Validation and Application of Empirical Liquefaction Models

Thomas Oommen[1]; Laurie G. Baise, M.ASCE[2]; and Richard Vogel, M.ASCE[3]

Abstract: Empirical liquefaction models (ELMs) are the standard approach for predicting the occurrence of soil liquefaction. These models 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. No objective and quantitative comparison of these models have been published. Similarly, no rigorous procedure has been published for choosing the threshold required for probabilistic models. This paper provides (1) a quantitative comparison of the predictive performance of ELMs; (2) a reproducible method for choosing the threshold that is needed to apply the probabilistic ELMs; and (3) an alternative deterministic and probabilistic ELM based on the machine learning algorithm, known as support vector machine (SVM). Deterministic and probabilistic ELMs have been developed for SPT and CPT data. For deterministic ELMs, we compare the "simplified procedure," the Bayesian updating method, and the SVM models for both SPT and CPT data. For probabilistic ELMs, we compare the Bayesian updating method with the SVM models. We compare these different approaches within a quantitative validation framework. This framework includes validation metrics developed within the statistics and artificial intelligence fields that are not common in the geotechnical literature. We incorporate estimated costs associated with risk as well as with risk mitigation. We conclude that (1) the best performing ELM depends on the associated costs; (2) the unique costs associated with an individual project directly determine the optimal threshold for the probabilistic ELMs; and (3) the more recent ELMs only marginally improve prediction accuracy; thus, efforts should focus on improving data collection.

DOI: 10.1061/(ASCE)GT.1943-5606.0000395

CE Database subject headings: Bayesian analysis; Validation; Empirical equations; Soil liquefaction; Data collection.

Author keywords: Bayesian updating method; Support vector machine; Simplified procedure; Model validation; CPT; SPT; Machine learning.

Introduction

Soil liquefaction is the loss of shear strength induced by shaking, which can lead to various types of ground failures. Liquefaction is most often evaluated with empirical liquefaction models (ELMs). ELMs have been developed for in situ index tests, such as standard penetration test (SPT), cone penetration test (CPT), and shear-wave velocity (V_s). These in situ data are used to estimate the potential for "triggering" or initiation of seismically induced liquefaction. In the context of the analyses of in situ data, the estimate of liquefaction potential derived from ELMs can be broadly classified as (1) deterministic (Seed and Idriss 1971; Iwasaki et al. 1978; 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, 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, 2004; Lee et al. 2003; Sonmez 2003; Lai et al. 2004; Sonmez and Gokceoglu 2005; Papathanassiou et al. 2005; Holzer et al. 2006; Moss et al. 2006; Juang and Li 2007). This paper attempts to improve liquefaction models by (1) quantitatively comparing the predictive performance of several ELMs; (2) identifying the threshold needed to apply the probabilistic ELMs; and (3) developing an alternative deterministic and probabilistic ELM based on the machine learning algorithm, known as support vector machine (SVM).

Currently, the most widely used ELM for the assessment of liquefaction potential is the "simplified procedure," originally recommended by Seed and Idriss (1971) based on SPT blow counts. Since 1971, this procedure has been revised and developed for other in situ tests, such as the CPT and V_s (Seed et al. 1985; Youd and Noble 1997). Simplified methods that follow the general format of the Seed-Idriss procedure were reviewed in a workshop report edited by Youd and Noble (1997). Youd et al. (2001), Cetin et al. (2004), and Moss et al. (2006) provided recent updates to this method. In addition, Cetin et al. (2004) and Moss et al. (2006) presented liquefaction models that use the Bayesian updating method for SPT and CPT data, respectively. The recent work represents an update to the data sets combined with the use of the Bayesian updating method for probabilistic evaluation of liquefaction potential. A thorough comparison of these competing ELMs for assessing liquefaction potential has yet to be published in the literature.

In recent years, innovative computing techniques such as artificial intelligence and machine learning have gained popularity in geotechnical engineering. For example, Goh (1994) and Goh (2002) introduced the artificial neural networks for liquefaction

[1]Postdoctoral Associate, Dept. of Civil and Environmental Engineering, Tufts Univ., 113 Anderson Hall, Medford, MA 02155; presently, Assistant Professor, Dept. of Geological Engineering, Michigan Tech., Houghton, MI 49931 (corresponding author). E-mail: toommen@mtu.edu

[2]Associate Professor, Dept. of Civil and Environmental Engineering, Tufts Univ., 113 Anderson Hall, Medford, MA 02155.

[3]Professor, Dept. of Civil and Environmental Engineering, Tufts Univ., 113 Anderson Hall, Medford, MA 02155.

Note. This manuscript was submitted on July 30, 2009; approved on June 2, 2010; published online on November 15, 2010. Discussion period open until May 1, 2011; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Geotechnical and Geoenvironmental Engineering***, Vol. 136, No. 12, December 1, 2010. ©ASCE, ISSN 1090-0241/2010/12-1618–1633/$25.00.

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

potential, Cetin et al. (2004) and Moss et al. (2006) applied the Bayesian updating method for probabilistic assessment of liquefaction, and Hashash (2007) used the genetic algorithms for geomechanics. An important advantage of artificial intelligence techniques is that the nonlinear behavior of multivariate dynamic systems is computed efficiently with no a priori assumptions regarding the distribution of the data. The SVM algorithm combines the principles of structural risk minimization and the statistical learning theory pioneered by Cortes and Vapnik (1995) and Vapnik (1995). SVM has been successfully employed in solving classification problems (Goh and Goh 2007; Oommen et al. 2008; Pal 2006; Sahoo et al. 2007) and functional regression problems (Oommen and Baise 2010; Samui et al. 2008) in geotechnical engineering. However, the applicability of a probabilistic approach, which integrates the regularized likelihood with the SVM, has not been explored for geotechnical engineering applications.

The deterministic method provides a "yes/no" response to the question of whether or not a soil layer at a specific location 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 1980s by Liao et al. (1988). However, 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 soil at a particular location to be classified as either liquefiable or nonliquefiable. Thus, a soil where $P_L < TH_L$ is classified as nonliquefiable and a soil 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. In this study, we provide a thorough and reproducible approach to interpret P_L and to compute the optimal TH_L that incorporates the costs associated with the risk of liquefaction and the costs associated with risk mitigation.

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; Moss et al. 2006). We also compare the SVM based deterministic and probabilistic ELMs to previous models to (1) evaluate an independent ELM; (2) quantify model improvement; and (3) quantify the independent limitations stemming from model selection and sampling bias. By using a model validation framework with relevant validation statistics, we can evaluate the performance of each model and identify any inherent bias. Model bias occurs because a model can be optimized for either precision or recall of one class over another. Without an open model validation, this inherent model bias is hidden from the user.

In the following section of the paper, we describe the data used for comparing the different ELMs. Next, we describe the methodology that we used for model validation and the validation statistics that we adopted and the details on the different ELMs that we consider in this paper. Finally, we present an objective method for identifying optimal TH_L between the instances of liquefaction and nonliquefaction.

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) reevaluation 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 the 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 nonliquefied sites. The CPT database has 182 case histories of which 139 are from liquefied sites and 43 are from nonliquefied sites. The ratio of liquefaction to nonliquefaction 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.

Figs. 1 and 2 use box plots to compare the different predictive variables for the liquefied and nonliquefied sites from the SPT and CPT databases. The variables that exhibit the largest difference between instances of liquefaction and nonliquefaction for the SPT data are (1) the moment magnitude and (2) the overburden corrected blow count. The largest differences for the CPT data are (1) moment magnitude and (2) tip resistance. Note that Fig. 2 shows that the tip resistance exhibits more separability than the friction ratio. Comparing the predictive variables from SPT and CPT databases we observe that the total and effective stress values have very poor separability between the liquefaction and nonliquefaction instances in both the databases. The separability between the liquefaction and nonliquefaction instances increase with variable peak ground acceleration and CSR in both the databases. However, it is observed that the CSR values in the database of Moss et al. (2006) (Fig. 2) have better separability between liquefaction and nonliquefaction instances compared to the CSR values in the database of Cetin et al. (2004) (Fig. 1). Moreover, the range of CSR values of the nonliquefaction instances in the database of Moss et al. (2006) is much smaller compared to the database of Cetin et al. (2004). We believe that this is due to the difference in approach adopted by Cetin et al. (2004) and Moss et al. (2006) in calculating the shear mass participation factor (r_d) which is used in calculating the CSR. The major difference in the r_d relationship used to develop the two databases is that the database of Cetin et al. (2004) uses the shear-wave velocity (V_s) over the top 12 m (40 ft), whereas Moss et al. (2006) does not. Cetin et al. (2004) stated that the V_s values for the r_d relationship can be obtained using measured data or approximated from appropriate empirical correlations. However, Moss et al. (2006) stated that most CPT case histories did not have V_s values and the use of inferred V_s was avoided to prevent introducing associated uncertainties.

