

## Research Article

# Soft Schemes for Earthquake-Geotechnical Dilemmas

**Silvia García**

*Geotechnical Department, Institute of Engineering, National University of Mexico, Mexico*

Correspondence should be addressed to Silvia García; [sgab@pumas.iingen.unam.mx](mailto:sgab@pumas.iingen.unam.mx)

Received 27 November 2012; Revised 17 March 2013; Accepted 25 March 2013

Academic Editor: Marek Grad

Copyright © 2013 Silvia García. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Models of real systems are of fundamental importance in virtually all disciplines because they can be useful for gaining a better understanding of the organism. Models make it possible to predict or simulate a system's behavior; in earthquake geotechnical engineering, they are required for the design of new constructions and for the analysis of those that exist. Since the quality of the model typically determines an upper bound on the quality of the final problem solution, modeling is often the bottleneck in the development of the whole system. As a consequence, a strong demand for advanced modeling and identification schemes arises. During the past years, soft computing techniques have been used for developing unconventional procedures to study earthquake geotechnical problems. Considering the strengths and weaknesses of the algorithms, in this work a criterion to leverage the best features to develop efficient hybrid models is presented. Via the development of schemes for integrating data-driven and theoretical procedures, the soft computing tools are presented as reliable earthquake geotechnical models. This assertion is buttressed using a broad history of seismic events and monitored responses in complicated soils systems. Combining the versatility of fuzzy logic to represent qualitative knowledge, the data-driven efficiency of neural networks to provide fine-tuned adjustments via local search, and the ability of genetic algorithms to perform efficient coarse-granule global search, the earthquake geotechnical problems are observed, analyzed, and solved under a holistic approach.

## 1. Introduction

There are significant challenges for the future development and application of earthquake-geotechnical engineering that requires innovative approaches within a multidisciplinary framework. Very useful and up-to-date information on the occurrence frequency and impact of earthquake disasters is being assessed and analyzed by a number of organizations around the world. The earthquake-geotechnical engineering is an important bridge between geology, geomorphology, seismology, and civil engineering and serves as the environment where integrated and multidisciplinary approaches can be developed. In such applications, regarding specialized geotechnical engineering merely as a subset of civil engineering will lead to incomplete understanding of problems and the development of inadequate or incomplete solutions. Narrow perspectives can also suffocate progress and innovation. Links between the geosciences, seismology, mathematics, computing, and geotechnical engineering in terms of common concerns and needs such as obtaining, organizing, validating, displaying, and interpreting surface and subsurface must be considered.

One must look at the big picture for understanding past developments and present practices and for developing valid perspectives of the future. This should not mean simply following well-worn paths and considering progress only in terms of improvements, adjustments, and modifications of the current elements of what is regarded as good practice. Adopting new paradigms may be desirable or even necessary. The inclusion of innovative archetype of knowledge and skills concerning computing, creativity, and globalization in addition to factual and analytical knowledge seems mandatory. Innovative *computing* refers to the importance of exploiting the emerging processing resources to improve basic concepts and logic rather than the present emphasis only on faster applications and the blind using of software developed by others.

A competent modeling of engineering systems, when they are affected by seismic activity, has to be flexible enough to handle various degrees of complexity and uncertainty, and at the same time be sufficiently powerful to deal with situations in which the input signal may or may not be controllable. Mathematically based models are developed using scientific

theories and concepts that just apply to particular conditions. Thus, the core of the model comes from assumptions that for complex systems usually lead to simplifications (perhaps oversimplifications) of the problem phenomena. It is fair to argue that the representativeness of a particular theoretical model largely depends on the degree of comprehension the developer has on the behavior of the actual engineering problem. Predicting natural-phenomena characteristics like those of earthquakes, and thereupon their potential effects at particular sites, certainly belongs to a class of problems we do not fully understand. Accordingly, analytical modeling often becomes the bottleneck in the development of more accurate procedures. Soft computing (SC) technologies have provided us with a unique opportunity to establish coherent seismic analysis environments in which uncertainty and partial data knowledge are systematically handled.

By seamlessly combining learning, adaptation, evolution, and fuzziness, SC complements current engineering approaches allowing us to develop a more comprehensive and unified framework to the effective management of earthquake phenomena. Each SC algorithm has well-defined labels and could usually be identified with specific scientific communities. Lately, as we improved our understanding of these algorithms' strengths and weaknesses, we began to leverage their best features and develop hybrid algorithms that indicate a new trend of coexistence and integration between many scientific communities to solve a specific task. In this paper, geotechnical aspects of earthquake engineering under a soft examination are covered. Via the development of reinterpretations of the following selected topics: (i) spatial variation of soil dynamic properties, (ii) attenuation laws for rock sites (seismic input), (iii) generation of artificial-motion time histories, and (iv) evaluation of liquefaction susceptibility, SC techniques are presented as appealing alternatives for integrated data-driven and theoretical procedures to generate reliable and intelligent geoseismic models.

The author of this document is well aware that standards for geotechnical seismic design are under development worldwide. While there is no need to "reinvent the wheel", there is a requirement to adapt such initiatives to fit the emerging safety philosophy and demands. This investigation also strongly endorses the view that "guidelines" are far more desirable than "codes" or "standards" disseminated all over seismic regions. Flexibility in approach is a key ingredient of geotechnical engineering and the cognitive technology in this area is rapidly advancing. The science and practice of geotechnical earthquake engineering is far from mature and needs to be expanded and revised periodically in the coming years. It is important that readers and users of the computational models presented here familiarize themselves with the latest advances and amend the recommendations herein appropriately.

The following sections are not intended to be a detailed treatise of the latest research in geotechnical earthquake engineering, but to provide sound guidelines to support rational cognitive approaches. While every effort has been made to make the material useful in a wider range of applications, applicability of the material is a matter for the user to judge. The main aim of this guidance document is to promote

consistency of cognitive approach to everyday situations and, thus, improve geotechnical-earthquake aspects of the performance of the built safe environment.

## 2. Computational Intelligence: Soft Computing Technologies

The *computational intelligence* is a synergistic integration of essentially three computing paradigms, namely, neural networks, fuzzy logic, and evolutionary computation entailing probabilistic reasoning (belief networks, genetic algorithms, and chaotic systems) [1]. This synergism provides a framework for flexible information processing applications designed to operate in the real world and is commonly called *soft computing* (SC) [2]. Soft computing technologies are robust by design and operate by trading off precision for tractability. Since they can handle uncertainty with ease, they conform better to real world situations and provide lower cost solutions.

The three components of soft computing differ from one another in more than one way. Neural networks operate in a numeric framework and are well known for their learning and generalization capabilities. Fuzzy systems [3] operate in a linguistic framework, and their strength lies in their capability to handle linguistic information and perform approximate reasoning. The evolutionary computation techniques provide powerful search and optimization methodologies. All the three facets of soft computing differ from one another in their time scales of operation and in the extent to which they embed *a priori* knowledge.

Figure 1 shows a general structure of soft computing technology. The following main components of SC are known by now: fuzzy logic (FL), neural networks (NN), probabilistic reasoning (PR), genetic algorithms (GA), and chaos theory (ChT) (Figure 1). In SC, FL is mainly concerned with imprecision and approximate reasoning, NN with learning, PR with uncertainty and propagation of belief, GA with global optimization and search, and ChT with nonlinear dynamics. Each of these computational paradigms (emerging reasoning technologies) provides us with complementary reasoning and searching methods to solve complex, real-world problems. In large scope, FL, NN, PR, and GA are complementary rather than competitive [4, 5]. The interrelations between the components of SC, shown in Figure 1, make the theoretical foundation of hybrid intelligent systems. As noted by Zadeh: "... the term hybrid intelligent systems is gaining currency as a descriptor of systems in which FL, NC, and PR are used in combination. In my view, hybrid intelligent systems are the wave of the future" [6]. The use of hybrid intelligent systems leads to the development of numerous manufacturing system, multimedia system, intelligent robots, and trading systems, which exhibits a high level of MIQ (machine intelligence quotient).

The constituents of SC can be used independently (fuzzy computing, neural computing, evolutionary computing, etc.) and more often in combination [7–11]. Based on independent use of the constituents of soft computing, fuzzy technology, neural technology, chaos technology, and others have been

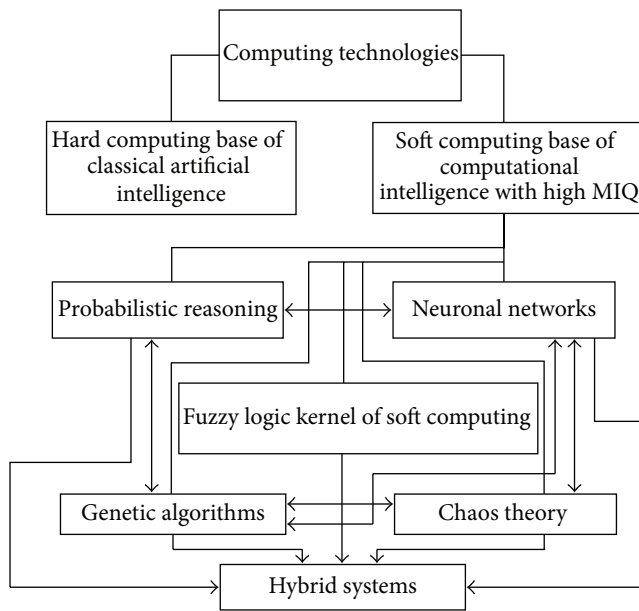


FIGURE 1: Soft computing components.

recently applied as emerging technologies to both industrial and nonindustrial areas.

Fuzzy logic is the leading constituent of soft computing. In soft computing, fuzzy logic plays a unique role. FL serves to provide a methodology for computing [7]. It has been successfully applied to many industrial spheres, robotics, complex decision making and diagnosis, data compression, and many other areas. To design a system processor for handling knowledge represented in a linguistic or uncertain numerical form, we need a fuzzy model of the system. Fuzzy sets can be used as a universal approximator, which is very important for modeling unknown objects. If an operator cannot tell linguistically what kind of action he or she takes in a specific situation, then it is quite useful to model his/her control actions using numerical data. However, fuzzy logic in its so called *pure form* is not always useful for easily constructing intelligent systems. For example, when a designer does not have sufficient prior information (knowledge) about the system, development of acceptable fuzzy rule base becomes impossible. As the complexity of the system increases, it becomes difficult to specify a correct set of rules and membership functions for describing adequately the behavior of the system. Fuzzy systems also have the disadvantage of not being able to extract additional knowledge from the experience and correcting the fuzzy rules for improving the performance of the system.

Another important component of soft computing is neural networks. Neural networks (NN), viewed as parallel computational models, are parallel fine-grained implementation of nonlinear static or dynamic systems. A very important feature of these networks is their adaptive nature, where “learning by example” replaces traditional “programming” in problems solving. Another key feature is the intrinsic parallelism that allows fast computations. Neural networks

are viable computational models for a wide variety of problems including pattern classification, speech synthesis and recognition, curve fitting, approximation capability, image data compression, associative memory, and modeling and control of nonlinear unknown systems [11, 12]. NN are favorably distinguished for efficiency of their computations and hardware implementations. Another advantage of NN is generalization ability, which is the ability to classify correctly new patterns. A significant disadvantage of NN is their poor interpretability. One of the main criticisms addressed to neural networks concerns their black box nature [6].

Evolutionary computing (EC) is a revolutionary approach to optimization. One part of EC—genetic algorithms—are algorithms for global optimization. Genetic algorithms (GA) are based on the mechanisms of natural selection and genetics [13]. One advantage of genetic algorithms is that they effectively implement parallel multicriteria search. The mechanism of genetic algorithms is simple. Simplicity of operations and powerful computational effect are the two main advantages of genetic algorithms. The disadvantages are the problem of convergence and the absence of strong theoretical foundation. The requirement of coding the domain of the real variables into bit strings also seems to be a drawback of genetic algorithms. It should be also noted that the computational speed of genetic algorithms is low. Table 1 presents the comparative characteristics of the components of soft computing. For each component of soft computing, there is a specific class of problems, where the use of other components is inadequate.

As it was shown above, the components of SC complement each other, rather than compete. It becomes clear that FL, NC, and GA are more effective when used in combinations. Lack of interpretability of neural networks and poor learning capability of fuzzy systems are similar problems that limit the application of these tools. Neurofuzzy systems are hybrid systems which try to solve this problem by combining the learning capability of connectionist models with the interpretability property of fuzzy systems. As it was noted above, in case of dynamic work environment, the automatic knowledge base correction in fuzzy systems becomes necessary. On the other hand, artificial neural networks are successfully used in problems connected to knowledge acquisition using learning by examples with the required degree of precision.

The cooperation between these formalisms gives a useful tool for modeling and reasoning under uncertainty in complicated real-world problems. Such cooperation is of particular importance for constructing perception-based intelligent information systems. We hope that the mentioned intelligent combinations will develop further, and the new ones will be proposed. These SC paradigms will form the basis for creation and development of computational intelligence.

### 3. Cognitive Models of Ground Motions

The existence of numerous databases in the field of civil engineering, and in particular in the field of geotechnical earthquake, has opened new research lines through the

TABLE 1: Central characteristics of soft computing technologies.

	Fuzzy sets	Artificial neural networks	Evolutionary computing, GA	Probabilistic reasoning	Chaotic computing
Weaknesses	(i) Knowledge acquisition (ii) Learning	Black box interpretability	(i) Coding (ii) Computational speed	(i) Limitation of the axioms of probability theory (ii) Lack of complete knowledge (iii) Computational complexity	(i) Computational complexity (ii) Chaos identification complexity
Strengths	(i) Interpretability (ii) Transparency (iii) Plausibility (iv) Graduality (v) Modeling (vi) Reasoning (vii) Tolerance to imprecision	(i) Learning (ii) Adaptation (iii) Fault tolerance (iv) Curve fitting (v) Generalization ability (vi) Approximation ability	(i) Computational efficiency (ii) Global optimization	(i) Rigorous framework (ii) Good understanding	(i) Nonlinear dynamics simulation (ii) Discovering chaos in observed data (with noise) (iii) Determining the predictability (iv) Prediction strategies formulation

introduction of analysis based on soft computing. Three methods are mainly applied in this emerging field: the ones based on the neural networks (NN), the ones created using fuzzy sets (FS) theory, and the ones developed from the evolutionary computation [14].

The SC hybrids used in this investigation are directed to tasks of prediction (classification and/or regression). The central objective is obtaining numerical and/or categorical values that mimic input-output conditions from experimentation and *in situ* measurements and, then, through the recorded data and accumulated experience to predict future behaviors.

The examples presented herein have been developed by an engineering committee that works for generating useful guidance to geotechnical practitioners in geotechnical seismic design. This effort could help to minimize the perceived significant and undesirable variability within geotechnical earthquake practice. Some urgency in producing the alternative guidelines was seen, after the most recent earthquakes disasters, as being necessary with a desire to avoid a long and protracted process. To this end, a two-stage approach was suggested with the first stage being a cognitive interpretation of the well-known procedures with appropriate factors for geotechnical design, with a posterior step identifying the relevant philosophy for a new geotechnical seismic design.

**3.1. Spatial Variation of Soil Dynamic Properties.** When using considerable volumes of 3D geotechnical information, with nonlinear and multidimensional relations, its proper management and analysis with conventional tools are limited, reducing its exploitation in natural phenomena modeling. The spatial variability of subsoil properties represents a major challenge in both the design and construction phases of most geoengineering projects. Subsoil investigation is an imperative step in any civil engineering project and it is also considered a prerequisite to the economical design of substructures. In general, the purpose of an exploratory investigation is to infer accurate information about actual soil and rock conditions at the site.

The geotechnical investigation stage of any project should be carefully planned to obtain the most reliable soil parameters with the minimum expense. A detailed planned program of boring and sampling, as well as the utilization of an accurate interpretation technique (i.e., modeling), is the cornerstone of any reliable and precise exploratory investigation. It is impossible to determine the optimum spacing of borings before an investigation begins because the spacing depends not only on type of structure but also on uniformity or regularity of encountered soil deposits. Even the most detailed soil maps are not efficient enough for predicting a specific soil property because it might exhibit variations from place to place, even for the same soil type. Consequently, interpolation techniques have been extensively applied. Unfortunately, the most common used methods, kriging and cokriging, require a significant number of measurements for each soil type to obtain assertive estimations, which is generally unmanageable. Based on the high cost of collecting soil attribute data at many locations across landscape, new interpolation methods have to be tested. The tested methodologies must be able to properly organize the historical geotechnical into multiple databases for establishing spatial/chronological predictive models through interpreting properties (soil exploration) and behaviors (*in situ* measured).

In the following, the spatial variation of the cone tip resistance  $q_c$  and shear wave velocity  $V_s$  defined using NN and GA and a georeferenced 3D model of the soils underlying Mexico City is presented. The classification/prediction criterion for this very complex urban area is established according to two much related soils properties, a mechanical property and a dynamic property.

Cone penetration testing CPT and shear wave velocity information are frequently used to determine the vertical succession of different layers of soft soil and parameters related to these materials. Huijzer [15] worked first on the automatic interpretation of cone penetration tests—using linear regression—but a satisfactory recognition process based exclusively on resistance value turned out to be impossible.



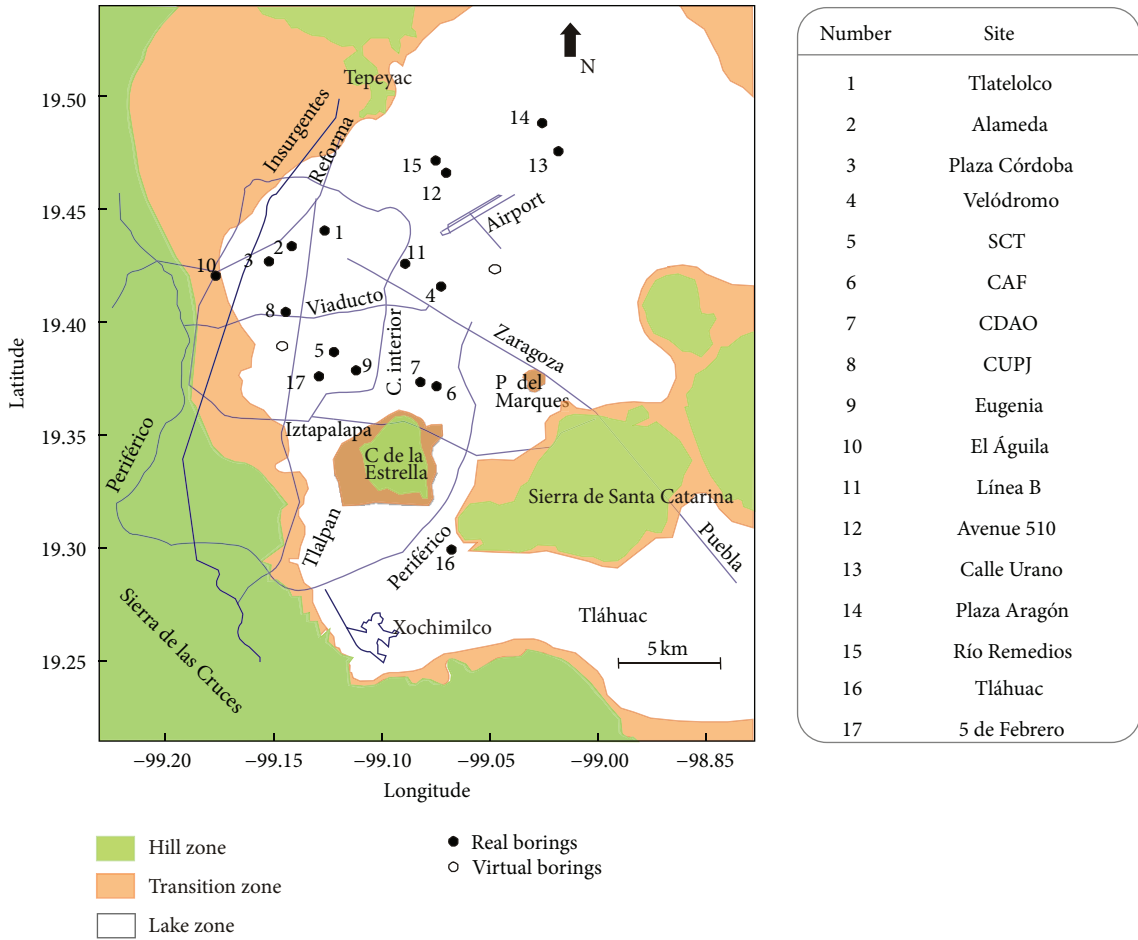


FIGURE 2: Geotechnical zones and exploration sites, Mexico City.

A later study by Coerts [16] started the interpretation from a predefined number of geotechnical units (number of boreholes) based on profiles published in geological maps whose parameter values determination was contaminated by many vague and uncertain sources. Additionally, a general regression neural approximation for site characterization, in terms of soil strengths, was developed by Juang et al. [17]. The authors presented the benefits of using NNs; however, their estimations exhibited very low confidence since the total number of boreholes available was used to train the network and no test patterns were presented for evaluating the capabilities for generalization of the network.

In this investigation, a proved methodology [18] was used to estimate  $V_s$  and  $q_c$  at locations on the half space where no measurements were performed. Therefore,  $V_s$  and  $q_c$  were generalized in any subdomain of the studied space. The estimations at any point within the half space represent a virtual boring so that an intricate grid of values has to be created to support the neural contour maps. Although of an NN model is not the most powerful interpolator, it represents an excellent classification/prediction function (parametric and structural related to geographical position and mechanical soil properties). In this investigation, the García and Romo's methodology [18] is reinterpreted and

improved by using a mechanical and a dynamical property coupled with linguistic and historical information to better reproduce the half-space of the lacustrine soil deposits of the Mexico Valley. In addition, the suggested NN hidden structure tuned by GA represents novelty.

