27 May 2010, 7:30 pm - 9:00 pm

A Practical Approach for Implementing the Probability of Liquefaction in Performance Based Design

Thomas Oommen
Tufts University, Medford, MA

Laurie G. Baise
Tufts University, Medford, MA

Fifth International Conference on
Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics
and Symposium in Honor of Professor I.M. Idriss
May 24-29, 2010 • San Diego, California

A PRACTICAL APPROACH FOR IMPLEMENTING THE PROBABILITY OF LIQUEFACTION IN PERFORMANCE BASED DESIGN

Thomas Oommen,
Ph.D. Candidate,
Department of Civil and Environmental Engineering,
Tufts University, 113 Anderson Hall,
Medford, MA 02155, USA.
thomas.oommen@tufts.edu,
Tel: 215-435-0867.

Laurie G. Baise,
Associate Professor,
Department of Civil and Environmental Engineering,
Tufts University, 113 Anderson Hall,
Medford, MA 02155, USA.
laurie.baise@tufts.edu,
Tel: 617-627-2211.

ABSTRACT

Empirical Liquefaction Models (ELMs) are the usual approach for predicting the occurrence of soil liquefaction. These ELMs are typically based on in situ index tests, such as the Standard Penetration Test (SPT) and Cone Penetration Test (CPT), and are broadly classified as deterministic and probabilistic models. The deterministic model provides a “yes/no” response to the question of whether or not a site will liquefy. However, Performance-Based Earthquake Engineering (PBEE) requires an estimate of the probability of liquefaction (P_L) which is a quantitative and continuous measure of the severity of liquefaction. Probabilistic models are better suited for PBEE but are still not consistently used in routine engineering applications. This is primarily due to the limited guidance regarding which model to use, and the difficulty in interpreting the resulting probabilities. The practical implementation of a probabilistic model requires a threshold of liquefaction (TH_L). The researchers who have used probabilistic methods have either come up with subjective TH_L or have used the established deterministic curves to develop the TH_L. In this study, we compare the predictive performance of the various deterministic and probabilistic ELMs within a quantitative validation framework. We incorporate estimated costs associated with risk as well as with risk mitigation to interpret P_L using precision and recall and to, compute the optimal TH_L using Precision-Recall (P-R) cost curve. We also provide the P-R cost curves for the popular probabilistic model developed using Bayesian updating for SPT and CPT data by Cetin et al. (2004) and Moss et al. (2006) respectively. These curves should be immediately useful to a geotechnical engineer who needs to choose the optimal TH_L that incorporates the costs associated with the risk of liquefaction and the costs associated with mitigation.

INTRODUCTION

Soil liquefaction is the loss of shear strength induced by shaking, which can lead to various types of ground failures. Empirical Liquefaction Models (ELMs) have been developed for in situ index tests, such as Standard Penetration Test (SPT), Cone Penetration Test (CPT), and Shear Wave Velocity (Vs). These in situ data are used to estimate the potential for “triggering” or initiation of seismically induced liquefaction. Different classes of ELMs include: (1) deterministic (Seed and Idriss 1971; Seed et al. 1983; Robertson and Campanella 1985; Seed and De Alba 1986; Shibata and Teparaksa 1988; Goh 1994; Stark and Olson 1995; Robertson and Wride 1998; Juang et al. 2000; Juang et al. 2003; Idriss and Boulanger 2006; Pal 2006; Hanna et al. 2007; Goh and Goh 2007) and (2) probabilistic (Liao et al. 1988; Toprak et al. 1999; Juang et al. 2002; Goh 2002; Cetin et al. 2002; Lai et al. 2004; Cetin et al. 2004; Moss et al. 2006).

Currently, the most widely used ELM for the assessment of liquefaction potential is the “simplified procedure,” recommended by Seed and Idriss (1971). Youd et al. (2001), Cetin et al. (2004), and Moss et al. (2006), provide recent updates to the method. In addition, Cetin et al. (2004) and Moss et al. (2006) have presented liquefaction models that use the Bayesian updating method for SPT and CPT data respectively. The recent work represents an update to the datasets combined with the use of the Bayesian updating method for probabilistic evaluation of liquefaction potential. Although there are several deterministic and probabilistic models to evaluate the liquefaction potential using SPT and

CPT data, most of these approaches neither provide a quantitative evaluation of the predictive performance nor critically compare with other approaches used in practice. As a result, the research community has not provided practitioners with objective quantifiable recommendations on which ELM to use for the evaluation of liquefaction potential.

The deterministic method provides a "yes/no" response to the question of whether or not a site will liquefy. However, Performance-Based Earthquake Engineering (PBEE) requires an estimate of the probability of liquefaction (P_L) rather than a deterministic (yes/no) estimate (Juang et al. 2008). P_L is a quantitative and continuous measure of the severity of liquefaction. Probabilistic methods were first introduced to liquefaction modeling in the late 1980's by Liao et al. (1988). But such methods are still not consistently used in routine engineering applications. This is primarily due to the limited guidance regarding which model to use, and the difficulty in interpreting the resulting probabilities. The implementation of probabilistic methods requires a threshold of liquefaction (TH_L). The need for a TH_L arises because engineering decisions require the site to be classified as either liquefiable or non-liquefiable. Thus, a site where $P_L < TH_L$ is classified as non-liquefiable and a site where $P_L > TH_L$ is classified as liquefiable. Juang et al. (2002) provided a subjective TH_L and Cetin et al. (2004) and Moss et al. (2006) used deterministic curves to determine TH_L. However, the importance of the probabilistic approach warrants objective guidelines for the determination of TH_L.

The primary goal of this study is to provide a critical, objective, and quantitative comparison of the predictive performance of the "simplified procedure" as presented by Youd et al. (2001) and the Bayesian updating method (Cetin et al. 2004; and Moss et al. 2006). We also provide a thorough and reproducible approach to interpret P_L using precision and recall and to, compute the optimal TH_L that incorporates the costs associated with the risk of liquefaction and the costs associated with mitigation using a new metric that we developed called the Precision-Recall (P-R) cost curve. In the first section of the paper, we describe the data used for comparing the different ELMs. Then we introduce the model validation statistics, the different ELMs that we consider in this paper, and the objective method of identifying TH_L.

DATA

In this study, we use the SPT and CPT data compiled by Cetin et al. (2004) and Moss et al. (2006). These databases were created in three steps: (1) re-evaluation of the Seed et al. (1983) data to incorporate the new field case studies; (2) screen data to remove questionable observations; and (3) account for recent advances in SPT and CPT interpretation and evaluation of in situ Cyclic Stress Ratio (CSR).

