. Mech. Publication, 251-262.

Andrus, R. D., & Youd, T. L. (1987). “Subsurface Investigation of a Liquefaction-Induced Lateral Spread, Thousand Springs Valley, Idaho.” Brigham Young University, Provo, UT, Department of Civil Engineering.

Anscombe, F. J. (1973). “Graphs in Statistical Analysis.” American Statistician. 27 (1), 17–21.
Andrus, R. D., & Youd, T. L. (1987). “Subsurface investigation of a liquefaction-induced lateral spread, Thousand Springs Valley, Idaho.”

ASTM. (2008). “D4943, Standard test method for shrinkage factors of soils by the wax method.” Annual book of ASTM standards.

Bardet, J. P., Tobita, T., Mace, N., & Hu, J. (2002). “Regional modeling of liquefaction-induced ground deformation.” Earthquake Spectra, 18(1), 19-46.

Boulanger, R. W., & Idriss, I. M. (2004). “State normalization of penetration resistance and the effect of overburden stress on liquefaction resistance.” Proceedings 11th SDEE and 3rd ICEGE, Uni of California, Berkeley, CA.

Bartlett, S. F., & Youd, T. L. (1990). “State-of-the-Art for Assessing Earthquake Hazards in the United States.” Report 27. Evaluation of Ground Failure Displacement Associated with Soil Liquefaction: Compilation of Case Histories. Brigham Young University, Provo UT, Dept of Civil Engineering.

Bartlett, S. F., & Youd, T. L. (1992). “Empirical analysis of horizontal ground displacement generated by liquefaction-induced lateral spreads.” Brigham Young University, Ph.D. dissertation, 138 p.

Bartlett, S. F., & Youd, T. L. (1992a). “Empirical prediction of lateral spread displacement.” Technical Report, US National Center for Earthquake Engineering Research (NCEER), 1(92-0019), 351-365.

Bartlett, S. F., & Youd, T. L. (1992b). “Case histories of lateral spreads caused by the 1964 Alaska earthquake. In Case studies liquefaction and lifeline performance during past earthquakes: United states case studies.” US National Center for Earthquake Engineering Research (NCEER), 1-127.

Bartlett, S. F., & Youd, T. L. (1995). “Empirical prediction of liquefaction-induced lateral spread.” Journal of Geotechnical Engineering, 121(4), 316-329.

Baziar, M. H., Dobry, R., & Elgamal, A. W. (1992). “Engineering evaluation of permanent ground deformations due to seismically induced liquefaction.”

Baziar, M. H., & Ghorbani, A. (2005). “Evaluation of lateral spreading using artificial neural networks.” Soil Dynamics and Earthquake Engineering, 25(1), 1-9.

Baziar, M. H., & Saeedi Azizkandi, A. (2013). “Evaluation of lateral spreading utilizing artificial neural network and genetic programming.” International Journal of Civil Engineering, (2), 100-111.

Bengio, Y. (2012). “Practical recommendations for gradient-based training of deep architectures.” Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, 437-478.

Bennett, M. J. (1989). “Liquefaction analysis of the 1971 ground failure at the San Fernando Valley Juvenile Hall, California.” Bulletin of the Association of Engineering Geologists, 26(2), 209-226.

Bennett, M. J., McLaughlin, P. V., Sarmiento, J. S., & Youd, T. L. (1984). “Geotechnical investigation of liquefaction sites, Imperial Valley, California.” US Geological Survey Open File Report, 84-252.

Boulanger, R. W., & Idriss, I. M. (2006). “Liquefaction susceptibility criteria for silts and clays.” Journal of geotechnical and geoenvironmental engineering, 132(11), 1413-1426.

Bottou, L. (2012). “Stochastic gradient descent tricks.” Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, 421-436.

Bray, J. D., & Sancio, R. B. (2006). “Assessment of the liquefaction susceptibility of fine-grained soils.” Journal of geotechnical and geoenvironmental engineering, 132(9), 1165-1177.

Brooks, S. P., & Gelman, A. (1998). “General methods for monitoring convergence of iterative simulations.” Journal of computational and graphical statistics, 7(4), 434-455.

Cetin, K. O., Der Kiureghian, A., & Seed, R. B. (2002). “Probabilistic models for the initiation of seismic soil liquefaction.” Structural safety, 24(1), 67-82.

Cetin, K. O., Seed, R. B., Der Kiureghian, A., Tokimatsu, K., Harder Jr, L. F., Kayen, R. E., & Moss, R. E. (2004). “Standard penetration test-based probabilistic and deterministic assessment of seismic soil liquefaction potential.” Journal of geotechnical and geoenvironmental engineering, 130(12), 1314-1340.

Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). “Deep learning-based crack damage detection using convolutional neural networks.” Computer-Aided Civil and Infrastructure Engineering, 32(5), 361-378.

Chu, D. B., Stewart, J. P., Lee, S., Tsai, J. S., Lin, P. S., Chu, B. L., ... & Wang, M. C. (2004). “Documentation of soil conditions at liquefaction and non-liquefaction sites from 1999 Chi–Chi (Taiwan) earthquake.” Soil Dynamics and Earthquake Engineering, 24(9-10), 647-657.

Chung, R. M. (1996). “January 17, 1995 Hyogoken-Nanbu (Kobe) earthquake: Performance of structures, lifelines, and fire protection systems (NIST SP 901).”

Civil Engineering. (1997). “Deep soil mix wall - a first for Wisconsin.” American Society of Civil Engineers, Reston, Va., 67(6), 18.

Coffman, J. L., Von Hake, C. A., & Stover, C. W. (Eds.). (1982). “Earthquake history of the United States.” US Department of Commerce, National Oceanic and Atmospheric Administration and US Department of the Interior, Geological Survey, 55.

Cortwright, C. (1975). “Effects of the San Fernando Earthquake on the Van Norman Reservoir Complex.” San Fernando, California Earthquake of 9 February, 1971 Bulletin 196, California Division of Mines and Geology, Sacramento, Ca, 395-406

Deterling, O. C. (2015). “Factors influencing the lateral spread displacement from the 2011 Christchurch, New Zealand earthquake.” ⟨https://repositories.lib.utexas.edu/bitstream/handle/2152/31790/⟩ (Jun. 19, 2017)

Dietterich, T. (1995). “Overfitting and undercomputing in machine learning.” ACM computing surveys (CSUR), 27(3), 326-327.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). “Hybrid monte carlo.” Physics letters B, 195(2), 216-222.

Dobry, R., Thevanayagam, S., Medina, C., Bethapudi, R., Elgamal, A., Bennett, V., ... & Mercado, V. M. (2010). “Mechanics of lateral spreading observed in a full-scale shake test.” Journal of geotechnical and geoenvironmental engineering, 137(2), 115-129.

Dubois P.F. (2007). “Python: Batteries Included,” Computing in Science & Engineering. IEEE/AIP, May 2007, 9.

Durante, M. G., & Rathje, E. M. (2021). “An exploration of the use of machine learning to predict lateral spreading.” Earthquake Spectra, 37(4), 2288-2314.

Edwards, W., Lindman, H., & Savage, L. J. (1963). “Bayesian statistical inference for psychological research.” Psychological review, 70(3), 193.

El Shamy, U., Zeghal, M., Dobry, R., Thevanayagam, S., Elgamal, A., Abdoun, T., ... & Bennett, V. (2010). “Micromechanical aspects of liquefaction-induced lateral spreading.” International Journal of Geomechanics, 10(5), 190-201.

Elliott, J. R., Nissen, E. K., England, P. C., Jackson, J. A., Lamb, S., Li, Z., ... & Parsons, B. (2012). “Slip in the 2010–2011 Canterbury earthquakes, New Zealand.” Journal of Geophysical Research: Solid Earth, 117(B3).

Finn, W. D. (1991). “Assessment of liquefaction potential and post-liquefaction behavior of earth structures: Developments 1981-1991.”