**Database.** The application of the NN-GAs methodology is illustrated by an example dealing with the spatial variability of the  $q_c$  and  $V_s$  in the lacustrine zone of Mexico valley. The result is a 3D model of the soils underlying the city area. Cone-tip penetration resistances and shear wave velocities have been measured along 19 bore holes spread throughout the clay deposits of Mexico City. Borehole locations are marked with dots in Figure 2. This information was used as the set of examples inputs, which are georeferenced in latitude, longitude, and depth and the output will be the ratio CPT resistance on shear wave velocity. It is important to point out that 25% of these patterns (sample points and complete depth-property information) are not used in the training stage; they will be presented for corroborating the generalization capabilities of the closed system components. Once the training is stopped, the ratio test/validation set is presented and the correlation between measured and NN-estimated values is checked and the error is assessed.

ID number	Site	Geo coord		Z (m)
		Lat.	Long.	
11	Línea B	19.32	99.01	8.3
				10
				12
				14
				17
				20
				22
				24
				26
				28.5
				31
6	CAF	19.366	99.086	3
				6.5
				9
				12
				14
				17
				19
				21
				23
				25

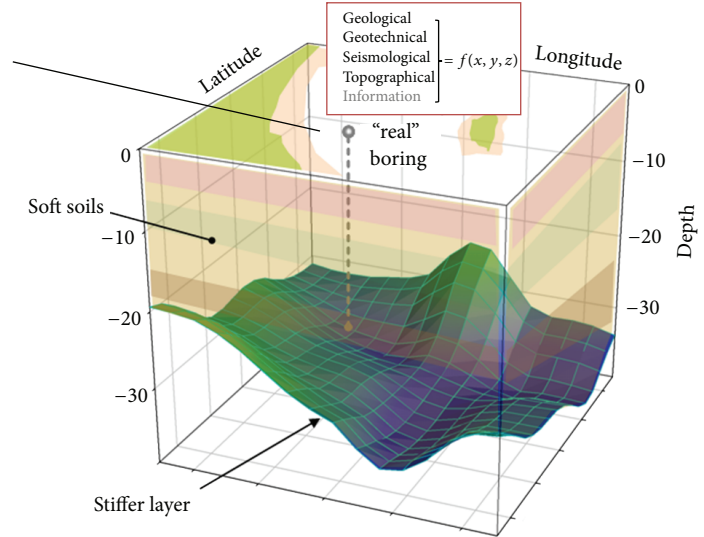


FIGURE 3: Schematic representation of Table 1: "real" borings situation.

**Spatial Model Construction.** The figure of merit used to compare the NN architectures trained (those trails for acquiring the optimum neural design) was percent root mean squared error defined as

$$\text{RMSE}\% = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - x_i}{\max(x)} \right)^2}, \quad (1)$$

where  $x_i$  are the correct values on the test set and  $y_i$  are the values obtained by interpolation. It is important to note that the definition of spatial variation of a soil property using neurogenetic tools does not need a priori identification or selection of homogeneous soil layers, which is mandatory in many procedures. In the 3D neurogenetic analysis, the functions  $q_c = \{q_c(X, Y, Z)\}/V_s = \{V_s(X, Y, Z)\}$  are to be approximated, with which the  $q_c/V_s$  values at any half-space point can be determined. For predicting and defining soil properties and layers geometries, the 3D neuroenvironment is generated using the procedure outlined below.

- (1) The generation of the database including identification of the boring of the site ( $X$  and  $Y$  geographical coordinates,  $Z$  depth, and a CODE-ID number), elevation reference (meters above sea level, m.a.s.l.), thickness of predetermined structures (layers), and additional information related to geotechnical zoning that could be useful for results interpretation. The database is organized in three tables. The first contains the CODE,  $X$ ,  $Y$ , and  $Z$  data for each sounding (Figure 3). In a second table (Figure 4) any previous knowledge of strata structure is included for all soundings (parametric, geometrical, or/and linguistic descriptions). This "strata" table contains the properties arranged as columns related to depth. It can be extended through additional columns (mechanical, physical, or geometrical soil parameters) to generate a wide-ranging characterization. A third table

(Figure 5) contains the NN results (crisp evaluations) and it is constructed once the NN model is labeled as "optimum." This table contains the training/testing information (the "real" *in situ* measurements) plus the "virtual" borings, the neural information that supports the 3D grid and completes the media with materials and properties neuroinferred.

- (2) The first two tables (Figures 3 and 4) are used to train an initial neural topology whose weights and layers are adjusted by an evolutive algorithm, until the minimum RMSE between calculated and measured  $q_c/V_s$  values is achieved (Figure 6). In the case of 2D prediction, the  $q_c/V_s$  vectors separated for testing the generalization capabilities of the NN are compared with those obtained during the direct phase of the NN (Figure 7). In three dimensional analyses, all (training) data points from whole volume were considered at once. Through examining the neurogenetic results for unseen measurements, it can be concluded that the procedure works extremely well in identifying the general trend in materials resistance (stiffness).
- (3) The construction of the 3D soils environment is done using the real/virtual parametric vectors from Table 3 (Figure 5) but it is tuned considering the additional information from Table 2. A clustering fuzzy subroutine is applied to the database in order to find the incongruences, analyze them, and decide about its possible occurrence or not. This 3D view of the studied zone represents an easier and more understandable engineering frame. Some examples of the kind of information obtained from this assembly are shown in Figure 8.

Many advantages has the 3D NN exploration, for example, determining drillhole spacing requires, for obvious reasons, to obtain the maximum benefit at minimum cost, meaning that the number of boreholes has to be sufficient to

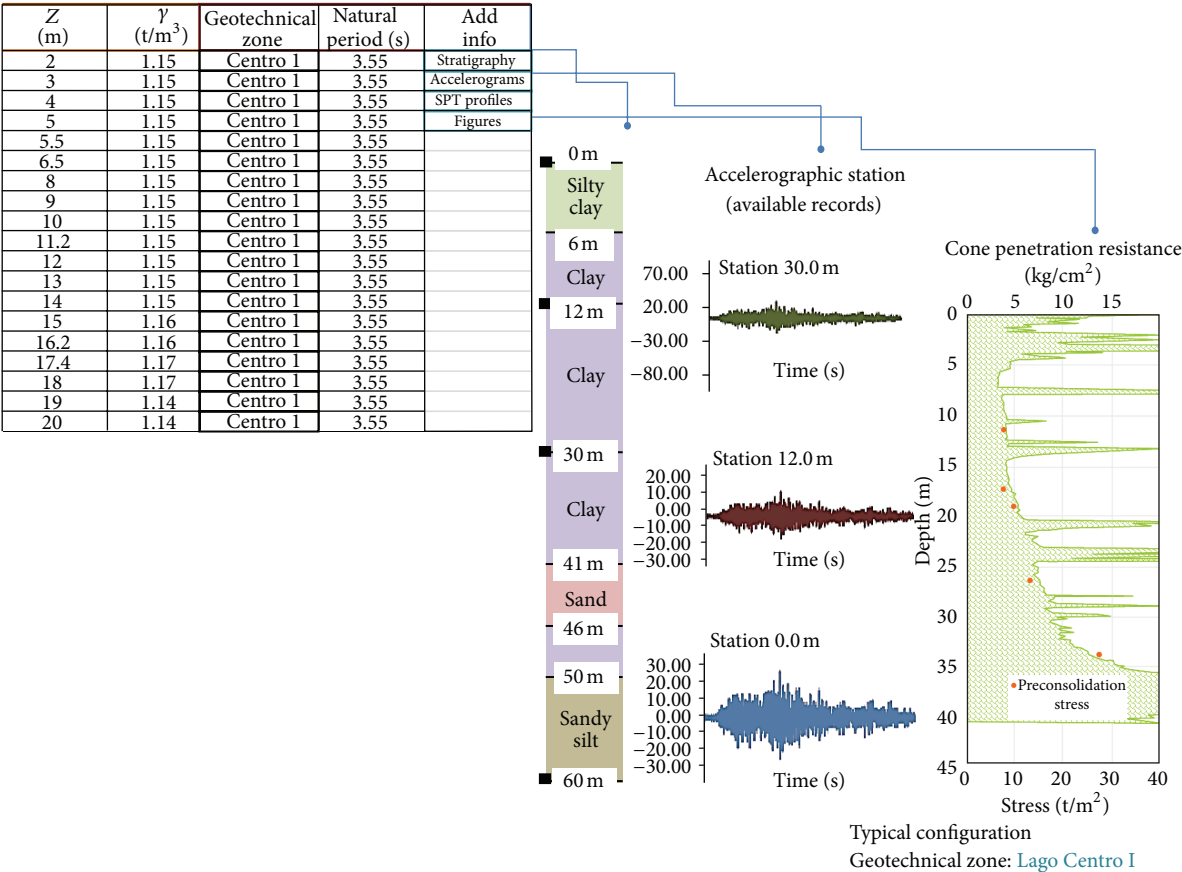


FIGURE 4: Schematic representation of Table 2: “real” borings geo- and seismic measurements.

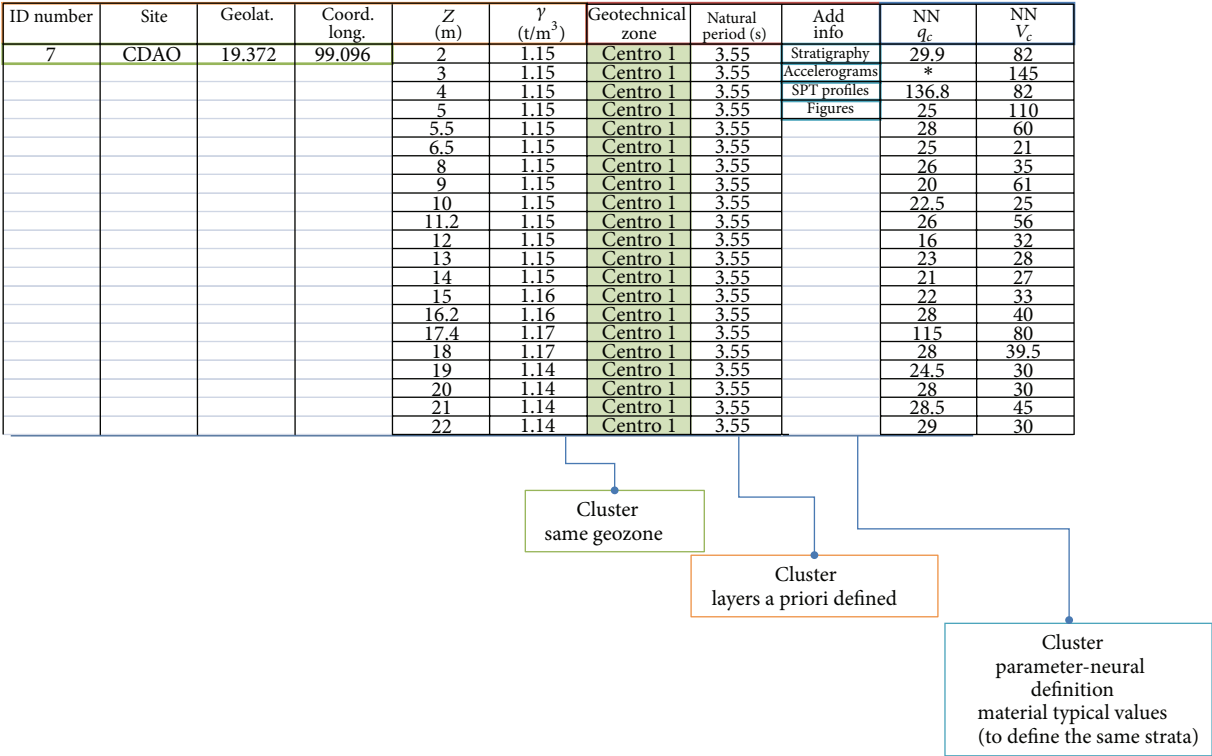


FIGURE 5: Schematic representation of Table 3: “real” and “virtual” borings to construct the stratigraphies or the 3D environment.

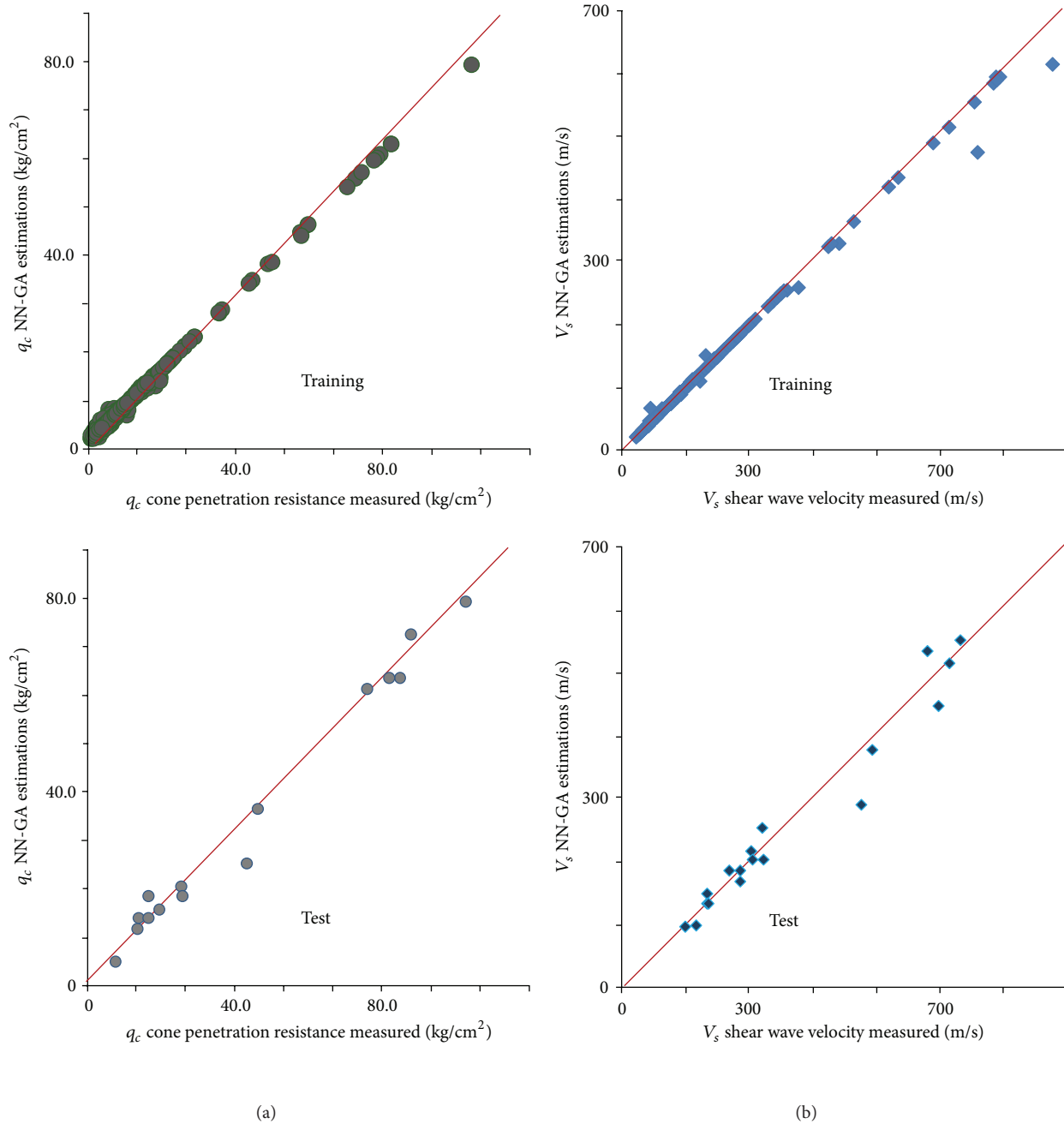


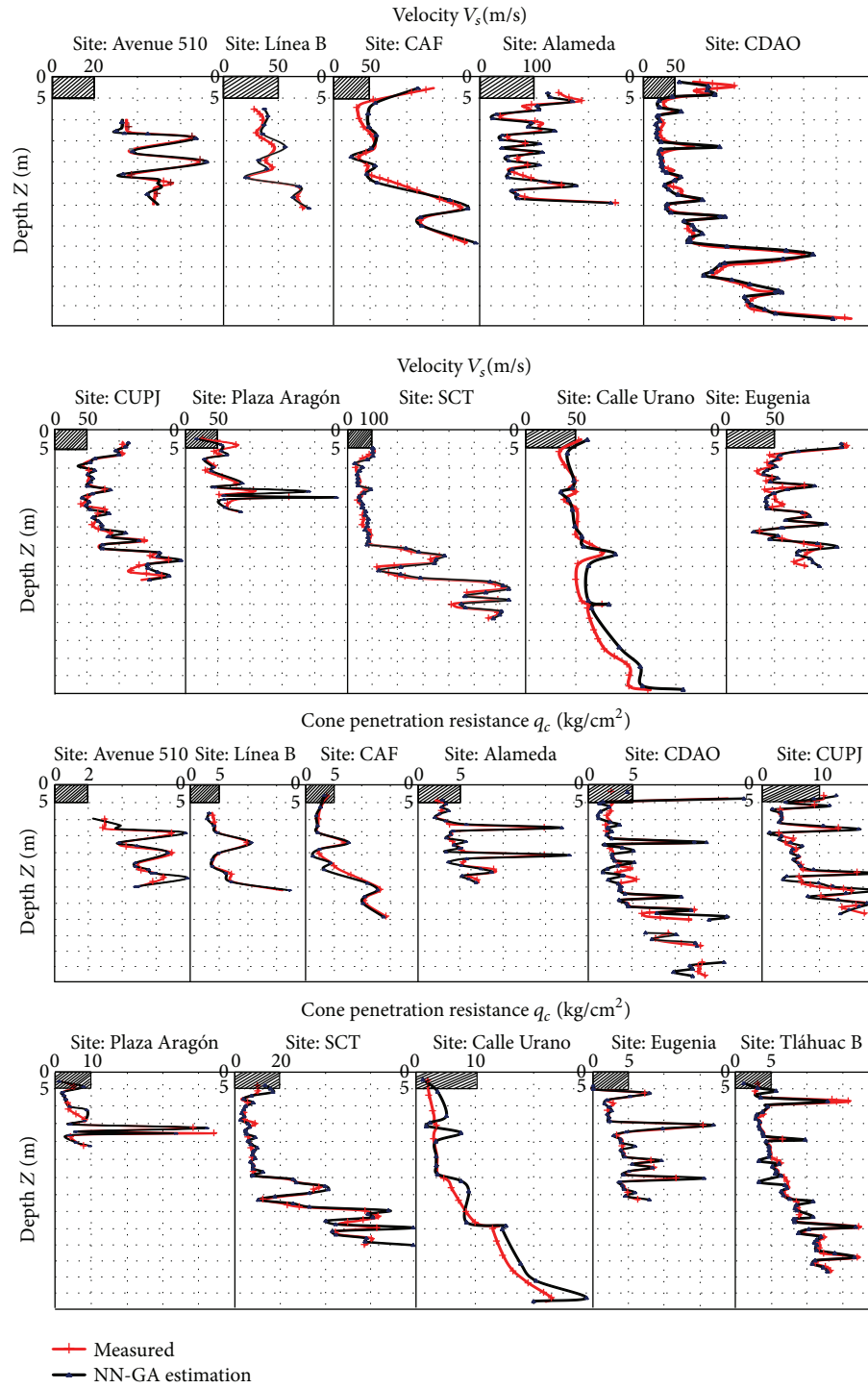
FIGURE 6: Measured *versus* evaluated values of  $q_c$  and  $V_s$ .

ensure continuity, without costing more than necessary. The inevitable question about the optimum drillhole spacing and the location of the boreholes can be answered using the 3D neural representation taking into account the prospecting, the experience, and the engineers' criteria. The level of risk that the management is willing to accept controls how rigorous should be the data collection at the start of a project as well as during subsequent developments. The cost of collecting information has to be weighed up against the potential cost of uncertainty. Neuronal estimations combined with geological/geotechnical evidences permit to infer and

in some cases to verify the geological and/or geotechnical continuity. Therefore, *in situ* exploration could be conducted with very high level of confidence and economic efficiency during further exploration steps (new borings).

To generate a reliable model of the soils underlying, it is necessary to review the horizontal continuity of the layers defined, from sounding to sounding. The definition of the stratigraphy for Mexico City clays is presented for showing that the neurogenetic capabilities are well suited for defining the intrusions and the lenses in the Mexican subsoil, agreeing with the formation and origin of these peculiar materials. It



FIGURE 7: Neural  $V_s$  and  $q_c$  vectors compared with *in situ* profiles.

is necessary to perform comparisons between layers at every site, in order to classify differently those layers that fulfill the same classification but do not belong to the same continuous layer. The procedure of identification and selection of layers can be easily conducted by using the neurogenetic tool that permits to manage, inter, and extrapolate information from massive databases. If a specific pattern is repeated in nearby

soundings, then it is likely the condition of a layer intruding another. Contrarily, it might be the presence of a lens.