The SPT database has 196 field case histories of which 109 are from liquefied sites and 87 are from non-liquefied sites. The CPT database has 182 case histories of which 139 are from liquefied sites and 43 are from non-liquefied sites. The ratio of liquefaction to non-liquefaction instances in the SPT database is 56:44, whereas, in the CPT database it is 76:24. Thus, the CPT database has higher class imbalance than the SPT database. The class imbalance is defined as the difference in the number of instances of occurrences of two different classes. Class imbalance is particularly important for comparing the performance of different models. Class imbalance issues for model validation are discussed later in this paper.

METHODOLOGY

We calculated the liquefaction potential for the SPT and CPT databases by using the "simplified procedure" (Youd et al. 2001), and the Bayesian updating method (Cetin et al. 2004; Moss et al. 2006). The following subsections provide a brief description of the fundamental principles of these approaches/classifiers and the equations used. We validate and quantify the different deterministic classifiers by using overall accuracy, precision, recall (i.e., True Positive Rate (TPR)), and F-score. And for the probabilistic classifiers, we use Receiver Operating Characteristic (ROC) curves, and Precision-Recall (P-R) curves. Then, we present a new objective method for combining the precision and recall with cost curves to determine the optimal TH_L triggering for probabilistic assessment of liquefaction potential.

Model Validation

Model development (i.e., model "training") should be followed by a model validation to assess predictive capability. The models that we consider in this paper were trained on the complete datasets so we have to validate these classifiers on the same dataset used for model development. As a result, the validation statistics for these methods will likely overestimate the prediction accuracy of the models. In an ideal situation, we would have a both a training and testing dataset.

For deterministic models, useful validation statistics include: overall accuracy, precision, recall, and F-score. These metrics are all computed from elements of the confusion matrix. A confusion matrix is a table used to evaluate the performance of a classifier. It is a matrix of the observed versus the predicted classes, with the observed classes in columns and the predicted classes in rows as shown in Table 1. The diagonal elements (where the row index equals the column index) include the frequencies of correctly classified instances and non-diagonal elements include the frequencies of misclassifications.

The overall accuracy is a measure of the percentage of correctly classified instances

$$\text{Overall Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where the True Positive (*TP)* is the sum of instances of liquefaction correctly predicted, the True Negative *(TN)* is the sum of instances of non-liquefaction correctly predicted, the False Positive (*FP)* is the sum of instances of non-liquefaction classified as liquefaction, and the False Negative (*FN)* is the sum of instances of liquefaction classified as non-liquefaction. Overall accuracy is a common validation statistic that is used and an accuracy of 0.75 means that 75% of the data have been correctly classified. However, it doesn't mean that the 75% of each class (e.g., liquefaction and non-liquefaction class) has been correctly predicted. Therefore, the evaluation of the predictive capability based on the overall accuracy alone can be misleading when class imbalance exists (e.g., for the CPT dataset 76% of the data are liquefaction instances and 24% are non-liquefaction instances).

Table 1: Confusion matrix, presenting the observed classes in rows and the predicted classes in columns where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative.

		Observed	
		Yes	No
Predicted	Yes	TP	FP
	No	FN	TN

Precision and recall are common metrics applied separately to each class in the dataset. This is particularly valuable when the class imbalance in the dataset is significant. Precision measures the accuracy of the predictions for a single class, whereas recall measures accuracy of predictions only considering predicted values.

$$\text{Precision} = P = TP/(TP + FP), \tag{2}$$

$$\text{Recall} = R = TP/(TP + FN). \tag{3}$$

In the context of liquefaction potential assessment, a precision of 1.0 for the liquefaction class means that every case that is predicted as liquefaction experienced liquefaction, but this does not account for instances of observed/actual liquefaction that are misclassified. Analogously, a recall of 1.0 means that every instance of observed liquefaction is predicted correctly by the model, but this does not account for instances of observed non-liquefaction that are misclassified. An inverse relationship exists between precision and recall: it is possible to increase one at the expense of the other.

The F-score is a measure that combines the precision and recall value to a single evaluation metric. The F-score is the weighted harmonic mean of the precision and recall

$$F_\beta = (1 + \beta^2)(P \cdot R)/(\beta^2 \cdot P + R). \tag{4}$$

where β is a measure of the importance of recall to precision and can be defined by the user for a specific project.

In order to evaluate a probabilistic classifier, we must choose a probability threshold value that marks the liquefaction/non-liquefaction boundary to apply deterministic metrics such as given in equations 1 through 4. When a probability threshold is defined, the subsequent validation is specific to that threshold value. Therefore, for the comprehensive evaluation of a probabilistic classifier we use P-R and ROC curves. P-R and ROC curves provide a measure of the classification performance for the complete spectrum of probability thresholds (i.e., "operating conditions"). The P-R and ROC curves are developed by calculating the precision, the recall, and the False Positive Rate (FPR) by varying the threshold from 0 to 1. The FPR is

$$FPR = FP/(FP + TN). \tag{5}$$

Thus, any point on either the P-R or ROC curve corresponds to a specific threshold. Fig. 1 presents a basic ROC curve, where the dashed line is the idealized best possible ROC curve. The area under the ROC curve (AUC) is a scalar measure that quantifies the accuracy of the probabilistic classifier. The AUC varies from 1.0 (perfect accuracy) to 0. Randomly selecting a class produces the diagonal line connecting (0, 0) and (1, 1) (shown as dotted diagonal line Fig. 1). This gives AUC=0.5, thus it is unrealistic for a classifier to have an AUC less than 0.5.

Fig. 2 presents a basic P-R curve. The dashed line represents the best P-R curve with point A marking the best performance. Unlike ROC curves, P-R curves are sensitive to the influence of sampling bias in a dataset. Sampling bias is the misrepresentation of a class in the samples compared to the actual ratio of occurrences in the population. Often class imbalance and sampling bias are misrepresented and it is important to understand that they represent two distinct issues. Example, if the true population of the data has a class ratio of 80:20 and a sample has a class ratio of 50:50, then the sample has no class imbalance but it has a sampling bias because the proportion of the classes in the sample is different from the original population. Oommen et al. (2009a) have demonstrated that sampling bias can significantly influence model development and performance.