Methodology

We calculated the liquefaction potential for the SPT and CPT databases by using the simplified procedure (Youd et al. 2001), the Bayesian updating method (Cetin et al. 2004; Moss et al. 2006), and the SVM method. The following sections provide a brief description of the fundamental principles of these approaches/classifiers and the equations used. We validate and

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

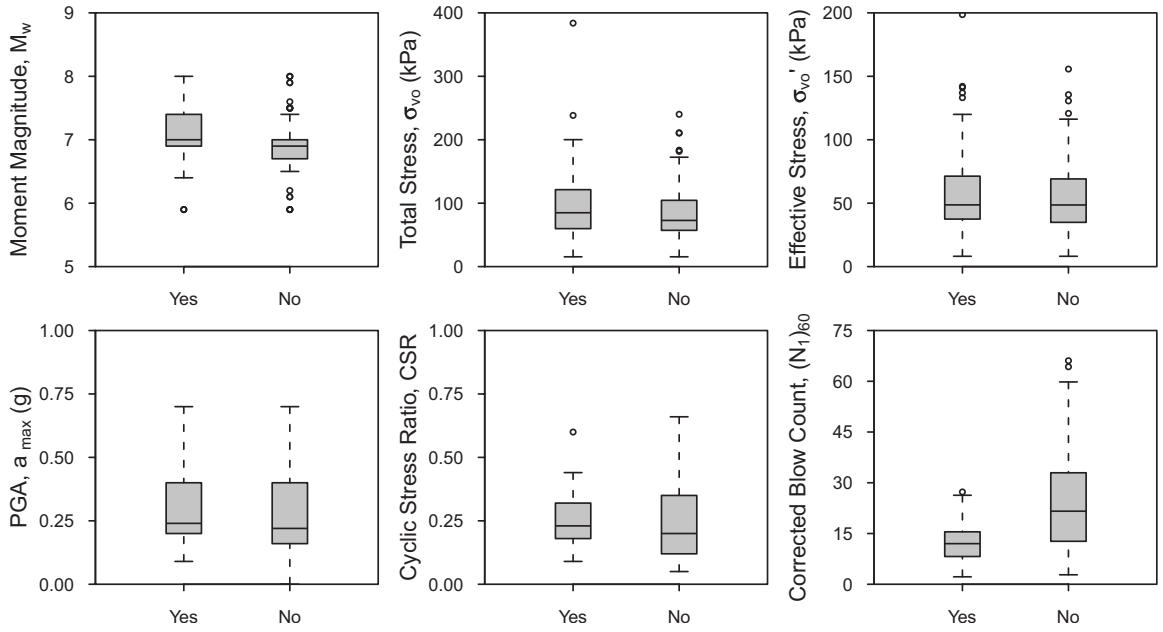

Fig. 1. Box plots of various predictor variables based on the SPT data set where "yes" represents liquefaction and "no" represents nonliquefaction instances. Legend explaining the key elements of a box plot is provided in Fig. 2.

quantify the different deterministic classifiers by using overall accuracy (OA), precision, recall [i.e., true positive (TP) rate], and F score. 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 the probabilistic assessment of liquefaction potential.

Model Validation

Model development (i.e., model "training") should be followed by a model validation to assess predictive capability. When a nonlinear classifier is trained and validated on the same data set, its predictive performance can often be optimistically biased due to overfitting (Hawkins 2004; Oommen and Baise 2010). Therefore, the SVM classifier was validated using a *K*-fold cross validation approach. For the *K*-fold cross validation, we randomly break the data set into *K* partitions. Then, the model is trained using $K-1$-folds and the remaining onefold is used for validation. This is repeated *K* times, so that each time a different partition is used for validation. Finally, the predictability of the model is estimated by calculating error estimates on the test instances of each *K* split. The advantage of *K*-fold cross validation is that all the examples in the data set are eventually used for both training and validation, yet for each example in the data set, training and validation are implemented independently. We use $K=5$ for the *K*-fold cross validation. However, the previously developed models that we consider in this paper were trained on the complete

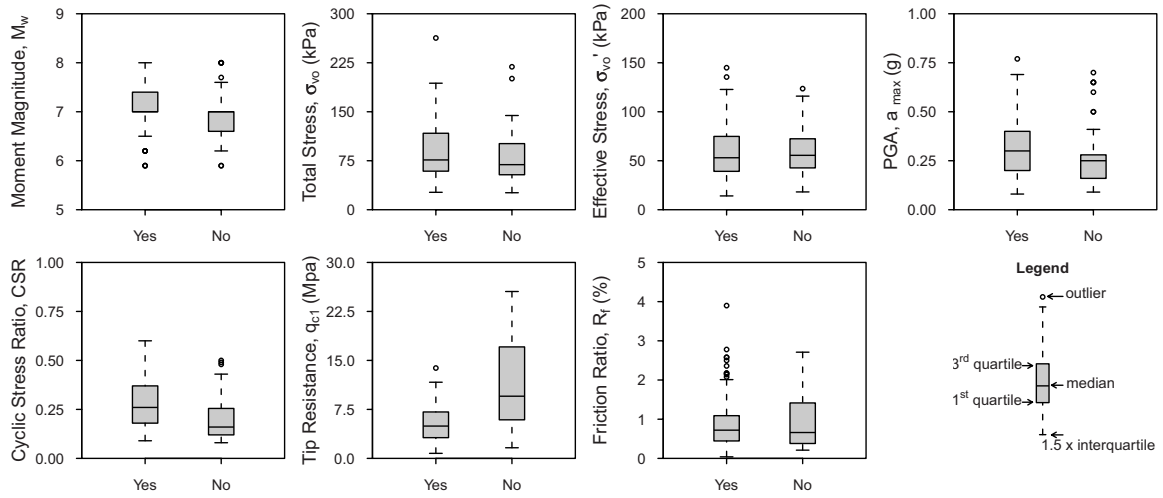

Fig. 2. Box plots of various predictor variables based on the CPT data set where "yes" represents liquefaction and "no" represents nonliquefaction instances

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

Table 1. Confusion Matrix, Presenting the Observed Class in Rows and the Predicted Class 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

data sets so we have to validate these classifiers on the same data set used for model development. As a result, the validation statistics for these methods will likely overestimate the prediction accuracy of the models.

For deterministic models, useful validation statistics include OA, 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 nondiagonal elements include the frequencies of misclassifications. The OA is a measure of the percentage of correctly classified instances

$$\text{overall accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (1)$$

where the TP=sum of instances of liquefaction correctly predicted; the true negative (TN)=sum of instances of nonliquefaction correctly predicted; the false positive (FP)=sum of instances of nonliquefaction classified as liquefaction; and the false negative (FN)=sum of instances of liquefaction classified as nonliquefaction. OA 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 does not mean that the 75% of each class (e.g., liquefaction and nonliquefaction classes) has been correctly predicted. Therefore, the evaluation of the predictive capability based on the OA alone can be misleading when a class imbalance exists or the number of instances from each class is not equal in the data set (e.g., for the CPT data set 76% of the data are liquefaction instances and 24% are nonliquefaction instances).

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

$$\text{precision} = P = \text{TP}/(\text{TP} + \text{FP}) \quad (2)$$

$$\text{recall} = R = \text{TP}/(\text{TP} + \text{FN}) \quad (3)$$

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 nonliquefaction that are misclassified. An inverse relationship exists between precision and recall: it is possible to increase one at the expense of the other.

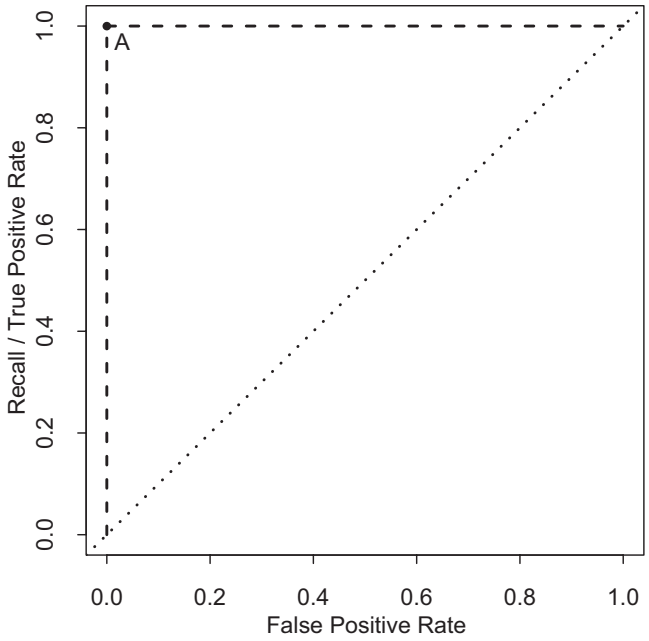

Fig. 3. ROC curve illustrating its basic elements. Dashed line indicates a near perfect probability prediction, whereas dotted line indicates predictions which result from random guessing.

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

$$F_\beta = (1 + \beta^2)(P \cdot R)/(\beta^2 \cdot P + R) \quad (4)$$

where β=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/nonliquefaction boundary to apply deterministic metrics such as given in Eqs. (1)–(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 FP rate (FPR) by varying the threshold from 0 to 1. The FPR is

$$\text{FPR} = \text{FP}/(\text{FP} + \text{TN}) \quad (5)$$

Thus, any point on either the P-R or ROC curve corresponds to a specific threshold. Fig. 3 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 in Fig. 3). This gives AUC=0.5; thus, it is unrealistic for a classifier to have an AUC less than 0.5.

