12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP12
Vancouver, Canada, July 12-15, 2015

Bayesian Methods and Liquefaction

John T. Christian
Consulting Engineer, 36E Seven Springs Lane, Burlington, MA 01803, USA

Gregory B. Baecher
Prof., Dept. of Civil and Env. Engineering, University of Maryland, College Park, Maryland, USA

ABSTRACT: Since the beginning of serious study of liquefaction in the 1960s, uncertainties in both observed data and in processing field and laboratory tests have been major concerns. Assigning reasonable coefficients of variation to the parameters and correction factors in the conventional deterministic analyses indicates that a site with SPT N values between 12 and 22 and deterministic factors of safety of 1.5 can actually have liquefaction probability above 20%. About a third of the variance in the factor of safety comes from uncertainty in the load (magnitude scaling, stress reduction factor, etc.), which is independent of the method used to estimate the resistance. Researchers have traditionally presented the results of case studies in the form of charts showing instances in which liquefaction did and did not occur and have developed relations to separate the two. Although the original researchers developed the separation lines informally, recent work has applied statistical methods, such as discriminant analysis and logistic regression or combinations of them. In their original form, these methods give the sampling distributions of the observed data (i.e., the probability of observing the data given the hypothesis) rather than the probability of the hypothesis given the data, but the engineer needs the latter, that is, the probability of liquefaction given a set of observations. Researchers have addressed this issue using Bayesian methods, adopting non-informative priors to develop the results. Published curves of liquefaction probabilities can thus be interpreted as likelihood ratios. Other, independent work demonstrates that geological, meteorological, and historical data can be used to develop prior liquefaction probabilities that are not non-informative, so it may not be necessary to assume a non-informative prior. The actual prior can then be combined with the previously developed likelihood ratios to provide rational probabilities of liquefaction.

Although the earthquake literature (e. g., Richter 1958) describes many historical instances of liquefaction of soils during earthquakes, the scientific study of the phenomenon can be said to begin with the work of H. B. Seed and his colleagues (Seed and Lee 1966, Seed and Idriss 1971). Since the 1960s liquefaction has been a major concern of the geotechnical engineering community, and the literature on the subject has become vast. There have been at least two professional workshops attempting to elucidate the state-of-the-art (NRC/NAE 1985, Youd et al. 2001) and two editions of an EERI monograph on the subject (Seed and Idriss 1982, Idriss and Boulanger 2008). These four publications contain extensive bibliographies on the subject, which, in the interest of saving space, will not be repeated here. As this is being written, a new National Research Council study of the subject is under way.

Most engineering practice employs a form of the simplified method proposed by Seed and Idriss (1971), modified many times since, and presented in the most widely available modern form by Idriss and Boulanger (2008). The approach has two parts. In one, the soil's resistance to liquefaction is evaluated from in situ measurements such as the Standard Penetration Test (SPT), Cone Penetration Test (CPT), or shear wave velocity (vs), modified to account for effects that cannot be controlled during the test, and expressed as a Cyclic Resistance Ratio (CRR). In the other, the cyclic loading is described by a Cyclic Stress Ratio (CSR), which is calculated from parameters describing the earthquake acceleration, patterns of dynamic stress distribution, magnitude, and other features of the imposed loading. In practice the required values of CSR are usually found by estimating from observed field cases of liquefaction and non-liquefaction the CRR values required to prevent liquefaction. For a particular location and a particular expected earthquake loading, the ratio between the estimated values of CRR and CSR is defined as the Factor of Safety (FS), or FS = CRR / CSR.

There is a great deal of uncertainty in this process. This papr addresses two aspects of the uncertainty: (a) If the published critical relations between CSR and CRR are accepted as valid, how much uncertainty is introduced into FS by the uncertainty in the parameters that enter into the estimates of CSR and CRR? (b) When relations between CSR and CRR are stated in probabilistic terms, what do the results mean and how can they be used in practice?