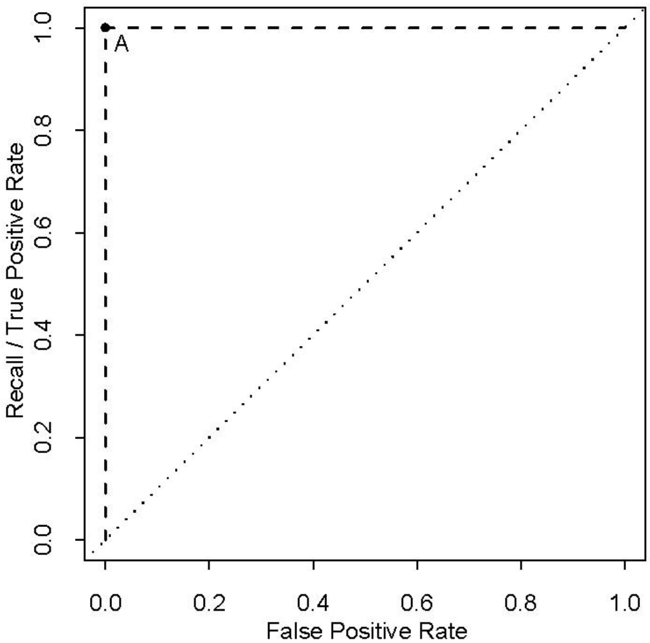

Fig. 1: Receiver Operating Characteristic (ROC) curve illustrating its basic elements. The dashed line indicates a near perfect probability prediction whereas, the dotted line indicates predictions which result from random guessing.

Simplified Procedure (Youd et al. 2001)

Following the disastrous earthquakes in Alaska and in Nigata, Japan in 1969, Seed and Idriss (1971) developed the "simplified procedure" which uses empirical evaluations of field observations for estimating liquefaction potential. A series of publications revised the procedure (Seed and Idriss 1971; Seed et al. 1983; Youd et al. 2001). Youd et al. (2001) state that the periodic modifications have improved the "simplified procedure", however, these improvements are not quantified and hence remain unknown for practicing engineers.

The evaluation of liquefaction potential using the "simplified procedure" requires estimation of two variables: (1) the seismic demand on a soil layer, expressed in terms of the cyclic stress ratio (CSR); and (2) the capacity of the soil to resist liquefaction expressed in terms of the cyclic resistance ratio (CRR). The latter variable depends upon the type of in situ measurement (*i.e.* SPT or CPT). CSR is

$$CSR = 0.65(a_{max}/g)(\sigma_{vo}/\sigma'_{vo})r_d, \quad (6)$$

where a_{max} = peak horizontal acceleration at the ground surface generated by the earthquake; g = acceleration of gravity; σ_{vo} and σ'_{vo} are total and effective vertical overburden stresses, respectively; and r_d = stress reduction coefficient

$$r_d = \begin{cases} 1.0 - 0.00765z & \text{for } z \leq 9.15m \\ 1.174 - 0.0267z & \text{for } 9.15m < z \leq 23m, \end{cases} \quad (7)$$

where z = depth beneath ground surface in meters.

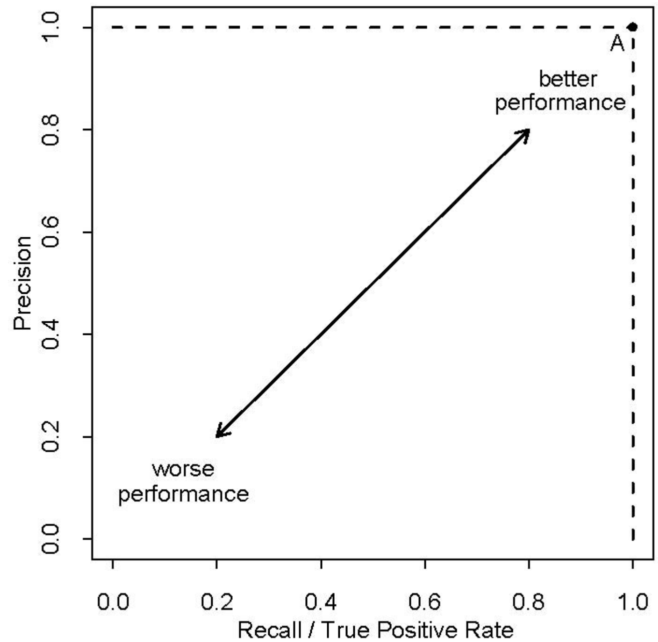

Fig. 2: Precision-Recall (P-R) curve illustrating its basic elements. The dashed line represents the best P-R curve.

The CRR for fines contents <0.05 is the basic penetration criterion for the "simplified procedure" and is referred to as the clean sand base curve, calculated for a magnitude of 7.5

$$CRR_{7.5}^{SPT} = \frac{1}{34-(N_1)_{60}} + \frac{(N_1)_{60}}{135} + \frac{50}{[10\cdot(N_1)_{60}+45]^2} - \quad (8)$$

$$CRR_{7.5}^{CPT} = \begin{cases} 0.833[(q_{c1N})_{cs}/1000] + 0.05 & \text{for } (q_{c1N}) \\ 93[(q_{c1N})_{cs}/1000]^3 + 0.08 & \text{for } 50 \leq (q_{c1N}) \end{cases} \quad (9)$$

where $CRR_{7.5}^{SPT}$ = CRR for SPT, $CRR_{7.5}^{CPT}$ =CRR for CPT, $(N_1)_{60}$ = corrected SPT blow count and is <30, $(q_{c1N})_{cs}$ = clean sand cone penetration resistance normalized to approximately 100 kPa. Finally, liquefaction hazard is estimated in terms of Factor of Safety (FS) against liquefaction by scaling the CRR to the appropriate magnitude and is given as

$$FS = (CRR_{7.5}/CSR)MSF, \quad (10)$$

where MSF = magnitude scaling factor.

Bayesian Updating Method

Cetin et al. (2004) and Moss et al. (2006) formulated the Bayesian updating method for the probabilistic evaluation of liquefaction potential using SPT and CPT data, respectively. The development of a limit state model for the initiation of soil liquefaction using the Bayesian approach begins with the selection of a mathematical model. The general form of the limit state function is g = g(x, θ) + ε, where x = the set of predictive variables; θ = the set of unknown model parameters and ε = the random model correction term to account for the influences of the missing variables and possible incorrect model forms. The limit state function assumes that the liquefaction potential is completely explained by the set of predictive variables and the model corrections ε are normally distributed with zero mean and standard deviation of σ_ε.