Finn, W. D., Ledbetter, R. H., & Wu, G. (1994, October). “Liquefaction in silty soils: design and analysis.” In Ground failures under seismic conditions, ASCE, 51-76.

Franke, K. W., & Kramer, S. L. (2013). “Procedure for the empirical evaluation of lateral spread displacement hazard curves.” Journal of Geotechnical and Geoenvironmental Engineering, 140(1), 110-120.

Garcia, S. R., Romo, M. P., & Botero, E. (2008). “A neurofuzzy system to analyze liquefaction-induced lateral spread.” Soil Dynamics and Earthquake Engineering, 28(3), 169-180.

Garson, G. D. (1991). “Interpreting neural-network connection weights.” AI expert, 6(4), 46-51.

Gelman, A., & Rubin, D. B. (1992). “Inference from iterative simulation using multiple sequences.” Statistical science, 7(4), 457-472.

Géron, A. (2017). “Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems.” O'Reilly Media, Inc.

Gillins, D. T. (2012). “Mapping the probability and uncertainty of liquefaction-induced ground failure.” The University of Utah.

Gillins, D. T., & Bartlett, S. F. (2013). “Multilinear regression equations for predicting lateral spread displacement from soil type and cone penetration test data.” Journal of Geotechnical and Geoenvironmental Engineering, 140(4), 04013047.

Goh, A. T., & Zhang, W. G. (2014). “An improvement to MLR model for predicting liquefaction-induced lateral spread using multivariate adaptive regression splines.” Engineering geology, 170, 1-10.

Golesorkhi, R. (1989). “Factors influencing the computational determination of earthquake-induced shear stresses in sandy soils.” University of California, Berkeley.

Gowda, G. B., Dinesh, S. V., Govindaraju, L., & Babu, R. R. (2022). “Effect of Liquefaction Induced Lateral Spreading on Seismic Performance of Pile Foundations.” Civil Engineering Journal, 7, 58-70.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). “Recent advances in convolutional neural networks.” Pattern Recognition, 77, 354-377.

Guan, Z., & Wang, Y. (2022). “Assessment of Liquefaction-Induced Differential Ground Settlement and Lateral Displacement Using Standard Penetration Tests with Consideration of Soil Spatial Variability.” Journal of Geotechnical and Geoenvironmental Engineering, 148(5), 04022018.

Hadush, S., Yashima, A., & Uzuoka, R. (2000). “Importance of viscous fluid characteristics in liquefaction induced lateral spreading analysis.” Computers and Geotechnics, 27(3), 199-224.

Hadush, S., Yashima, A., Uzuoka, R., Moriguchi, S., & Sawada, K. (2001). “Liquefaction induced lateral spread analysis using the CIP method.” Computers and Geotechnics, 28(8), 549-574.

Hamada, M., Yasuda, S., Isoyama, R., & Emoto, K. (1986). “Study on liquefaction induced permanent ground displacements.” Report for the Association for the Development of Earthquake Prediction, 87.

Hansen, W. R. (1966). “The Alaska earthquake, March 27, 1964: Field investigations and reconstruction effort.” US Government Printing Office, 541.

Holzer, T.L., Bennett M.J., Youd T.L. (1988). “U.S. case histories of liquefaction-induced ground displacement.” Proceedings of 1st Japan- U.S. Workshop on Liquefaction, Large Ground Deformation and Their Effects on Lifeline Facilities, 22-31.

Holzer, T. L., Hanks, T. C., & Youd, T. L. (1989). “Dynamics of liquefaction during the 1987 Superstition Hills, California, earthquake.” Science, 244(4900), 56-59.

Hovland, H. J., & Darragh, R. D. (1981, August). “Earthquake-induced ground movements in the Mission Bay area of San Francisco in 1906.” In Lifeline Earthquake Engineering: The Current State of Knowledge, 1981, ASCE, 293-309.

Honegger, D. G., Wijewickreme, D., & Youd, T. L. (2014, July). “Regional pipeline vulnerability assessment based upon probabilistic lateral spread hazard characterization.” In Proc., 10th National Conf. on Earthquake Engineering.

Huang, Y., & Wen, Z. (2015). “Recent developments of soil improvement methods for seismic liquefaction mitigation.” Natural Hazards, 76(3), 1927-1938.

Idriss, I. M. (1999). “An update to the Seed-Idriss simplified procedure for evaluating liquefaction potential.” Proc., TRB Worshop on New Approaches to Liquefaction, Pubbl. n. FHWA-RD-99-165, Federal Highway Administation.

Idriss, I. M., & Boulanger, R. W. (2006). “Semi-empirical procedures for evaluating liquefaction potential during earthquakes.” Soil Dynamics and Earthquake Engineering, 26(2-4), 115-130.

Idriss, I. M., & Boulanger, R. W. (2008). “Soil liquefaction during earthquakes.” Earthquake Engineering Research Institute.

Ishihara, K., Acacio, A. A., & Towhata, I. (1993). “Liquefaction-induced ground damage in Dagupan in the July 16, 1990 Luzon earthquake.” Soils and Foundations, 33(1), 133-154.

Javadi, A. A., Rezania, M., & Nezhad, M. M. (2006). “Evaluation of liquefaction induced lateral displacements using genetic programming.” Computers and Geotechnics, 33(4-5), 222-233.

Jefferies, M., & Been, K. (2015). “Soil liquefaction: a critical state approach.” CRC press.
Kanamori, H. (1978). “Quantification of earthquakes.” Nature, 271(5644), 411-414.

Kawasumi, H. (1968). “General report on the Niigata Earthquake of 1964.” Tokyo, Japan: Electrical Engineering College Press, 550.

Kaya, Z. (2016). “Predicting liquefaction-induced lateral spreading by using neural network and neuro-fuzz y techniques.” International Journal of Geomechanics, 16(4), 04015095.

Khoshnevisan, S., Juang, H., Zhou, Y. G., & Gong, W. (2015). “Probabilistic assessment of liquefaction-induced lateral spreads using CPT—focusing on the 2010–2011 Canterbury earthquake sequence.” Engineering geology, 192, 113-128.

Kutter, B. L., Gajan, S., Manda, K. K., & Balakrishnan, A. (2004). “Effects of layer thickness and density on settlement and lateral spreading.” Journal of Geotechnical and Geoenvironmental Engineering, 130(6), 603-614.

Kuwano, J. and Ishihara, K. (1988). “Analysis of permanent deformation of earth dams due to earthquakes.” Soils and Foundations 28, 41–55.

LeCun, Y. A., Bottou, L., Orr, G. B., & Müller, K. R. (2012). “Efficient backprop.” Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, 9-48.

LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E., & Jackel, L. D. (1990). “Handwritten digit recognition with a back-propagation network.” Advances in neural information processing systems, 396-404.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). “Gradient-based learning applied to document recognition.” Proceedings of the IEEE, 86(11), 2278-2324.

Lee, P. M. (1989). “Bayesian statistics.” London: Oxford University Press, 54-55.

Liao, S. S., & Whitman, R. V. (1986). “A catalog of liquefaction and non-liquefaction occurrences during earthquakes.” Department of Civil Engineering, MIT.

Liao, S. S., Veneziano, D., & Whitman, R. V. (1988). “Regression models for evaluating liquefaction probability.” Journal of Geotechnical Engineering, 114(4), 389-411.

Liao SSC (1986). “Statistical modeling of earthquake-induced liquefaction.” PhD dissertation. Department of Civil Engineering, Massachusetts Institute of Technology. 470p.

Little, M., & Rathje, E. (2021). “Key Trends Regarding the Effects of Site Geometry on Lateral Spreading Displacements.” Journal of Geotechnical and Geoenvironmental Engineering, 147(12), 04021142.

Liu, F., Li, Z., Jiang, M., Frattini, P., & Crosta, G. (2016). “Quantitative liquefaction-induced lateral spread hazard mapping.” Engineering Geology, 207, 36-47.