Fig. 4 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 data set. 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

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

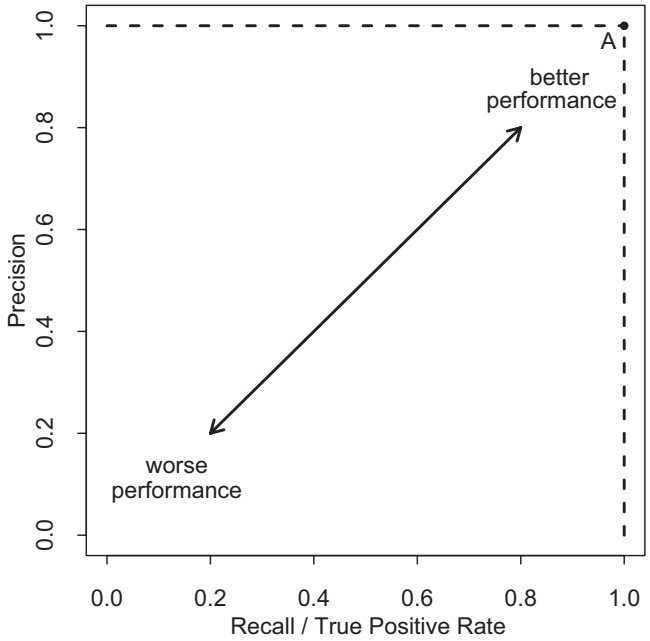

Fig. 4. P-R curve illustrating its basic elements. Dashed line represents the best P-R curve.

are misrepresented. For 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. (2010a) demonstrated that sampling bias can significantly influence model development and performance.

Simplified Procedure (Youd et al. 2001)

Following the disastrous earthquakes in Alaska and in Nigata, Japan in 1964, Seed and Idriss (1971) developed the simplified procedure using empirical evaluations of field observations for evaluating liquefaction potential. A series of publications revised the procedure (Seed et al. 1983; Youd et al. 2001). Youd et al. (2001) stated 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 values: (1) the seismic demand on a soil layer, expressed in terms of the CSR, and (2) the capacity of the soil to resist liquefaction expressed in terms of the cyclic resistance ratio (CRR). The latter variable depends on the type of in situ measurement (i.e., SPT or CPT). Here

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

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

$$r_d = \begin{cases} 1.0 - 0.007\,65z & \text{for } z \le 9.15 \text{ m} \\ 1.174 - 0.0267z & \text{for } 9.15 \text{ m} < z \le 23 \text{ m} \end{cases} \tag{7}$$

where z=depth beneath ground surface in meters.

The CRR for fine 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,

$$\mathrm{CRR}_{7.5}^{\mathrm{SPT}} = \frac{1}{34 - (N_1)_{60}} + \frac{(N_1)_{60}}{135} + \frac{50}{[10 \cdot (N_1)_{60} + 45]^2} - \frac{1}{200} \tag{8}$$

$$\mathrm{CRR}_{7.5}^{\mathrm{CPT}} = \begin{cases} 0.833[(q_{c1N})_{CS}/1{,}000] + 0.05 \\ \quad \text{for } (q_{c1N})_{CS} < 50 \\ 93[(q_{c1N})_{CS}/1{,}000]^3 + 0.08 \\ \quad \text{for } 50 \le (q_{c1N})_{CS} < 160 \end{cases} \tag{9}$$

where $\mathrm{CRR}_{7.5}^{\mathrm{SPT}}$=CRR for SPT; $\mathrm{CRR}_{7.5}^{\mathrm{CPT}}$=CRR for CPT; $(N_1)_{60}$=corrected SPT blow count and <30; and $(q_{c1N})_{CS}$=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

$$\mathrm{FS} = (\mathrm{CRR}_{7.5}/\mathrm{CSR})\mathrm{MSF} \tag{10}$$

where MSF=magnitude scaling factor.

Bayesian Updating Method (Cetin et al. 2004; Moss et al. 2006)

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 $g = g(x,\theta) + \varepsilon$, where x is the set of predictive variables, θ is the set of unknown model parameters, and ε is 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_z.

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

$$L(\theta,\sigma_z) = \prod P[g(x_i,\theta) + \varepsilon_i \le 0]^{W_{liq}} \cdot \prod P[g(x_i,\theta) + \varepsilon_i > 0]^{W_{nonliq}} \tag{11}$$

where W_{liq} and W_{nonliq}=correction terms to account for the class imbalance in the field case history database due to the disproportionate number of liquefied versus nonliquefied 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) recommended using a P_L value of >0.15 as liquefiable, otherwise all remaining as nonliquefiable. 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) provided similar recommendations for the probability values as in Cetin et al. (2004).

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

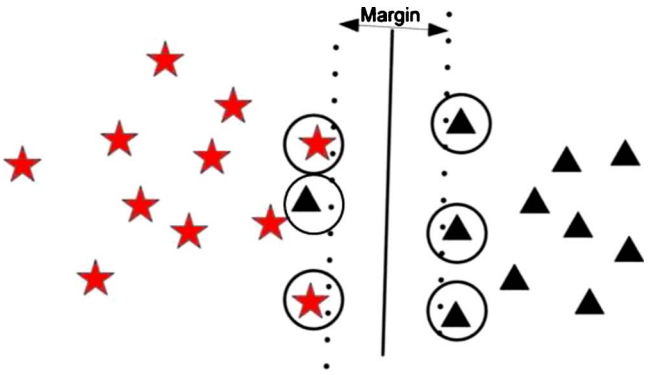

Fig. 5. Illustrates a linearly separable two-class SVM problem, with the center line representing the optimal hyperplane (classification line) that separates the two classes. The dotted lines represent the maximum margin, and the instances that fall along the maximum margin, and the instances that are misclassified represent the support vectors. Support vectors are circled.

SVMs

SVM is an artificial intelligence algorithm used for classification, regression, and novelty/outlier detection. The model development using SVM involves a training/development phase in which a series of instances of related inputs and target output values are provided. The SVM implicitly maps these training instances into a higher-dimensional feature space defined by a kernel function. The kernel function helps to reduce the computational complexity of explicitly projecting x and x' into the higher-dimensional space. For example, a kernel function represents the dot product $\langle \Phi(x), \Phi(x') \rangle$ of a projection $\Phi: X \rightarrow H$

$$k(x,x') = \langle \Phi(x), \Phi(x') \rangle \tag{12}$$

This is computationally simpler than explicitly projecting x and x' into the higher-dimensional space (Scholkopf and Smola 2002). During the training phase, a hyperplane separates the two classes (i.e., liquefiable and nonliquefiable) of interest in the higher-dimensional feature space, $\langle w, \Phi(x) \rangle + b = 0$, corresponding to the decision function, $f(x) = (\langle w, \Phi(x) \rangle + b)$.

The vector w determines the orientation of the hyperplane and the scalar b determines the offset of the hyperplane from its origin. In the higher-dimensional space, an infinite number of hyperplanes may exist that can separate the two classes. However, one hyperplane has the maximum margin of separation between the two classes as shown in Fig. 5. This hyperplane that has the maximum margin is called the optimal hyperplane and is constructed by solving the following constrained quadratic optimization problem:

$$\text{minimize} \quad t(w,\xi) = \frac{1}{2}\|w\|^2 + \frac{c}{m}\sum_{i=1}^{m} \xi_i$$

$$\text{subject to} \quad y_i(\langle \Phi(x_i), w \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (i = 1, \ldots, m) \tag{13}$$

where m=number of instances in the training data set; y_i=target output values (i.e., liquefiable and nonliquefiable); x_i=input parameters; c=magnitude of penalty associated with incorrectly classified training instances; and ξ_i=slack variable that indicates the distance of the incorrectly classified instances from the optimal hyperplane. Minimization of the first term in the objective function results in maximization of the margin, while the second term tends to penalize the instances that are incorrectly classified. The penalty term c allows the model to strike a balance between these two competing criteria of margin maximization and error minimization. A high value of c will force SVM to create a complex prediction function with minimum error, while a lower c will lead to a simpler prediction function with higher error. The instances that fall along the maximum margin of the optimal hyperplane and the instances that are misclassified are called the support vectors, and they carry all the relevant information about the model.

In this research, we used the ksvm algorithm in the kernlab package (Karatzoglou et al. 2006) of the R programming language (R Development Core Team 2009) to solve the optimization problem in Eq. (12). The ksvm algorithm uses the sequential minimal optimization algorithm to solve the quadratic optimization problem. The ksvm algorithm provides several kernel functions such as Gaussian radial basis, polynomial, linear, and sigmoid. Previous research has documented that a linear kernel is a special case of the radial basis and the sigmoid kernel behaves like a radial basis for certain parameters (Keerthi and Lin 2003). Hence, we chose the Gaussian radial basis kernel, which is given by

$$k(x,x') = e^{-\gamma\|x - x'\|^2} \tag{14}$$

where γ controls the width of the Gaussian kernel. When a Gaussian radial basis function for the SVM classification is employed, we optimize the two parameters: c and γ. The SVM classification model was developed by using the methodology recommended by Hsu et al. (2003) which can be used to predict the label of any test example. However, many applications require a posterior class probability $P(y=1/x)$ instead of predicting the class label of a test example. Platt (2000) proposed a method for approximating the posterior class probability that leaves the SVM function unchanged and instead adds a trainable postprocessing with regularized binomial maximum likelihood (Lin et al. 2007). A two parameter sigmoid function was used to approximate the posterior class probability, where

$$P(y = 1/x) \approx P_{AB}(f) \equiv \frac{1}{1 + e^{(Af+B)}} \tag{15}$$

with $f=f(x)$ and each f_i is an estimate of $f(x_i)$. Optimal values for A and B are determined by solving the regularized likelihood optimization problem (with L for the liquefaction class and NL for nonliquefaction class) which is given by minimizing $\min_{z=(A,B)}$, where

$$F(z) = -\sum_{i=1}^{m} [t_i \log(p_i) + (1 - t_i)\log(1 - p_i)] \tag{16}$$

where $p_i = P_{A,B}(f_i)$, and

$$t_i = \begin{cases} \dfrac{L+1}{L+2} & \text{if } y_i = 1 \\ \dfrac{1}{NL+2} & \text{if } y_i = -1 \end{cases} \qquad i = 1, \ldots, m \tag{17}$$

The method of Platt (2000) integrating the regularized likelihood with the SVM preserves the sparseness of the SVM while producing probabilistic outputs.