The limit state function together with the field case histories are used to develop the likelihood function. If the i[th] term in the field case history is a liquefaction case $g(x_i, \theta) + \varepsilon_i \le 0$ and conversely if the i[th] term in the field case history is a non-liquefaction case $g(x_i, \theta) + \varepsilon_i > 0$. Thus, the likelihood function can be expressed as

$$L(\theta, \sigma_\varepsilon) = \prod P[g(x_i, \theta) + \varepsilon_i \le 0]^{W_{liq}} \cdot \prod P[g(x_i, \theta) \cdot \quad (11)$$

where W_{liq} and W_{nonliq} is a correction term to account for the class imbalance in the field case history database due to the disproportionate number of liquefied *vs.* non-liquefied field instances. In order to determine the unknown model parameters θ, the multifold integrals over the Bayesian kernel evaluate the likelihood function and the prior distributions of the model parameters.

The Bayesian updating method formulation to calculate the P_L using SPT data is presented in Eq. 19 of Cetin et al. (2004). For a deterministic assessment, Cetin et al. (2004) recommend using a P_L value of >0.15 as liquefiable otherwise all remaining as non-liquefiable. For the CPT data, the Bayesian formulation for the P_L is presented in Eq. 20 of Moss et al. (2006). For the deterministic analysis Moss et al. (2006) provide similar recommendations for the probability values as in Cetin et al. (2004).

Thresholds for Liquefaction Triggering

In this section we present a new approach by combining project cost information with the precision and recall (P-R curve) to determine the optimal TH_L triggering. Here we assume that for a given project, the expected misclassification cost for the FP (C_{FP}) and the cost for the FN (C_{FN}) are known. The P-R cost curve is a tool that practicing engineers can use to find the optimal TH_L triggering for a given project and to determine the uncertainty associated with that decision. Figure 3 presents a typical P-R cost curve, which consists of two plots. Figure 3a illustrates the choice of the threshold vs. precision and recall. For a given probabilistic approach, Fig. 3a is developed by varying the threshold from 0 to 1 and calculating the corresponding precision and recall values for each of these thresholds. Figure 3b presents the optimal TH_L vs. the ratio of the C_{FP} (C_{FP} = cost of predicting a true non-liquefaction instance as liquefaction) to the C_{FN} (C_{FN} = cost of predicting a true liquefaction instance as non-liquefaction) abbreviated as CR. The optimal TH_L is approximated by minimizing the cost

$$\mathrm{Optimal}[TH_L]_j = \min[FP_i \cdot CR_j + FN_i] \quad (18)$$

where i= entire range of threshold from 0 to 1, FP_i and FN_i are number of false positive and false negative values corresponding to i, $CR_j = (C_{FP})_j / C_{FN}$ assuming that $C_{FN} = 1$, and the index j takes on the range of the values of *CR* under consideration. We used a range of *CR* from 0 to 1.2 (i.e. C_{FP} = 0 to C_{FP} = 1.2 X C_{FN}*)*. In practice, the C_{FP} and C_{FN} can be computed based on the Performance Based Earthquake Engineering (PBEE) recommended decision variables such as dollar losses, downtime and deaths sometimes referred to as the three D's (Krawinkler, 2004).

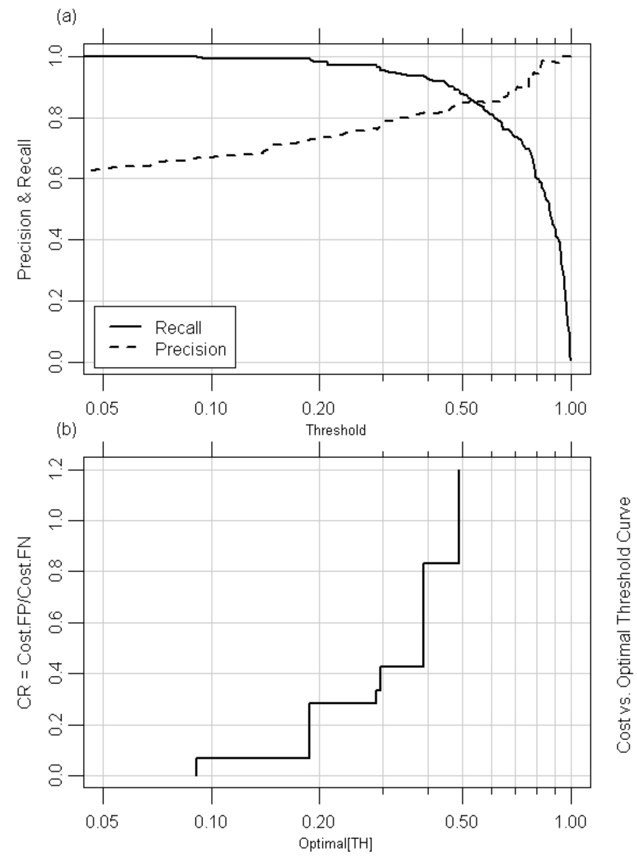

Fig. 3: P-R cost curve used to determine the optimal threshold of liquefaction (TH_L) triggering for probabilistic evaluation (a) precision and recall vs. threshold (b) cost ratio vs. optimal TH_L.

RESULTS AND DISCUSSION

Performance of Deterministic Approaches

Using the validation statistics described above, we evaluated the predictive performance of the deterministic approaches for the assessment of liquefaction potential based on the SPT and CPT data. For the deterministic case, Cetin et al. (2004) and Moss et al. (2006) use assigned TH_L values (0.15) in their probabilistic analysis. Table 2 presents the comparison of the SPT based Youd et al. (2001), and the Cetin et al. (2004) approaches. Comparing the overall accuracy for both the approaches, it is evident that the Cetin et al. 2004 has higher overall accuracy. Since the SPT database has a class imbalance of 56:44 (liquefaction: non-liquefaction), the overall accuracy alone cannot be used as an indicator of the predictive performance of the approaches. Therefore, liquefaction and non-liquefaction classes are analyzed separately using recall, precision and F-score.

Table 2: Various estimates of the predictive performance of the SPT based deterministic models: (1) overall accuracy (O.A), and (2) recall, precision, and F-score for both liquefaction and non-liquefaction occurrences.

Cetin et al. 2004 Dataset

Approach		Youd et al. 2001	Cetin et al. 2004
Overall Accuracy		0.826	0.831
Liquefaction	Recall	0.816	0.789
	Precision	0.864	0.895
	F-score	0.839	0.839
Non-liquefaction	Recall	0.839	0.885
	Precision	0.784	0.77
	F-score	0.811	0.823

In the case of the liquefaction class, we see that the Youd et al. (2001) model has the highest recall whereas, the Cetin et al. (2004) model has the highest precision. However, when we compute the F-score, which is the harmonic mean of precision and recall using equal weights for both, we see that both Cetin et al. (2004) and Youd et al. (2001) have similar F-score values with the latter being slightly higher.

