Evaluating model uncertainty of a CPT-based model for earthquake-induced soil liquefaction

Y.F. Lee
Department of Leisure Recreation and Travel Management, Toko University, Pu Tzu, Chiayi, Taiwan

Y.Y. Chi
Department of Land Management and Development, Chang Jung Christian University, Kway Jen, Tainan, Taiwan

C.H. Juang
Department of Civil Engineering, Clemson University, Clemson, SC, USA

ABSTRACT: To account for uncertainties in the environmental parameters and model that are required in the calculation of soil liquefaction potential, the probabilistic approaches may be employed. In this paper a new CPT-based Liquefaction Triggering Model (LTM) is arranged using field performance case histories of Moss et al (2006). The model uncertainty, \( \mu_{c1} \) and COV\((c_1)\) of the LMT is determined by the sampling process taking proportional cases from the group of liquefied cases and the group of non-liquefied cases. However, the probability of liquefaction is calculated considering uncertainties of model and parameters by the Hasofer–Lind reliability index and the KCP (Lee et al, 2007). It is principally when evaluating probability of liquefaction that all (the model and parameter) uncertainty considered differs significantly from earlier and widely used only parameter uncertainty considered.

1 INTRODUCTION

The earthquake-induced liquefaction potential evaluation is one of the important topics in geotechnical engineering. In practice, the engineer most often employs simplified procedure to evaluate liquefaction potential of a soil. The simplified procedure was created by Seed and Idriss (1971) in which the seismic loading required to initiate liquefaction was expressed as a Cyclic Stress Ratio (CSR) and the soil resistance against liquefaction (CRR) was measured by the energy-corrected blow count \( (N_{60}) \) from the Standard Penetration Test (SPT). Following this general framework, many simplified models were later developed based on different in situ tests such as the cone penetration test (CPT), Becker Penetration Test (BPT), and shear wave velocity (Vs) measurements. Youd et al. (2001) documented an excellent summary of the latest versions of the SPT-, CPT-, and Vs-based models. The factor of safety \( (F_S) \), defined as \( F_S = \frac{CRR}{CSR} \), is most often employed to express the liquefaction potential of a soil. It is essential to characterize, even empirically, the uncertainty (or bias) of these simplified models because these models were primarily a result of calibration with field observations, and because some degree, albeit unknown, of conservatism was typically adopted in the development of these models.

To account for uncertainties in the seismic, soil parameters and model that are required in the calculation of CRR and CSR, the probabilistic approaches such as those described by Liao et al. (1988), Juang et al. (2000), Cetin et al. (2004), Moss et al. (2006), Juang et al. (2006), Jha and Suzuki (2009a,b) and Lee et al (2007, 2010) may be employed. The most part of these approaches, however, yield the probabilities of liquefaction for considering the uncertainties of parameters. Juang et al. (2006) suggested that this model uncertainty can be characterized with a single model bias factor (or model factor for short) that is applied to CRR (or the traditional “boundary curve” that separates liquefied cases from nonliquefied cases). Therefore, the \( F_S \) in a deterministic evaluation of liquefaction potential should be rewritten as:

\[
F_S = c_1 \times \frac{CRR}{CSR} = c_1 \times \frac{f(\bar{x})}{f(\bar{y})}
\]  

(1)

where \( c_1 \) is the model factor of the adopted simplified model, CRR is the cyclic resistance ratio composed of soil parameters \( \bar{x} = (x_1, x_2, x_3, ..., x_n) \)
and CSR is the cyclic stress ratio composed of earthquake parameters \( \varepsilon = (y_1, y_2, y_3, ..., y_k) \). The adjustment with a model factor is seen herein as a way to compensate the unknown degree of conservatism in the adopted deterministic model so that it can return to the “un-biased state.”

It should be noted that from the perspective of a deterministic evaluation, a conservative CRR model, which would lead to a smaller computed factor of safety, may be preferred by the engineer for design purposes. However, this paper focuses on the evaluation of the model uncertainty of the adopted deterministic model for the “un-biased” estimate of the probability of liquefaction. Thus, the task of estimating the model uncertainty of a given model is essential to determine the model factor \( c_1 \). Specifically, the mean value and the Coefficient of Variation (COV) of the model factor \( c_1 \), denoted as \( \mu_1 \) and COV(\( c_1 \)), need to be determined for a model.