In this study, soundings on a line along the East-West direction (E-W) of the city were selected to estimate the neural variations in the distribution of soils following the procedure to create georeferenced profiles (Figure 9). Four “real” exploration sites: SCT, Eugenia, Velódromo, and Línea

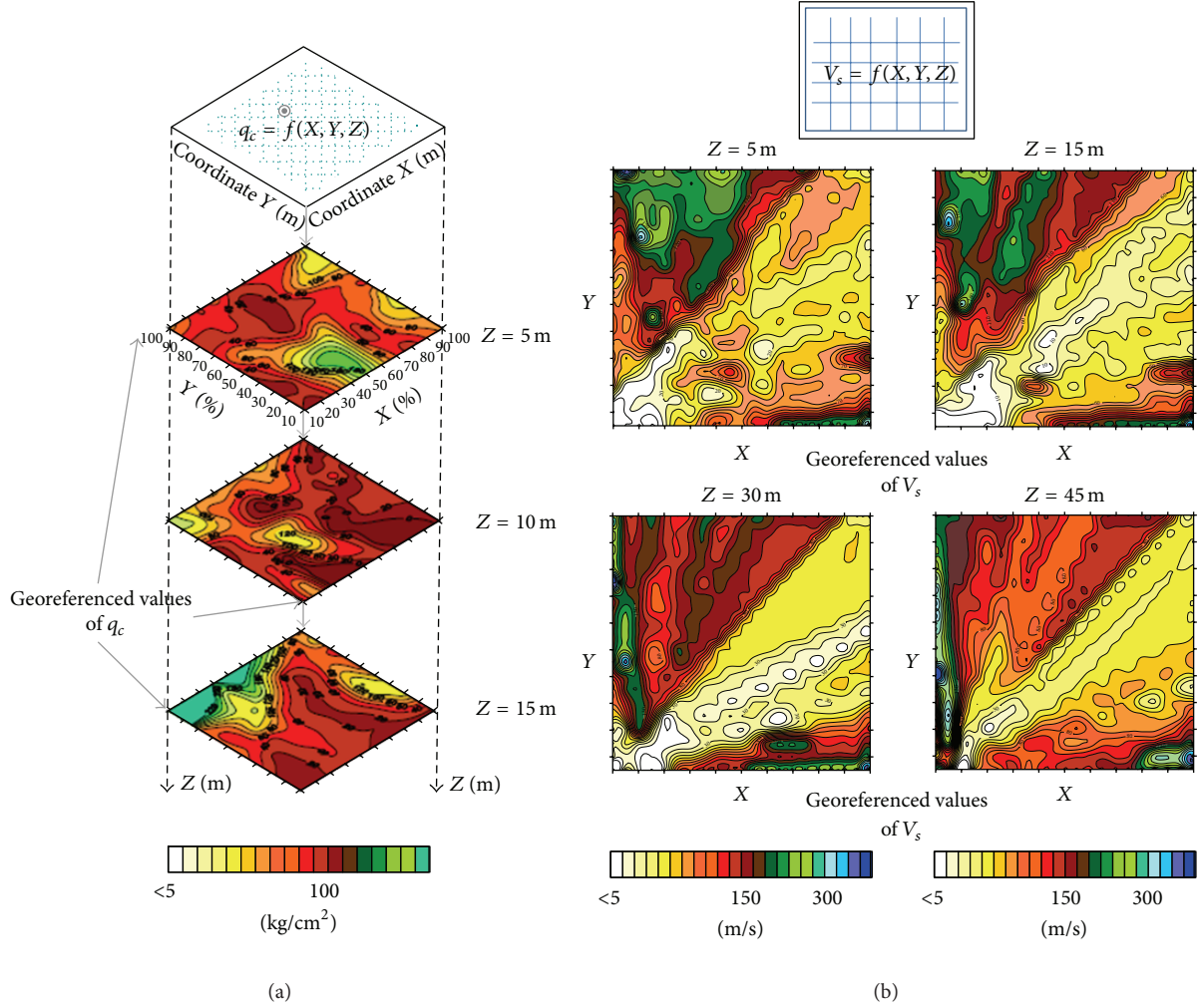
FIGURE 8: Isovalues curves for  $q_c$  and  $V_s$ .

TABLE 2: Database used to construct the ML model.

Dataset	Input attributes	Number of selected instances
1	$Z_{NAF}$ (m), $Z_{TOP LAYER}$ (m), $H$ (m), $\sigma'_V$ (KPa), $\sigma_V$ (KPa),	181 (Andrus and Stokoe, 1997, 2000) [20, 21]
	$S_T$ , $V_s$ (m/s), $M$ , $a_{max}$ (g)	
2	$Z_{TOP LAYER}$ (m), $R_f$ (%), $\sigma'_V$ (KPa), $q_c$ (MPa),	226 (Juang et al., 1999; Juang, 2003) [9, 22]
	$M$ , $a_{max}$ (g)	

B constitute the line of soundings and the schematic grid suggests that is the “virtual” information used to complete the layers successions. There are four well-defined soil layers but the shallowest one is not considered for the analyses. The variation of  $q_c$  and  $V_s$  (in Figure 9 the  $q_c$  profile is not included to make it more understandable) with depth from about 5 to 35 m is presented at Eugenia, Velódromo, and Línea B sites,

while the maximum depth is 50 m at the site SCT. Clearly, Velódromo site exhibits the lowest values of  $V_s$  and  $q_c$ , and shear wave velocities range approximately from 25 to 50 m/s, while  $q_c$  varies from 24.5 to 49 MPa. Otherwise, the site with the highest  $V_s$  and  $q_c$  correspond to SCT site.  $V_s$  and  $q_c$  are in the range from about 50 to 625 m/s and 49 and 613 MPa, respectively.

The neurogenetic model has been proven to be useful in the interpretation of natural resource information. The above methodology can be automated to produce geotechnical maps as a software system: a package of programs for conducting analyses of the spatial variability of one or more interrelated variables. Based on the results presented here, it can be concluded that a soft system would be menu driven and simple to use and visualize, requiring a minimum number of input data with dynamic allocation of memory so that datasets with a wide range of variables and positions can be used without altering the basic structure’s program.

3.2. Estimation of Peak Ground Accelerations for Mexican Subduction Zone Earthquakes. Earthquake ground motions

TABLE 3: Seismic events contained in the database.

Earthquake	Magnitude $M_w$	PGA* (g)	Epicentral distance (km)	Number of patterns
1906 San Francisco	7.9	0.28–0.26	13–27	4
1964 Alaska	9.2	0.21–0.39	35–100	7
1964 Niigata, Japan	7.5	0.19	21	299
1971 San Fernando	6.4	0.55	1	28
1979 Imperial Valley	6.5–6.6	0.21–0.51	2–6	32
1983 Nihonkai-Chubu, Japan	7.7	0.25	27	72
1987 Superstition Hills	6.6	0.21	23	6

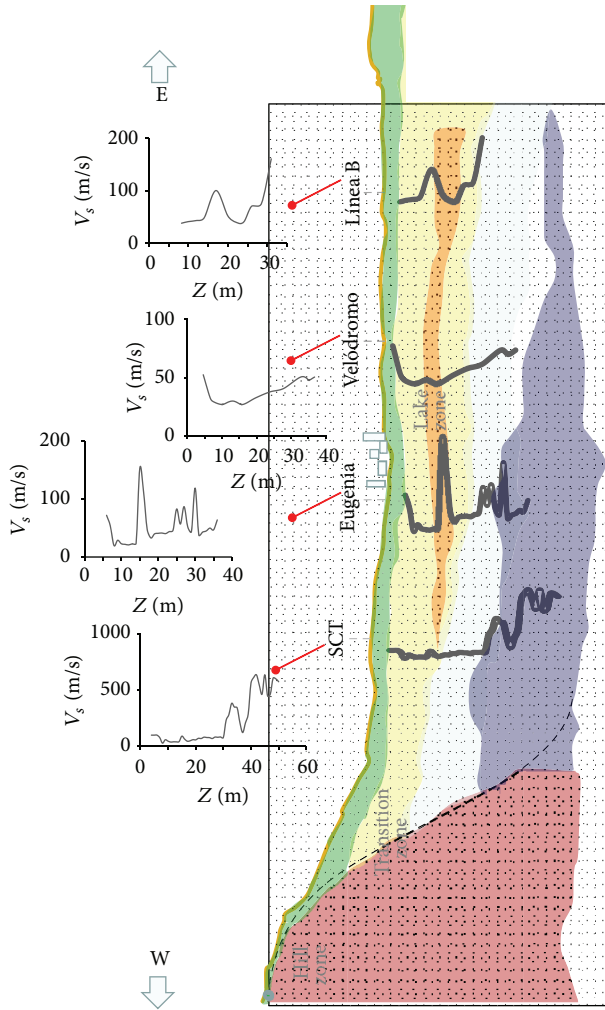


FIGURE 9: Layer definition using “virtual” profiles (every vertical dotted line is a “virtual” axis boring) and “real” exploration sites.

are affected by several factors including source, path, and local site response. These factors should be considered in engineering design practice using seismic hazard analyses that normally use attenuation relations derived from strong motion recordings to define the occurrence of an earthquake with a specific magnitude at a particular distance from the site.

These relations are typically obtained from statistical regression of observed ground motion parameters. Because of the uncertainties inherent in the variables describing the source (e.g., magnitude, epicentral distance, focal depth, and fault rupture dimension), the difficulty to define broad categories to classify the site (e.g., rock or soil), and our lack of understanding regarding wave propagation processes and the ray path characteristics from source to site, commonly the predictions from attenuation regression analyses are inaccurate.

As an effort to recognize these aspects, multiparametric attenuation relations have been proposed by several researchers (e.g., [19, 23–28]). However, most of these authors have concluded that the governing parameters are still source, ray path, and site conditions.

In this investigation, an empirical NN formulation that uses the minimal information about magnitude, epicentral distance, and focal depth for subduction-zone earthquakes is developed to predict the three components (two horizontal, one vertical) of peak ground acceleration PGA at rock sites (consisting of at most a few meters of stiff soil over weathered or sound rock). The NN model was obtained from existing information compiled in the Mexican strong motion database.

Events with poorly defined magnitude or focal mechanism, as well as recordings for which site-source distances are inadequately constrained, or recordings for which problems were detected with one or more components were removed from the data. It uses earthquake moment magnitude  $M_w$ , epicentral distance ED, and focal depth FD. The obtained results indicate that the proposed NN is able to capture the overall trend of the recorded PGAs.

This approach seems to be a promising alternative to describe earthquake phenomena despite of the limited observations and qualitative knowledge of the recording stations geotechnical site conditions, which leads to a reasoning of a partially defined behavior. Based on the procedure for achieving PGAs, spectral ordinates for any particular period can be estimated in the same manner with similar confidence levels.

**Database.** The database used in this study consists of 1058 records. These events were recorded at rock and rock-like sites during Mexican subduction earthquakes (Figure 10). Event dates range from 1964 to 2006. Events with poorly defined magnitude or focal mechanism, as well as recordings for which site-source distances, are inadequately constrained,

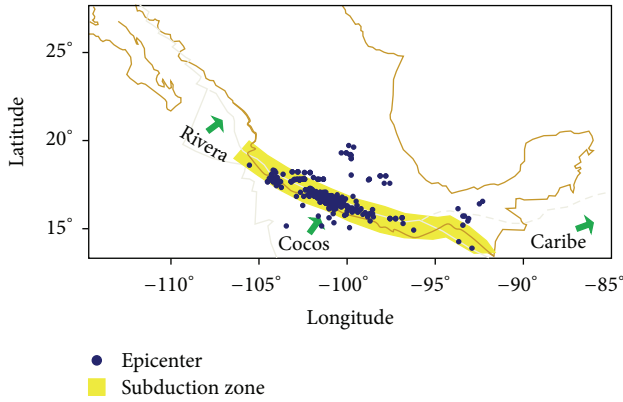


FIGURE 10: Motion array and events included in the data base.

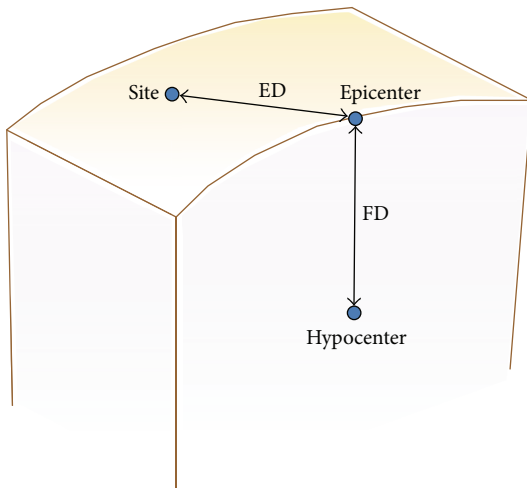


FIGURE 11: Schematic representation of parameters: focal depth, FD, and epicentral distance, ED.

or recordings for which problems were detected with one or more components were removed from the data. To test the predicting capabilities of the neuronal model, 186 records were excluded from the dataset used in the learning phase. One was the September 19, 1985,  $M_w = 8.1$  earthquake, and the other 185 events were randomly selected, making sure that a broad spectrum of cases were included in the testing database.

The moment magnitude scale  $M_w$  is used to describe the earthquakes size, resulting in a uniform scale for all intensity ranges. If the user has another magnitude scale, the empirical relations proposed by Scordilis [29] can be used. In this paper the used ED is considered to be the length from the point where fault rupture starts in the recording site, as indicated in Figure 11. The third input parameter, FD, does not express mechanism classes; it is declared as a nominal variable which means that the NN identifies if there is an event type (broad class) or the value impacts on the result through its crisp quantity. Some studies have led to consider that subduction-induced earthquakes may be classified as interface events ( $FD < 50$  km) and intraslab events ( $FD > 50$  km) ([19, 30]). This is a rough classification because crustal and interface earthquakes

would be mixed [31], and therefore it was considered not relevant for this model. The dynamic range of  $M_w$  goes from 3 to 8.1 approximately and the events were recorded at near (a few km) and far field stations (about 690 km). The depth of the zone of energy release ranged from very shallow to about 360 km.

**Neural-Attenuation Model Construction.** Modeling of the database has been performed using the quick propagation QP learning algorithm [32]. Horizontal (mutually orthogonal PGAh1, N-S component, and PGAh2, E-W component) and vertical components (PGAv) are included as outputs for neural mapping into three units. The neural modules that met the convergence criterion (mean square error  $\leq 5\%$ ) have a total of 72 and 126 hidden nodes for PGAh1 and PGAh2, respectively, while the PGAv module behavior was quite acceptable using a simple alternative (QP, 2 layers/15 units or nodes each). Details of the topology-selection process can be found in [33].

The neuronal attenuation model for  $\{M_w, ED, FD\} \rightarrow \{PGAh1, PGAh2, PGAv\}$  was evaluated by performing testing analyses. The predictive capabilities of the NNs were verified by comparing the PGAs estimated to those induced by the 186 events excluded from the original database that was used to develop the NN architectures (training stage). In Figure 12(a), the PGAs computed during the training and testing stages are compared to the measured values. The relative correlation factors ( $R^2 \approx 0.93$ ), obtained in the training phase, indicate that those topologies selected as optimal behave consistently within the full range of intensity, distances, and focal depths depicted by the patterns. Once the networks converge to the selected stop criterion, learning is finished and each of these black boxes becomes a nonlinear multidimensional functional. Every functional is then assessed (testing stage) by comparing their predictions to the separated 186 PGA values of the database. As new conditions of  $M_w$ , ED, and FD are presented to the neural functional,  $R^2$  decreases. This drop is more appreciable for the horizontal components. Nonetheless, as indicated by the upper and lower boundaries included in the Figure 12(b), forecasting of all three seismic components is reliable enough for practical applications.

A sensitivity study for the input variables was conducted for the three neuronal modules. The results are strictly valid only for the data base utilized; nevertheless, after several sensitivity analyses conducted changing the database composition, it was found that the following trend prevails; the  $M_w$  would be the most relevant parameter (presents larger relevance) then would follow the epicentral distance, ED, and the less influential parameter was the focal depth, FD. However, for near site events the epicentral distance could become as relevant as the magnitude, particularly, for the vertical component. The selected functional forms incorporate the results of analyses into specific features of the data, such as the PGAh dependence of the geometrical ED-FD description.

At this stage it is convenient to note that traditionally the PGAs that are used to develop attenuation relationships are defined as randomly oriented (e.g., [26]), mean (e.g., [34]), and the larger value from the two horizontal components



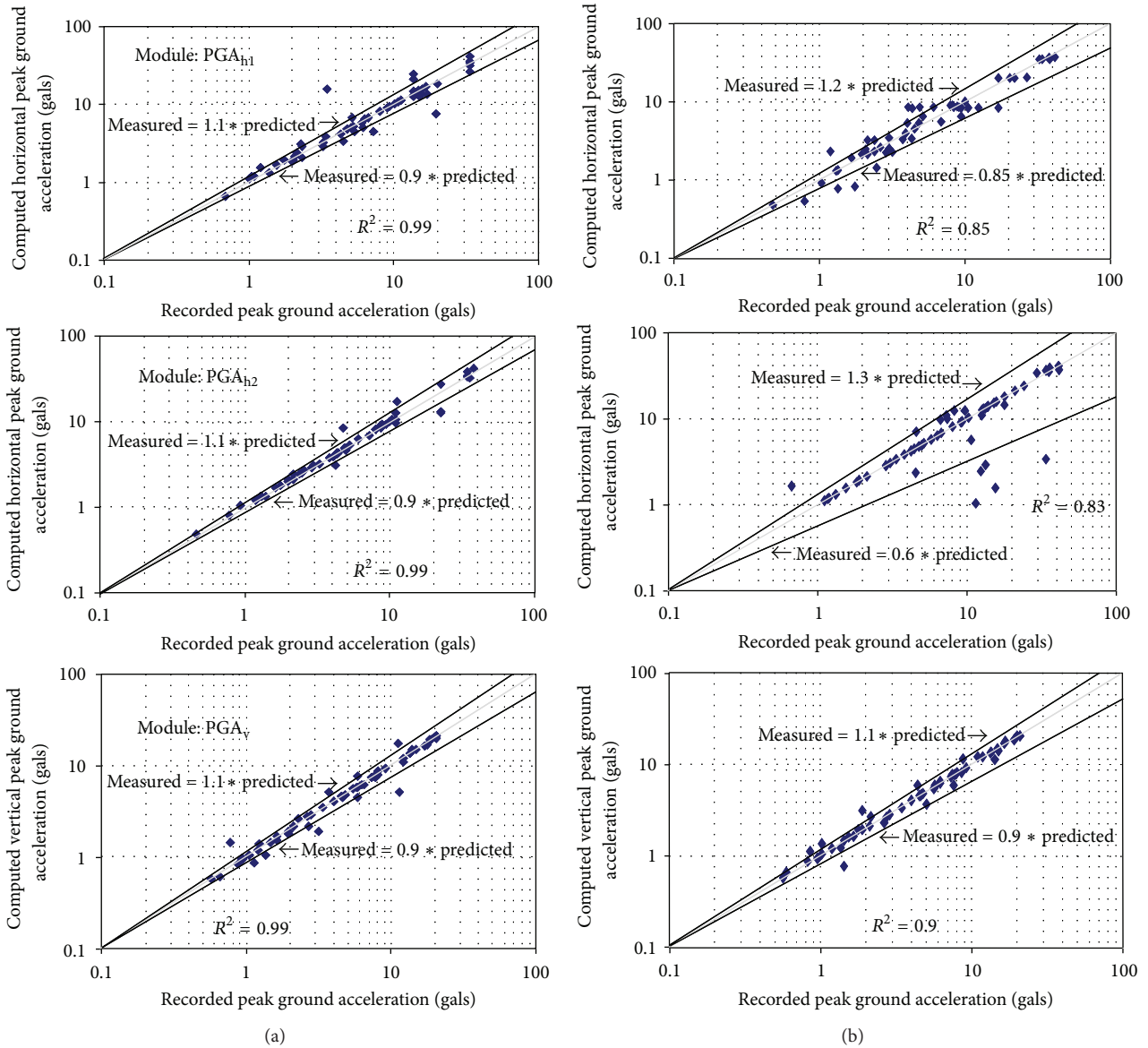


FIGURE 12: Comparisons between estimated and recorded peak ground accelerations for each component independently. (a) Left: training phase and (b) right: testing phase.

(e.g., [35]). Accordingly, considering the variety of PGA definitions used in most existing attenuation relations, it was deemed fairer to use the PGA<sub>h1-h2</sub> prediction modules for comparison purposes. Figure 13 compares five fitted relationships to PGA data from interface earthquakes recorded on rock and rock-like sites. The two case histories correspond to a large and a medium size events (the September 19, 1985, Michoacán earthquake and the July 4, 1994, event, resp.).

The estimated values obtained for these events using the relationships proposed by Gómez et al. [36], Youngs et al. [19], Atkinson and Boore [20]—proposed for rock sites—and Crouse [24]—proposed for stiff soil sites—and the predictions obtained with the PGA<sub>h1-h2</sub> modules are shown in Figure 13. It can be seen that the estimation obtained with Gómez et al. [36] seems to underestimate the response for

the large magnitude event. However, for the lower magnitude event both the measured responses and NN predictions follow closely. Youngs et al. [19] attenuation relationship follows closely the overall trend but tends to fall sharply for long epicentral distances. Although, as mentioned previously, the PGA<sub>h1-h2</sub> modules yielded important differences in the testing phase, its predictions follow closely the trends and yield a better behavior, in the full range of epicentral distances included in the data base, than traditional attenuation relations applied to the Mexican subduction zone.