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

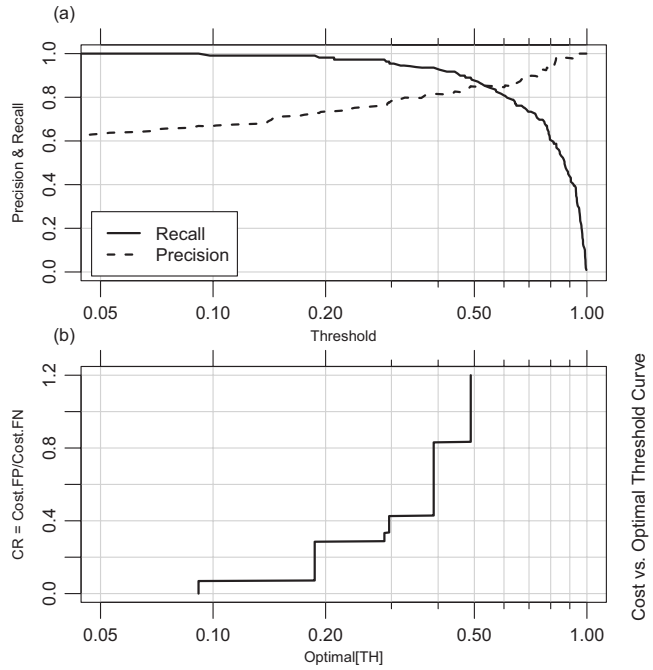

Fig. 6. P-R cost curve used to determine the optimal threshold of liquefaction (TH_L) triggering for probabilistic evaluation: (a) precision and recall versus threshold; (b) cost ratio versus optimal TH_L

Thresholds for Liquefaction Triggering

In this section we present a new approach by combining project cost information with the precision and recall (P-R cost 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. Fig. 6 presents a typical P-R cost curve, which consists of two plots. Fig. 6(a) illustrates the choice of the threshold versus precision and recall. For a given probabilistic approach, Fig. 6(a) is developed by varying the threshold from 0 to 1 and calculating the corresponding precision and recall values for each of these thresholds. Fig. 6(b) presents the optimal TH_L versus the ratio of the C_{FP} (C_{FP}=cost of predicting a true nonliquefaction instance as liquefaction) to the C_{FN} (C_{FN}=cost of predicting a true liquefaction instance as nonliquefaction) abbreviated as *CR*. The optimal TH_L is approximated by minimizing the cost

$$\text{optimal}[\text{TH}_L]_j = \min|\text{FP}_i \cdot CR_j + \text{FN}_i| \quad (18)$$

where i=entire range of threshold from 0 to 1; FP_i and FN_i =number of FP and FN 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 $C_{FP}=1.2\times C_{FN}$). In practice, the C_{FP} and C_{FN} can be computed based on the PBEE recommended decision variables such as dollar losses, downtime, and deaths (Krawinkler 2004).

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) assigned TH_L values (0.15) in their probabilistic analysis. Table 2 presents the comparison of the SPT-based approaches from Youd et al. (2001) and Cetin et al. (2004) and the current study using SVM. Comparing the OA for the three approaches, it is evident that the SVM has the highest OA, whereas the approach of Youd et al. (2001) has the least OA. Since the SPT database has a class imbalance of 56:44 (liquefaction:nonliquefaction), the OA alone cannot be used as an indicator of the predictive performance of the approaches. Therefore, liquefaction and nonliquefaction classes are analyzed separately using recall, precision, and F score. In the case of the liquefaction class, we see that the SVM has the highest recall, whereas the model of Cetin et al. (2004) 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 the SVM has the highest F score, whereas Cetin et al. (2004) has the least F score. Moreover, both Youd et al. (2001) and Cetin et al. (2004) have similar F-score values with the former having a slightly improved score than the latter. In the case of the nonliquefaction class, we observe that the model of Cetin et al. (2004) has the highest recall, whereas the SVM has the highest precision. In addition, a comparison of the F scores indicates the SVM and the model of Cetin et al. (2004) as having comparable F-score values for the nonliquefaction case with the former having slightly better performance. From Table 2, we observe using the F-score measure that the SVM approach has improved the predictive performance of both the liquefaction and nonliquefaction instances compared to the approaches of Cetin et al. (2004) and Youd et al. (2001). The optimal values of c and γ for the SVM approach for the SPT data

Table 2. Various Estimates of the Predictive Performance of the SPT-Based Deterministic Models: (1) OA and (2) Recall, Precision, and F Score for Both Liquefaction and Nonliquefaction Occurrences

	Data set of Cetin et al. (2004)						
Approach	OA	Recall liquefaction	Precision liquefaction	F-score liquefaction	Recall nonliquefaction	Precision nonliquefaction	F-score nonliquefaction
Youd et al. 2001	0.826	0.816	0.864	0.8396	0.8391	0.784	0.811
Cetin et al. 2004 (TH_L=0.15)	0.831	0.789	0.895	0.839	0.885	0.77	0.823
SVM	0.852	0.899	0.844	0.871	0.793	0.862	0.826
Cetin et al. 2004 (TH_L=0.5)	0.831	0.724	0.963	0.827	0.965	0.736	0.835

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

Table 3. Various Estimates of the Predictive Performance of the CPT-Based Deterministic Models: (1) OA and (2) Recall, Precision, and F Score for Both Liquefaction and Nonliquefaction Occurrences

	Data set of Moss et al. 2006						
Approach	OA	Recall liquefaction	Precision liquefaction	F-score liquefaction	Recall nonliquefaction	Precision nonliquefaction	F-score nonliquefaction
Youd et al. 2001	0.846	0.877	0.917	0.897	0.744	0.653	0.695
Moss et al. 2006 (TH_L=0.15)	0.879	0.985	0.872	0.925	0.534	0.92	0.676
SVM	0.89	0.978	0.888	0.931	0.604	0.896	0.722
Moss et al. 2006 (TH_L=0.5)	0.857	0.913	0.9	0.907	0.674	0.7	0.69

are 5.75 and 0.030 415, respectively. It is important to note that although the approach of Cetin et al. (2004) has slightly improved predictive capability compared to that of Youd et al. (2001) in the nonliquefaction 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) and the current study using SVM. Comparing the OA for the three approaches, we see a similar trend to that of the SPT approaches with SVM having the highest OA, whereas the approach of Youd et al. (2001) has the least OA. However, the CPT database has greater class imbalance (76:24, liquefaction:nonliquefaction) than the SPT database. Hence again, the OA alone cannot be used as an indicator to compare the predictive performance. Analyzing the predictive performance based on the individual classes (liquefaction and nonliquefaction) using precision, recall, and F score, we observe that for the liquefaction class, the approach of Moss et al. (2006) has the highest recall whereas the approach of Youd et al. (2001) has the highest precision. A comparison of the F-score measures shows that although Moss et al. (2006) had the best recall and Youd et al. (2001) had the highest precision, SVM has the best F-score measure for the liquefaction case using equal weights for both precision and recall. We also see a very similar result in the case of nonliquefaction instances where although Youd et al. (2001) has the highest recall and Moss et al. (2006) has the best precision, the SVM has the best F-score measure. The optimal values of c and γ for the SVM approach for the CPT data are 371.25 and 0.001 71, respectively. The reason for the SVM having a better F score than the approaches of Youd et al. (2001) and Moss et al. (2006) is SVM has a more balanced predictive performance in comparison to both liquefaction and nonliquefaction instances as well as in comparison to precision and recall. Whereas, in the case of Moss et al. (2006) the increase in F score for liquefaction is at the cost of a poor F score for nonliquefaction and vice versa in the case of Youd et al. (2001). It is noted from Tables 2 and 3 that compared the precision and recall values for the liquefaction and nonliquefaction classes of the SPT and CPT data set, the CPT nonliquefaction class has a large difference in the precision recall values in all three approaches. Oommen et al. (2010a) demonstrated that such a large difference in the precision and recall values indicates that the data set has high sampling bias and the predicted probabilities have large deviations from the actual probabilities.

In addition to analyzing the predictive capability of the Bayesian updating method (Cetin et al. 2004; Moss et al. 2006) for a TH_L=0.15, we also analyzed its predictive capability for a TH_L=0.5. The predictive performances for the TH_L of 0.15 and 0.5 are compared in Tables 2 and 3. It is observed that in the case of SPT (Table 2) and CPT (Table 3) changing the TH_L from 0.15 to 0.5 does not improve the overall predictive performance or the F-score value of liquefaction whereas the F-score value for nonliquefaction has improved. It is also observed that as the TH_L changes from 0.15 to 0.5 the recall decreases for liquefaction whereas it increases for nonliquefaction. Similarly, the precision increases for liquefaction whereas it decreases for nonliquefaction. This result is as expected and indicates that several instances of liquefaction are classified as nonliquefaction as the TH_L changes from 0.15 to 0.5.

Performance of Probabilistic Approaches

We analyzed the predictive performance of the probabilistic evaluation of liquefaction potential using ROC and P-R curves. Figs. 7 and 8 present the evaluation of the SPT-based probabilistic approaches using ROC and P-R curves, respectively. We observe both the approach of Cetin et al. (2004) and the SVM approach as having similar predictive performance with the former having slightly improved AUC for both liquefaction and nonliquefaction instances. Fig. 8 shows the P-R curve for the liquefaction case as falling closer to the (1, 1) point than for the nonliquefaction case. This indicates that both probabilistic approaches have better predictive capability for the liquefaction instances compared to the nonliquefaction instances.