In the case of the non-liquefaction class, we observe that the Cetin et al. (2004) model has the highest recall, whereas the Youd et al. (2001) has the highest precision. In addition, a comparison of the F-scores indicates that the Cetin et al. (2004) and Youd et al. (2001) have comparable F-score values for the non-liquefaction case with the former having slightly better performance.

From Table 2, we observe using the F-score measure that both Youd et al. (2001) and Cetin et al. (2004) approaches have similar predictive performance for the liquefaction and non-liquefaction instances. However, it is important to note that although the Cetin et al. (2004) approach has slightly improved predictive capability compared to the Youd et al. (2001) in the non-liquefaction case, it has a lower predictive performance in the liquefaction case.

Table 3 presents the comparison of the CPT based approaches from Youd et al. (2001), and Moss et al. (2006). Comparing the overall accuracy for both the approaches, we see that the Moss et al. (2006) has higher overall accuracy than Youd et al. (2001). However, the CPT database has greater class imbalance (76:24, liquefaction: non-liquefaction) than the SPT database. Hence again, the overall accuracy alone cannot be used as an indicator to compare the predictive performance.

Table 3: Various estimates of the predictive performance of the CPT based deterministic models: (1) overall accuracy (O.A), and (2) recall, precision, and F-score for both liquefaction and non-liquefaction occurrences.

Moss et al. 2006 Dataset

Approach		Youd et al. 2001	Moss et al. 2006
Overall Accuracy		0.846	0.879
Liquefaction	Recall	0.877	0.985
	Precision	0.917	0.872
	F-score	0.897	0.925
Non-liquefaction	Recall	0.744	0.534
	Precision	0.653	0.92
	F-score	0.695	0.676

Analyzing the predictive performance based on the individual classes (liquefaction and non-liquefaction) using precision, recall and F-score, we observe that for the liquefaction class, the Moss et al. (2006) approach has the highest recall whereas the Youd et al. (2001) approach has the highest precision. A comparison of the F-score measures shows that Moss et al. (2006) has improved predictive performance for the liquefaction class over Youd et al. 2001.

In the case of non-liquefaction instances, Youd et al. (2001) has the highest recall and Moss et al. (2006) has the best precision. Comparing both the approaches for non-liquefaction instances using F-score it is evident Youd et al. (2001) has an improved predictive capability than Moss et al. (2006).

It is noted from Tables 2 and 3 that the difference between the precision and recall values are higher for the CPT data compared to the SPT. Oommen et al. (2009a) has demonstrated that such a large difference in the precision and recall values indicates that the dataset has high sampling bias and the predicted probabilities have large deviations from the actual probabilities.

Performance of Probabilistic Approaches

We analyzed the predictive performance of the probabilistic evaluation of liquefaction potential using ROC and P-R curves. Figures 4 and 5 present the evaluation of the SPT and CPT based probabilistic approaches using ROC and P-R curves, respectively. We observe both the Cetin et al. (2004) and the Moss et al. (2006) approaches as having similar predictive performance with the latter having slightly improved AUC for both liquefaction and non-liquefaction instances. Figure 5 shows the P-R curve for the liquefaction case as falling closer to the (1, 1) point than for the non-liquefaction case. This indicates that both probabilistic approaches have better predictive capability for the liquefaction instances compared to the non-liquefaction instances. The difference in the predictive performance between liquefaction and non-liquefaction has increased for Moss et al. (2006) approach compared to the Cetin et al. (2004). This difference in the predictive performance is indicative of the sampling bias in the SPT and CPT dataset. As the sampling bias is increased from the SPT to CPT dataset the predictive performance of the minority class is decreased. This clearly indicates that model development using the Bayesian updating method (Cetin et al. 2004; Moss et al. 2006) is sensitive to the sampling bias in the dataset.

Comparing the probabilistic approaches based on the SPT and CPT datasets, we conclude that considering both liquefaction and non-liquefaction instances the SPT based probabilistic approaches have a slight advantage over the CPT based probabilistic approaches.

Choice of the Optimal Threshold of Liquefaction

In this section we use the P-R cost curves to determine the optimal TH_L. Figures 6 and 7 present the P-R cost curves for the SPT and CPT based datasets. In Figs. 6 and 7, plot a presents the optimal TH_L vs. the ratio of the C_{FP} to the C_{FN} for a given project (CR) and plot b represents the precision and recall for the liquefaction case using the “Bayesian updating” probabilistic approach. For the deterministic evaluation, the recommended TH_L using “Bayesian updating” is 0.15 for both

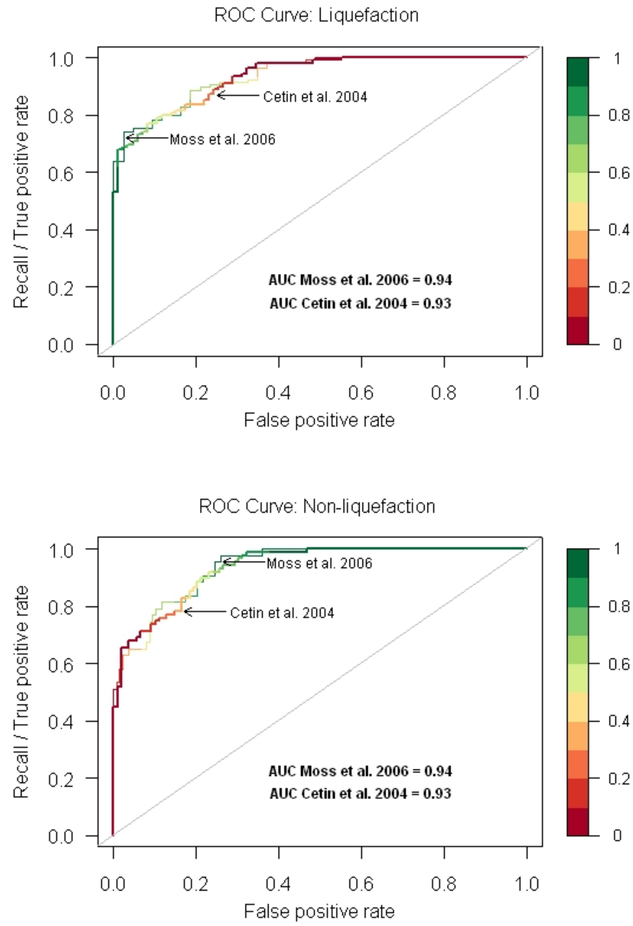

Fig. 4: ROC curve for the Moss et al. (2006) and Cetin et al. (2004) probabilistic approaches based on the SPT dataset.