Furthermore, it should be stressed the fact that the September 18, 1985, earthquake was not included in the database used in the development of the neural networks and that this event falls well outside the range of values in such database, and hence it is an example of the extrapolation

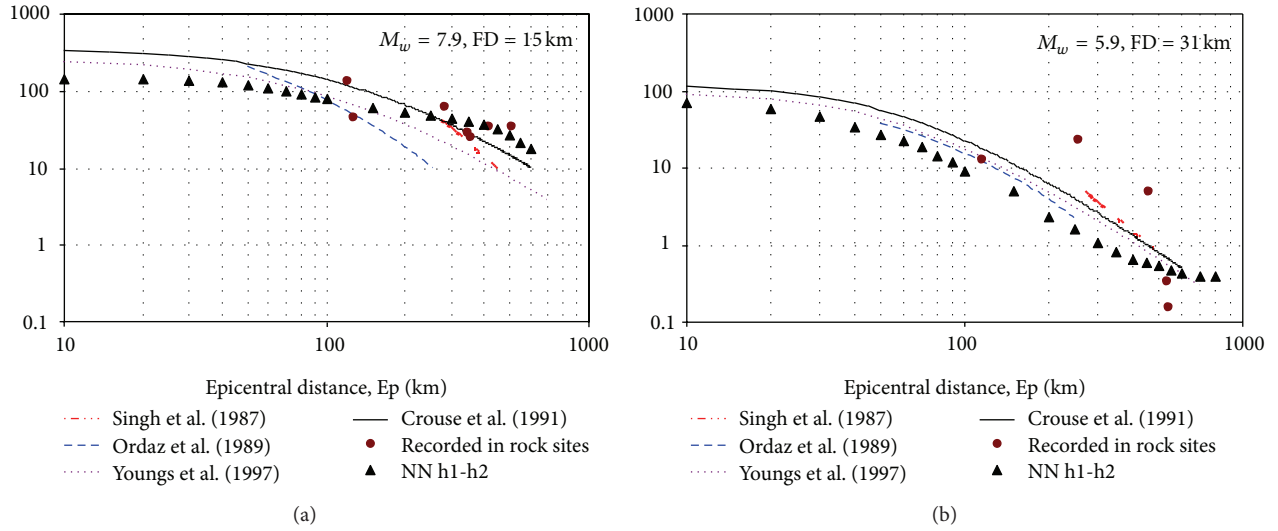


FIGURE 13: Attenuation relationship obtained with the proposed neural attenuation model PGAh1-h2 for two subduction-related events: (a) September 19, 1985, earthquake and (b) July 4, 1994, Event.

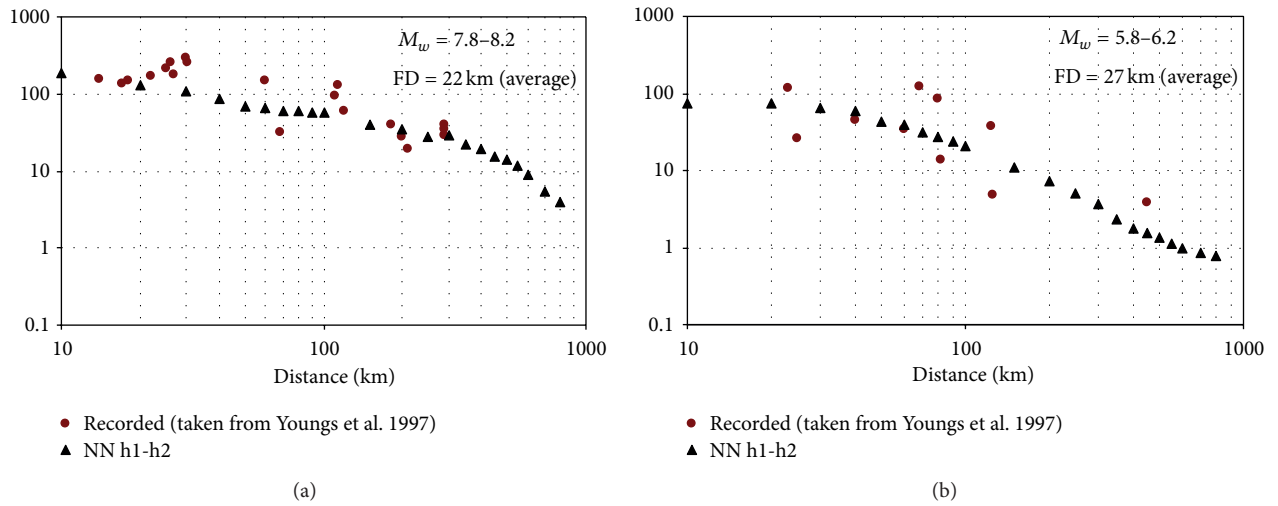


FIGURE 14: Comparison of NN model PGAh1-h2 with data published by Youngs et al. (1997) [19].

capabilities of the networks developed in this paper. It is worth to note that while the NN trend follows the general behavior of the measure data, the traditional functional approaches have predefined extreme boundaries. On the other hand, when the intensity of the earthquake is moderate, most of the PGAs measured in rock sites are within a narrow band, and thus generally the NN and traditional functionals follow similar patterns. The generalization capabilities of the PGAh1-h2 module can be explored even more by simulating other subduction zones events. Measured random horizontal PGAs taken from Youngs et al. [19] belonging to Japan and North America for two magnitude intervals ( $M_w$ : 7.8–8.2 and  $M_w$ : 5.8–6.2) were compared to the NN predictions. These results are plotted in Figure 14. It can be seen that the NN prediction agrees well with the general trend even considering averages of both earthquake magnitude and focal depth.

**3.3. Artificial Generation of Time Series: Accelerograms Application.** For nonlinear seismic response analysis, where the superposition techniques do not apply, earthquake acceleration time histories are required as inputs. Virtually all seismic design codes and guidelines require scaling of selected ground motion time histories so that they match or exceed the controlling design spectrum within a period range of interest (e.g., [21]). After many years of strong motion recording programs, there are now more significant accelerograms that have been recorded, digitized, processed, and analyzed. While the available data represents a unique and invaluable collection for studies and research of strong earthquake ground motion, it does not cover all the need ranges of the parameters commonly used in empirical scaling laws (e.g., earthquake magnitude and focal depth, source to station distance, percentage of rock along the wave paths, and recording geological and soil conditions).

Considerable variability in the characteristics of the recorded strong motions under similar conditions may still require a characterization of future shaking in terms of an ensemble of accelerograms rather than in terms of just one or two “typical” records. This situation has thus created a need for the generation of synthetic (artificial) strong-motion time histories that simulate realistic ground motions from different points of views and/or with different degrees of sophistication. To provide the ground motions for analysis and design, various methods have been developed: (i) frequency-domain methods where the frequency content of recorded signals is manipulated (e.g., [22, 37–39]) and (ii) time-domain methods where the recorded ground motions amplitude is controlled (e.g. [40, 41]). Regardless of the method, it is separated the selection of the representative earthquake and the scaling to match the design spectrum. First, one or more time histories are selected subjectively, and then scaling mechanisms for spectrum matching are applied.

In this research a Genetic Generator of Signals GENES is presented. GENES is a tool for finding the coefficients of a pre-specified functional form, which fit a given sampling of values of the dependent variable associated with particular given values of the independent variable(s). When GENES is applied to synthetic accelerograms construction, the proposed tool is capable of (i) searching, under specific soil and seismic conditions, between thousands of earthquake records and recommending a desired subset that better match a target design spectrum, and (ii) through processes that mimic *mating*, *natural selection*, and *mutation*, of producing new *generations* of accelerograms until an optimum *individual* is obtained. The procedure is fast and reliable and results in time series that match any type of target spectrum with minimal tampering and deviation from recorded earthquakes characteristics.

The objective of GENES, when applied to synthetic earthquakes construction, is to produce compatible artificial signals with specific design spectra. GENES introduces specific seismic and site characteristics taking into consideration that (i) a typical strong motion record consists of a variety of waves whose contribution depends on the earthquake source mechanism (wave path) and its particular characteristics are influenced by the distance between the source and the site, some measure of the size of the earthquake, and the surrounding geology and site conditions; and (ii) the design spectra can be an envelope or integration of many expected ground motions that are possible to occur in certain period of time, or the result of a formulation that involves earthquake magnitude, distance and soil conditions.

The input data consist of the ordinates of the target acceleration design spectrum, the period range for the matching, lower- and upper-bound acceptable values for scaling signal shape, and a set of seismic/soil and GA parameters. The output is the more success chromosome in terms of an accelerations vector (or a set of). Additionally some GAs parameters are required: a population size, number of generations, crossover ratio, and mutation ratio.

The algorithm (see Figure 15) is started with a set of solutions (each solution is called a *chromosome*). A solution

is composed of thousands of components or *genes* (accelerations  $a_i$  recorded at the time  $t_i$ ), each one encoding a particular trait. The initial solutions (original population) are selected based on seismic parameters at a site (defined by the user): moment magnitude, epicentral distance, geotechnical and geological site classification, depth of sediments and component direction. If the designer does not have *a priori* seismic/site knowledge, GENES selects the initial population randomly (Figure 16). One of the GENES advantages is the possibility of modifying *on line* the image of the expected earthquake. While the GAs is running the user interface shows the individual per epoch and its response spectra in the same window, if the duration time, the highest intensities interval or the  $\Delta t$  are not convenient for the designer's interests, these values can be modified without GENES retraining or a change on its structure (Figure 17).

Proposing an entire recorded accelerogram as a *chromosome*, the space of all feasible solutions can be called *accelerograms space* (state space). Each point in this search space represents one feasible solution and can be “marked” by its value or fitness for the problem. The looking for a solution is then equal to a looking for some extreme (minimum or maximum) in the space. The *accelerograms space* can be whole known by the time of solving a problem, but usually only a few points from it are known and other points as the process of finding solution continues are generated.

According to the individuals' fitness, expressed by difference between the target design spectrum and the *chromosome* response spectrum, the problem is formulated as the minimization of the error function,  $Z$ , between the actual and the target spectrum in a certain period range. Solutions with highest fitness are selected to form new solutions (*offspring*). During *reproduction*, the recombination (or crossover) and mutation permits to change the *genes* (accelerations) from *parents* (earthquake signals) in some way that the whole new chromosome (synthetic signal) contains the *older* organisms attributes that assure success. This is repeated until some user's condition (e.g., number of populations or improvement of the best solution) is satisfied (Figure 18). The program is very fast and it takes only few minutes to converge to an optimum solution on a PC.

In Figure 19 three examples of signals recovered following this methodology are shown. The examples illustrate the application of the GENES to select any number of records to match a given target spectrum (only the more successful individuals for each target are shown in the figure). It can be noticed the stability of the genetic algorithm in adapting itself to smooth, code, or scarped spectrum shapes.

In GENES, contrary to other techniques, the characteristics of seismic source, path attenuation, and local soil conditions have been taken into account explicitly when generating synthetic ground motions. Given seismic/site conditions and a target (design response spectra in this accelerograms application), the processes inspired in Darwin's theory about evolution (GENES mimics *mating*, *natural selection*, and *mutation*) generate accelerations time histories following a very simple method. The procedure is fast and reliable as the results in records match the target spectrum with minimal

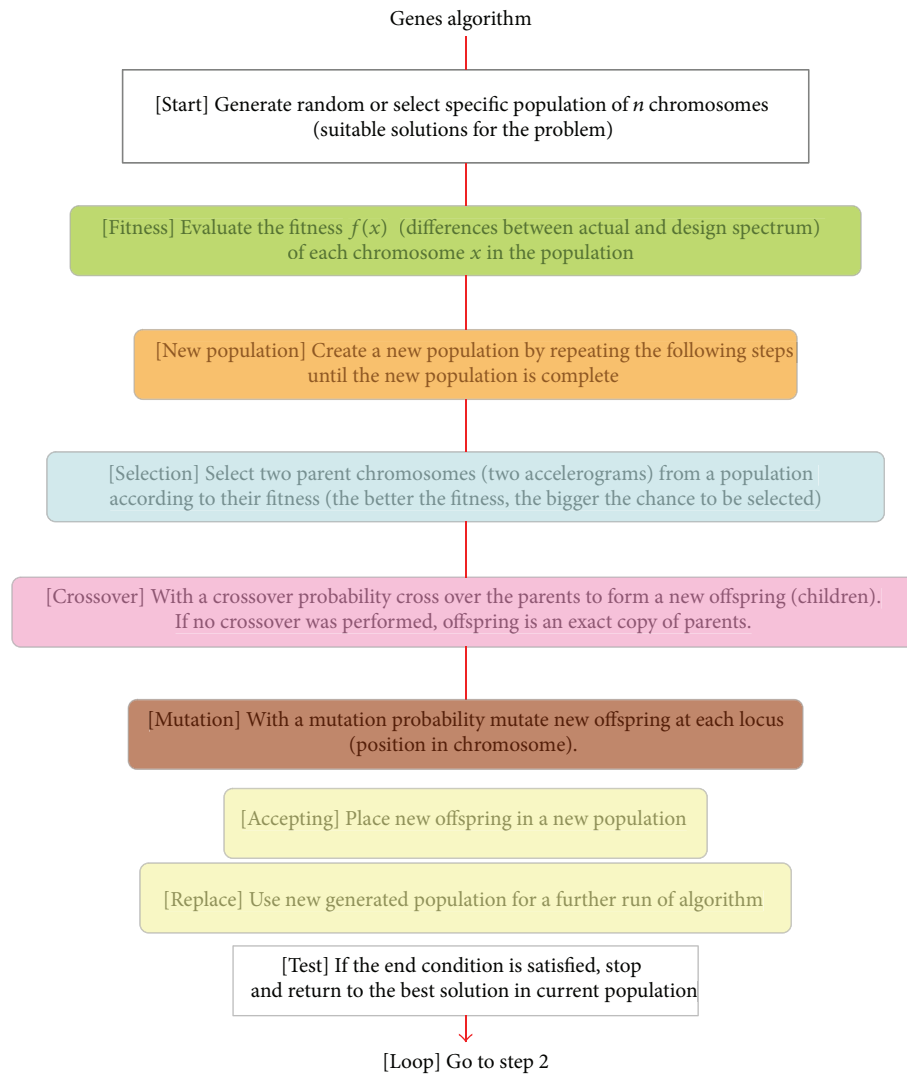


FIGURE 15: Genetic Generator: flow diagram.

deviation. GENES has been applied successfully to generate synthetic ground motions having different amplitudes, duration, and combinations of moment magnitude and epicentral distance. Although the variations in the target spectra, the resulted signals preserve the nonlinear and nonstationary characteristics of real earthquakes.

As an advantage, in GENES it is possible to incorporate the uncertainties related with geotechnical and/or seismological effects on the earthquake wave forming process. An additional toolbox is still under development that will permit the use of advanced signal analysis instruments because, as it has been demonstrated (e.g., [42, 43]), monitoring nonstationary signals through Fourier or response spectra is not the most convenient feat. Supplementary guidelines for practicing engineers will be implemented in order to make GENES a self-contained kit for risk seismic analyses.

**3.4. Liquefaction Phenomena.** Soil liquefaction and related ground failures are commonly associated with large earthquakes. In common usage, liquefaction refers to the loss of

strength in saturated, cohesionless soils due to the build-up pore water pressures during dynamic loading. The losses are attributed to the earthquake-induced liquefaction phenomenon bill of up to hundreds of millions dollars all over the world each year. Therefore, the assessment of the liquefaction potential and the associated damages is an imperative task in earthquake geotechnical engineering. In this section, two models for analyzing this important topic are presented: (i) a classification tree for predicting the occurrence of liquefaction and (ii) a neurofuzzy model for estimating the lateral spreading induces by liquefaction.

**3.4.1. Classification Tree for Liquefaction Occurrence Prediction.** Over the past forty years, scientists have conducted extensive research and have proposed many methods to predict the occurrence of liquefaction. In the beginning, undrained cyclic loading laboratory tests had been used to evaluate the liquefaction potential of a soil [44] but due to difficulties in obtaining undisturbed samples of loose sandy soils, many researchers have preferred to use *in situ*



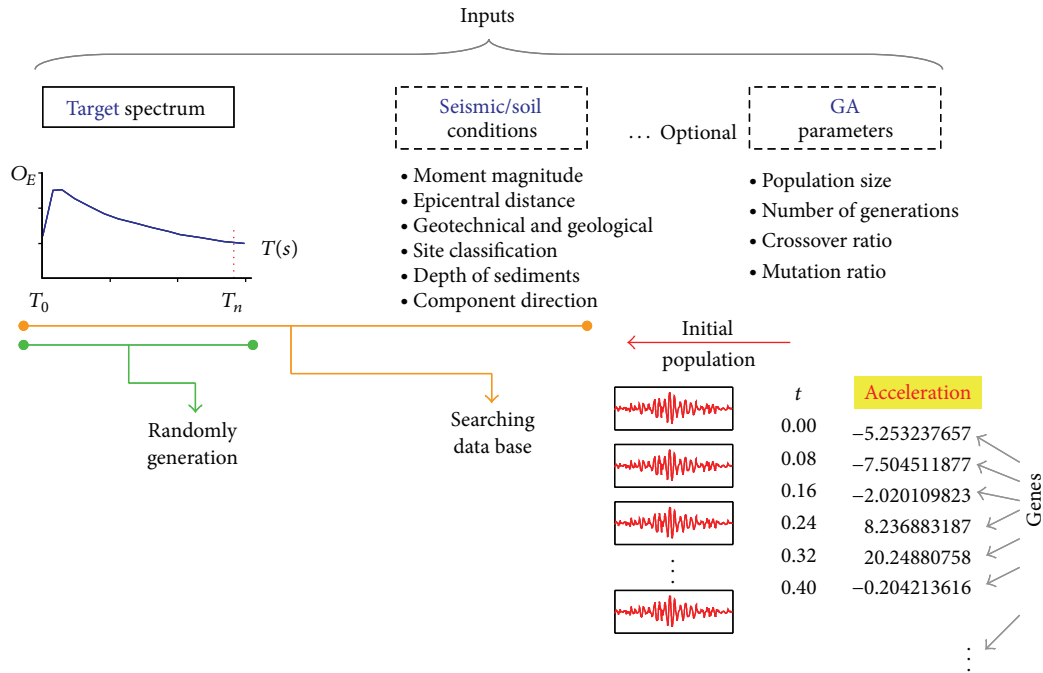
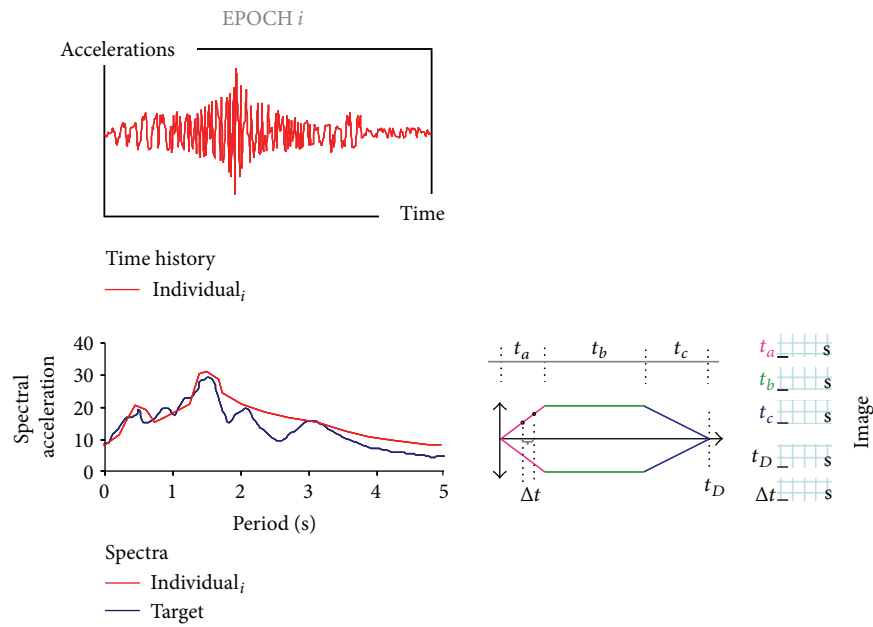


FIGURE 16: Genetic Generator: working phase diagram.

FIGURE 17: Genetic generation: modifications *on line*.

tests [45]. Empirical field-based procedures for determining liquefaction potential have two critical constituents: (i) the analytical framework to organize past experiences, and (ii) an appropriate *in situ* index to represent soil liquefaction characteristics. In a semiempirical approach the theoretical considerations and experimental findings provide the ability to make sense out of the field observations, tying them together, and thereby having more confidence in the validity

of the approach as it is used to interpolate or extrapolate to areas with insufficient field data to constrain a purely empirical solution.

The original simplified procedure [46] for estimating earthquake-induced cyclic shear stresses continues to be an essential component of the analysis framework. The refinements to the various elements of this context include improvements in the *in situ* index tests (e.g., standard

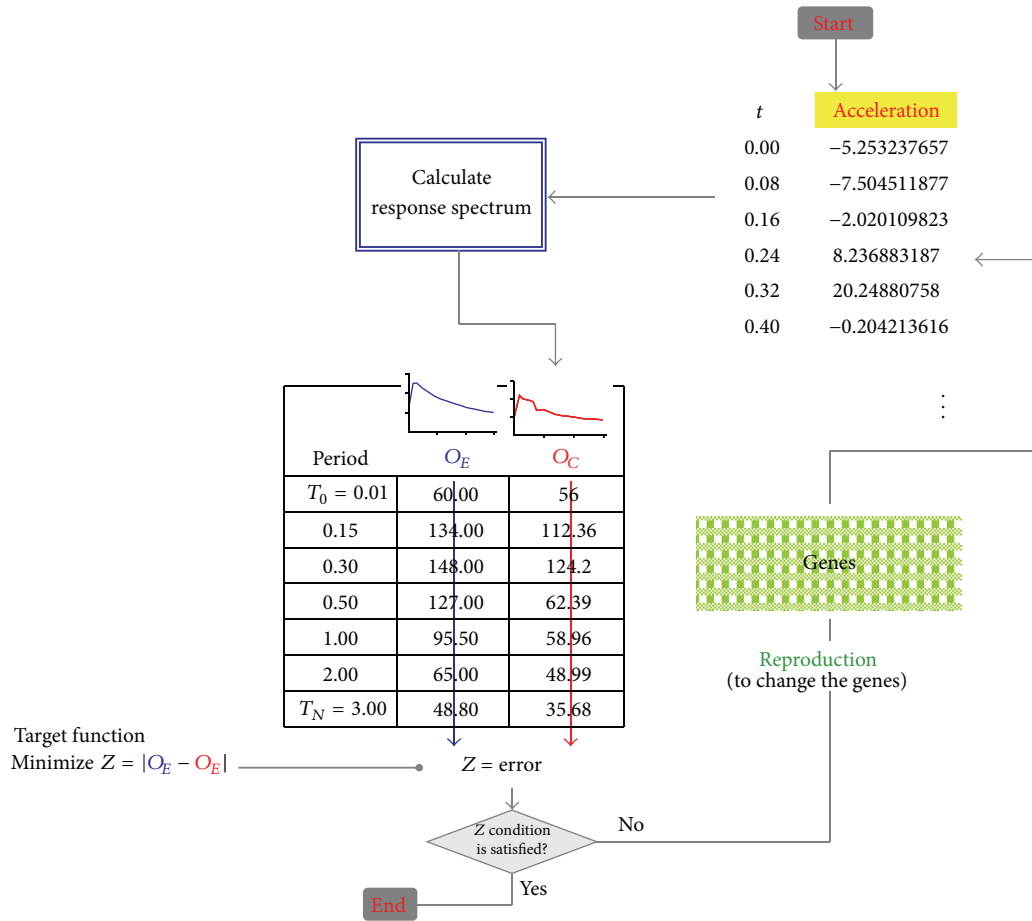


FIGURE 18: Iteration process of the genetic generator GENES.

penetration test SPT, cone penetration test CPT, self-boring pressure meter tests BPT, and shear wave velocities  $V_s$ ). Unfortunately, as new liquefaction cases from recent earthquakes have become available, these empirical plots have to be modified for calibrating the boundary between states [47]. In recent years, a powerful computing tool, artificial neural networks (NN), has been introduced for solving the problem of assessing the liquefaction potential (two classes pattern recognition). Many researchers have reported similar or superior accuracy to that of simplified methods using NN in discriminating between liquefaction and nonliquefaction cases [48–54]. They adopted different types of NN architectures and various combinations of input variables for evaluating liquefaction potential from field records (both CPT- $q_c$  and SPT-N datasets), concluding that NN are simpler than and as reliable as conventional simplified methods. Despite the notable performing of these neural approximations, these “black box” models have central shortcomings: (i) their impractical knowledge interpretation, (ii) their slight power to determine the strategic parameters (relative importance), and (iii) the existing NN models cannot be used for others than the model designers.