Figs. 9 and 10 present the evaluation of the CPT-based probabilistic approaches using ROC and P-R curves, respectively. The ROC curves for the liquefaction and nonliquefaction instances show a similar performance to the SPT-based probabilistic approaches with the SVM having slightly lower AUC values than for the method of Moss et al. (2006). The comparison of the P-R curves (Fig. 10) indicates a better predictability of the probabilistic approaches for liquefaction compared to nonliquefaction. The difference in the predictive performance between liquefaction and nonliquefaction has increased for CPT-based data over the SPT data. This difference in the predictive performance is indicative of the sampling bias in the SPT and CPT data set. As the sampling bias increases from the SPT to CPT data set the predictive performance of the minority class decreases. This clearly indicates that model development using both the Bayesian updating method (Cetin et al. 2004; Moss et al. 2006) and SVM based approach is sensitive to the sampling bias in the data set.

Comparing the probabilistic approaches based on the SPT and CPT data sets, we conclude that considering both liquefaction and nonliquefaction instances the SPT-based probabilistic approaches have a slight advantage over the CPT-based probabilistic approaches. We hypothesize that this difference in performance is at least in part due to the larger sampling bias in the CPT data set.

Choice of the Optimal Threshold of Liquefaction

In this section we use the P-R cost curves to determine the optimal TH_L. Figs. 11 and 12 present the P-R cost curves for the SPT- and CPT-based data sets. In Figs. 11 and 12 Plots a and c repre-

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

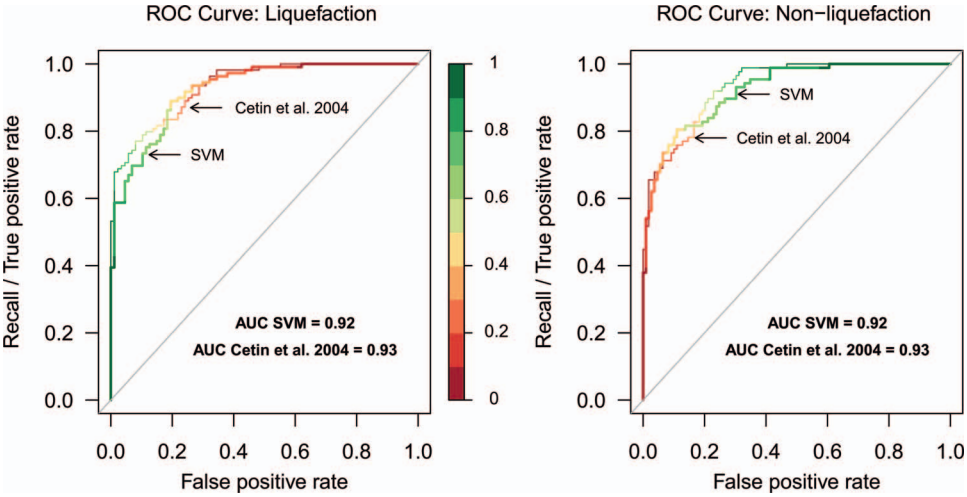

Fig. 7. (Color) ROC curve for the SVM and probabilistic approaches of Cetin et al. (2004) based on the SPT data set

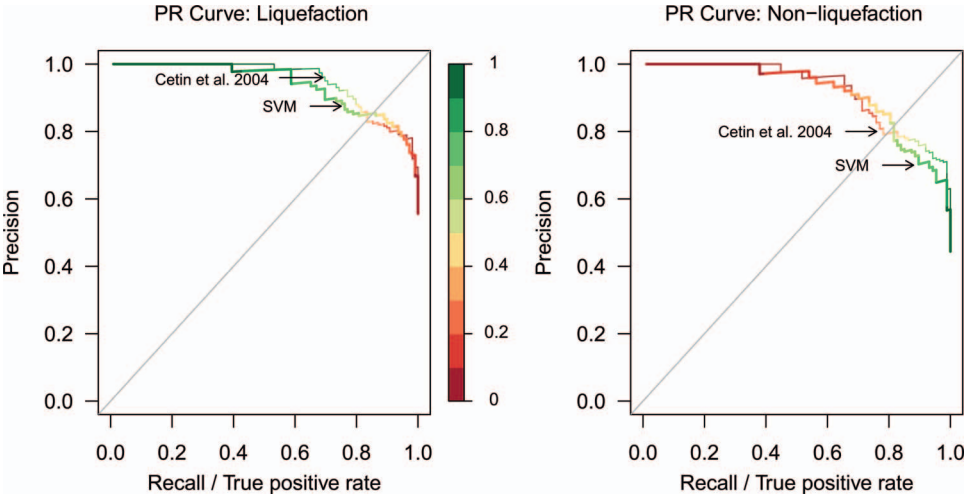

Fig. 8. (Color) P-R curve for the SVM and probabilistic approaches of Cetin et al. (2004) based on the SPT data set

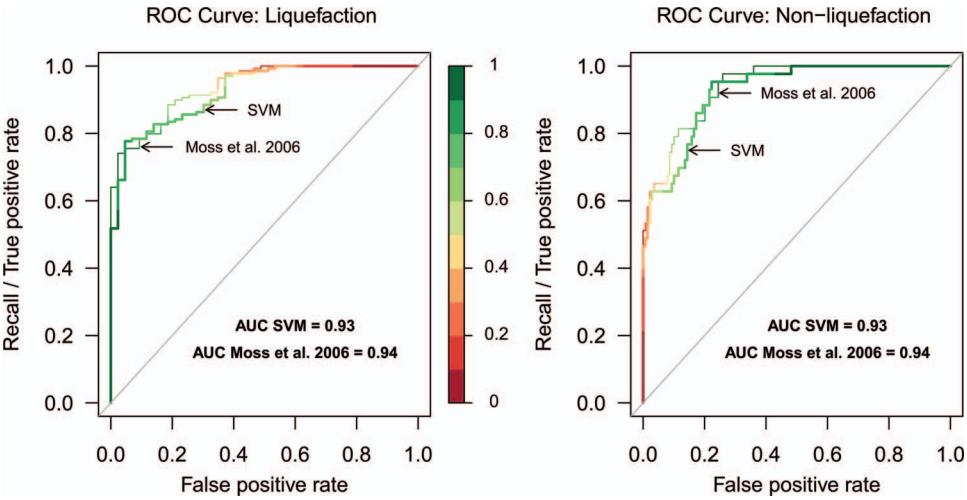

Fig. 9. (Color) ROC curve for the SVM and probabilistic approaches of Moss et al. (2006) based on the CPT data set

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

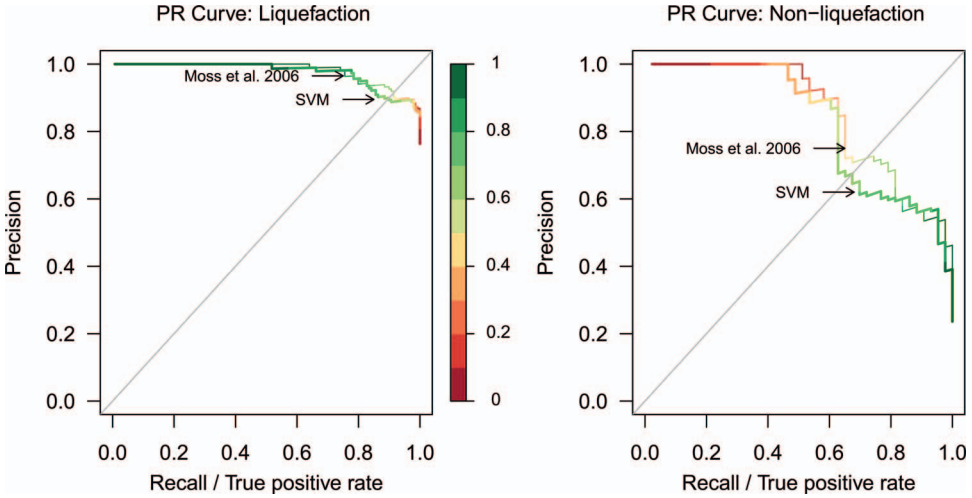

Fig. 10. (Color) P-R curve for the SVM and probabilistic approaches of Moss et al. (2006) based on the CPT data set

sent the precision and recall for the liquefaction case using the SVM and the "Bayesian updating" probabilistic approaches, respectively, whereas Plot b presents the optimal TH_L versus the ratio of the C_{FP} to the C_{FN} for a given project (CR). For the deterministic evaluation, the recommended TH_L using the Bayesian updating is 0.15 for both the SPT and CPT data sets (Cetin et al. 2004; Moss et al. 2006). In the case of SPT (Fig. 11), a TH_L of 0.15 corresponds to a $CR \approx 1$ using the approach of Cetin et al. (2004), which implies that the $C_{FN}=C_{FP}$ (cost of predicting a true liquefaction instance as nonliquefaction=cost of predicting a true nonliquefaction instance as liquefaction). In the case of CPT (Fig. 12), a TH_L of 0.15 corresponds to a $CR \approx 0.6$ using the approach of Moss et al. (2006), which implies that the $C_{FN}=0.6$ times the C_{FP}. We also observe from Fig. 11 that using any TH_L value in the range of 0.05–0.60 will have the same cost as using the 0.15 recommended by Cetin et al. (2004).