the SPT and CPT datasets (Cetin et al. 2004; Moss et al. 2006). In the case of SPT (Fig. 6), a TH_L of 0.15 corresponds to a $CR \approx 1$ using the Cetin et al. (2004) approach, which implies that the C_{FN} = C_{FP} (cost of predicting a true liquefaction instance as non-liquefaction = cost of predicting a true non-liquefaction instance as liquefaction). Whereas, in the case of CPT (Fig. 7), a TH_L of 0.15 corresponds to a $CR \approx 0.6$ using the Moss et al. (2006) approach, which implies that the C_{FN} = 0.6 times the C_{FP}. We also observe from Fig. 6 that using any TH_L value in the range of 0.05 to 0.60 will have same cost as using the 0.15 recommended by Cetin et al. (2004).

Case Study on the Applicability of P-R Cost Curve

In the case of new projects/buildings, the geotechnical engineer must present the level of liquefaction risk, so that the owner/investor can decide whether or not to make the investment, or to increase the level of investment to improve its seismic performance and thus decrease the level of potential losses.

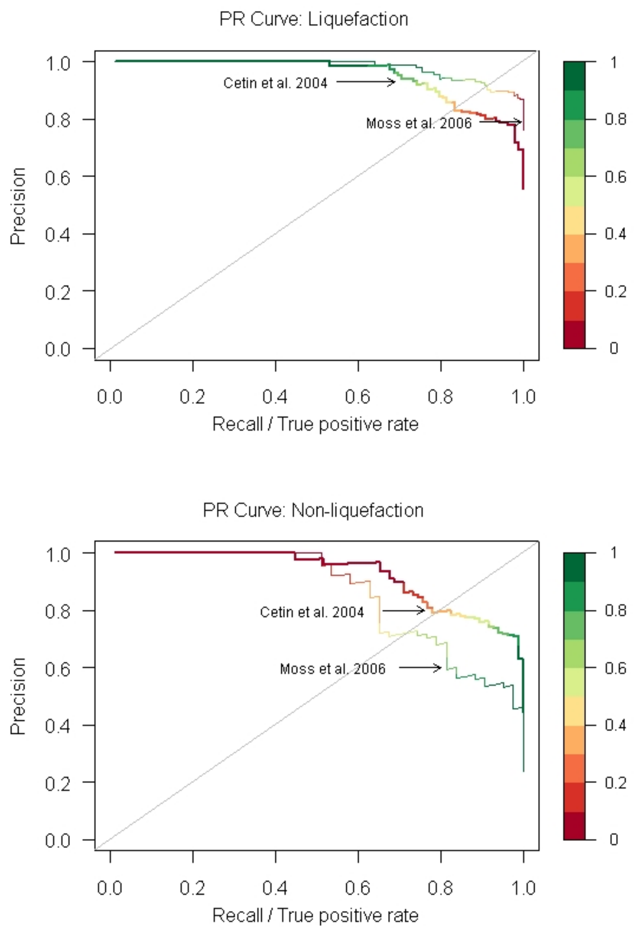

Fig. 5: P-R curve for the Moss et al. (2006) and Cetin et al. (2004) probabilistic approaches based on the SPT dataset.

Considering five hypothetical cases (H-1, H-2, H-3, H-4, and H-5), we illustrate how the P-R cost curve can be used by a geotechnical engineer in practice for determining the optimal TH_L for probabilistic assessment, and thereby quantitatively account for the costs associated with that decision. For the above hypothetical cases we calculated the CR (the ratio of the C_{FP} to the C_{FN}). The C_{FP} is equivalent to the cost of making the mistake of classifying a site that wouldn't liquefy as liquefiable. This includes the extra cost that is incurred on the project for site remediation, design, and construction. The C_{FN} is equivalent to the cost of making the mistake of classifying a site that would liquefy as nonliquefiable. This includes the cost of the building, the cost of lives and the cost of downtime, which includes the time, cost, and the business that was lost during the time to fix the building in the event of liquefaction. In the case of H-1, H-2, H-3, H-4, and H-5 we assume that the C_{FP} = \$35 million, whereas the C_{FN} = \$50 million. Thus the resulting CR is equal to

$$CR = 35/50 = 0.7$$

We also assume that the P_L for the cases H-1, H-2, H-3, H-4, and H-5 are 0.20, 0.25, 0.40, 0.60, and 0.30 respectively, calculated using the Bayesian updating method (Moss et al. 2006) with CPT data. From Fig. 7a or Table 5 we observe that the optimal threshold for CR = 0.7 using Bayesian updating method (Moss et al. 2006) with CPT data is 0.308, which means a P_L value > 0.308 should be classified as liquefiable.

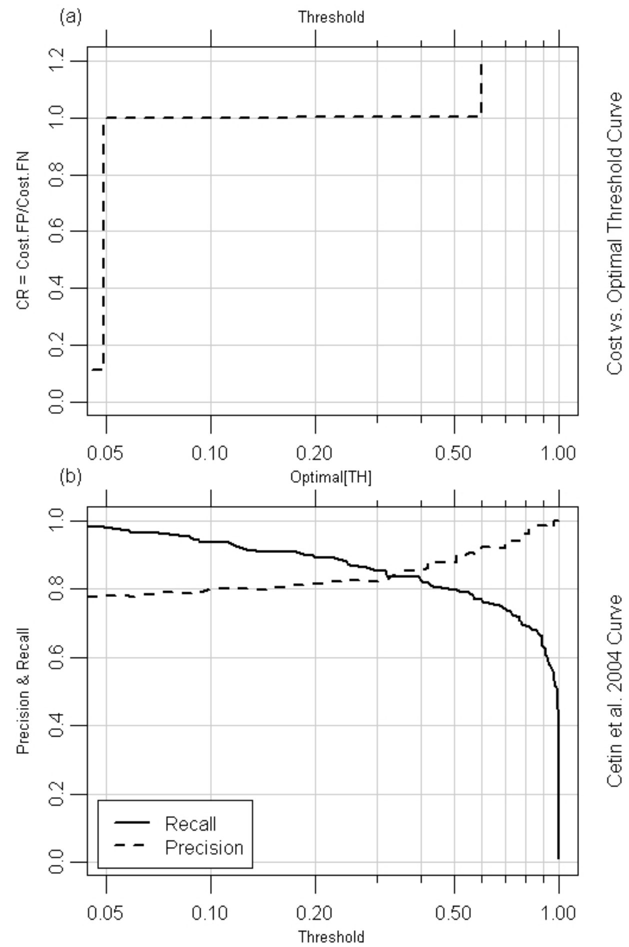

Fig. 6: P-R cost curve for the Cetin et al. (2004) probabilistic approach based on the SPT dataset.