Liu, K., Zhou, Y. G., She, Y., Xia, P., Chen, Y. M., Ling, D. S., & Huang, B. (2018). “Revisit of the empirical prediction methods for liquefaction-induced lateral spread by using the LEAP centrifuge model tests.” In Physical Modelling in Geotechnics (pp. 407-412). CRC Press.

Makdisi, A. J., & Kramer, S. L. (2019). “Applicability of sliding block analyses for lateral spreading problems.” Soil Dynamics and Earthquake Engineering, 124, 374-388.

Milmann K.J. and Avaizis M., editors (2011). “Scientific Python,” Computing in Science & Engineering. IEEE/AIP, March 2011, 11

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). “Efficient estimation of word representations in vector space.” arXiv preprint arXiv:1301.3781.

Nair, V., & Hinton, G. E. (2010, January). “Rectified linear units improve restricted boltzmann machines.” ICML.

Narazaki, Y., Hoskere, V., Hoang, T. A., & Spencer Jr, B. F. (2018). “Automated vision-based bridge component extraction using multiscale convolutional neural networks.” arXiv preprint arXiv:1805.06042.

National Academies of Sciences, Engineering, and Medicine (2016). “State of the Art and Practice in the Assessment of Earthquake-Induced Soil Liquefaction and Its Consequences.” Washington, DC: The National Academies Press. https://doi.org/10.17226/23474.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). “Applied linear statistical models.” Chicago: Irwin.

Olson, S. M., & Johnson, C. I. (2008). “Analyzing liquefaction-induced lateral spreads using strength ratios.” Journal of geotechnical and geoenvironmental engineering, 134(8), 1035-1049.

Oommen, T., & Baise, L. G. (2010). “Model development and validation for intelligent data collection for lateral spread displacements.” Journal of Computing in Civil Engineering, 24(6), 467-477.

O'Rourke, T. D., & Hamada, M. (1992). “Large ground deformations and their effects on lifeline facilities: 1971 San Fernando earthquake.” Case studies liquefaction and lifeline performance during past earthquakes: United states case studies. US National Center for Earthquake Engineering Research (NCEER), 1-85.

O’Rourke, T. D., & Hamada, M. (1992). “Case Studies of Liquefaction and Lifeline Performance During Past Earthquakes: Volume 2, United States Case Studies.” National Center for Earthquake Engineering Research, State University of New York, Buffalo, NY.

O'Rourke, T. D., & Lane, P. A. (1989). “Liquefaction hazards and their effects on buried pipelines.” Technical Report NCEER-89-0007. National Center for Earthquake Enineering Research, Buffalo, NY.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). “Glove: Global vectors for word representation.” In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 1532-1543.

Potter, S. H., Becker, J. S., Johnston, D. M., & Rossiter, K. P. (2015). “An overview of the impacts of the 2010-2011 Canterbury earthquakes.” International Journal of Disaster Risk Reduction, 14, 6-14.

Rashid, T. (2016). “Make your own neural network.” CreateSpace Independent Publishing Platform.

Rauch, A. F., & Martin III, J. R. (2000). “EPOLLS model for predicting average displacements on lateral spreads.” Journal of Geotechnical and Geoenvironmental Engineering, 126(4), 360-371.

Reid, S., Tibshirani, R., Freidman, J. (2016). “A study of error variance estimation in lasso regression.” Statistica Sinica, 35-67.

Rezania, M., Faramarzi, A., & Javadi, A. A. (2011). “An evolutionary based approach for assessment of earthquake-induced soil liquefaction and lateral displacement.” Engineering Applications of Artificial Intelligence, 24(1), 142-153.

Robertson, P. K., Wride, C. E., Boulanger, R. W., Mejia, L. H., & Idriss, I. M. (1999). “Liquefaction at Moss Landing during Loma Prieta Earthquake.” Journal of Geotechnical and Geoenvironmental Engineering, 125(1), 91-95.

Ross, G., Seed, H., & Migliacio, R. (1964). “Performance of highway bridge foundations in the great Alaska earthquake of 1964.” Committee on the Alaskan Earthquake of the Division of Earth Sciences National Research Council. The Great Alaska Earthquake of 1964.

Roth, R. A., & KAVAZANJIAN JR, E. D. W. A. R. D. (1984). “Liquefaction Susceptibility Mapping for San Francisco, California.” Bulletin of the Association of Engineering Geologists, 21(4), 459-478.

Salgado, R. (2008). “The engineering of foundations.” New York: McGraw-Hill, 888.

Santosa, F. and Symes, W. W. (1986). “Linear inversion of band-limited reflection seismograms.”

SIAM Journal on Scientific and Statistical Computing 7, 1307–1330.

Seed, H. B., & Idriss, I. M. (1971). “Simplified procedure for evaluating soil liquefaction potential.” Journal of Soil Mechanics & Foundations Div.

Seed, H.B., Tokimatsu, K., Harder, L. F., & Chung, R. M. (1985). “Influence of SPT procedures in soil liquefaction resistance evaluations.” Journal of Geotechnical Engineering, 111(12), 1425-1445.

Sharp, M. K., Dobry, R., and Abdoun, T. (2003). “Liquefaction centrifuge modeling of sands of different permeability.” Journal of geotechnical and geoenvironmental engineering 129, 1083–1091.

Soroush, A., & Koohi, S. (2004). “Liquefaction-induced lateral spreading- an overview and numerical analysis.” International Journal of Civil Engineering, 2(4).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). “Dropout: a simple way to prevent neural networks from overfitting.” The journal of machine learning research, 15(1), 1929-1958.

Taboada-Urtuzuastegui, V. M., & Dobry, R. (1998). “Centrifuge modeling of earthquake-induced lateral spreading in sand.” Journal of geotechnical and geoenvironmental engineering, 124(12), 1195-1206.

Team, T. T. D., Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., ... & van Tulder, G. (2016). “Theano: A Python framework for fast computation of mathematical expressions.” arXiv preprint arXiv:1605.02688.

Theil, Henri (1961). “Economic Forecasts and Policy. Holland, Amsterdam: North.” 213.

Tibshirani, R. (1996). “Regression shrinkage and selection via the lasso.” Journal of the Royal Statistical Society. Series B (Methodological) pp. 267–288.

Tibshirani, R., and Friedman, J. (2016). “A study of error variance estimation in lasso regression.” Statistica Sinica, 35–67.

Tokida, K., Matsumoto, H., Azuma, T., & Towhata, I. (1970). “Simplified procedure to estimate lateral ground flow by soil liquefaction.” WIT Transactions on The Built Environment, 3.

Tokimatsu, K., Kojima, H., Kuwayama, S., Abe, A., & Midorikawa, S. (1994). “Liquefaction-induced damage to buildings in 1990 Luzon earthquake.” Journal of Geotechnical Engineering, 120(2), 290-307.

Towhata, I., Sasaki, Y., Tokida, K., Matsumoto, H., Tamari, Y., & Yamada, K. (1992). “Prediction of permanent displacement of liquefied ground by means of minimum energy principle.” Soils and Foundations, 32(3), 97-116.

Valsamis, A. I., Bouckovalas, G. D., and Papadimitriou, A. G. (2010). “Parametric investigation of lateral spreading of gently sloping liquefied ground.” Soil Dynamics and Earthquake Engineering 30, 490–508.

Van der Walt, S., Colbert, S.C., and Varoquaux, G. (2011). “The NumPy array: A structure for efficient numerical computation.” Computing in Science and Engineering, 11.

Vellido, A., Martín-Guerrero, J. D., & Lisboa, P. J. (2012, April). “Making machine learning models interpretable.” ESANN, 12, 163-172.

Vargas, R. R., Tang, Z., Ueda, K., & Uzuoka, R. (2022). “Validation of numerical predictions for liquefaction phenomenon–Lateral spreading in clean sands.” Soils and Foundations, 62(1), 101101.

Wang, J., & Rahman, M. S. (1999). “A neural network model for liquefaction-induced horizontal ground displacement.” Soil Dynamics and Earthquake Engineering, 18(8), 555-568.