Tables 4 and 5 summarize the results illustrated in Figs. 11 and 12. We observe from Tables 4 and 5 that for the SPT data, the approach of Cetin et al. (2004) has higher precision and recall compared to the SVM for a CR range of 0.28–1.0, whereas for the CPT data, the approach of Moss et al. (2006) and SVM approach have comparable precision and recall values.

By comparing multiple liquefaction models for both SPT and CPT data with validation statistics, we have illustrated that the predictive capabilities of the three modeling approaches are comparable in general; however, each model has distinct advantages or disadvantages in terms of precision or recall for the different classes. The individual model differences in terms of precision and recall are at least in part a result of the model developer's decisions. By using model validation statistics, the models' strengths and weaknesses are more transparent to the user. Using robust and quantitative validation methods 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, given information on the costs associated with both outcomes of liquefaction and nonliquefaction.

Case Study on the Applicability of P-R Cost Curve

In the case of new projects/buildings, the geotechnical engineer faces challenges to 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 losses. Considering two hypothetical cases (H-1 and H-2), 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 cases we calculated 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 would not 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 include the time, the cost, and the business that would be lost during the time to fix the building in the event of liquefaction. In the cases of H-1 and H-2, 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's for Cases H-1 and H-2 are 0.30 and 0.40, respectively, calculated using the Bayesian updating method (Moss et al. 2006) with CPT data. From Fig. 12(b) or Table 5 we observe that the optimal threshold for $CR=0.7$ using the 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.

Since H-1 has P_L value of less than 0.308, it is classified as nonliquefiable, and since H-2 has P_L value greater than 0.308, it is classified as liquefiable. The P-R curve [Fig. 12(c)] helps us to determine how confident we can be with this decision that they are liquefiable and nonliquefiable. We observe from Fig. 12(c) that the precision and recall values corresponding to the P_L for each case are H-1 (precision=0.894, recall=0.978) and H-2 (precision=0.897, recall=0.942). Recall gives the chance that concluding the site will not liquefy is wrong. Moreover, precision gives the chance that concluding the site will liquefy is wrong. In the case of H-1, a recall=0.978 means that there is 2.2% chance for the decision that the site will not liquefy is wrong. Whereas, in the case of H-2, a precision=0.89 means that there is 11% chance that concluding the site will liquefy is wrong.

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

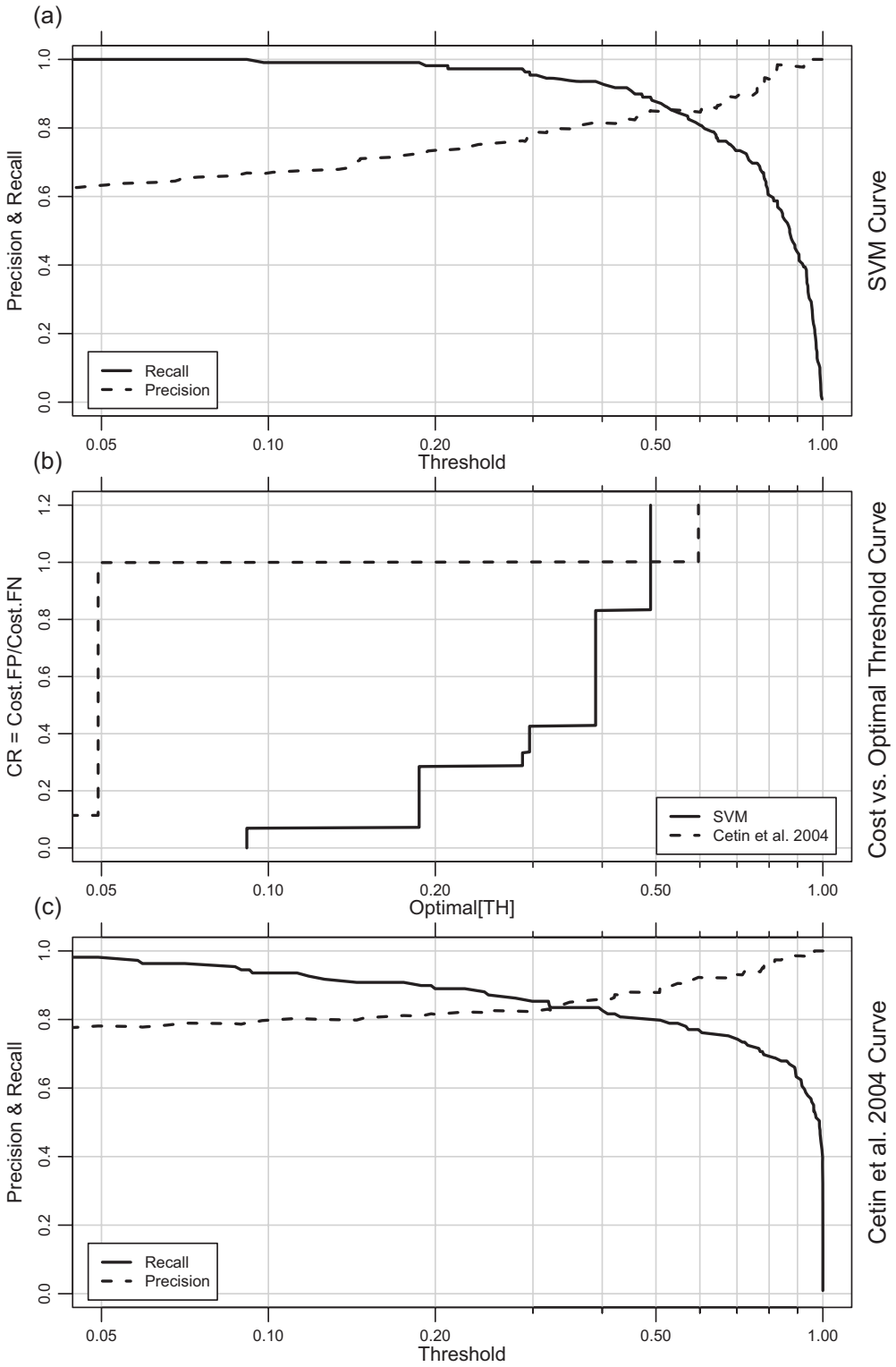

Fig. 11. P-R cost curve for the SVM and probabilistic approaches of Cetin et al. (2004) based on the SPT data set

Database Limitations

The validation analyses indicated that the sampling bias is higher in the CPT data set compared to the SPT data set and it can have a significant impact on model performance. Therefore, future data collection efforts should be focused on reducing the sampling bias by developing a more representative sample of the actual population. Using the SVM model, one can also more closely examine the model and data space to identify other gaps in the data set. Identifying gaps in data sets is extremely important for improving empirical models because such gaps essentially amount in practice to extrapolation of an empirical model. The conventional empirical approaches use the entire training data for the development of the model, whereas the SVM uses a subset of the training data known as support vectors (as shown in Fig. 5). Therefore, identi-

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

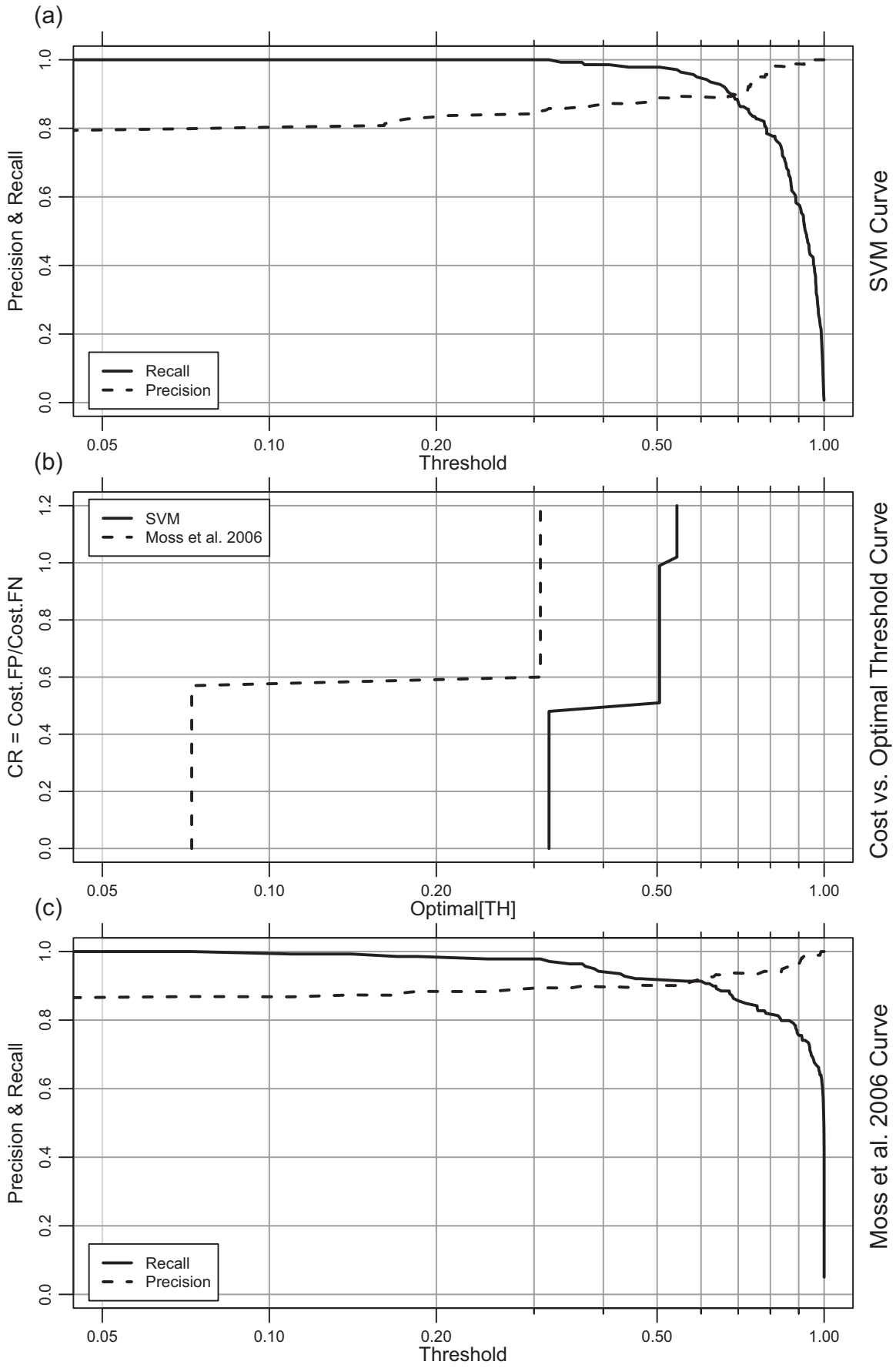

Fig. 12. P-R cost curve for the SVM and probabilistic approaches of Moss et al. (2006) based on the CPT data set