Based on the recognized weaknesses when using NN, this study presents an empirical machine learning ML model for evaluating liquefaction potential. ML, a branch

of soft computation, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviours based on empirical data (numerical/categorical). Between the ML representations, a classification tree (CT) was selected to determine the liquefaction occurrence and to uncover the character of driving parameters, including earthquake and soil conditions [55–58], being a crucial concern the representation as the more direct way to understand the information structures and relate them to the data from which they came.

The CT for liquefaction prediction establishes a natural connection between experimental and theoretical findings. The CT is presented as a feasible tool for discovering unknown, valid patterns and relationships between geotechnical, seismological, and engineering descriptions through the relevant available information about liquefaction occurrence around the world (expressed as empirical prior knowledge or input-output data).

**Database.** The database used in this study was constructed using the information compiled by Juang and Chen [51], Juang and Jiang [59], and Andrus et al. [60]. A summary of the parameters included in these datasets is presented in Table 1. From the 407 patterns, the 53% are cases where liquefaction occurred and the other 47% cases are nonliquefied ones. The 80% of the lines were selected as training

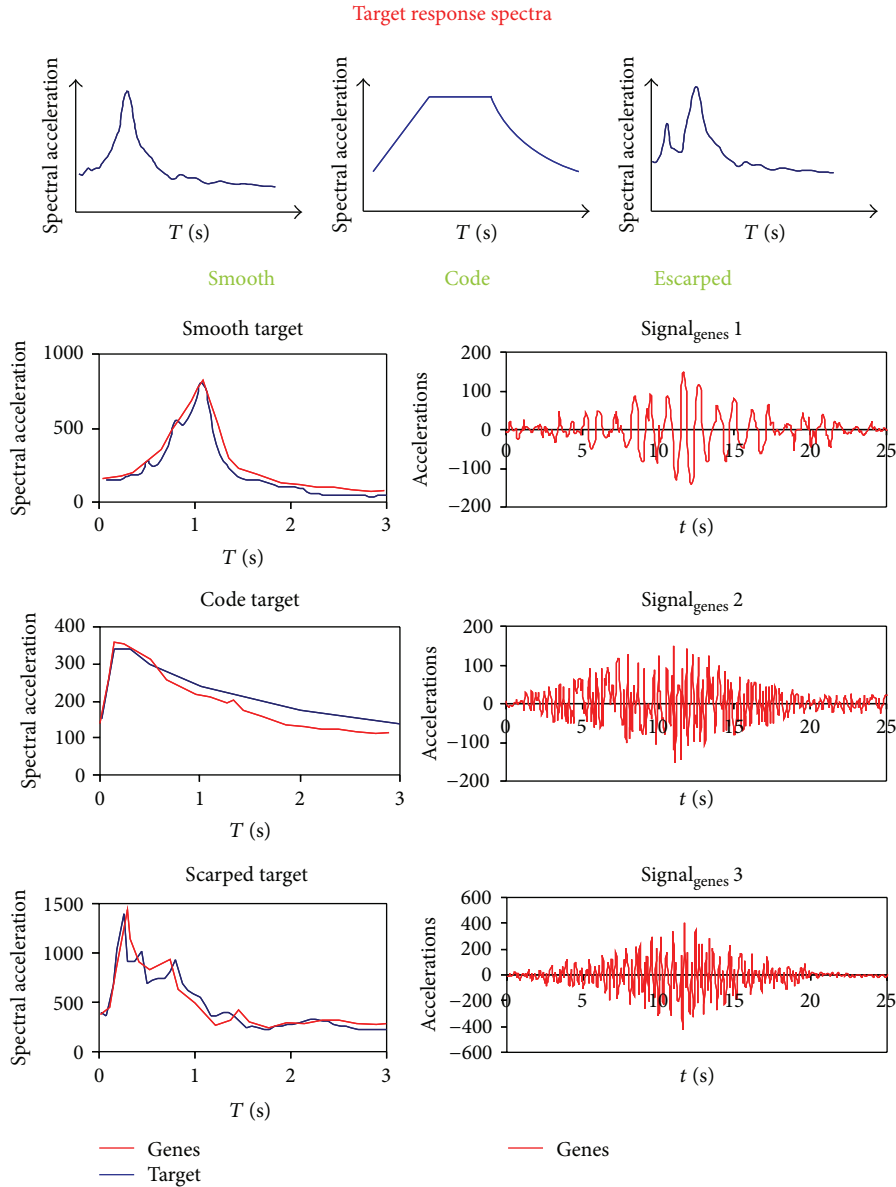


FIGURE 19: Some generator results: accelerograms application.

patterns (used during the model construction) and the 20% was separated for testing the CLIP generalization capabilities. The information is derived from CPT and  $V_s$  measurements and different seismic conditions (USA, China, Taiwan, and Japan). The soils types range from clean sand and silty sand to silt mixtures (sandy and clayey silt). Diverse geological and geomorphological characteristics are included. The reader is referred to the citations in Table 2 for details.

In Table 2, according to nomenclature in each original database,  $Z_{\text{TOP LAYER}}$  is the top layer depth,  $Z_{\text{NAF}}$  is the water table depth,  $H$  is the layer thickness,  $a_{\text{max}}$  is the maximum registered acceleration (peak ground acceleration (PGA)),  $q_c$  is the cone penetration resistance,  $R_s$  is the fine content,  $\sigma_v$  is the effective vertical stress and  $\sigma_v$  is the total one,  $M$  is the magnitude,  $V_s$  is the shear wave velocity, and  $S_T$  is the soil type.

*CT Construction.* The general CSR and CRR formulations by Seed and Idriss [61] and Youd et al. [62] are adopted in the CT proposal. The cyclic stress ratio definition includes

$$\text{CSR} = f_{\text{ML}}(M, \sigma_v, \sigma'_v, a_{\text{max}}) \quad (2)$$

while the cyclic resistance ratio the CT follows the expression

$$\text{CRR} = f_{\text{ML}}\left(\frac{q_c}{V_s}, \sigma'_v, S_T\right). \quad (3)$$

For the CT construction, none of these variables are adjusted, preclassified, or transfigured as in other approximations that must be done (i.e., the correction of  $q_c$  for overburden stress). This implies that any restrictive or subjective hypothesis needs to be included to predict the liquefaction occurrence. It should be noted that the soil type  $S_T$  is a parameter

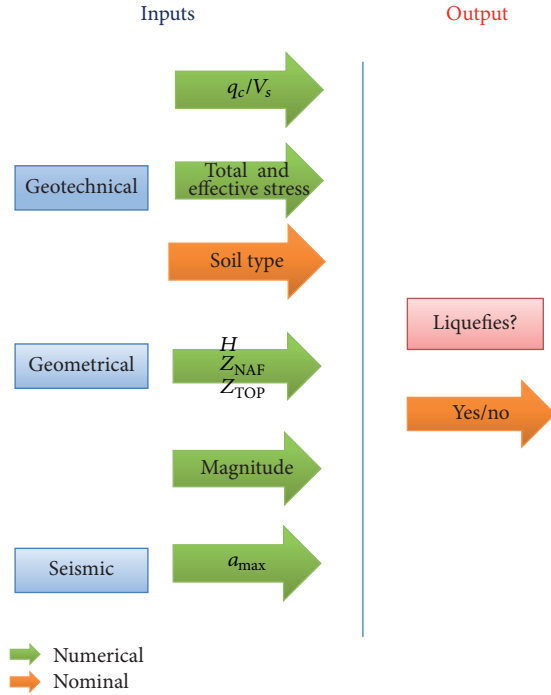


FIGURE 20: Inputs-output schema.

that has not been properly defined [63] and it is still a debate point between researchers. In the CT pursuing the classification reported in the original data base is included as categorical instance. Although there are more sophisticated soil definitions, they require multiple steps of subjective calculation and the proposed instance is easier to use and is more suitable for training the machine learner model and does not need to be calibrated along with other formulas, methods, or the model results.

The final schematic representation of the liquefaction tree model is shown in Figure 20. The following input variables were booked.

- (1) Geotechnical: cone penetration resistance " $q_c$ ", shear wave velocity " $V_s$ ", soil type " $S_T$ ", fines content  $R_f$ , and total and effective stresses " $\sigma_v$ ", " $\sigma'_v$ ".
- (2) Geometrical: layer thickness " $H$ ", water level depth " $Z_{NAF}$ ", and top of layer depth " $Z_{TOP LAYER}$ ".
- (3) Seismic: maximum magnitude " $M$ ", and peak ground acceleration " $a_{max}$ ".

The output variable is "Liquefies?" and it can take the categorical linguistic values "YES"/"NO".

The objective of the tree training is to determine the more suitable representation for acquiring knowledge and structures and relate them to the data from which they came. The discovered patterns are accepted as successful as the prediction from the tree for the given set of inputs ( $M, \sigma_v, \sigma'_v, a_{max}, q_c/V_s, S_T$ ) matches the target (Liquefies? YES/NO), based on field observations. The CT was built through a process known as binary recursive partitioning. After the iterative process of splitting the data into partitions, and splitting it up further on each of the branches, all of

the records in the training set (the preclassified records that are used to determine the structure of the tree) initially put together in one big box are breaking up using every possible binary split on every field. The resulting classification tree is a connected acyclic graph with remarkable advantages over regression, discriminant analysis, and other procedures based on algebraic models.

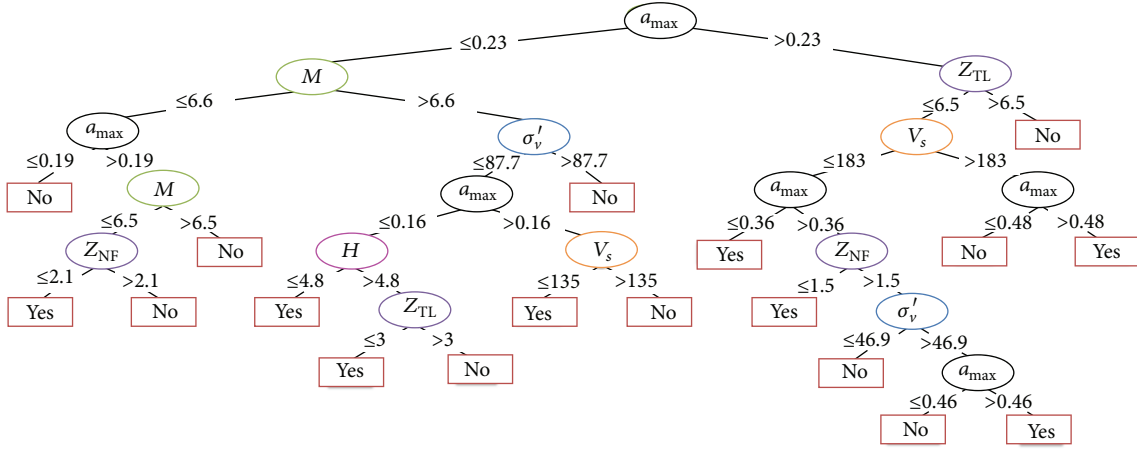
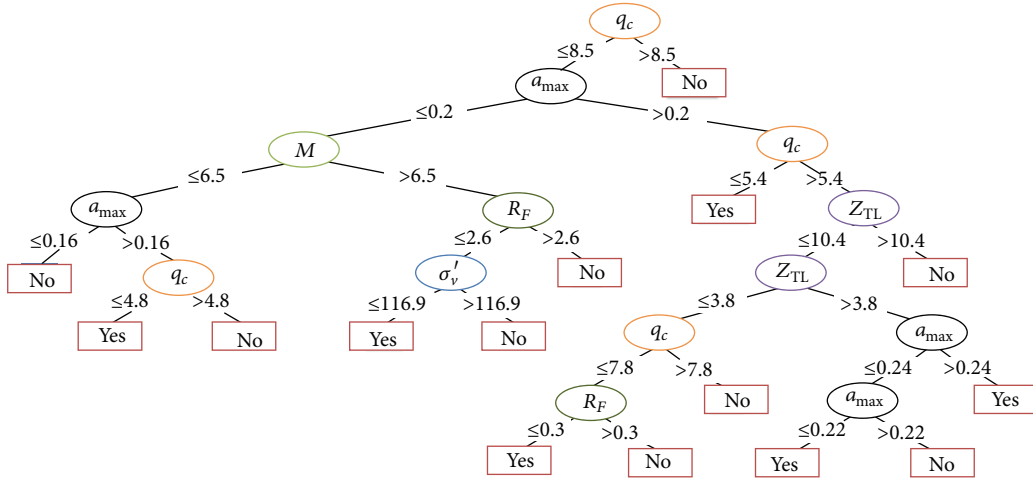
The CT for liquefaction analysis was pruned to minimize the sum of (1) the output variable variance in the validation data, taken a terminal node at a time, and (2) the product of the cost complexity factor and the number of terminal nodes. After working with global liquefaction data, an automatic procedure for detecting interactions among variables (related to classical cluster analysis) generates the bulky trees shown in Figures 21 and 22.

**CT Interpretation.** The input data for the liquefaction occurrence is complex and contains many different categories and many possible predictors for performing the classification (deficient taxonomy) and then the resulting tree is large. This is not so much a computational problem as it is a problem of presenting the trees in a manner that is easily accessible to the analyst or for presentation to the "consumers" of the research. Figures 21 and 22 show the pruned model trees (after CCF operation) using  $q_c$  and  $V_s$  as geotechnical inputs, respectively, and sharing the terminal class nodes (YES/NO). The trees for predicting the liquefaction occurrence by concerns about seismic, geotechnical, and geometrical parameters can serve as a basis for structuring the discussions about phenomena-parameterization strategy. Notice that the classes predicted are at the bottom of the tree and any further step is not necessary. The practical exploit of this tool is straight forward: the user comes into the CT and presents basic parameters for defining the event and site conditions (even if there are missed attributes a CT can be exploited) then each branch and node of the tree is tracked for offering, in the terminal node, a simple conclusion about liquefaction occurrence.

To use the tree structure, the analyst has to tag on the branches in line with the instance being analyzed, and when it reaches a terminal node a simple class expression is given for determine the concept according to the attributes and values contained in the example. In engineering practice, it is very uncommon to get a full description of soil conditions (*in situ* testing results); therefore, the CT was developed for evaluating the liquefaction occurrence having  $V_s$  or  $q_c$  data information.

When the dynamic variable  $V_s$  is being used, the CT has the following configuration (Figure 21): the driven variable is  $a_{max}$  and the partitioning algorithm found a behavior boundary in 0.23 g. For seismic events where the intensity is under this value, the following driving parameter is  $M$ ; lower magnitudes ( $<6.6$ ) need the definition of  $a_{max}$  and  $Z_{NAF}$  for offering the prediction, while higher magnitudes ( $>6.6$ ) require the numerical description of  $\sigma'_v$ ,  $H$ ,  $Z_{TOP LAYER}$ , and  $V_s$ . This structure and the relations contained are very important for understanding the "occurrence" phenomena. For minor cyclic loading, the CT uses information about water level depth and the maximum accelerations as its driven



FIGURE 21: CLIP: Geotechnical description  $V_s$ .FIGURE 22: CLIP: Geotechnical description  $q_c$ .

parameters, while for greater loads, the criteria for selecting the unsafe situations is based on practically the whole set of geotechnical, geometrical, and seismological variables.

If the geotechnical description available is the mechanical variable  $q_c$ , the CT has the configuration depicted in Figure 22. The primary categorical variable is  $q_c$  with a boundary around 8.5 MPa for the first split. When the analyzed soil layer has a  $q_c > 8.5$  MPa, there is no possibility of liquefaction under the dynamic range of  $a_{\max}$  and  $M$  included in the training examples, and this indicates that the differences in the amplitudes, magnitudes, and geometries are critical when analyzing soils with resistances under 8.5 MPa. Then, to predict the liquefaction occurrence for soils with  $q_c < 8.5$  MPa, it is necessary (i) to declare  $M$  value, (ii) to define the fine content percent, and (iii) to declare  $\sigma'_v$  value. As a consequence of this tree separation, appealing conclusions can be achieved, for example,  $Z_{\text{TOP LAYER}}$  related to  $Z_{\text{NAF}}$  explanation is important for accelerations above 0.2 g

while for those under this frontier the percent of fine particles contained in the soil mass is more important.

As it was exposed, besides the huge size of the CT, important features about the physics of the problem can be easily detected. For example, between the databases used in this investigation, many input soil classes are declared (gravel, sand, gravel and sand, sandy silt, silty sand, and clayey silt) but when these labels are studied in a multi-dimensional environment, the relationships derived from experience registered in the databases eliminate some of them (broader categories are detected) and still having satisfactory prediction results. The CT shows that the refinement in numerous labels for  $S_T$  is not in benefit of assertive prediction of liquefaction occurrence; it seems that the *in situ* soil properties ( $V_s$  or  $q_c$ ) are better classifiers for soil masses and its responses. The pruned classification trees use a total geometrical description ( $H$ ,  $Z_{\text{TOP LAYER}}$ ,  $Z_{\text{NAF}}$ ) as a driven condition to get the final nodes in the set of examined

conditions. Observing the seismological descriptions, to use the CT it is not necessary to separate the instances from minor and severe magnitudes to develop two models (as is required in many published models). The whole dynamic range of  $M$  and  $a_{\max}$  is integrated in the training set being a general conclusion of the learning algorithm; the behavior boundary is at  $M = 6.5$  and  $a_{\max}$  has a direct impact in the liquefaction occurrence only for  $M$  values above 6.5.

The impact of  $R_f$  on the number and direction of tree branches evidences the distracting effect of using a fusion of *specialist* and *numerical* criteria for defining this “indefinable” parameter. The fines content expression seems very subjective, between the selected inputs and output parameters, when trying to discover the numerical relationships.  $R_f$  coupled with  $q_c$  generates a better (smaller and efficient) tree and consequently improved knowledge; on the other hand, when the soil description is through  $V_s$ , its inclusion is not in profit of efficient predictions. This can be explained based on the property nature: one is mechanical ( $q_c$ ) and needs more information for a response under seismic forces and the other one is directly a dynamic property ( $V_s$ ) and it seems sufficient for elaborate useful relations.

Without potentially confusing regression strategies, the attenuation tree validation results (cases not used during the model construction and used for proving the CT generalization capabilities) shown in Figures 23 and 24 can be seen as an indication of how effectively the CT maps the assigned predictor variables to the response parameter. The success rate of the trained CT in predicting the occurrence of liquefaction (i.e., distinguishing liquefied cases from nonliquefied cases) is 99% during the training phase of network development. The success rate of the trained model in predicting the occurrence of liquefaction in the cases in the testing data subset and the overall success rate of the trained network in predicting liquefaction in all cases in the entire dataset is 96%. Using data from many regions and broad ranges of inputs, the prediction capabilities of the tree are superior to many other approximations used in common practice, but the most important remark is the generation of meaningful clues about the reliability of physical parameters, measurement and calculation process, and practice recommendations.

In 1999, two major earthquakes, namely, Chi-Chi, Taiwan, earthquake (magnitude  $M_w = 7.6$ ) and Kocaeli, Turkey, earthquake, (magnitude  $M_w = 7.4$ ) triggered ground failure principally induced by soil liquefaction throughout the city of Adapazari (Turkey) and the cities of Wufeng, Nantou, and Yuanlin (Taiwan). These case records are used for validating the proposed tree model. The multidimensional CT permits a very efficient prediction of the liquefaction occurrence (Figures 25 and 26). If the analyst has uncertainties in the inputs definition, or even some of them are missed, CLIP can be used to understand the effect of the fuzziness on behavior path and to use intelligently the information in the branches and leaves.