1. PARAMETRIC UNCERTAINTY

The CSR at some location in the potentially liquefiable soil is conventionally defined by:

$$CSR = 0.65 \frac{\sigma_{vc}}{\sigma'_{vc}} \frac{a_{\max}}{g} r_d \tag{1}$$

The terms are as follows:

- 0.65 is a factor to account for the average effective acceleration in an accelerogram being less than the peak value. The value 0.65 has been generally accepted as a reasonable estimate of the combined effects of multiple cycles, and it is usually simply stated as the numerical value without any symbol. However, there is clearly some uncertainty in this value.
- σ_{vc} is total vertical stress on the horizontal plane.
- σ'_{vc} is effective vertical stress on the horizontal plane.
- amax is peak ground acceleration.
- g is the acceleration of gravity.
- rd is a stress reduction factor to account for the flexibility of the soil profile, which causes the shear stress transmitted through the profile to decrease with depth.

Different methods of evaluating the in situ conditions lead to different formulas for estimating CRR. If the SPT is used, the expression is

$$CRR = (f((N_1)_{60cs} \times NSF \times K_\sigma)$$

$$N_{1_{60cs}} = C_N C_E C_B C_R C_S N_m + \Delta N_{1_{60cs}} \tag{2}$$

In these equations:

- $f(N_1)_{60cs}$ is a function based on observed field behavior.
- MSF is a magnitude-scaling factor accounting for the fact that earthquakes of larger magnitude have more cycles of loading.
- Kσ brings all the empirical observations to the same overburden stress.
- CN accounts for the effect of depth on SPT.
- CE accounts for the effect of the transmitted energy.
- CB accounts for the effects of the diameter of the borehole.
- CR accounts for the effects of the length of the drill rod.
- CS accounts for the effects of the configuration of the sampler.
- $\Delta(N_1)_{60}$ accounts for the presence of fines.

If the cone penetration test (CPT) is used, the corresponding relations are

$$CRR = f(q_{c1Ncs}) \times MSF \times K_\sigma$$

$$q_{c1Ncs} = C_N q_{cN} + \Delta q_{c1N} \tag{3}$$

For the CPT tests:

- f((qc1Ncs) is a function of the corrected CPT based on observed field behavior.
- qcN is the measured tip resistance of the cone corrected for various factors such as layer

thickness, near surface values, and cone friction effects.

- Δq_{c1N} is an additional term to account for the presence of fines.

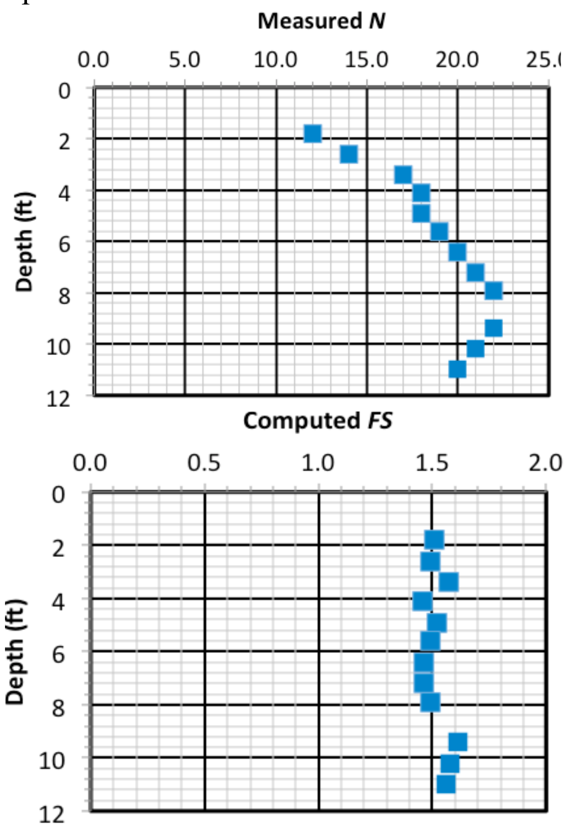


Figure 1: Adjusted N and FS from example of Idriss and Boulanger (2008): (a) Revised values of N, (b) Factors of Safety from revised N values

The corrections to the measured tip resistance involve a number of empirical relations that are not simple multiplication factors. The SPT and CPT cannot be used in coarse-grained soils, so correlations have been proposed for tests such as the Becker hammer test in those materials. The measured in situ shear wave velocity has also been used as an alternative to the SPT and CPT, and corrections similar to those for the SPT and CPT have been proposed.