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

Table 4. P-R Cost Curve Summarized for the Approach of Cetin et al. (2004) and SVM Approach Based on the SPT Data Set

	Cetin et al. 2004		
Cost ratio (CR) range	Optimal threshold	Precision	Recall
$0<CR<0.11$	0.002	0.692	0.99
$0.11<CR<1.0$	0.049	0.781	0.981
>1.0	0.596	0.923	0.77
	SVM		
$0<CR<0.06$	0.091	0.668	1
$0.06<CR<0.28$	0.187	0.729	0.99
$0.28<CR<0.33$	0.287	0.762	0.972
$0.33<CR<0.42$	0.295	0.777	0.963
$0.42<CR<0.83$	0.389	0.816	0.935
$CR>0.83$	0.488	0.85	0.889

fying the characteristics of these support vectors a priori can improve the design associated with data collection efforts and in turn result in improvements in the resulting empirical model using SVM.

In the SVM model for classification, the support vectors define the shape of the hyperplane that separates the two classes (liquefaction and nonliquefaction) and they are the instances that fall along the maximum margin of the optimal hyperplane and the instances that are misclassified (Fig. 5) during the training phase of the model development. The instances that are not support vectors do not contribute to the model development. Support vectors being points on the maximum margin closest to the hyperplane and instances that are misclassified, thus support vectors have the highest uncertainty in terms of the class they belong to.

Table 5. P-R Cost Curve Summarized for the Approach of Moss et al. (2006) and SVM Approach Based on the CPT Data Set

	Moss et al. 2006		
Cost ratio (CR) range	Optimal threshold	Precision	Recall
$0<CR<0.6$	0.072	0.868	1
$CR>0.6$	0.308	0.894	0.978
	SVM		
$0<CR<0.49$	0.319	0.858	1
$0.49<CR<0.99$	0.505	0.888	0.978
$CR>0.99$	0.542	0.894	0.971

Having the highest uncertainty, the region close to the support vectors requires further data collection to better constrain the empirical model. Here, we have identified such regions by analyzing the range of the predictor variables and the quantity of support vectors in each of these ranges. One can divide the predictor variable into different ranges based on equal intervals or based on their probability distribution. Ideally, one would like to have a low percentage of support vectors from each range of the predictor variable (i.e., indication that there is low uncertainty in these ranges). The specific range in the predictor variable that needs future data collection will be the one that has the highest percentage of support vector or the one with no data at all.

The five splits of the K-fold cross validation of the SPT data had percentages of support vectors varying from 46 to 50%. Fig. 13 presents the distribution of the support vectors in the range of the predictor variables for the training instances in split-1 of the SPT data set. Split-1 has 156 training instances, of which 73 are support vectors and the remaining 83 are nonsupport vectors (47% support vectors). The range of the predictor variables is divided into equal intervals of 10. Fig. 13 illustrates how the model uses the data and identifies regions of data space that are not well constrained by the SVM model. We observe that of the predictor variables, $(N_1)_{60}$ has the highest uncertainty in the range of 15–28, CSR in the range of 0.05–0.17, M_w in the range of 6.5–7.5, and a_{max} in the range of 0.01–0.2. In general, the lower ranges of the predictor variables had higher uncertainty compared to the higher ranges.

In the case of CPT data, the percentage of support vectors for the five splits of the K-fold cross validation varied from 29 to 32%. Fig. 14 presents the distribution of the support vectors in the range of the predictor variables for the training instances in split-1 of the CPT data set. Split-1 has 145 training instances, of which 46 are support vectors and the remaining 99 are nonsupport vectors (31% support vectors). Similar to the SPT data set, the range of the predictor variables of the CPT data set was divided into equal intervals of 10. Using Fig. 14, we can see regions of the CPT model space that are not well constrained and require additional data collection. We observe that of the predictor variables CSR has the least uncertainty and lowest percentage of support vectors from the entire range of the data, whereas R_f has the highest uncertainty. The specific ranges of the predictor variables that need higher priority for data collection are values from 5.7 to 8.2 for qc1, <0.19 for CSR, 5.9–6.5 for M_w, >118 for σ'_{vo}, and >2.3 for R_f.

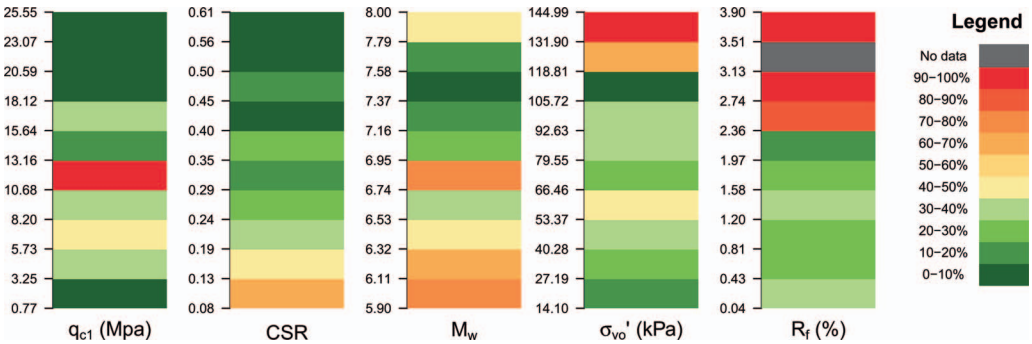

Fig. 13. (Color) Percentage of support vectors in the range of the predictor variables for the liquefaction evaluation based on the SPT data set. y-axis represents the range of the predictor variable and the colors represent the percent of support vector. Low percentage of support vector specifies that of all the instances available from the particular region only a few instances are support vectors, which in turn indicates that there is less error and uncertainty in this region of the predictor variable.

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

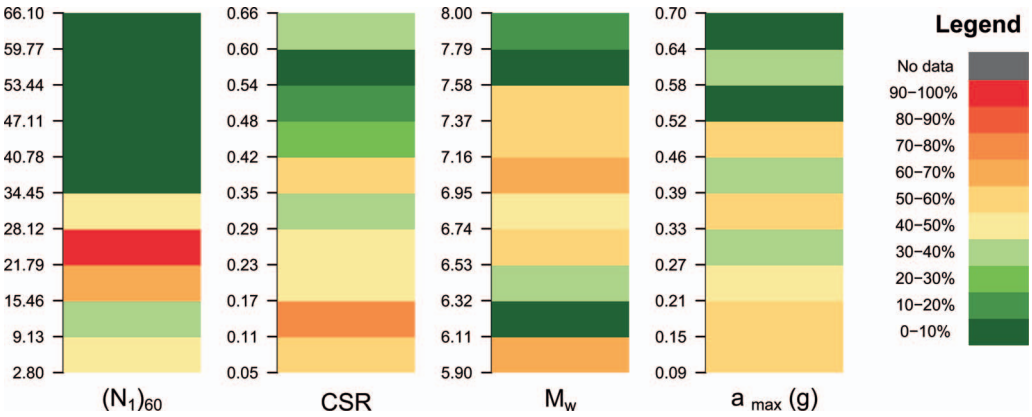

Fig. 14. (Color) Percentage of support vectors in the range of the predictor variables for the liquefaction evaluation based on the CPT data set. *y*-axis represents the range of the predictor variable and the colors represent the percent of support vector. Gray indicates lack of data from that region.

A comparison of the percentage of support vectors in the training phase of the SPT (47%) and CPT (31%) data set shows that the CPT data set provides better coverage or support for the model. This is also substantiated with the higher OA, precision, recall, and F score for liquefaction class with the approaches based on CPT data compared to the SPT data (Tables 2 and 3). However, the approaches based on the SPT data have higher precision, recall, and F score for nonliquefaction class due to the lower sampling bias. Therefore, data collection to improve the CPT-based approaches should emphasize reducing the sampling bias. Data collection for SPT-based approaches should try to fill the identified data gaps.

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, (2) the Bayesian updating method, and (3) the SVM models, whereas for the probabilistic ELMs we compared the (1) Bayesian updating method and (2) SVM. We also presented a new optimization criterion 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 and artificial intelligence 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 nonliquefaction 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 nonliquefaction class.
- Based on the F score and OA, the SVM approach has a slightly better predictive capability than the simplified procedure and the Bayesian updating method for the deterministic evaluation of both SPT and CPT data.
- The probabilistic evaluation of the liquefaction potential indicates comparable performance for both SVM and Bayesian updating method 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 either based on the Bayesian updating method (Cetin et al. 2004; Moss et al. 2006) or the SVM approach based on the SPT or CPT data.