Using randomly selected samples of the database, Lee et al (2010) characterized the model factor \( c_1 \) based on the analysis of the accuracy of the Seed model (SPT-based, Youd et al, 2001) and RW model (CPT-based, Youd et al, 2001). Their process yields \( \mu_1 = 1.06 \) and COV(\( c_1 \)) = 0.06 for the Seed model and \( \mu_1 = 1.16 \) and COV(\( c_1 \)) = 0.12 for the RW model (Robertson and Wride, 1998). Moss et al (2006) selected 188 CPT-based liquefied/nonliquefied case histories from over 500 possible case histories. They estimated the probability of liquefaction conditional on parameters uncertainty by a Bayesian framework. In this paper, using the procedure of Lee et al (2010), a CPT-based simplified model based on Moss database was developed as an example, the model of \( \mu_1 \) and COV(\( c_1 \)) is determined. Additionally, the probability of liquefaction was calculated considering uncertainties of model and parameters by the Hasofer–Lind reliability index (Hasofer and Lind, 1974; Ditlevsen, 1981). The results are compared to Moss et al (2006).

2 METHODOLOGY

Lee et al (2010) described the performance criterion is based on the concept that the “most accurate and precise” \( c_1 \) should be the one that produces the highest success rate for predicting the occurrence of liquefaction/no liquefaction among all cases in the database. Here, a “success” is recorded for each liquefied case if the \( F_\text{s} \) computed with Eq. (1) is less than 1; likewise, for each non-liquefied case, a “success” is recorded if the computed \( F_\text{s} \) is greater than 1. Thus, a counter can be set up such that an incorrect prediction yields 1 while a correct (successful) prediction yields 0. After all cases in the database have been evaluated using a given simplified model, the counter shows the total number of incorrect predictions. Thus, the total error in the prediction based on a given model, denoted as \( \varepsilon(\mu_1) \), is simply the error count. If the “population” of all cases is given, the most accurate and precise \( c_1 \) can be determined by minimizing \( \varepsilon(\mu_1) \) with respect to \( c_1 \). In summary, the performance criterion is to achieve the highest success rate as realized by a minimization of \( \varepsilon(\mu_1) \).

For simplicity, \( c_1 \) is assumed to be a real number with two decimal digits, ranging from 0.50 to 2.50. Then, the total error \( \varepsilon(\mu_1) \) is computed for each \( c_1 \) in this range using all cases in the database, and the optimized \( c_1 \) is determined by finding the minimum \( \varepsilon(\mu_1) \) value. To consider the effect of the size and composition of a database (i.e., the effect of sampling), it is decided to create and evaluate different samples of the database. In the adopted database of case histories, the ratio of the number of liquefied cases to the number of nonliquefied cases is determined. This ratio is maintained in the sampling process by taking proportional cases from the group of liquefied cases and the group of non-liquefied cases.

To randomize the sampling process, the sample size is randomly selected from a range defined by a prescribed minimum percentage of the database and the full database. For example, if the minimum percentage is set at 30%, then the sample size will be between 30% and 100% of the database. In other words, the sample will consist of at least 56 cases (30% of the database of 188 cases) but no more than 188 cases (100% of the database). In this paper, the actual sample size is randomly selected from the range of 56 to 188 (a uniform distribution is conveniently assumed here). Once a sample size is selected, the actual samples are created by randomly selecting the intended number of cases from the database; for example, if the sample size is 56, then by randomly selecting 56 cases from the set of 188 cases would yield a sufficiently large number of samples with different compositions. However, the results of a sensitivity analysis show that there is practically no change in the computed model factor when the number of samples has reached the level of 5,000.

The above process for the simplified model is repeated with different chosen minimum percentages, such as 40%, 50%, ..., and 90%, in lieu of 30%. It should be noted that a smaller minimum percentage (say, less than 20%) is not feasible for creating a suitable sample given the size of the adopted database. The results of these analyses show that practically, there is no significant change in the resulting \( \mu_1 \) value; however, COV(\( c_1 \)) increases as the minimum percentage decreases. The trend line may be extrapolated to cross the ordinate suggests that the largest variation COV(\( c_1 \)) where the sample size ranges from 0 to 100% of the database.
The Hasofer–Lind Reliability Index (RI) is defined as (Hasofer and Lind, 1974; Ditlevsen, 1981):

\[
RI = \min_{x \in \Psi} \sqrt{\sum_{i=1}^{n} \left( \frac{Z_i - \mu_i}{\sigma_i} \right)^2} \left[ R^T \right]^{-1} \left[ Z_i - \mu_i \right] \]