Tables 4 and 5 summarize the results illustrated in Figs. 6 and 7.

Table 4: P-R cost curve summarized for the Cetin et al. (2004) approach based on the SPT dataset.

Cost Ratio (CR) Range	Cetin et al., 2004		
	Optimal Threshold	Precision	Recall
0 < CR < 0.11	0.002	0.692	0.99
0.11 < CR < 1	0.049	0.781	0.981
> 1.0	0.596	0.923	0.77

Table 5: P-R cost curve summarized for the Moss et al. (2006) approach based on the CPT dataset.

Cost Ratio (CR) Range	Moss et al., 2006		
	Optimal Threshold	Precision	Recall
0 < CR < 0.6	0.072	0.868	1
CR > 0.6	0.308	0.894	0.978

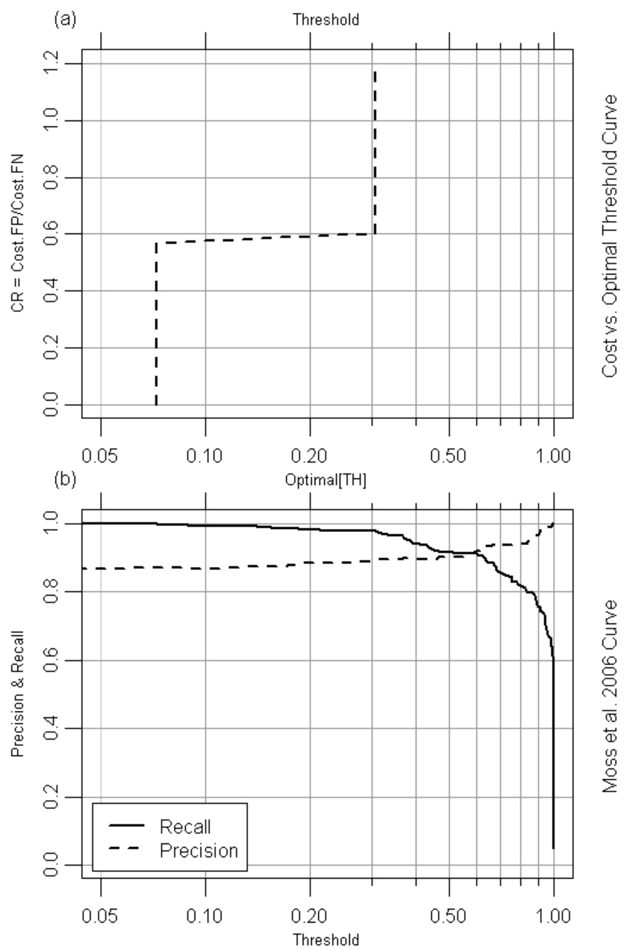

Fig. 7: P-R cost curve for the Moss et al. (2006) probabilistic approach based on the CPT dataset.

Therefore, since H-1, H-2, and H-5 have P_L value < than 0.308, they are classified as non-liquefiable, whereas H-3 and H-4 have P_L value > than 0.308, and they are classified as liquefiable. The P-R curve (Fig. 7b) helps us to determine how confident we can be with this decision that they are non-liquefiable or liquefiable. We observe from Fig. 7b that the precision and recall values corresponding to the P_L for each case are H-1 (precision = 0.883, recall = 0.985), H-2 (precision = 0.883, recall = 0.978), H-3 (precision = 0.897, recall = 0.942), H-4 (precision = 0.920, recall = 0.913), and H-5 (precision = 0.894, recall = 0.978). Recall gives the chance that concluding the site will not liquefy is wrong. And precision gives the chance that concluding the site will liquefy is wrong.

In the case of H-1, a recall = 0.985 means that there is 1.5% chance for the decision that the site will not liquefy is wrong. We observe that although H-2 and H-3 had different P_L values (0.25 and .30), both cases have the same recall values, or in other words, both cases have 2.2% chance that concluding the site will not liquefy is wrong. In the case of H-3, a precision = 0.89 means that there is 11% chance that concluding the site will liquefy is wrong. Similarly, in the case of H-4, a precision = 0.92 means that there is 8% chance that concluding the site will liquefy is wrong.

CONCLUSIONS

In this study, we have critically compared the deterministic and probabilistic ELMs based on SPT and CPT data to provide an objective and quantitative validation framework to evaluate the predictive performance and to inform the use of ELMs. For the deterministic ELMs we compared (1) the "simplified procedure", and (2) Bayesian updating method, whereas for the probabilistic ELMs we compared the (1) Cetin et al. (2004), and (2) Moss et al. (2006) Bayesian updating methods. We also presented a new optimization approach for choosing the optimal TH_L for implementation of the probabilistic assessment of liquefaction, which minimizes the overall costs associated with a particular project design.

By comparing multiple liquefaction models for both SPT and CPT data with validation metrics that are commonly used in statistics yet are uncommon in the geotechnical literature, we have illustrated that the predictive capabilities are comparable in general. However, each model has distinct advantages or disadvantages in terms of precision or recall for the different classes. These validation metrics will better inform geotechnical users and allow them to choose the method and optimal TH_L (for probabilistic methods) that best suits a particular project. The following specific conclusions arise from the model validation results in this study:

- For the deterministic evaluation of liquefaction potential using SPT data, the "simplified procedure" has a slightly better predictive capability than the Bayesian updating method for the liquefaction class, whereas, the latter has a better predictive capability for the non-liquefaction class based on an overall metric termed the F-score.
- For the deterministic evaluation of CPT data, the Bayesian updating method has a better predictive capability than the "simplified procedure" for the liquefaction class, and vice versa for the non-liquefaction class.
- The probabilistic evaluation of the liquefaction potential indicates comparable performance for both Cetin et al. (2004), and Moss et al. (2006) with the latter having slightly improved AUC.
- The P-R cost curve is an efficient and objective approach to determine the optimal TH_L and the associated risks associated with the decision in the case

of probabilistic evaluation. Practicing geotechnical engineers can use tables 4 and 5 to determine the optimal TH_L when they evaluate the P_L based on the Bayesian updating methods (Cetin et al. 2004; Moss et al. 2006).