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. It is in such future data collection efforts that the use of support vectors may find particular value. Sampling bias in both the SPT and CPT data sets results in a difference in predictive performance for the three approaches between the liquefaction and nonliquefaction classes. Comparing the P-R and ROC curves for the different classes of the SPT and CPT data, it is evident that the P-R curves are sensitive to the sampling bias within the data set, whereas the ROC curves are not. Further data collection efforts should aim to reduce such sampling bias. In addition, support vectors can improve our understanding of the ranges in data that tend to result in a high degree of prediction uncertainty. Thus, support vectors can be useful for the design of future data collection efforts.

Acknowledgments

The work of the first and second writers is funded by National Science Foundation Grant No. CMMI-0547190. This financial

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

support is greatly appreciated. The writers also gratefully acknowledge the editorial comments and suggestions to the first draft of this paper provided by Dr. Robb E. S. Moss of California Polytechnic State University, Dr. Robert Kayen of the United States Geologic Survey, Dr. Eric M. Thompson of Tufts University, and reviewers at JGGE.

References

Cetin, K. O., et al. (2004). "Standard penetration test-based probabilistic and deterministic assessment of seismic soil liquefaction potential." *J. Geotech. Geoenviron. Eng.*, 130(12), 1314–1340.

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

Cortes, C., and Vapnik, V. (1995). "Support vector networks." *Mach. Learn.*, 20(3), 273–297.

Goh, A. T. C. (1994). "Seismic liquefaction potential assessed by neural networks." *J. Geotech. Engrg.*, 120(9), 1467–1480.

Goh, A. T. C. (2002). "Probabilistic neural network for evaluating seismic liquefaction potential." *Can. Geotech. J.*, 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." *Comput. Geotech.*, 34, 410–421.

Hanna, A. M., Ural, D., and Saygili, G. (2007). "Evaluation of liquefaction potential of soil deposits using artificial neural networks." *Eng. Comput.*, 24(1), 5–16.

Hashash, Y. M. A. (2007). "Special issue on biologically inspired and other novel computing techniques in geomechanics." *Comput. Geotech.*, 34(5), 329–329.

Hawkins, D. M. (2004). "The problem of overfitting." *J. Chem. Inf. Comput. Sci.*, 44(1), 1–12.

Holzer, T. L., Bennett, M. J., Noce, T. E., Padovani, A. C., and Tinsley, J. C., III. (2006). "Liquefaction hazard mapping with LPI in the greater Oakland, California, area." *Earthquake Spectra*, 22(3), 693–708.

Hsu, C. W., Chang, C. C., and Lin, C. J. (2003). *A practical guide to support vector classification*, Vol. 12, National Taiwan University, Taipei, Taiwan.

Idriss, I. M., and Boulanger, R. W. (2006). "Semi-empirical procedures for evaluating liquefaction potential during earthquakes." *Soil Dyn. Earthquake Eng.*, 26, 115–130.

Iwasaki, T., Tatsuoka, F., Tokida, K. I., and Yasuda, S. (1978). "A practical method for assessing soil liquefaction potential based on case studies at various sites in Japan." *Proc., 2nd Int. Conf. on Microzonation for Safer Construction—Research and Application*, Vol. II, 885–896.

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." *J. Geotech. Geoenviron. Eng.*, 128(7), 580–589.

Juang, C. H., and Li, D. K. (2007). "Assessment of liquefaction hazards in Charleston quadrangle, South Carolina." *Eng. Geol. (Amsterdam)*, 92(1–2), 59–72.

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." *J. Geotech. Geoenviron. Eng.*, 134(8), 1050–1058.

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." *J. Geotech. Geoenviron. Eng.*, 129(1), 66–80.

Karatzoglou, A., Meyer, D., and Hornik, K. (2006). "Support vector machines in R." *J. Stat. Software*, 15(9).

Keerthi, S. S., and Lin, C. J. (2003). "Asymptotic behaviors of support vector machines with Gaussian kernel." *Neural Comput.*, 15(7), 1667–1689.

Krawinkler, H. (2004). "Exercising performance-based earthquake engineering." *Proc., 3rd Int. Conf. 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." *J. Geotech. Geoenviron. Eng.*, 130(12), 1271–1282.

Lee, D.-H., Ku, C.-S., and Yuan, H. (2003). "A study of the liquefaction risk potential at Yuanlin, Taiwan." *Eng. Geol. (Amsterdam)*, 71(1–2), 97–117.

Liao, S. S. C., Veneziano, D., and Whitman, R. V. (1988). "Regression-models for evaluating liquefaction probability." *J. Geotech. Engrg.*, 114(4), 389–411.

Lin, H. T., Lin, C. J., and Weng, R. C. (2007). "A note on Platt's probabilistic outputs for support vector machines." *Mach. Learn.*, 68(3), 267–276.

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." *J. Geotech. Geoenviron. Eng.*, 132(8), 1032–1051.

Oommen, T., and Baise, L. G. (2010). "Model development and validation for intelligent data collection for lateral spread displacements." *J. Comput. Civ. Eng.*, 24(6), 467–477.

Oommen, T., Baise, L. G., and Vogel, R. (2010a). "Sampling bias and class imbalance in maximum likelihood logistic regression." *Mathematical Geosciences*, in press.

Oommen, T., Misra, D., Twarakavi, N. K. C., Prakash, A., Sahoo, B., and Bandopadhyay, S. (2008). "An objective analysis of support vector machine based classification for remote sensing." *Mathematical Geosciences*, 40(4), 409–424.

Pal, M. (2006). "Support vector machines-based modelling of seismic liquefaction potential." *Int. J. Numer. Analyt. Meth. Geomech.*, 30(10), 983–996.

Papathanassiou, G., Pavlides, S., and Ganas, A. (2005). "The 2003 Lefkada earthquake: Field observations and preliminary microzonation map based on liquefaction potential index for the town of Lefkada." *Eng. Geol. (Amsterdam)*, 82(1), 12–31.

Platt, J. (2000). "Probabilistic outputs for support vector machines and comparison to regularised likelihood methods." *Advances in large margin classifiers*, A. Smola, B. Bartlett, B. Scholkopf, and D. Schuurmans, eds., Cambridge, Mass.

R Development Core Team. (2009). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Robertson, P. K., and Campanella, R. G. (1985). "Liquefaction potential of sands using the CPT." *J. Geotech. Engrg.*, 111(3), 384–403.

Robertson, P. K., and Wride, C. E. (1998). "Evaluating cyclic liquefaction potential using the cone penetration test." *Can. Geotech. J.*, 35(3), 442–459.

Sahoo, B. C., Oommen, T., Misra, D., and Newby, G. (2007). "Using the one-dimensional S-transform as a discrimination tool in classification of hyperspectral images." *Can. J. Remote Sens.*, 33(6), 551–560.

Samui, P., Sitharam, T. G., and Kurup, P. U. (2008). "OCR prediction using support vector machine based on piezocone data." *J. Geotech. Geoenviron. Eng.*, 134(6), 894–898.

Scholkopf, B., and Smola, A. (2002). *Learning with kernels*, MIT Press, Cambridge, Mass.

Seed, H. B., and De Alba, P. (1986). *Use of SPT and CPT tests for evaluating the liquefaction resistance of sands*, Blacksburg, Va., 281–302.

Seed, H. B., and Idriss, I. M. (1971). "Simplified procedure for evaluating soil liquefaction potential." 97(SM9), 1249–1273.

Seed, H. B., Idriss, I. M., and Arango, I. (1983). "Evaluation of liquefaction potential using field performance data." *J. Geotech. Engrg.*, 109(3), 458–482.

Seed, H. B., Tokimatsu, K., Harder, L. F., and Chung, R. M. (1985). "Influence of SPT procedures in soil liquefaction resistance evaluations." *J. Geotech. Engrg.*, 111(12), 1425–1445.

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org

Shibata, T., and Teparaksa, W. (1988). "Evaluation of liquefaction potentials of soils using cone penetration tests." *Soils Found.*, 28(2), 49–60.

Sonmez, H. (2003). "Modification of the liquefaction potential index and liquefaction susceptibility mapping for a liquefaction-prone area (Inegol, Turkey)." *Environ. Geol.*, 44(7), 862–871.

Sonmez, H., and Gokceoglu, C. (2005). "A liquefaction severity index suggested for engineering practice." *Environ. Geol.*, 48(1), 81–91.

Stark, T. D., and Olson, S. M. (1995). "Liquefaction resistance using CPT and field case-histories." *J. Geotech. Engrg.*, 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.

Vapnik, V. (1995). *The nature of statistical learning theory*, Springer-Verlag, New York.

Youd, T. L., et al. (2001). "Liquefaction resistance of soils: Summary report from the 1996 NCEER and 1998 NCEER/NSF Workshops on evaluation of liquefaction resistance of soils." *J. Geotech. Geoenviron. Eng.*, 127(10), 817–833.

Youd, T. L., and Noble, S. K. (1997). "Liquefaction criteria based on statistical and probabilistic analyses." *Proc., NCEER Workshop on Evaluation of Liquefaction Resistance of Soils*, *NCEER Technical Rep. No NCEER-97-0022*, 201–205.

Downloaded 16 Jun 2011 to 130.64.82.126. Redistribution subject to ASCE license or copyright. Visit http://www.ascelibrary.org