Woldesellasse, H., & Tesfamariam, S. (2022). “Prediction of lateral spreading displacement using conditional Generative Adversarial Network (cGAN).” Soil Dynamics and Earthquake Engineering, 156, 107214.

Yang, Y., & Kavazanjian Jr, E. (2021). “Numerical evaluation of liquefaction-induced lateral spreading with an advanced plasticity model for liquefiable sand.” Soil Dynamics and Earthquake Engineering, 149, 106871.

Yegian, M. K., Marciano, E. A., & Ghahraman, V. G. (1991). “Earthquake-induced permanent deformations: probabilistic approach.” Journal of Geotechnical Engineering, 117(1), 35-50.
Youd, T. L. (1973). “Ground Movements in Van Norman Lake Vicinity During San Fernando Earthquake.” San Fernando, California Earthquake of February 9, 3, 197-206.

Youd, T. L. (2018). “Application of MLR procedure for prediction of liquefaction-induced lateral spread displacement.” Journal of Geotechnical and Geoenvironmental Engineering, 144(6), 04018033.

Youd, T. L., & Hoose, S. N. (1978). “Historic ground failures in northern California triggered by earthquakes.” US Government Printing Office, 993.

Youd, T. L., & Perkins, D. M. (1978). “Mapping liquefaction-induced ground failure potential.” Journal of the Soil Mechanics and Foundations Division, 104(4), 433-446.

Youd, T. L., & Bennett, M. J. (1983). “Liquefaction sites, Imperial Valley, California.” Journal of Geotechnical Engineering, 109(3), 440-457.

Youd, T. L., Harp, E. L., Keefer, D. K., & Wilson, R. C. (1985). "The Borah Peak, Idaho Earthquake of October 28, 1983—Liquefaction." Earthquake spectra, 2(1), 71-89.

Youd, T. L., and Perkins D. M. (1987). "Mapping of Liquefaction Severity Index," Journal of Geotechnical Engineering, ASCE, Vol. 113, No. 11 pp. 1374-1392.

Youd, T. L., & Bartlett, S. F. (1988, November). "US case histories of liquefaction-induced ground displacement." 1st Japan-United States Workshop on Liquefaction, Large Ground Deformation and Their Effects on Lifeline Facilities, 22-31.

Youd, T. L., DeDen, D. W., Bray, J. D., Sancio, R., Cetin, K. O., & Gerber, T. M. (2009). "Zero-displacement lateral spreads, 1999 Kocaeli, Turkey, earthquake." Journal of Geotechnical and Geoenvironmental Engineering, 135(1), 46-61.

Youd, T. L., Hansen, C. M., & Bartlett, S. F. (2002). "Revised multilinear regression equations for prediction of lateral spread displacement." Journal of Geotechnical and Geoenvironmental Engineering, 128(12), 1007-1017.

Youd, T. L., & Perkins, D. M. (1978). "Mapping liquefaction-induced ground failure potential." Journal of the Soil Mechanics and Foundations Division, 104(4), 433-446.

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2016). "Understanding deep learning requires rethinking generalization." arXiv preprint arXiv:1611.03530

Zhang, G., Robertson, P. K., & Brachman, R. W. I. (2004). "Estimating liquefaction-induced lateral displacements using the standard penetration test or cone penetration test." Journal of Geotechnical and Geoenvironmental Engineering, 130(8), 861-871.

Zhang, J., Yang, C., Zhao, J. X., & McVerry, G. H. (2012). "Empirical models for predicting lateral spreading considering the effect of regional seismicity." Earthquake Engineering and Engineering Vibration, 11(1), 121-131.

Zhang, J. and Zhao, J. X. (2005). "Empirical models for estimating liquefaction-induced lateral spread displacement." Soil Dynamics and Earthquake Engineering 25, 439–450.

APPENDIX A. EVALUATION METRICS

Confusion Matrix

Confusion matrix or error matrix is a specific table layout in the field of machine learning that allows visualization of the performance of a supervised algorithm specifically for the problem of classification. Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class or vice versa. Table A.1 shows an example of confusion matrix for a classification problem with two classes. It has four main components to it

- True positive (TP): actual class is positive and model correctly predicts it as positive,
- False negative (FN): actual class is positive but model falsely predicts it as negative; also known as Type II error,
- False positive (FP): actual class is negative but model falsely predicts it as positive; also known as Type I error, and
- True negative (TN): actual class is positive and model correctly predicts it as positive.

Table A.1. An example confusion matrix

		Predicted class	
		Positive	Negative
Actual class	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

Accuracy

Accuracy of a model is the fraction of true predictions divided by count of all samples

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{A.1}$$

Precision

Precision is the fraction of relevant instances among the retrieved instance

$$Precision = \frac{TP}{TP + FP} \tag{A.2}$$

Recall

Recall, also known as sensitivity, is the fraction of retrieved relevant instances among all relevant instances

$$Recall = \frac{TP}{TP + FN} \tag{A.3}$$

F1 score

F-measure or F1 score is a measure that combines precision and recall i.e. F1 score is the harmonic mean of precision and recall

$$F = 2 \times \frac{precision \times recall}{precision + recall} \tag{A.4}$$

APPENDIX B. LIQUPY: OPEN-SOURCE PYTHON LIBRARY FOR SOIL LIQUEFACTION AND LATERAL SPREAD ANALYSIS

Introduction

The prediction of soil liquefaction and its consequences (i.e. ground deformation) is among the most complex problems in geotechnical engineering. Many researchers have conducted research to evaluate various aspects of this phenomena. Several analytical, empirical, numerical, and semi-empirical methods have been developed for liquefaction triggering and prediction of horizontal ground displacements, settlement, etc.

Python was selected as the programming language because it is one of the most popular programming languages among the scientific community. It has been a very appealing choice for exploratory data analysis and algorithmic development because of its high-level interactive nature and its matured ecosystem of scientific libraries (Dubois, 2007; Milmann and Avaizis, 2011).

LiquPy uses this rich environment to provide state-of-the-art implementations of many well-known soil liquefaction and lateral spread methods, while maintaining an easy-to-use interface tightly integrated with the Python programming language. This answers the growing need for soil liquefaction and lateral spread analysis by non-specialists.

LiquPy is distributed under the New BSD License, details of which are available at https://github.com/LiquPy/LiquPy/blob/master/license.txt

Project Vision

We envisioned to develop consistent quality code with comprehensive documentation to which the community could contribute on an ongoing basis and release it under BSD licensing.

- **Code quality** - Emphasis is put on providing solid implementations rather than as many features as possible. We strive to use consistent naming for the methods, variables, and arguments used throughout a strict adherence to the Python coding guidelines and numpy style documentation.
- **BSD licensing** - Most of the Python ecosystem is licensed with non-copyleft licenses. While such policy is beneficial for adoption of these tools by commercial projects, it does impose some restrictions.
- **Community-driven development** - We base our development on collaborative tools such as git, github and public mailing lists. External contributions are welcome and encouraged.
- **Documentation** - LiquPy provides documentation for all objects, methods, functions, and variables as well as examples of real-world applications if applicable.

Underlying Technologies

Despite that it is strived to develop LiquPy as independent as possible, it is based on top of some of existing extensively developed and widely used libraries of Python namely Numpy, Pandas, Scikit-learn and Matplotlib.

- **Numpy** - The base data structure used for data and model parameters. Input data is

presented as numpy arrays, thus, integrating seamlessly with other scientific Python libraries. Numpy's view-based memory model limits copies, even when binding with compiled code (Van der Walt et al., 2011). It also provides basic arithmetic operations.