Because the values of the coefficients and input parameters are not known precisely, there is uncertainty in applying current procedures to estimate liquefaction potential. A numerical study was carried out using the spreadsheet provided in Appendix A of Idriss's and Boulanger's (2008) monograph. The raw values of N in the example provided in that report are so low for most of the samples that the computed factors of safety are much less than 1.0. Thefore, values of N were raised so that all the factors of safety for the sand samples were close to 1.5. Figure 1 shows the revised values of N and the new factors of safety.

It is reasonable to assume that the parameters and the resulting values of CSR, CRR, and FS are log-normally distributed. If the mean and standard deviation of the values of a log-normally distributed variable are designated μ and σ, respectively, and the mean and standard deviation of the logarithms of those values are designated λ and ζ, respectively, and COV stands for the coefficient of variation of the data, then the well-known relations are

$$COV = \frac{\sigma}{\mu}$$

$$\varsigma^2 = ln1 + COV^2$$

$$\lambda = ln\mu + \tfrac{1}{2}\varsigma^2 \tag{4}$$

The variance of the logarithm of the product or ratio of several independent log-normally distributed variables can be found by simply adding the variances of the input. Since equations (1), (2), and (3) are largely multiplications and divisions, the calculations are straightforward. The major exception is the evaluation of the variances of the f () functions in equations (2) and (3). These are exponentials of polynomial expressions whose variances can be approximated by a first-order Taylor series.

The values of the COVs for the parameters were estimated from their published descriptions. The values selected for the present analysis are listed in Table 1. In order not to exaggerate the uncertainty, the values selected for most of the COVs are small. The two exceptions are the COVs of 0.20 for MSF and rd, but the range of proposed values for these parameters suggests that these parameters really are poorly known and incorporating large uncertainties is appropriate.

Table 1. COVs and values of ζ2 for parameters and coefficients

Parameter	COV	ζ2
N meas.	0.10	0.009950
CE	0.02	0.000400
CE	0.02	0.000400
CE	0.02	0.000400
CE	0.02	0.000400
CE	0.02	0.000400
MSF	0.20	0.039221
Kσ	0.07	0.004888
0.65 factor	0.05	0.002497
σ_{vc}/σ'_{vc}	0.05	0.002497
amax	0.10	0.009950
rd	0.20	0.039221

The spreadsheet was modified to incorporate the variances and to compute the variances of CRR and CSR, the reliability index, and the probability of liquefaction. For example, at a depth of 4.1m, N=18, and

$$\varsigma_{CSR}^2 = 0.054165 \quad \varsigma_{CRR}^2 = 0.084555$$

$$\varsigma_{FS}^2 = 0.138720 \quad \varsigma_{FS} = 0.372541$$

$$\lambda_{FS} = 0.305101 \; \beta_{\lambda_{FS}/\varsigma_{FS}} = 0.819170$$

$$p_{fliq} = \Phi - \beta = 0.206345 \approx 20\% \qquad (5)$$

Results for the other samples are comparable. Figure 2 gives the results for the entire profile.

In this example the values of N and the computed deterministic factors of safety of 1.5 are so large that most practicing geotechnical engineers would consider the site safe against liquefaction. Nevertheless, there is significant probability of liquefaction, in part because multiplying several uncertain parameters has a strong impact on overall uncertainty. Even when the calculated factor of safety is 1.5, there is 20% probability of liquefaction simply because of parametric uncertainty.

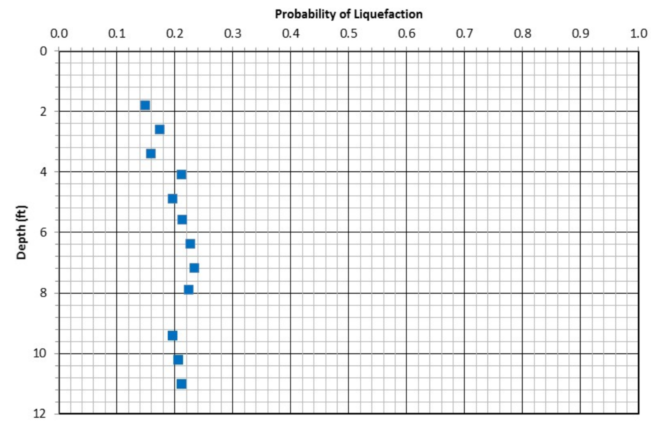


Figure 2. Probabilities of liquefaction computed from uncertainties in coefficients and input parameters, from the example of Idriss and Boulanger (2008)

Because there are fewer uncertain coefficients in the CPT-based approach, the estimated uncertainty is less. However, the uncertainty in CSR, which contributes at least one third of the overall uncertainty, is unchanged from the SPT-based analysis, so large uncertainty exists in the results, even when the resistance is based on in situ tests such as CPT or vs.