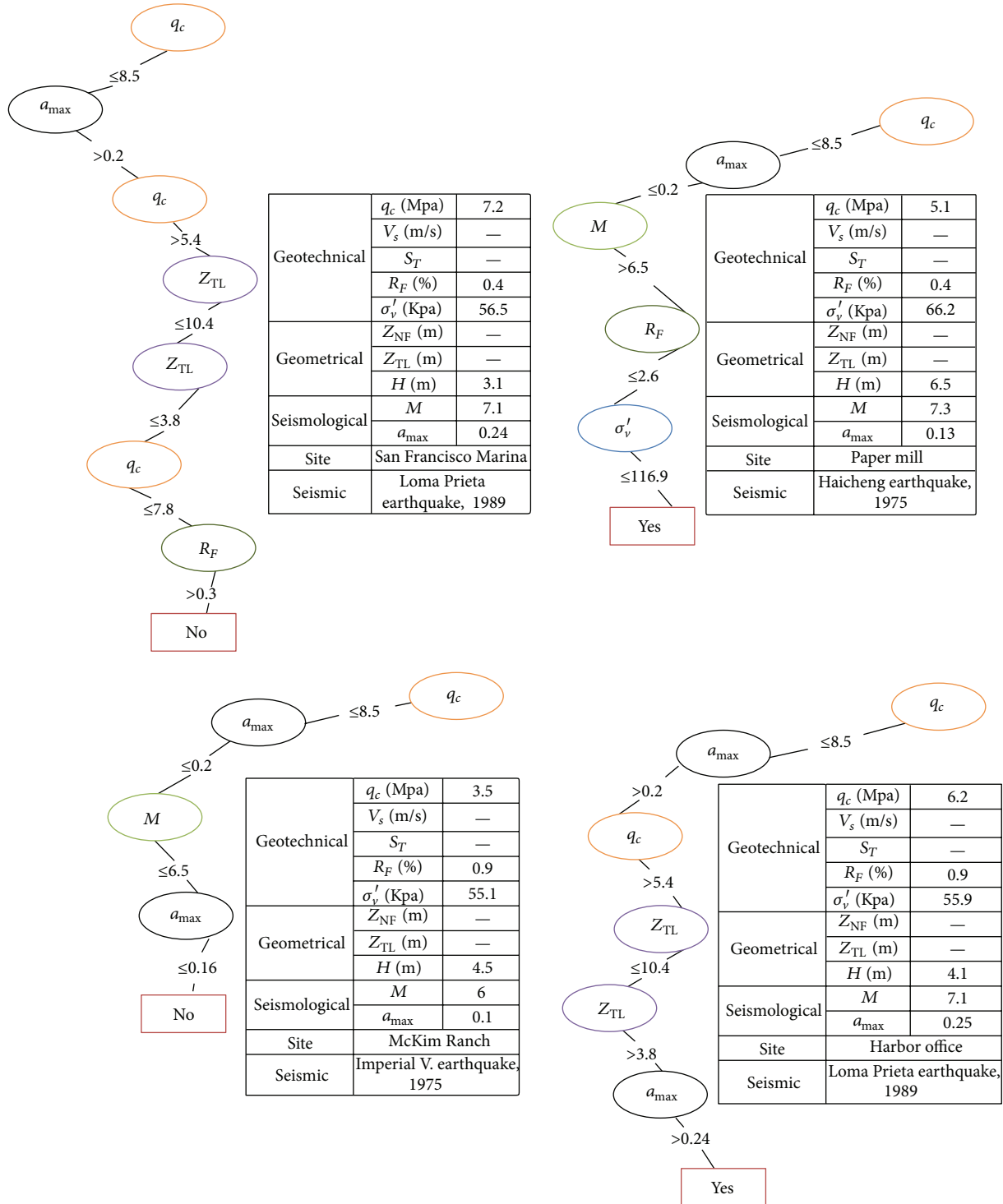
The classification tree is clearly an advantageous tool: practical, free to use, easy to understand, and with straightforward behaviors interpretation. The CT simplicity is useful not only for purposes of rapid classification (or prediction) of

new observations but also for yielding a much simpler model for explaining why observations are classified or predicted in a particular manner (e.g., to analyze input-output parameters importance, to present simple statements to management, or to eliminate elaborate and inaccurate equations). There is no doubt that ML represents a powerful alternative in predicting the liquefaction occurrence and phenomenon related.

**3.4.2. A Neurofuzzy System to Analyze Liquefaction-Induced Lateral Spread.** Lateral spreading is conceivably the most common type ground failure induced by the liquefaction of saturated fine granular materials. Horizontal spreads cause basically two loading mechanisms upon engineered structures. One is due to drag forces exerted, mainly on piles and piers, and the thrust that may be caused by a crust of nonliquefied soil riding on the liquefied layer as it is driven against buried structures. Indeed, piles and piers can be subjected to appreciable loading by the soil flowing past them, causing severe damage. Also, during lateral spreads, blocks of intact, superficial soil displace along a shear zone within the liquefied layer, either down slope or toward a free face (i.e., river, channel, or an abrupt topographical depression) driven by gravity or earthquake forces. The resulting ground deformation usually has extensional fissures at the head of the failure as well as shear deformations at the flanks and compression of the soil at the toe [64].

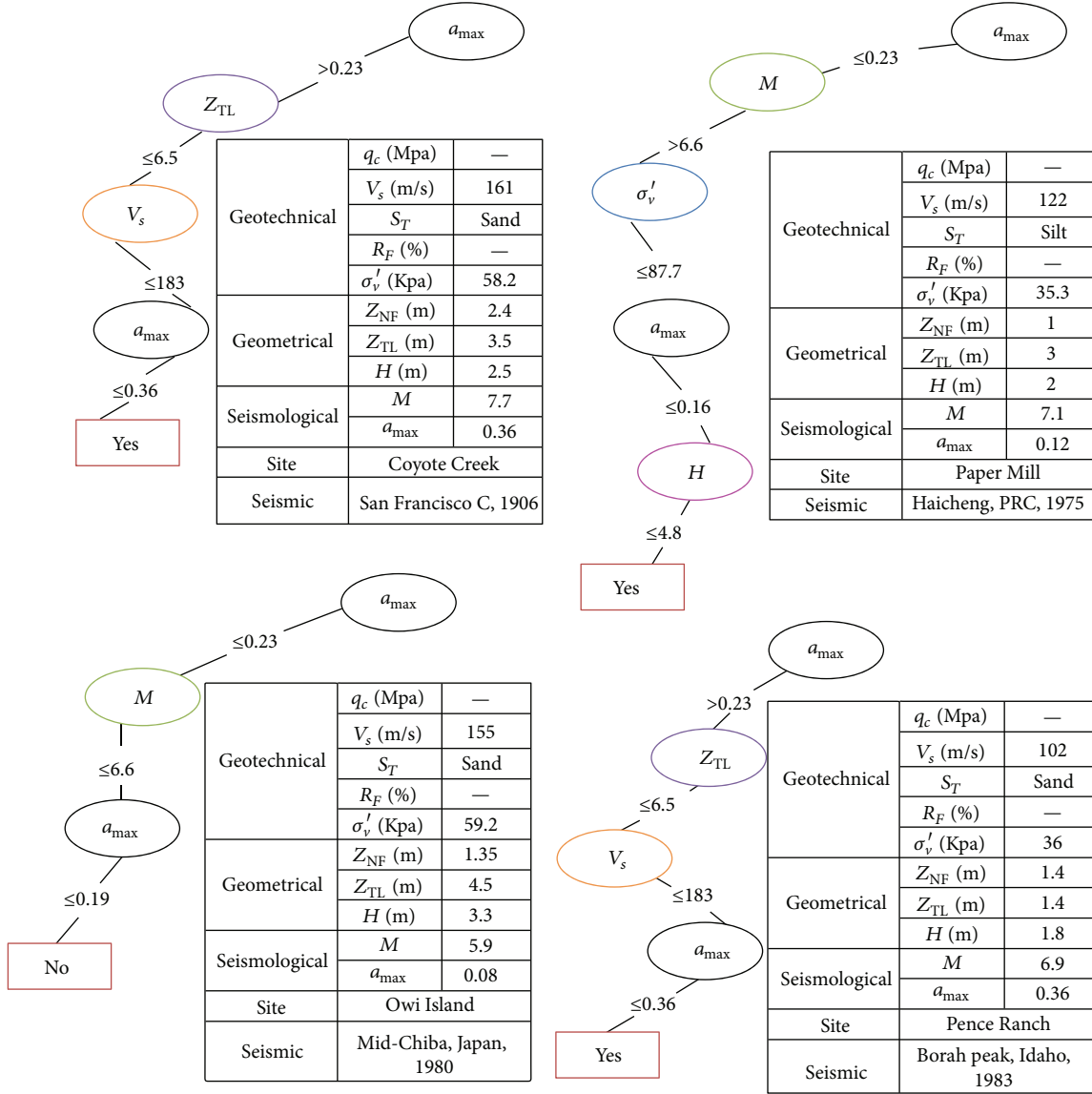
Several researchers (e.g., [65, 66]) have studied drag forces exerted on piles by liquefied soil and have found that such forces are often too small to inflict any damage. However, other studies of centrifuge tests have shown that large forces were applied by the flowing ground (e.g., [67]). This discrepancy on the potential effects may be due mainly to the amount of lateral spread computed by these researchers and thereupon the importance of having means to estimate as accurately as possible the magnitude of lateral spreads caused by seismic events. When the stiffer unliquefied stratum is carried along the underlying, the spreading sand may produce high enough pressures on buried structures to cause them severe damage and even their failure. In the limit, these forces reach the passive resistance of the unliquefied soil.

Displacements ranging from a few centimeters to several meters are commonly developed by this phenomenon [68]. Thus, it is not difficult to infer that engineered structures be severely marred when enduring lateral soil displacements of such a magnitude. Some examples that support this assertion are given by the recorded effects of the 1906 ( $M_w$  7.9) San Francisco earthquake on buildings, bridges, roads, and pipelines. Similarly, the 1964 Alaska ( $M_w$  9.2) and Niigata ( $M_w$  7.5) earthquakes caused extensive damage to engineered structures in cities such as Valdez and Anchorage, Alaska, and Niigata, Japan, respectively. In [69], it caused severe damage to pile bridge foundations induced by the 1987,  $M_w$  6.3, Edgecumbe, New Zealand, earthquake has been reported. The San Pedrito dock-pile foundation was also damaged during the 1995,  $M_w$  8.0, Manzanillo earthquake [70, 71]. Similarly, extensive destruction was induced during the Hyogoken-Nambu earthquake ( $M_w$  7.2) of 1995 to buildings and bridges in the reclaimed land areas along the coastline of Kobe, Japan.

FIGURE 23: Validation tests: using  $q_c$  as geotechnical descriptor.

These examples and many others reported in the technical literature clearly show that lateral spreads on gently sloping ground can severely impair engineered structures. Acknowledgement of the chaotic response of liquefying soil masses to earthquakes and having a clear evidence (as shown later) that the available neural and empirical methods do not adequately forecast lateral spreads, it was deemed important

to develop alternate procedures of analysis. In the following, a powerful computational paradigm, a neurofuzzy NF empirical model, is presented as a tool for the estimation of liquefaction-induced lateral spreads due to seismic loading. It takes into account parameters related to general earthquake characteristics, topographical, regional, geologic, and soil data.

FIGURE 24: Validation tests: using  $V_s$  as geotechnical descriptor.

*Database.* The information compiled by Bartlett and Youd [68] and extended later by Youd et al. [72] includes 448 entries corresponding to seven earthquakes. A summary of this database is given in Table 3. The data was divided (classified) in four main categories according to the known qualitative and quantitative information: (a) ground displacement amplitude; (b) borehole data; (c) boundary conditions, that is, ground-slope and free-face topographical data; (d) seismic knowledge. Since the raw database contains redundant and conflicting information, before being used in the training stage of the neurosystem development, it was preprocessed using a fuzzy clustering technique (see García and Romo [73] for technical details). The input-output data considered in the neurofuzzy system may be represented by the following function:

$$D_h = f(M_w, R^*, A, S, W, L, T_{15}, D50_{15}, F_{15}). \quad (4)$$

Notice that there are some differences with respect to the traditional equations published to calculate  $D_h$ . In the neurofuzzy case, the maximum ground acceleration,  $A$ , in units of gravity,  $g$ , and the length from the free-face to the point of displacement,  $L$ , in meters are included. On the other hand, N160S is not involved because it was found that its dynamic range was too narrow, thus its influence as classifier was found to be negligible [73]. In Figure 27 the main geometric aspects of the lateral spreading problem are depicted, and some of the neurofuzzy parameters  $S$ , and  $L$ , are depicted. The rest of the parameters in (4) are defined in Figure 27. According to (4), the general approach used in determining the horizontal displacements,  $D_h$ , caused by earthquake-induced liquefaction assumes that the pattern of ground displacements can be classified by their topographical variables ( $S$ ,  $W$ ,  $L$ ), geological and soil conditions ( $T_{15}$ ,  $D50_{15}$ ,  $F_{15}$ ), and the earthquake characteristics ( $M_w$ ,  $R^*$ ,  $A$ ).

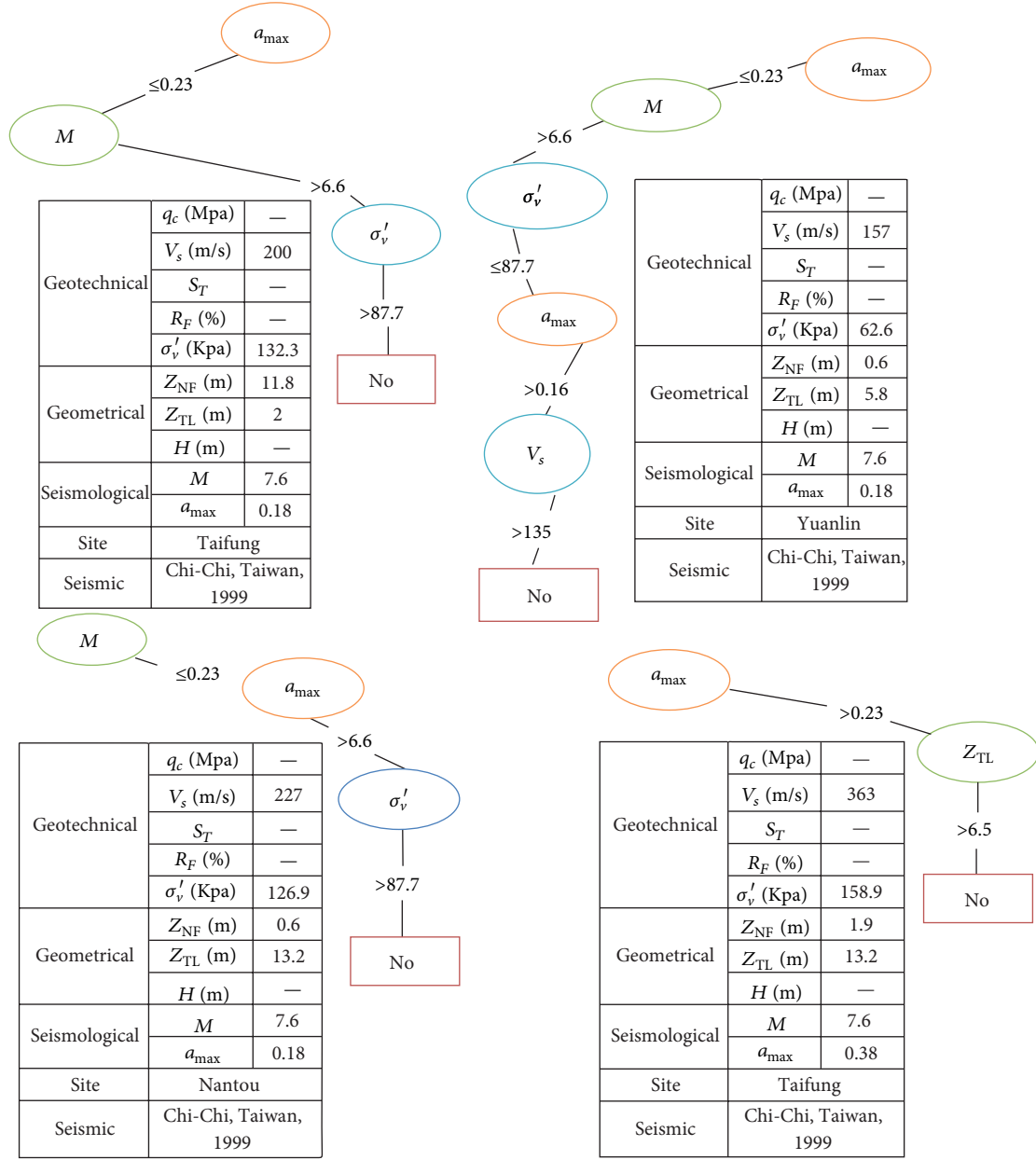


FIGURE 25: CLIP results for Chi-Chi, Taiwan, earthquake.

Therefore, the fuzzy system training process was carried out by considering the coupled effect of all dependent parameters (right hand side of (4)) on the independent variable  $D_h$ .

Prior to the training stage, all data points of the input-output function that showed some redundancy or antagonism were lumped with others of the database employing the fuzzy clustering technique described in García and Romo [73]. In this way, a single data point can replace each of the members of such a cluster to reduce the number of points. These remaining points, called typicals, constitute the patterns that combined with raw data (information points that did not fit any of the classifications defined by the fuzzy clustering technique used) and were utilized in the training of the NF system (see the example of  $S-D_h$  in Figure 28; these are

the originals examples distribution in the training set). After fuzzy preprocessing, the number of patterns was reduced to 337 (see Figure 29 where the typical and raw data for  $S-D_h$  are depicted). These patterns were found after many trials for searching the optimum size of the database (i.e., it represents the dataset having lower incongruence, thus leading to a better distribution in the input variables space).

Out of the 337 patterns, 257 were randomly selected for the training stage and the rest for testing. The training process used the “new” pattern distribution associated with the membership functions (here the popular triangles were considered) initially defined by experts. Now, to proceed with the training process, the membership degree of each membership function has to be obtained. To illustrate the



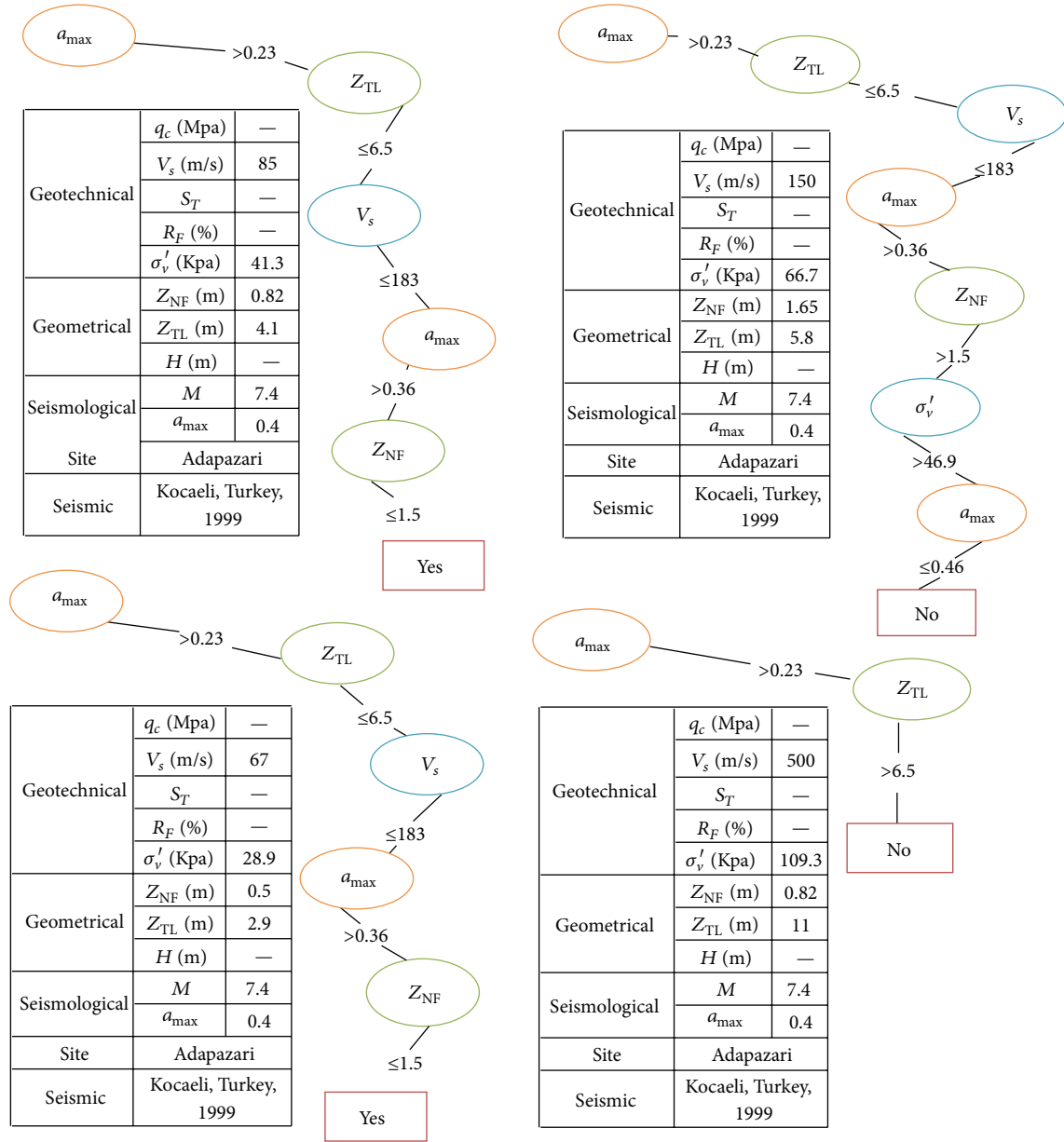


FIGURE 26: CLIP results for Kocaeli, Turkey, earthquake.