- **Pandas** - Tabular data such as subsurface data or other data sets are presented as Pandas DataFrame to make use of its high-performance, easy-to-use and interpret data structures and data analysis tools for Python programming language.
- **Scikit-learn** - With the advent of machine learning and new statistical predictive models, many of empirical models exploited them to predict horizontal ground motions due to lateral spreads. Though these methods seem promising in accuracy, they are not easy to implement or reproduce. The neural network presented in Wang and Rahman (1999) is an excellent example of such a problem. They evaluated applicability of artificial neural networks on prediction of horizontal ground displacement but their result is not reproducible. The same issue is associated with Garcia et al. (2008). Moreover, more complex models such as deep neural networks or random forests could not be elaborated in papers in a way that make the reproducable. Scikit-learn which is a Python library for machine learning could help save trained predictive models and use them to make new predictions.
- **Matplotlib** – Graphical representations of LiquPy uses Matplotlib, a Python plotting library which is used to produce publication quality figures in a variety of formats.

Benefits

The LiquPy have several benefits for the community including its transparency and reproducibility, enforcing limitations, ease of use, and verification and validation.

- **Transparency and reproducibility** - Being an open-source library, LiquPy enables researchers to go over the methods themselves and verify/reproduce the results of different approaches with ease. Most of the complex machine learning models cannot be described in papers properly. This could be the reason why all the artificial neural networks developed for horizontal ground displacements of lateral spreads have only one layer with a few nodes. The LiquPy library allows researchers to save their most complex trained models into LiquPy for others to make predictions and reproduce the exact results.
- **Enforcing methods' limitations** - In many applications, the end users (researchers and/or engineers) simply apply the methods without paying attention to the limitations of each method. Using LiquPy library, one can make sure that the limitations of each method is enforced and users will at least get a warning otherwise. Not paying attention to the limitations of each method could result in systematic underprediction/overprediction and consequently in disaster / high cost of ground modifications. One of such instances is the extrapolation of empirical methods.
- **Ease of use** - LiquPy provides researchers/engineers with prefabricated functions of the soil liquefaction and lateral spreads analysis methods which would save them the programming efforts and the time required to repeat the researches done earlier by others.
- **Validation and verification** - Since LiquPy is open-source and transparent to everyone it could be verified by everyone and will be updated over time based on the feedbacks from researchers who use it and those whose methods are implemented into LiquPy. Another benefit it contains is that erratum to papers could be easily included in the library while otherwise they may be missed by researchers.

Existing Features and Future Development Plan

Currently, the following simplified liquefaction triggering and lateral spread analyses are available in LiquPy:

Simplified liqufaction triggering analysis:

- Cyclic resistance ratio based on Boulanger and Idriss (2004) & Boulanger and Idriss (2008)
- Adjustments for fines content from (1) Boulanger and Idriss (2004), and (2) Cetin et al. (2004)
- Shear stress reduction factors available from (1) Golesorkhi 1989, (2) Idriss 1999, and (3) Liao & Whitman 1986
- Magnitude scaling factor from Idriss (1999)
- Overburden correction factor from Boulanger and Idriss (2004)
- Triggering correlation of liquefaction in clean sands from Idriss and Boulanger (2004)
- Probabilistic approaches from Cetin et al. (2004)

Lateral spread analysis:

- Multi Linear Regression (MLR) (Youd, Hansen, & Bartlett 2002)
- Multi Linear Regression (MLR) (Bardet et al. 2002)
- Genetic programming (Javadi et al. 2006)
- Evolutionary-based approach (Rezania et al. 2011)
- Artificial Neural Network & Genetic Algorithm (Baziar & Azizkani 2013)
- Multivariate Adaptive Regression Splines (MARS) (Goh et al. 2014)
- Lateral Displacement Index (LDI) and settlement (Zhang et al. 2004)

However, the authors are consistently adding more features. Researchers and engineers could also easily merge their codes into the library for the community to use.

Conclusion

LiquPy exposes a wide variety of soil liquefaction and lateral spread methods using a consistent, task-oriented interface and enables easy comparison of methods for a given data point or set. Since it is implemented in the high-level Python programming language, it can be easily integrated into other applications such as development of liquefaction potential hazard mapping. Besides, this library is transparent, easy-to-use and enforces methods' limitations.

APPENDIX C. MONTE-CARLO BASED SENSITIVITY ANALYSIS OF EMPIRICAL MODELS

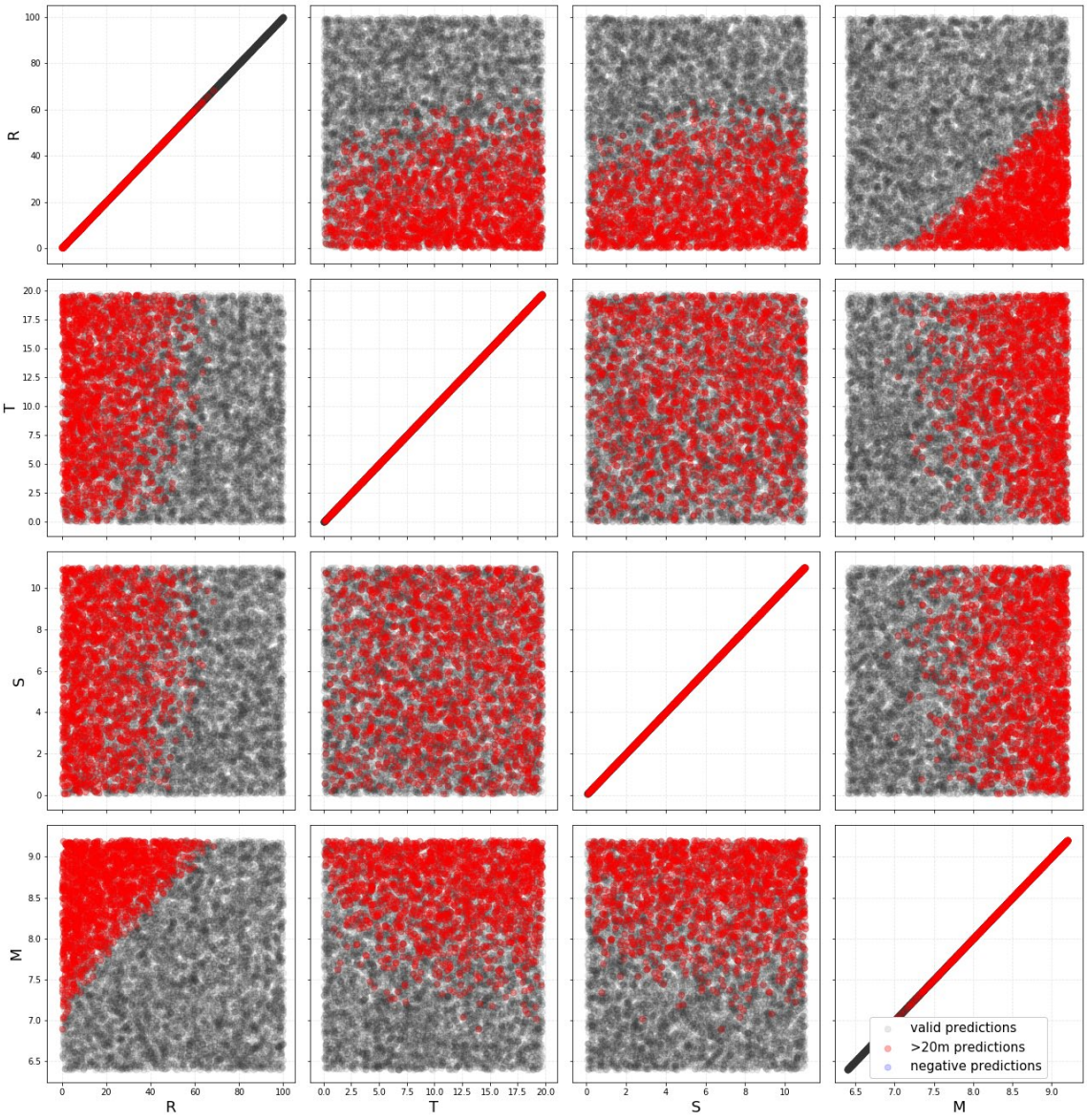


Figure C.1 Results of Monte Carlo analysis on sloping-ground case of Bardet et al. (2002)