Perhaps the most important implication of this study is that the recent improvements in liquefaction models have only marginally improved their prediction accuracy. Thus future efforts should instead be focused on strategic data collection to enhance model performance and reduce sampling bias and class imbalance.

REFERENCES

Cetin, K. O., Kiureghian, A. D., and Seed, R. B. [2002]. "Probabilistic models for the initiation of seismic soil liquefaction." *Structural Safety*, 24(1), 67-82.

Cetin, K. O., Seed, R. B., Kiureghian, A.D., Tokimatsu, K., Harder, L. F., Kayen, R. E., and Moss, R. E. S. [2004]. "Standard penetration test-based probabilistic and deterministic assessment of seismic soil liquefaction potential." *Journal of Geotechnical and Geoenvironmental Engineering*, 130(12), 1314-1340.

Goh, A. T. C. [1994]. "Seismic Liquefaction Potential Assessed by Neural Networks." *Journal of Geotechnical Engineering-ASCE*, 120(9), 1467-1480.

Goh, A. T. C. [2002]. "Probabilistic neural network for evaluating seismic liquefaction potential." *Canadian Geotechnical Journal*, 39(1), 219-232.

Goh, A. T. C., and Goh, S. H. [2007]. "Support vector machines: Their use in geotechnical engineering as illustrated using seismic liquefaction data." *Computers and Geotechnics*, 34, 410-421.

Hanna, A. M., Ural, D., and Saygili, G. [2007]. "Evaluation of liquefaction potential of soil deposits using artificial neural networks." *Engineering Computations*, 24[1-2), 5-16.

Idriss, I.M. and Boulanger, R.W. [2006], "Semi-empirical procedures for evaluating liquefaction potential during earthquakes," *Soil Dynamics and Earthquake Engineering*, Vol. 26, pp. 115-130.

Juang, C. H., Chen, C. J., Tang, W. H., and Rosowsky, D. V. [2000]. "CPT-based liquefaction analysis, Part 1: Determination of limit state function." *Geotechnique*, 50(5), 583-592.

Juang, C. H., Jiang, T., and Andrus, R. D. [2002]. "Assessing probability-based methods for liquefaction potential evaluation." *Journal of Geotechnical and Geoenvironmental Engineering*, 128(7), 580-589.

Juang, C. H., Yuan, H. M., Lee, D. H., and Lin, P. S. [2003]. "Simplified cone penetration test-based method for evaluating liquefaction resistance of soils." *Journal of Geotechnical and Geoenvironmental Engineering*, 129(1), 66-80.

Juang, C.H., Li, D.K., Fang, S.Y., Liu, Z., and Khor, E.H. [2008]. "Simplified procedure for developing joint distribution of a_{max} and M_w for probabilistic liquefaction hazard analysis." *Journal of Geotechnical and Geoenvironmental Engineering*, 134(8), 1050-1058.

Krawinkler, H. [2004]. "Exercising performance-based earthquake engineering." Proceedings of the Third International Conference on Earthquake Engineering, 212-218.

Lai, S.-Y., Hsu, S.-C., and Hsieh, M.-J. [2004]. "Discriminant model for evaluating soil liquefaction potential using cone penetration test data." *Journal of Geotechnical and Geoenvironmental Engineering*, 130(12), 1271-1282.

Liao, S. S. C., Veneziano, D., and Whitman, R. V. [1988]. "Regression-Models for Evaluating Liquefaction Probability." *Journal of Geotechnical Engineering-ASCE*, 114(4), 389-411.

Moss, R. E. S., Seed, R. B., Kayen, R. E., Stewart, J. P., Kiureghian, A. D., and Cetin, K. O. [2006]. "CPT-based probabilistic and deterministic assessment of in situ seismic soil liquefaction potential." *Journal of Geotechnical and Geoenvironmental Engineering*, 132(8), 1032-1051.

Oommen, T., Baise, L. G. and Vogel, R. [2009]. "A Cautionary Note to the Engineers and Geoscientist on the Use of Maximum Likelihood Logistic Regression." (In Prep).

Pal, M. [2006]. "Support vector machines-based modeling of seismic liquefaction potential." *International Journal for Numerical and Analytical Methods in Geomechanics*, 30(10), 983-996.

Robertson, P. K., and Campanella, R. G. [1985]. "Liquefaction Potential of Sands Using the CPT." *Journal of Geotechnical Engineering-ASCE*, 111(3), 384-403.

Robertson, P. K., and Wride, C. E. [1998]. "Evaluating cyclic liquefaction potential using the cone penetration test." *Canadian Geotechnical Journal*, 35(3), 442-459.

Seed, H. B., and De Alba, P. [1986]. "Use of SPT and CPT tests for evaluating the liquefaction resistance of sands." Blacksburg, VA, USA, 281-302.

Seed, H. B., and Idriss, I. M. [1971]. "Simplified procedure for evaluating soil liquefaction potential." 97(SM9), 1249-73.

Seed, H. B., Idriss, I. M., and Arango, I. [1983]. "Evaluation of Liquefaction Potential Using Field Performance Data." *Journal of Geotechnical Engineering-ASCE*, 109(3), 458-482.

Shibata, T., and Teparaksa, W. [1988]. "Evaluation of Liquefaction Potentials of Soils Using Cone Penetration Tests." *Soils and Foundations*, 28(2), 49-60.

Stark, T. D., and Olson, S. M. [1995]. "Liquefaction Resistance Using CPT and Field Case-Histories." *Journal of Geotechnical Engineering-ASCE*, 121(12), 856-869.

Toprak, S., Holzer, T. L., Bennett, M. J., and Tinsley, J. C. I., [1999]. "CPT and SPT-based probabilistic assessment of liquefaction." *Proc., 7th U.S.-Japan Workshop on Earthquake Resistant Design of Lifeline Facilities and Countermeasures against Liquefaction*, MCEER, Seattle, 69-86.

Youd, T. L., Idriss, I. M., Andrus, R. D., Arango, I., Castro, G., Christian, J. T., Dobry, R., Finn, W. D. L., Harder Jr., L. F., Hynes, M. E., Ishihara, K., Koester, J. P., Liao, S. S. C., Marcuson, W. F., Martin, G. R., Mitchell, J. K., Moriwaki, Y., Power, M. S., Robertson, P. K., Seed, R. B., and Stokoe, K. H. [2001]. "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*, 127(10), 817-833.