(2a)

where \( RI \) = reliability index; \( Z_i \) = the random variables, herein including \( \tilde{x}, \tilde{y} \) and \( \tilde{c}_i \); \( \mu_i \) = the mean values matrix of \( Z_i \); \( R \) = the correlation matrix of \( Z_i \); \( \Psi \) = the failure region (corresponding to the domain in which factor of safety, \( F_S \leq 1 \)). Determination of the reliability index RI is basically an optimization problem. Many solutions are available (Chowdhury and Xu, 1995; Low, 1997; Lee et al., 2007) and in this paper, the knowledge-based Cluster Portioning (KCP) approach (Lee et al., 2007) is adopted for its efficiency for conducting hundreds of reliability analyses at once. By the RI, the probability of liquefaction \( (p_r) \) can be calculated as:

\[
p_f = 1 - \Phi (RI), \]

(2b)

\[
\Phi (RI) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{RI} \exp \left( -\frac{1}{2} \xi^2 \right) d\xi,
\]

(2c)

where \( \Phi \) is the cumulative standard normal distribution function.

### 3 CPT-BASED LIQUEFACTION TRIGGERING MODEL

Moss et al (2006) estimated the conditional probability of liquefaction on parameters uncertainty by a Bayesian framework and 188 selected CPT-based liquefied/nonliquefied case histories. The equations can be expressed as follows:

\[
p_f = \Phi \left[ \frac{\left( q_{cl}^{1.045} + 0.11 \times q_{cl} \times R_f + 0.001 \times R_f + c \times (1 + 0.85 \times R_f) \right)}{1.632} \times \ln(CSR^*) - 0.848 \times \ln(M_w) - 0.002 \times \ln(\sigma_v) - 20.923 - 7.177 \right]
\]

(3)

\[
CRR_{p_f=0.5} = CSR^* = \exp \left[ \frac{\left( q_{cl}^{1.045} + 0.11 \times q_{cl} \times R_f + 0.001 \times R_f + c \times (1 + 0.85 \times R_f) \right)}{-0.848 \times \ln(M_w) - 0.002 \times \ln(\sigma_v) - 20.923} - 7.177 \right]
\]

(4)

\[
CSR^*_{M_w=7.5} = \frac{0.65 \times \frac{a_{\text{max}}}{g} \times \sigma_v}{1.78 \times M_w^{1.43}}
\]

(5)

results, a Liquefaction Triggering Model (LTM) of simplified model can be regressed as follows:

for \( R_f \leq 2\% \).
the mean value and standard deviation associated with the various environmental parameters. There are 188 CPT-based liquefied/nonliquefied case histories selected and quantified. An example of the distribution of coefficient of variation (COV) is shown as Figure 2. The mean value and standard deviation of the quasi-normal distribution of COV($q_c$) is equal to 0.13 and 0.046. This indicates that the variability of COV($q_c$) exits, however, is about 35% of the mean. In a similar manner, the mean value of COV($\mu_w$) = 0.016, COV($a_{\text{max}}$) = 0.162, COV($\sigma_v$) = COV($\sigma_{v'}$) = 0.08, COV($R_f$) = 0.173. The standard deviation of $\mu_w$, $a_{\text{max}}$, $\sigma_v$, $\sigma_{v'}$, and $R_f$ is equal to 0.003, 0.048, 0.04, 0.03, and 0.046, respectively.

5 QUANTIFY MODEL UNCERTAINTY FOR LTM

The above process for the LTM model is repeated with different chosen minimum percentages, such as $S = 40\%$, $50\%$, $\ldots$, and $90\%$, in lieu of $30\%$. An example of the distribution of the frequency density of $c_1$ for $S = 30\%$ is shown as Figure 3. The results of these analyses show that the $\mu_{c_1}$ value range from 0.79 to 0.85 and average value of $\mu_{c_1} = 0.81$; however, COV($c_1$) increases as the minimum percentage decreases, as shown in Figure 4. The trend shown in Figure 4 suggests that the largest variation can be extrapolated to be COV($c_1$) = 0.17 where the sample size ranges from 0 to 100% of the database. In summary, the model factor of the LTM model is characterized herein based on the analysis of the accuracy of the LTM model using randomly selected samples of the database. This is a data-driven characterization of

$$CRR_{p_f}=0.5 = \left(\frac{q_{c1}}{24.41}\right)^{1.878} + 0.054, \text{ for } q_{c1} \leq 5 \text{ Mpa},$$
$$R^2 = 0.92.$$  \hspace{1cm} (8a)