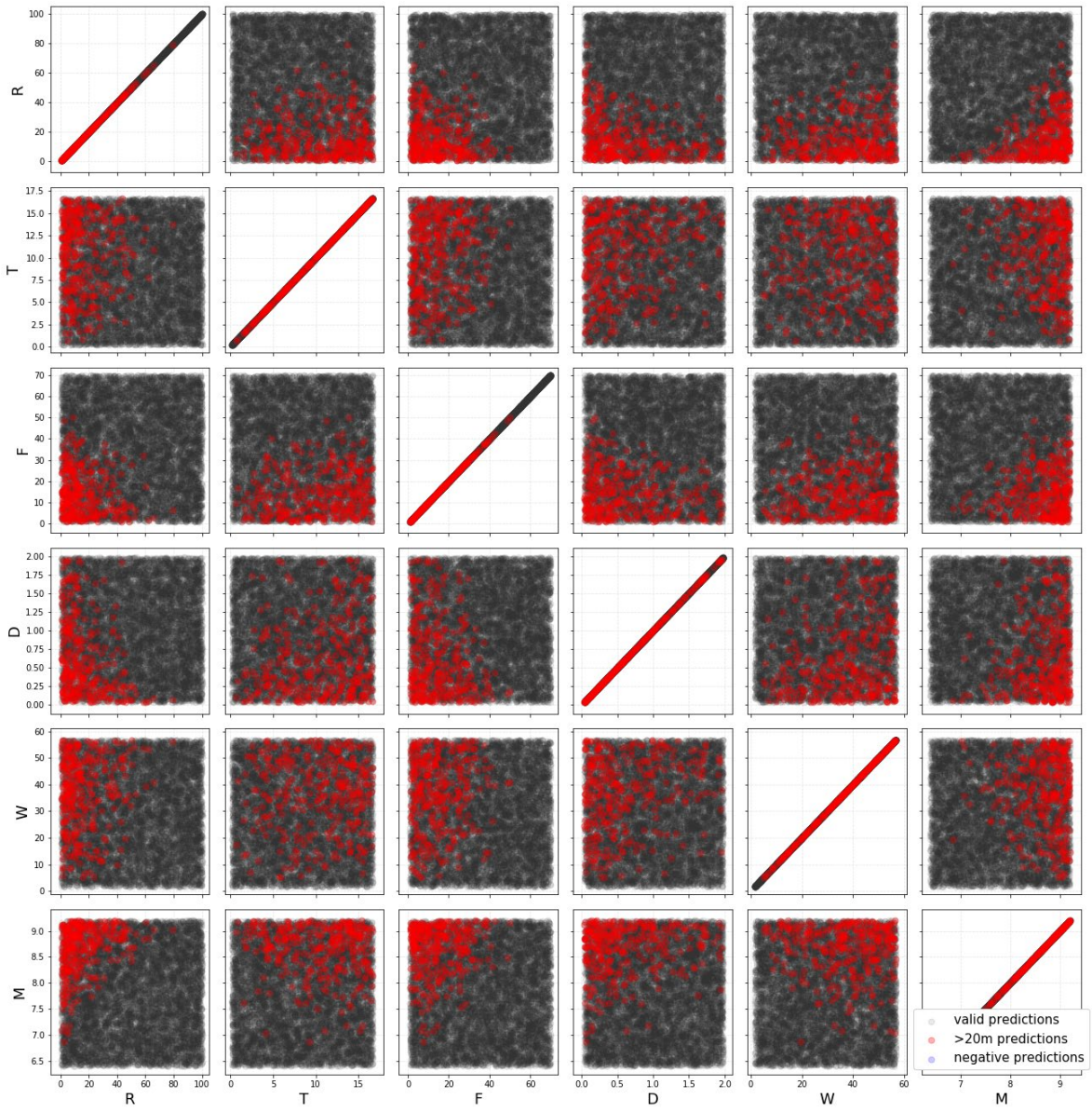


Figure C.2 Results of Monte Carlo analysis on Free-face case of Youd et al. (2002)

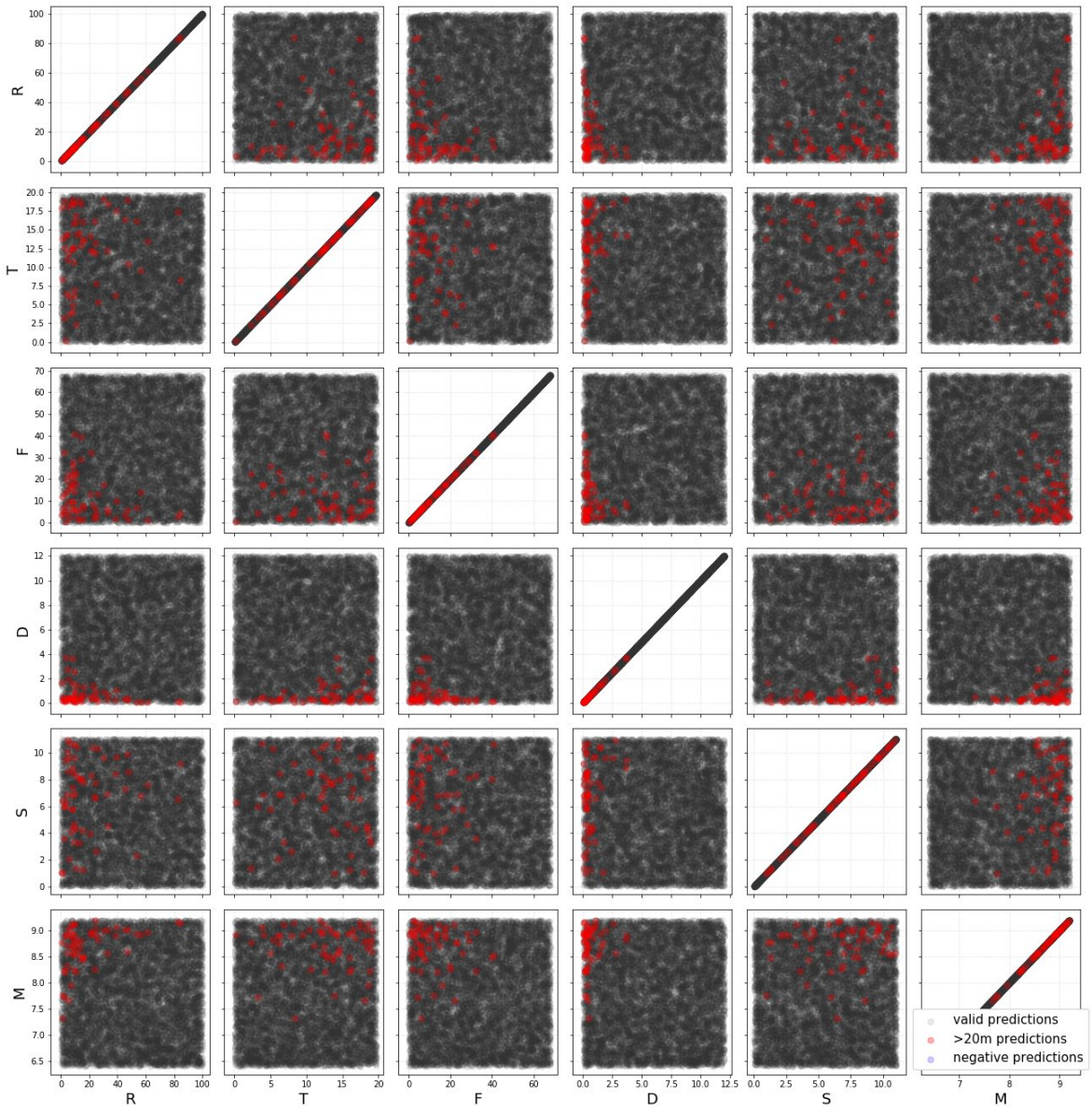


Figure C.3 Results of Monte Carlo analysis on sloping-ground case of Youd et al. (2002)