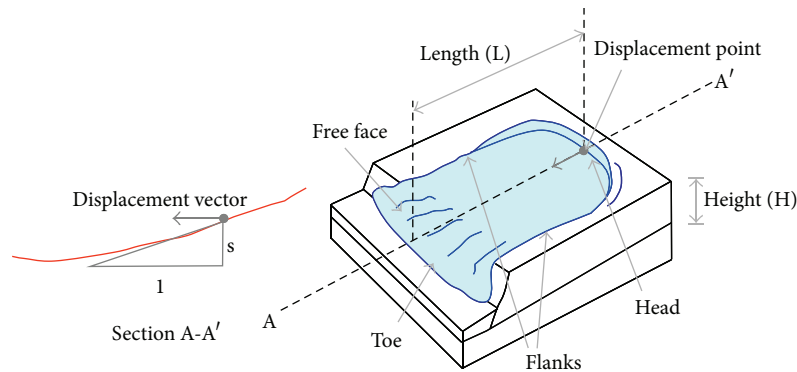


FIGURE 27: Sketch of lateral spread phenomenon.

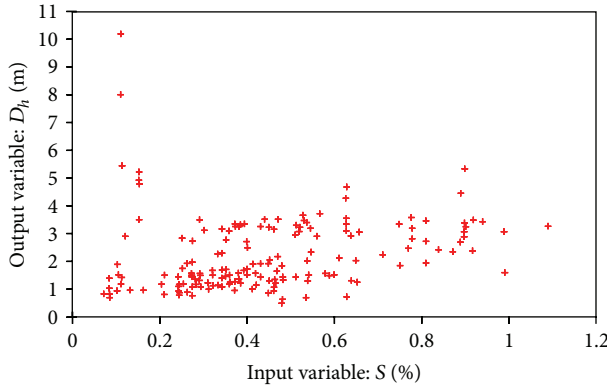


FIGURE 28: Lateral spreading sample dataset. One input-one output.

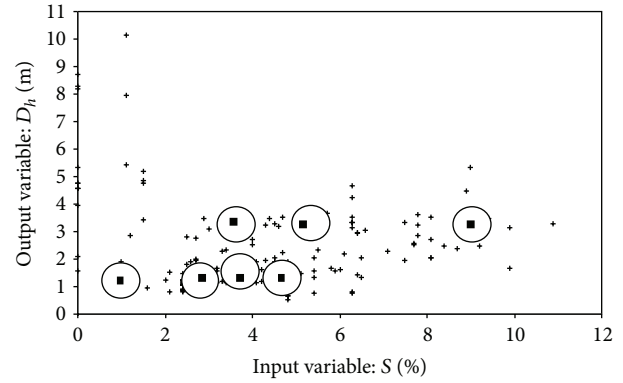


FIGURE 29: Clusters and typicals after fuzzy clustering (example with one input-one output).

procedure to define the membership degree (fuzzification), assume (see Figure 30) that  $S$ : 6.5%, thus HIGH: 0.22, MEDIUM: 0.40, and LOW: 0.00. Once all the fuzzification process is carried out for all the dependent variables ( $M_w$ ,  $R^*$ ,  $A$ , etc.), all possible fuzzy rules (if-then) are developed similarly. Afterwards, the training process begins and the fuzzy rules defined by the experts are modified according to the relationships between all the data patterns.

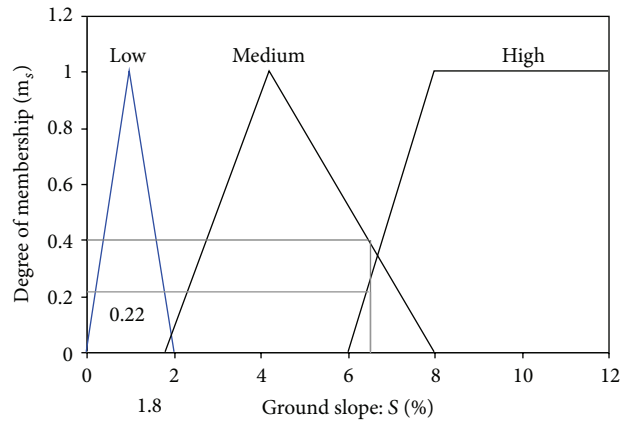
The membership functions defined by the experts are modified until the system response is optimized, which is achieved when the mean-squared error threshold specified is not exceeded. Once the optimization is reached, the system obtained is the fuzzy system that minimizes the  $D_h - f(M_w, R^*, A, S, L, W, D50_{15}, T_{15}, F_{15})$  error mapping.

One of the many resulting if-then rules that optimizes the  $D_h \rightarrow f(-)$  mapping is included below:

<b>IF</b>	
Earthquake	$M_w$ is HIGH and $R^*$ is MEDIUM and $A$ is LOW
Topography	and $S$ is LOW and $W$ is MEDIUM and $L$ is LOW
Soil	and $T_{15}$ is LOW and $D50_{15}$ is LOW and $F_{15}$ is MEDIUM
<b>THEN</b>	
$D_h$ is LOW	

Considering the character of components that influence the lateral-spreading problem, it was decided to build up the neurofuzzy system using the following three modules.

- Reg-neurofuzzy: appropriate for predicting horizontal displacements in geographic regions where seismic hazard surveys have been identified.
- Site-neurofuzzy: proper for predictions of horizontal displacements for site-specific studies with minimal data on geotechnical conditions.
- Geotech-neurofuzzy: more refined predictions of horizontal displacements when additional data is available from geotechnical soil borings.

FIGURE 30: Membership functions for  $S$ .

Since the training process is carried out in parallel for all parameters, the resulting fuzzy system is integrated by all the membership functions, given in the column farthest to the right in Figure 31. This figure also includes the experts' membership functions (fourth column in Figure 31), as well as the input/soft variables and their corresponding units. It may be seen that their base and slope are appreciably modified, depicting the importance of taking into account the numerical interrelationships among all the variables that are thought to dominate the physical phenomenon. This aspect is very important for nonlinear problems having high dimensionality, such as the lateral-spreading problem. The procedures for carrying out this training process and a detailed sketch of the neurofuzzy system are given in [74]. Briefly, this system is incorporated in a model to perform the fuzzy clustering of the raw data and to obtain the improved database which is fed into a NF module, which then performs the NF training that yields the output,  $D_h$ . The schematic representation of the fuzzy structure is depicted in Figure 32. The organization of the model components is such that it allows the prediction of lateral spreads,  $D_h$ , even when no full descriptions of the seismic, topographical, or soil information

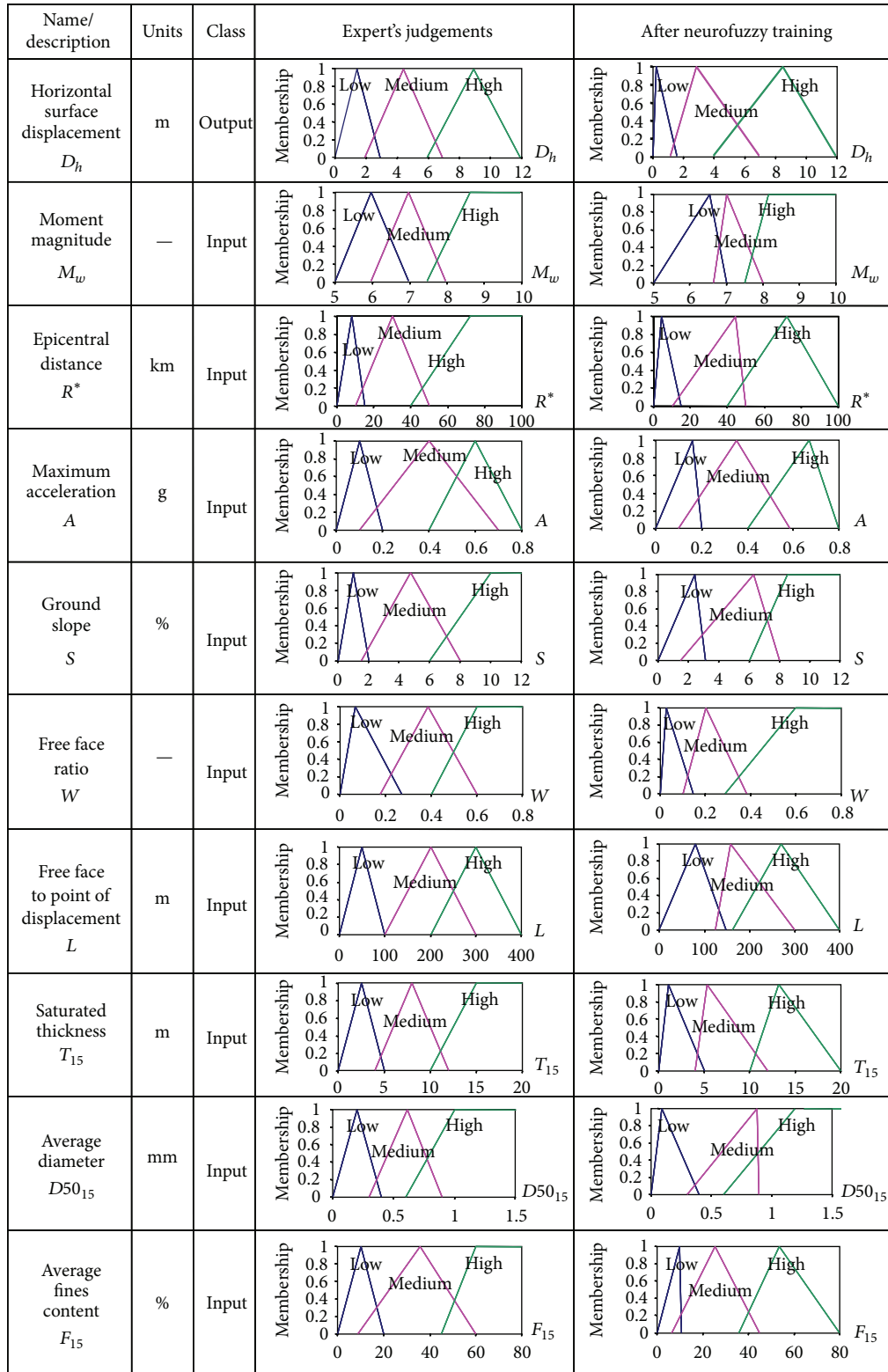


FIGURE 31: Membership functions used as neurofuzzy system variables.

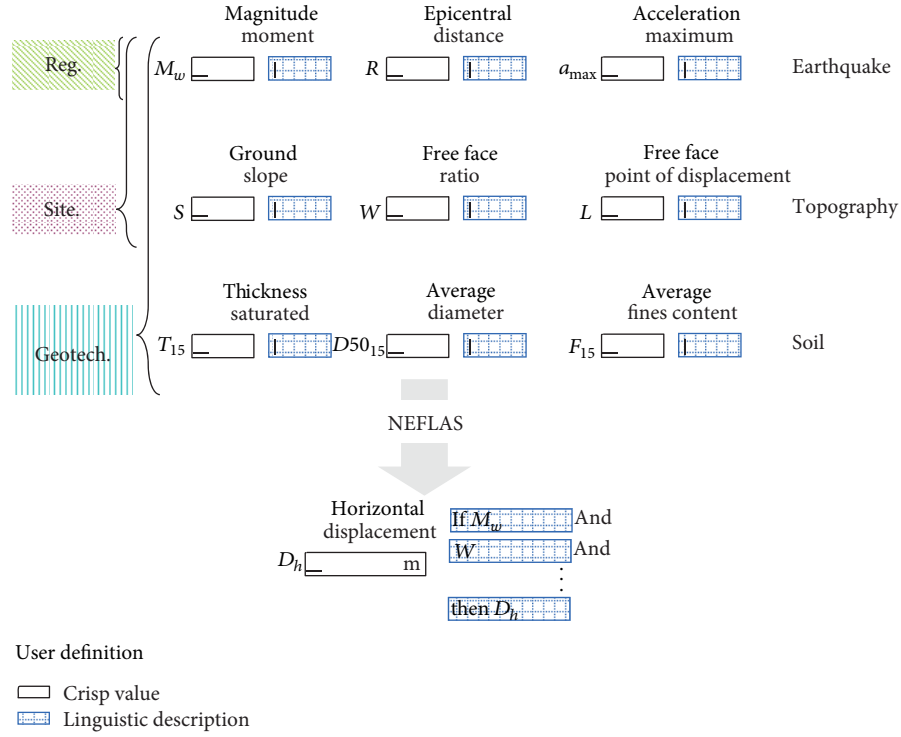


FIGURE 32: Neurofuzzy structure (NEFLAS is for neurofuzzy system to estimate lateral spread displacement).

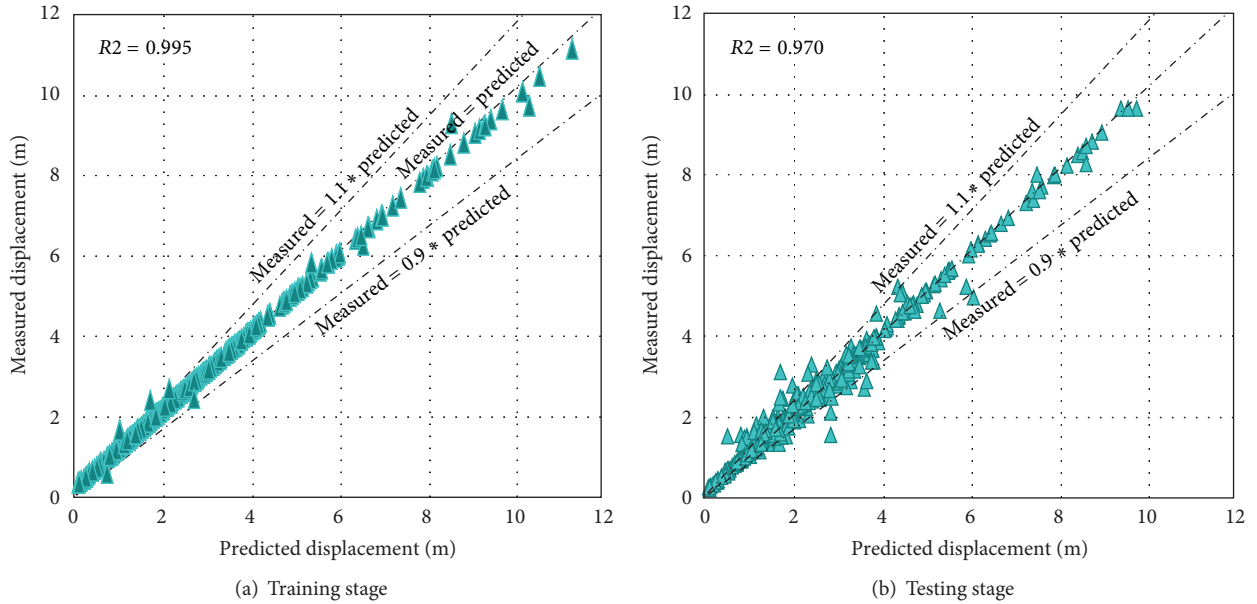


FIGURE 33: Comparison between neurofuzzy and measured lateral spread, whole dataset. (a) Training stage and (b) testing stage.

are available. Lexical assumptions about any input group can be implemented in terms of linguistic data for getting an approximate estimation. But, as would be expected, better predictions can be made as the quality (i.e., less vague and contradictory influences of  $M_w$ ,  $R^*$ ,  $A$ ,  $\gamma$ ,  $F_{15}$ ) of the input variables to the fuzzy system model improves.

The final step in the process followed to develop and assess the reliability of the lateral-spreads predictions by the fuzzy system consisted in feeding this NF model with the set of data (80 patterns) reserved from the whole database. This information was unknown and thus the results can be considered as predictions. Accordingly, by simply comparing fuzzy

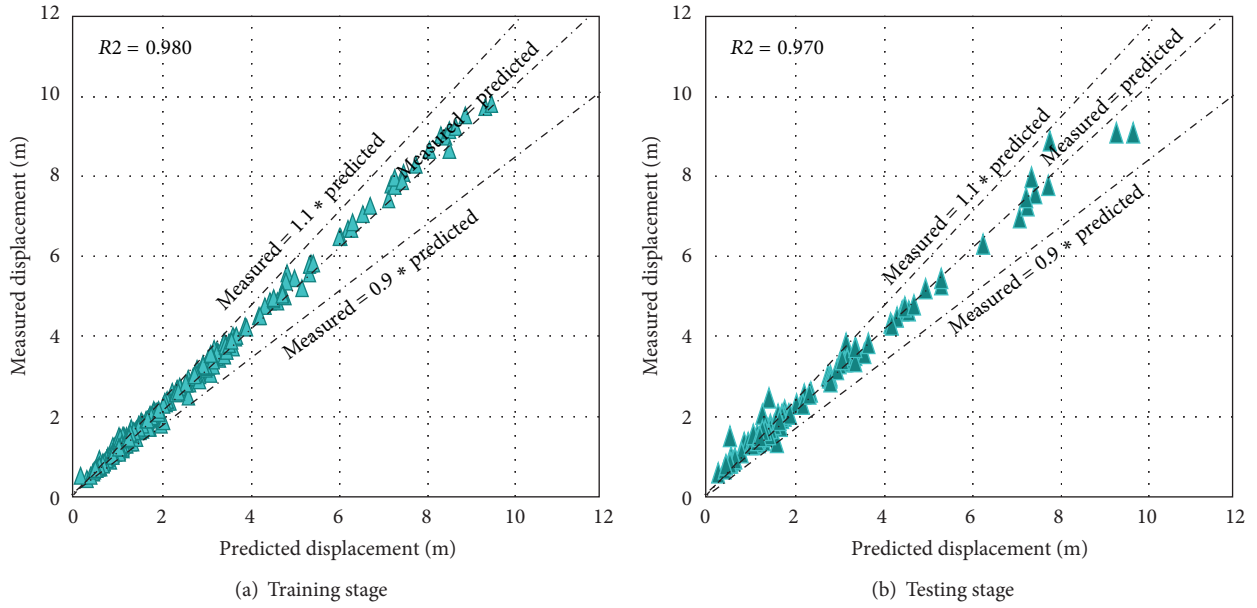


FIGURE 34: Comparison between NEFLAS and measured lateral spread, Niigata, Japan. (a) Training stage and (b) testing stage.

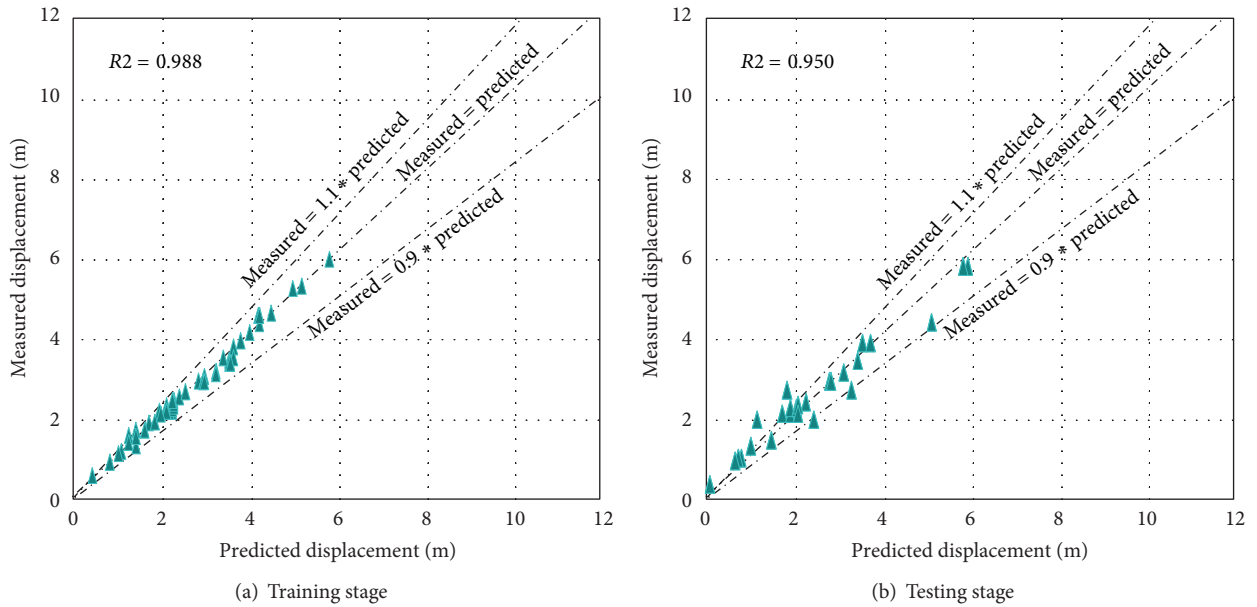


FIGURE 35: Comparison between NEFLAS and measured lateral spread, Nihonkai-Chubu, Japan. (a) Training stage and (b) testing stage.

computed values with the unknown horizontal displacements included in the database, the predicting capabilities can be evaluated. This comparison is depicted in Figure 33(b).

Several aspects stand out that (a) the  $R^2$  (0.97) is much higher than the values obtained with the previous procedures; (b) the  $R^2$  value remains practically unchanged in training and testing stages (see the graphs in Figures 33(a) and 33(b)), and thus it may be asserted that neurofuzzy system is a system that is able to predict lateral spreads with a high degree of confidence; (c) neurofuzzy system is a much more powerful tool than the previously proposed techniques; (d) while the

predicting capabilities of MRL and NNs procedures are rather poor, particularly for horizontal displacements smaller than about 3.0 m, neurofuzzy system shows that it is capable of containing its predictions within a narrow band even at very small horizontal displacements, which supports the accuracy of the proposed computing tool. To stress this assertion, Figures 34, 35, 36, and 37 show the predictions of  $D_h$  for a number of individual seismic events. Notice that the values of  $R^2$  remain practically unchanged when earthquakes and their corresponding dataset are considered independently, which demonstrates the robustness of neurofuzzy system.