$$CRR_{p_f}=0.5 = \left(\frac{q_{c1}}{20.05}\right)^{2.687} + 0.080, \text{ for } q_{c1} > 5 \text{ Mpa},$$
$$R^2 = 0.95.$$  \hspace{1cm} (8b)

for $R_f > 2\%$,

$$CRR_{p_f}=0.5 = 5.12 \times \left(\frac{q_{c1}}{30}\right)^{2.649} + 0.069, \ R^2 = 0.97.$$  \hspace{1cm} (8c)
the model factor, and the process yields $\mu_{c_1} = 0.81$ and $\text{COV}(c_1) = 0.17$ for the LTM model.

6 PROBABILITY OF LIQUEFACTION FOR CASE HISTORIES

Using the proposed LTM (Eq. (8)) and the reliability index $RI$ (Eq. (2)), the probability of liquefaction is calculated for case histories by the KCP (Lee et al., 2007). It is noted that the considering uncertainties of $\text{COV}(M_w), \text{COV}(a_{\text{max}}), \text{COV}(\sigma_v), \text{COV}(\sigma_v'), \text{COV}(q_c)$, and $\text{COV}(R_f)$ is equal to 0.016, 0.162, 0.08, 0.08, 0.130, and 0.173, respectively. Under the environmental parameter uncertainty, there are two scenarios of $\mu(c_1)$ and $\text{COV}(c_1)$ in this paper. The first scenario of $\mu(c_1)$ and $\text{COV}(c_1)$ is equal to 1 and 0, respectively. That indicates only environmental parameters uncertainty is considered, the model uncertainty is not included. The first scenario is similar to Moss et al. (2006).

The second scenario of $\mu(c_1)$ and $\text{COV}(c_1)$ is equal 0.81 and 0.17, respectively. As mentioned above, that includes the results of the LTM model uncertainty analysis. Herein, the calculated safety of factor by the mean values of parameter is defined as the central safety factor ($CFS = c_1 \times CRR/CSR$). The regression function of $p_f$ and $CFS$ can be presented as follows:

$$p_f = \frac{1}{1 + CFS^a} \quad (10)$$

where $a$ is the coefficient of regression, decreases as the total uncertainty increase (bigger COVs). Under two scenarios we obtained the regression results shown as Figure 5. It is noted that the abscissa is $CFS' = CRR/CSR$ in Figure 5. Based on the proposed LTM, the coefficient of regression is 6.954 ($R^2 = 0.99$) and 5.647 ($R^2 = 0.99$) under first and second scenario, respectively. The solid circles and hollow circles are presented the probability of liquefaction by Moss et al. (2006) (Eq. (3)). They are similar to the $p_f$ under the first scenario curve, distinctly. The results verified the proposed LTM. It can be seen that the $p_f$ by second scenario is higher than the $p_f$ by the first scenario under the same $CFS'$. In the same, the $CFS'$ by second scenario is higher than the $CFS'$ by the first scenario under the same $p_f$. For example, under $p_f = 0.5$, the $CFS'$ is equal to 1.23 and 1.0 for the second scenario and the first scenario, respectively.

The comparison of the results by Moss formula (Eq. (3)) and results by the proposed model for the second scenario is shown as Figure 6. During lower probability $p_f$, liquefaction ($p_f \leq 30\%$), the probability...
REFERENCES


7 SUMMARY

In this paper we present a new procedure for CPT-based assessments of earthquake induced soil liquefaction hazard. A new CPT-based liquefaction triggering model is arranged using a larger database of high quality field performance case histories by Moss et al (2006). The model uncertainty of μ, and COV(c) of the proposed CPT-based liquefaction triggering model is determined. Additionally, the probability of liquefaction is calculated considering uncertainties of model and parameters by the Hasofer–Lind reliability index and the KCP.

We have utilized new procedures for CPT case histories as presented in Moss et al. (2006). Overall, the results of the first scenario are in general agreement with the mean value of the probability of liquefaction by Moss equations. It is principally when evaluating probability of liquefaction that all (the model and parameter) uncertainty considered (as the second scenario) differs significantly from earlier and widely used only parameter uncertainty considered. To further study the variation of the model factor c, caused by sampling difference, a “sufficiently large” number of samples of different sizes and compositions are created and analyzed. We believe that the proposed process accurately capture the entire range of potentially liquefiable materials that the CPT can be used to measure.