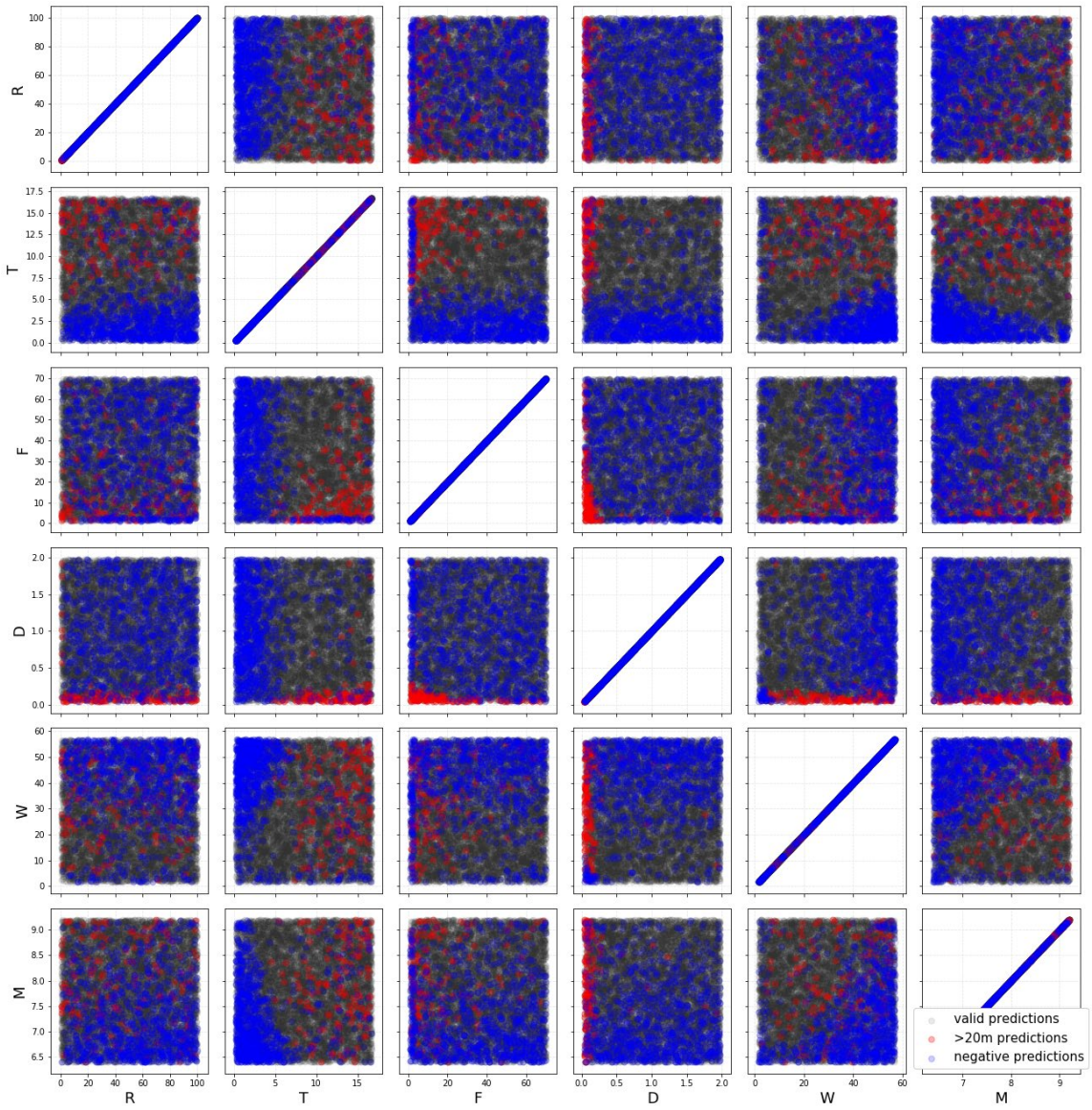


Figure C.4 Results of Monte Carlo analysis on Free-face case of Javadi et al. (2006)

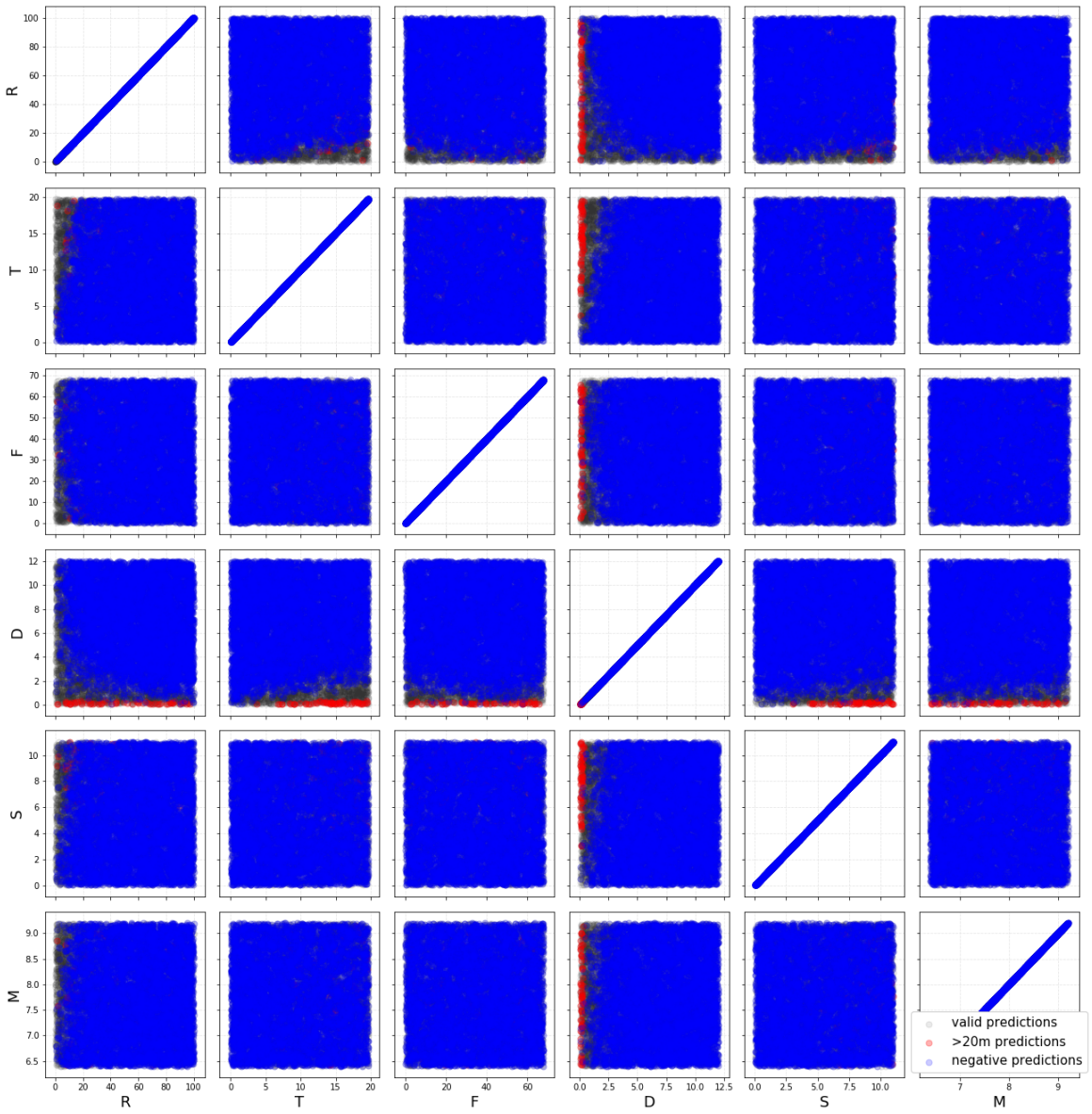


Figure C.5 Results of Monte Carlo analysis on sloping-ground case of Javadi et al. (2006)