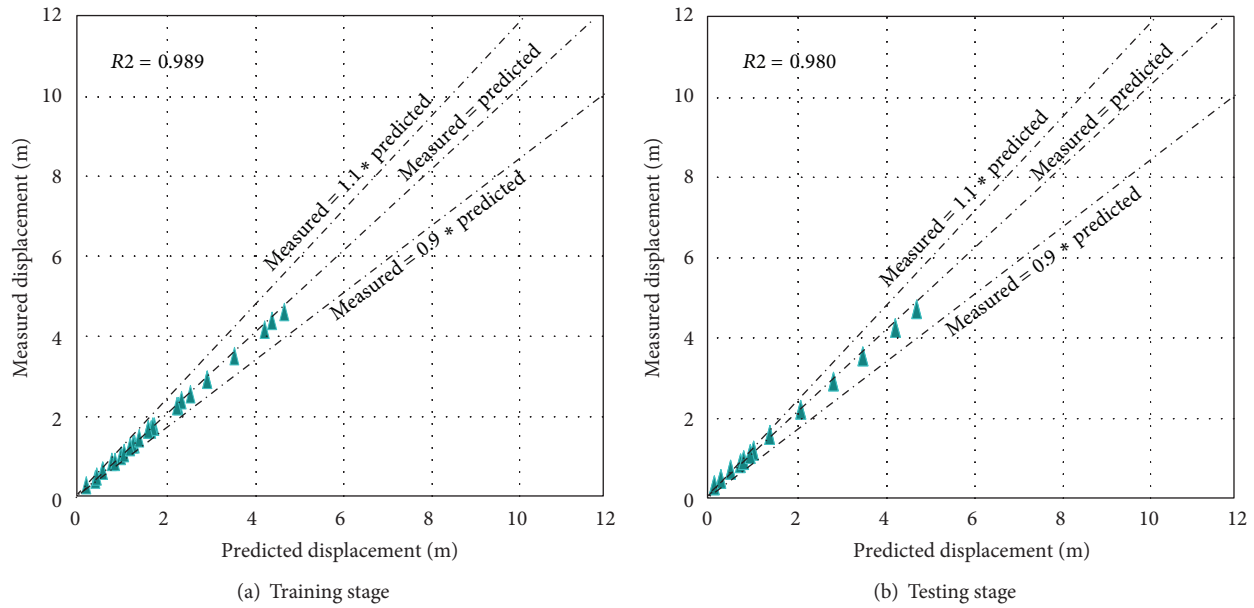


FIGURE 36: Comparison between NEFLAS and measured lateral spread, Imperial Valley, USA. (a) Training stage and (b) testing stage.

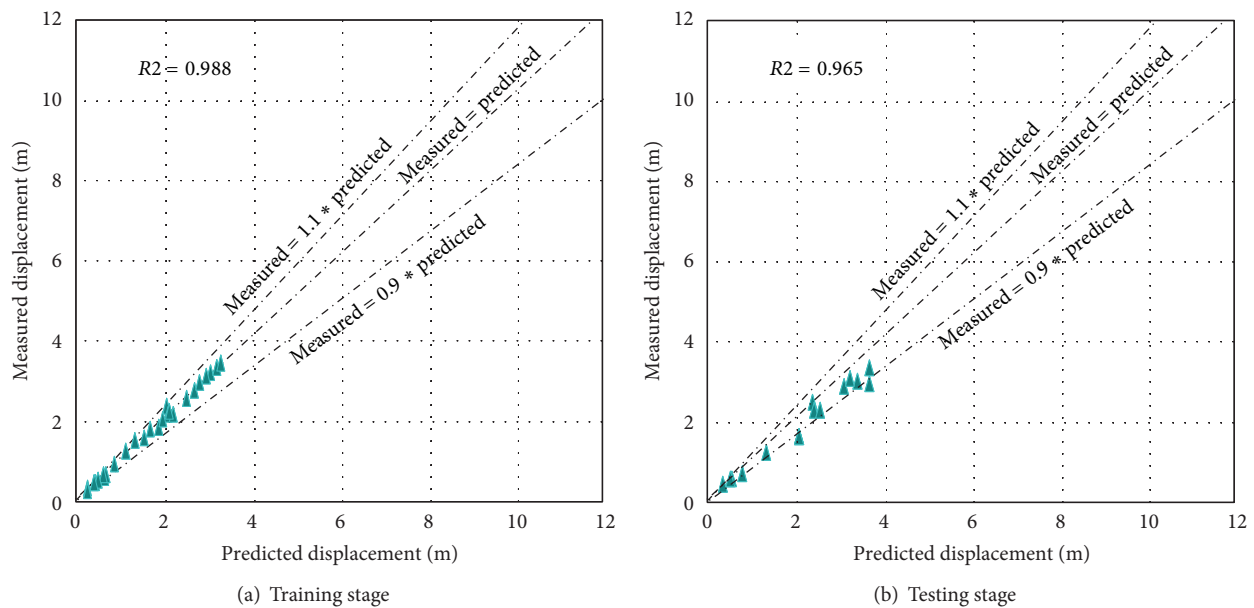


FIGURE 37: Comparison between NEFLAS and measured lateral spread, San Francisco, USA. (a) Training stage and (b) testing stage.

#### 4. Conclusions

Based on the results of the studies discussed in this paper, it is evident that cognitive techniques perform better than, or as well as, the conventional methods used for modeling complex and not well-understood geotechnical earthquake problems. Cognitive tools are having an impact on many geotechnical and seismological operations, from predictive modeling to diagnosis and control.

The hybrid *soft* systems leverage the tolerance for imprecision, uncertainty, and incompleteness, which is intrinsic

to the problems to be solved, and generate tractable, low-cost, robust solutions to such problems. The synergy derived from these hybrid systems stems from the relative ease with which we can translate problem domain knowledge into initial model structures whose parameters are further tuned by local or global search methods. This is a form of methods that do not try to solve the same problem in parallel but they do it in a mutually complementary fashion. The push for low-cost solutions combined with the need for intelligent tools will result in the deployment of hybrid systems that efficiently integrate reasoning and search techniques.

Traditional earthquake geotechnical modeling, as physically based (or knowledge-driven) models, can be improved using soft technologies because the underlying systems will be explained also based on data (CC data-driven models). Through the applications depicted here, it is sustained that cognitive tools are able to make abstractions and generalizations of the process and can play a complementary role to physically based models.

## References

- [1] A. P. S. Kumar, K. P. Sudheer, S. K. Jain, and P. K. Agarwal, "Rainfall-runoff modelling using artificial neural networks: comparison of network types," *Hydrological Processes*, vol. 19, no. 6, pp. 1277–1291, 2005.
- [2] J. C. Bezdek and R. J. Hathaway, "Numerical convergence and interpretation of the fuzzy c-shells clustering algorithm," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 787–793, 1992.
- [3] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [4] L. A. Zadeh, "The roles of fuzzy logic and soft computing in the conception, design and deployment of intelligent systems," *BT Technology Journal*, vol. 14, no. 4, pp. 32–36, 1994.
- [5] R. Aliev, K. Bonfig, and F. Aliew, *Soft Computing*, Technic, Berlin, Germany, 2000.
- [6] L. A. Zadeh, "Foreword," in *Proceedings of the 1st European Congress on Intelligent Techniques and Soft Computing (EUFIT '95)*, p. 7, 1995.
- [7] L. A. Zadeh, "Soft computing and fuzzy logic," *IEEE Software*, vol. 11, no. 6, pp. 48–56, 1994.
- [8] R. A. Aliev, "Fuzzy expert systems," in *Soft Computing: Fuzzy Logic, Neural Networks, and Distributed Artificial Intelligence*, F. Aminzadeh and M. Jamshidi, Eds., pp. 99–108, Prentice Hall, Englewood Cliffs, NJ, USA, 1994.
- [9] L. A. Zadeh, "Fuzzy logic, neural networks, and soft computing," *Communications of the ACM*, vol. 37, no. 3, pp. 77–84, 1994.
- [10] D. Nauck, F. Klawonn, and R. Kruse, *Foundations of Neuro-Fuzzy Systems*, John Wiley and Sons, New York, NY, USA, 1997.
- [11] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*, MIT Press, Cambridge, Mass, USA, 1995.
- [12] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Marmillau and IEEE Computer Society, 1994.
- [13] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, Mass, USA, 1989.
- [14] K. Murawski, T. Arciszewski, and K. de Jong, "Evolutionary computation in structural design," *Engineering with Computers*, vol. 16, no. 3–4, pp. 275–286, 2000.
- [15] G. P. Huijzer, *Quantitative penetrostratigraphic classification [Ph.D. thesis]*, Amsterdam Free University, 1992.
- [16] A. Coerts, *Analysis of the static cone penetration test data for subsurface modelling: a methodology [Ph.D. thesis]*, Utrecht University, The Netherlands, 1996.
- [17] C. H. Juang, T. Jiang, and R. A. Christopher, "Three-dimensional site characterisation: neural network approach," *Geotechnique*, vol. 51, no. 9, pp. 799–809, 2001.
- [18] S. R. García and M. P. Romo, "Sistema de información geográfica y redes neuronales," in *Proceedings of the Reunión Nacional de Mecánica de Suelos*, Guadalajara, Mexico, November 2004.
- [19] R. R. Youngs, S. J. Chiou, W. J. Silva, and J. R. Humphrey, "Strong ground motion attenuation relationships for subduction zone earthquakes," *Seismological Research Letters*, vol. 68, no. 1, pp. 58–73, 1997.
- [20] G. M. Atkinson and D. M. Boore, "Empirical ground-motion relations for subduction-zone earthquakes and their application to Cascadia and other regions," *Bulletin of the Seismological Society of America*, vol. 93, no. 4, pp. 1703–1729, 2003.
- [21] American Society of Civil Engineers (ASCE), *Prestandard and Commentary for the Seismic Rehabilitation of Buildings*, Federal Emergency Management Agency, Washington, DC, USA, 2000.
- [22] D. Gasparini and E. H. Vanmarcke, *SIMQKE: A Program for Artificial Motion Generation*, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Mass, USA, 1976.
- [23] J. G. Anderson, "Nonparametric description of peak acceleration above a subduction thrust," *Seismological Research Letters*, vol. 68, no. 1, pp. 86–93, 1997.
- [24] C. B. Crouse, "Ground motion attenuation equations for earthquakes on the Cascadia subduction zone," *Earthquake Spectra*, vol. 7, pp. 210–236, 1991.
- [25] S. K. Singh, M. Ordaz, M. Rodriguez et al., "Analysis of near-source strong-motion recordings along the Mexican subduction zone," *Bulletin of the Seismological Society of America*, vol. 79, no. 6, pp. 1697–1717, 1989.
- [26] C. B. Crouse, Y. K. Vyas, and B. A. Schell, "Ground motions from subduction-zone earthquakes," *Bulletin of the Seismological Society of America*, vol. 78, no. 1, pp. 1–25, 1988.
- [27] S. K. Singh, E. Mena, R. Castro, and C. Carmona, "Empirical prediction of ground motion in Mexico City from coastal earthquakes," *Bulletin of the Seismological Society of America*, vol. 77, pp. 1862–1867, 1987.
- [28] K. Sadigh, "Ground motion characteristics for earthquakes originating in subduction zones and in the western United States," in *Proceedings of the 6th Pan American Conference*, Lima, Peru, 1979.
- [29] E. M. Scordilis, "Empirical global relations converting MS and mb to moment magnitude," *Journal of Seismology*, vol. 10, no. 2, pp. 225–236, 2006.
- [30] B. W. Tichelaar and L. J. Ruff, "Depth of seismic coupling along subduction zones," *Journal of Geophysical Research*, vol. 98, no. 2, pp. 2017–2037, 1993.
- [31] G. M. Atkinson and D. M. Boore, "Empirical ground-motion relations for subduction-zone earthquakes and their application to Cascadia and other regions," *Bulletin of the Seismological Society of America*, vol. 93, no. 4, pp. 1703–1729, 2003.
- [32] S. E. Fahlman, "Fast learning variations on back-propagation: an empirical study," in *Proceedings of the Connectionist Models Summer School*, D. Touretzky, G. Hinton, and T. Sejnowski, Eds., pp. 38–51, Morgan Kaufmann, San Mateo, Calif, USA, 1988.
- [33] S. R. García, M. P. Romo, M. J. Mendoza, and V. M. Taboada-Urtuzuástegui, *Sand Behavior Modeling Using Static and Dynamic Artificial Neural Networks*, Series de Investigación y Desarrollo del Instituto de Ingeniería SID/631, 2002.
- [34] K. W. Campbell, "Predicting strong ground motions in Utah, given in the paper measurement, characterization, and prediction of strong ground motion," in *Earthquake Engineering and Soil Dynamics II-Recent Advances in Ground-Motion Evaluation*, W. B. Joyner and D. M. Boore, Eds., pp. 43–102, 1988.
- [35] F. Sabetta and A. Pugliese, "Attenuation of peak horizontal acceleration and velocity from Italian strongmotion records," *Bulletin of the Seismological Society of America*, vol. 77, pp. 1491–1513, 1987.

- [36] S. C. Gómez, M. Ordaz, and C. Tena, "Leyes de atenuación en desplazamiento y aceleración para el diseño sísmico de estructuras con aislamiento en la costa del Pacífico," in *Memorias del XV Congreso Nacional de Ingeniería Sísmica*, Morelia, Mexico, November 2005.
- [37] W. Silva and K. Lee, *State-of-the-Art for Assessing Earthquake Hazards in the United States, Report 24, ES RASCAL Code for Synthesizing Earthquake Ground Motions, Miscellaneous Paper S-73-1*, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Miss, USA, 1987.
- [38] B. A. Bolt and N. J. Gregor, "Synthesized strong ground motions for the seismic condition assessment of the eastern portion of the San Francisco Bay Bridge," Tech. Rep. UCB/EERC-93/12, University of California, Earthquake Engineering Research Center, Berkeley, Calif, USA, 1993.
- [39] J. E. Carballo and C. A. Cornell, "Probabilistic seismic demand analysis: spectrum matching and design," Tech. Rep. RMS-41, Department of Civil and Environmental Engineering, Stanford University, 2000.
- [40] C. Kircher, *Personal Communication with Farzad Naeim and Marshall Lew. Levy, S.*, 1992, Artificial Life, Vintage Books, New York, NY, USA, 1993.
- [41] F. Naeim and J. M. Kelly, *Design of Seismic Isolated Structures-from Theory to Practice*, John Wiley and Sons, New York, NY, USA, 1999.
- [42] N. E. Huang, S. Zheng, S. R. Long et al., "The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis," *Proceedings of the Royal Society A*, vol. 454, pp. 903–995, 1998.
- [43] S. Garcia, M. Romo, F. Correa, and M. Ortega, "Hilbert-Huang spectral analysis of seismic ground response," in *Proceedings of the 8th National Conference on Earthquake Engineering (NCEE '06)*, San Francisco, Calif, USA, 2006, paper no. 1647.
- [44] G. Castro, S. J. Poulos, J. W. France, and J. L. Enos, "Liquefaction induced by cyclic loading," Tech. Rep., Geotechnical Engineers Inc., National Science Foundation, Washington, DC, USA, 1982.
- [45] H. B. Seed, I. M. Idris, and I. Arango, "Evaluation of liquefaction potential using field performance data," *Journal of Geotechnical Engineering*, vol. 109, no. 3, pp. 458–482, 1983.
- [46] H. B. Seed and I. M. Idriss, "Simplified procedure for evaluating soil liquefaction potential," *Journal of the Geotechnical Engineering Division*, vol. 97, no. 9, pp. 1249–1273, 1971.
- [47] J. H. Hwang and C. W. Yang, "Verification of critical cyclic strength curve by Taiwan Chi-Chi earthquake data," *Soil Dynamics and Earthquake Engineering*, vol. 21, no. 3, pp. 237–257, 2001.
- [48] A. T. C. Goh, "Seismic liquefaction potential assessed by neural networks," *Journal of Geotechnical Engineering*, vol. 120, no. 9, pp. 1467–1480, 1994.
- [49] A. T. C. Goh, "Neural-network modeling of CPT seismic liquefaction data," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 122, no. 1, pp. 70–73, 1996.
- [50] A. T. C. Goh, "Probabilistic neural network for evaluating seismic liquefaction potential," *Canadian Geotechnical Journal*, vol. 39, pp. 219–232, 2002.
- [51] C. H. Juang and C. J. Chen, "CPT-based liquefaction evaluation using artificial neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 14, no. 3, pp. 221–229, 1999.
- [52] M. H. Baziar and N. Nilipour, "Evaluation of liquefaction potential using neural-networks and CPT results," *Soil Dynamics and Earthquake Engineering*, vol. 23, no. 7, pp. 631–636, 2003.
- [53] C. J. Lee and T. K. Hsiung, "Applying neural network model in seismic liquefaction case analysis: probabilistic neural network model v.s. multilayer perceptrons model," in *Proceedings of the 2nd Joint International Conference on Soft Computing and Intelligent Systems and 5th International Symposium on Advanced Intelligent Systems*, Yokohama, Japan, 2004.
- [54] C. H. Juang, C. J. Chen, T. Jiang, and R. D. Andrue, "Risk-based liquefaction potential evaluation using standard penetration tests," *Canadian Geotechnical Journal*, vol. 37, pp. 1195–1208, 2000.
- [55] E. Hunt, *Concept Learning: An Information Processing Problem*, Wiley, New York, NY, USA, 1962.
- [56] P. Winston, "Learning structural descriptions from examples," in *The Psychology of Computer Vision*, P. H. Winston, Ed., McGraw-Hill, 1975.
- [57] E. Feigenbaum and H. Simon, "Performance of a reading task by an elementary perceiving and memorizing program," *Behavioral Science*, vol. 8, no. 1, pp. 72–76, 1963.
- [58] B. G. Buchanan and T. M. Mitchell, "Model-directed learning of production rules," in *Pattern Directed Inference Systems*, D. A. Waterman and F. Hayes-Roth, Eds., Academic Press, 1978.
- [59] C. H. Juang and T. Jiang, "Assessing probabilistic methods for liquefaction potential evaluation," in *Soil Dynamics and Liquefaction 2000*, R. Y. S. Pak and J. Yamamura, Eds., Geotechnical Special Publication no. 107, pp. 148–162, American Society of Civil Engineers, 2000.
- [60] R. D. Andrus, K. H. Stokoe II, and R. M. Chung, *Draft Guidelines for Evaluating Liquefaction Resistance Using Shear Wave Velocity Measurements and Simplified Procedures*, NISTIR, 6277, National Institute of Standards and Technology, Gaithersburg, Md, USA, 1999.
- [61] H. B. Seed and I. M. Idriss, "Simplified procedure for evaluating soil liquefaction potential," *Journal of the Soil Mechanics and Foundations Division*, vol. 97, no. 9, pp. 1249–1273, 1971.
- [62] T. L. Youd, I. M. Idriss, R. D. Andrus et al., "Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 127, no. 10, pp. 817–833, 2001.
- [63] P. K. Robertson and C. E. Wride, "Evaluating cyclic liquefaction potential using the cone penetration test," *Canadian Geotechnical Journal*, vol. 35, no. 3, pp. 442–459, 1998.
- [64] T. L. Youd, "Geologic effects-liquefaction and associated ground failure," in *Proceedings of the Geologic and Hydraulic Hazards Training Program*, Open File Report 84-760, pp. 210–232, US Geological Survey, Menlo Park, Calif, USA, 1984.
- [65] K. Tokida, H. Iwasaki, H. Matsumoto, and T. Hamada, "Liquefaction potential and drag force acting on piles in flowing soils," *Soil Dynamics and Earthquake Engineering*, vol. 1, pp. 244–259, 1993.
- [66] W. Vargas and I. Towhata, "Measurement of drag exerted by liquefied sand on buried pipe," in *Proceedings of the 1st International Conference on Earthquake Geotechnical Engineering*, pp. 975–980, Tokyo, Japan, 1995.
- [67] S. K. Haigh and S. P. G. Madabhushi, "The effects of pile flexibility on pile-loading in laterally spreading slopes," in *Seismic Performance and Simulation of Pile Foundations in Liquefied and Laterally Spreading Ground*, Geotechnical Special Publication no. 145, pp. 24–37, ASCE and GEO-Institute, 2005.
- [68] S. F. Bartlett and T. L. Youd, "Empirical prediction of liquefaction-induced lateral spread," *Journal of Geotechnical Engineering*, vol. 121, no. 4, pp. 316–329, 1995.

- [69] J. B. Berrill, S. A. Christensen, R. J. Kenan, W. Okada, and J. R. Pettiga, "Lateral spreading loads on pile bridge foundation," in *Proceedings on Soil Behavior of Ground and Geotechnical Structures*, S. E. Pinto, Ed., vol. 2, pp. 173–187, Balkema, Rotterdam, The Netherlands, 1997.
- [70] E. Ovando, L. Vieitez, and J. D. Alemán, "Geotecnia, El Macrosismo de Manzanillo del 9 de octubre de 1995," in *Sociedad Mexicana de Ingeniería Sísmica*, A. Tena, Ed., pp. 82–133, Universidad de Colima, Colima, Mexico, 1997.
- [71] E. Ovando and M. P. Romo, "Three recent damaging earthquakes in Mexico, special lecture," in *Proceedings of the 5th International Conference on Case Histories in Geotechnical Engineering*, New York, NY, USA, April 2004.
- [72] T. L. Youd, C. M. Hansen, and S. F. Bartlett, "Revised multilinear regression equations for prediction of lateral spread displacement," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 128, no. 12, pp. 1007–1017, 2002.
- [73] S. R. García and M. P. Romo, "Analytical and neurogenetic methods to estimate in situ rockfill dam dynamic properties," in *Estado del Arte, Ingeniería Geosísmica, XXIV Reunión Nacional de Mecánica de Suelos*, pp. 45–55, Aguascalientes, Mexico, November 2008.
- [74] M. P. Romo and S. R. Garcia, *Prediction of Liquefaction-Induced Lateral Spread: A Neurofuzzy Procedure*, Serie Investigación y Desarrollo, Instituto de Ingeniería, UNAM, SID/651, Agosto, 2006, <http://www.iingen.unam.mx/>.