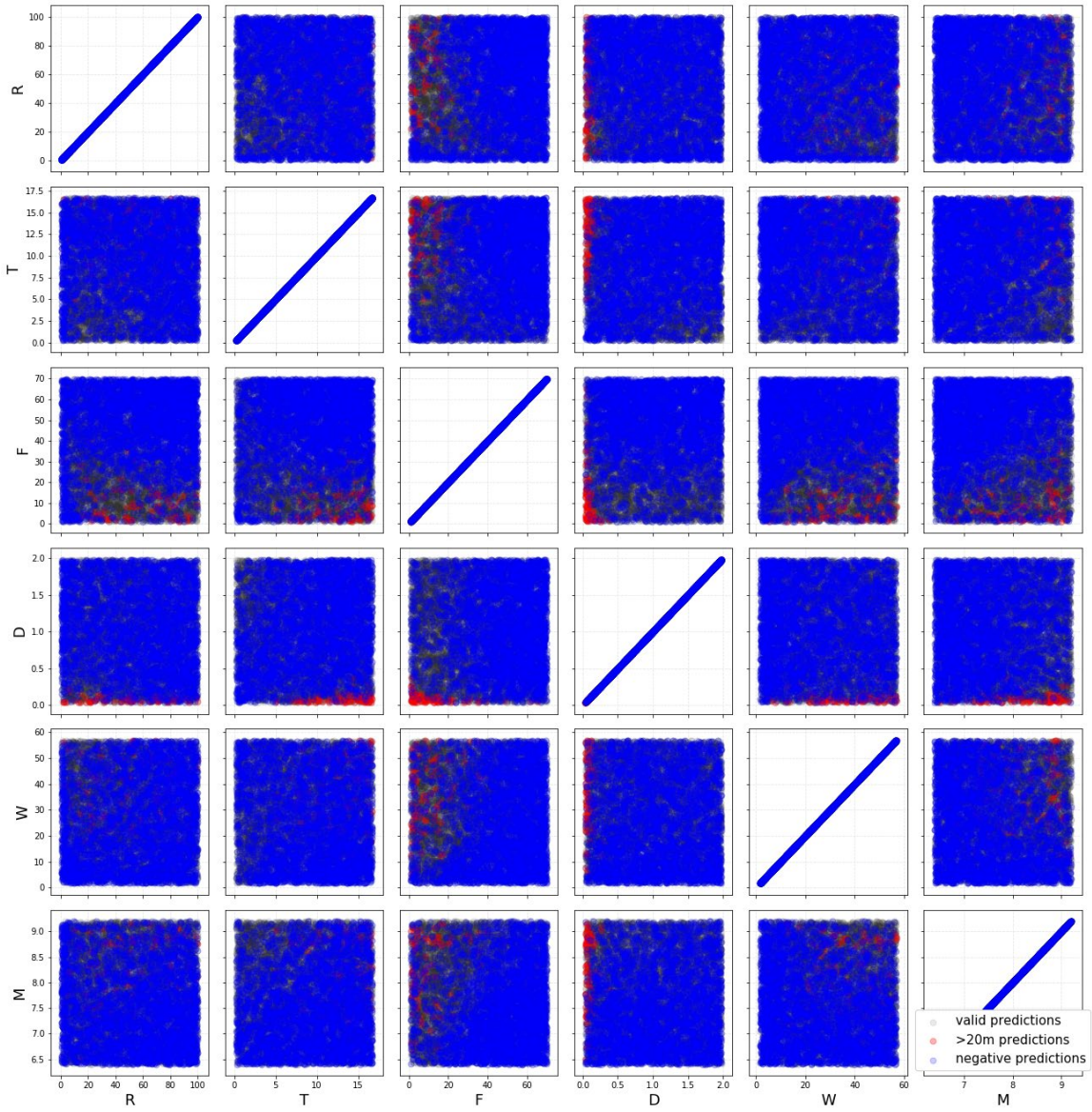


Figure C.6 Results of Monte Carlo analysis on Free-face case of Rezania et al. (2011)

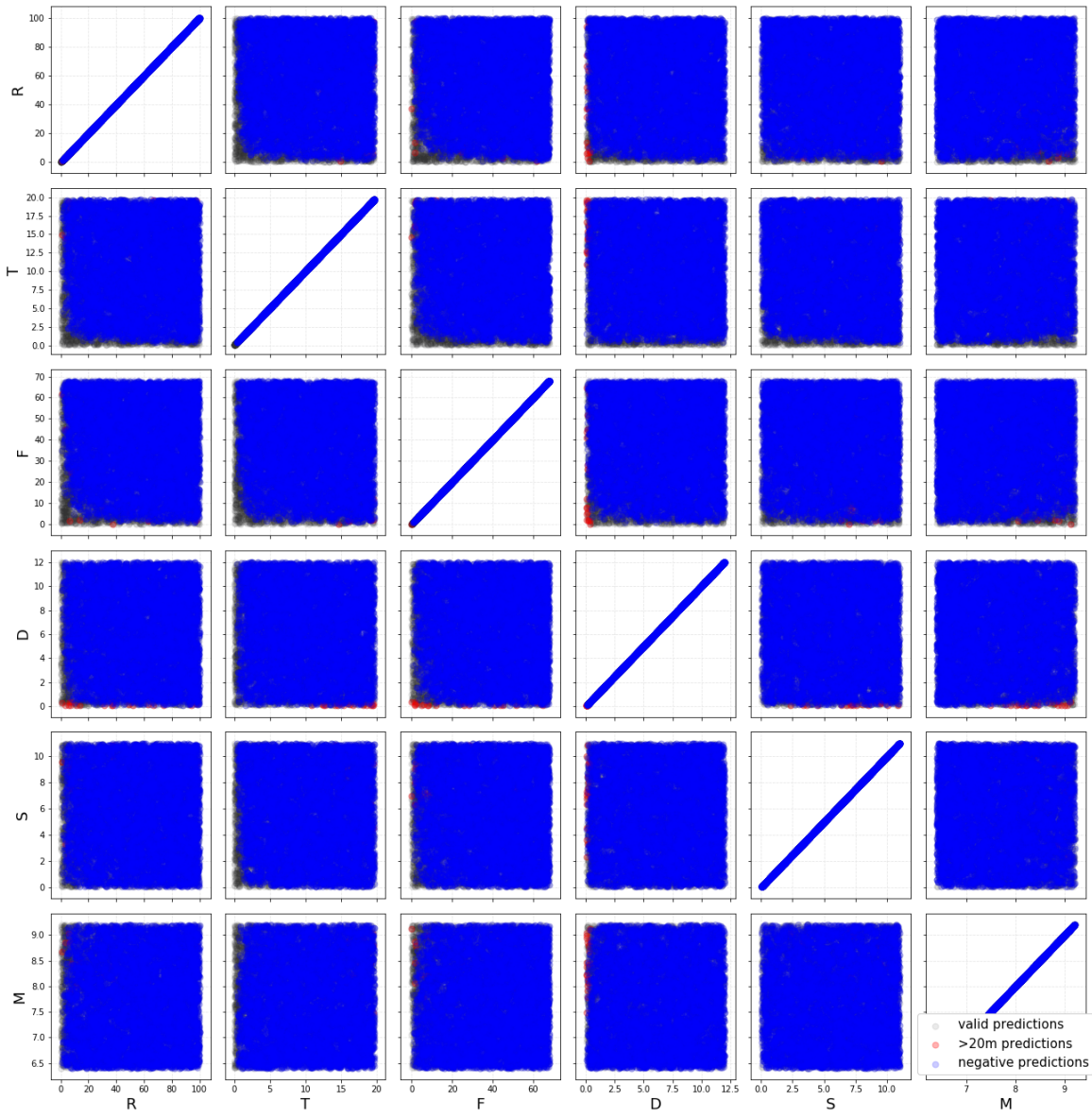


Figure C.7 Results of Monte Carlo analysis on sloping-ground case of Rezania et al. (2011)