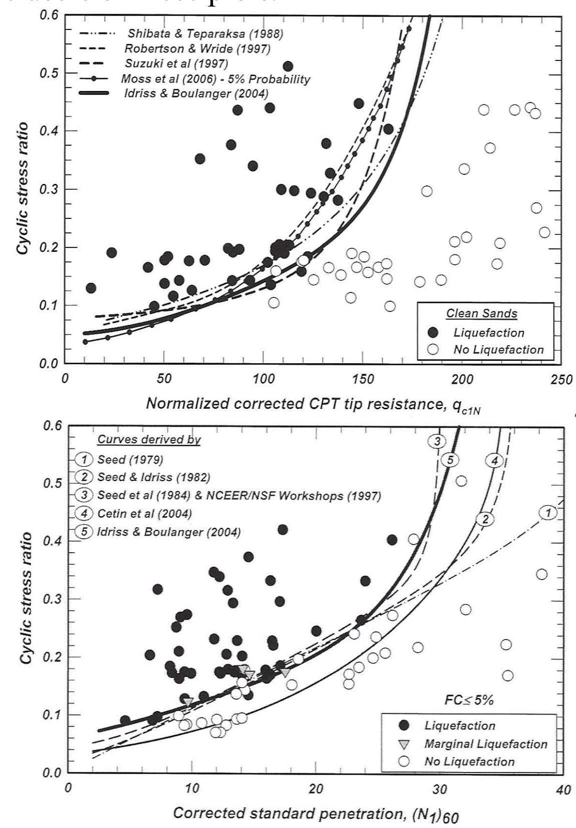
2. QUESTION (B) – PROBABILISTIC CRR – CSR RELATIONS

Almost all the criteria for relating CRR and CSR to in situ measurements are derived from observations at sites where liquefaction either did or did not occur during earthquakes. Such studies are inevitably beset the scarcity of data taken before the earthquake occurred, relatively fewer cases in which liquefaction did not occur, and difficulties distinguishing values representative of the site as a whole from a mass of data from individual borings.

The data are usually presented in a plot against two axes. The horizontal axis consists of units such as (N1)60, qc1, vs, or some other parameter representing the normalized measurement of soil strength. The vertical axis is expressed in terms of the estimated load imposed by the earthquake. Cases in which liquefaction occurred are plotted as filled-in circles, and cases without liquefaction are represented as open circles. Figure 3 contains two typical versions, as developed by Idriss and Boulanger (2008), one for SPT results the other for CPT results. The task is then to develop a function or line or surface that separates the closed from the open circles and that can be expressed as a function of the corrected strength. The lines drawn in Figure 3 are based largely on judgment, and it is not surprising that different researchers have come to different conclusions about where the division between the closed and open circles should be placed. There are essentially two reasons for the discrepancies: (a) argument over how to categorize particular case histories and what values to assign to the parameters and (b) disagreement over how to separate the two regions even after the locations of the pints have been established.

The problem of separating two regions described by empirical observations arises in many fields of science and engineering, so it is not surprising that, since Fisher (1936) addressed it, many approaches have been proposed, each having advantages and disadvantages, proponents and disparagers. There are essentially two ways to address the problem: discriminant analysis and logistic regression analysis. Discriminant analysis essentially rotaties the axes to find the orientation that maximizes the separation between the two sets of points. Logistic regression is an iterative procedure that adjusts the parameters in a function of the data (often a polynomial) so that the likelihood of observing the data is maximized. It can be shown that, under some not-too-unreasonable conditions, the two approaches give the same result. Although the first attempt to apply these methods to the liquefaction problem by Christian and Swiger (1975) employed linear discriminant analysis, Liao et al. (1988) used logistic regression, and that has become the preferred approach (Jha et al. 2009, Juang et al. 2002, 2006, Moss et al. 2006).

Regardless of the mathematical tool, the results depend strongly on choices made by the analyst, especially on the way the data are normalized and on the form of the function that is to separate the two classes of observations. It is remarkable that the published discrimination functions developed by different researchers differ by only about 30% in the principal region of interest. There are, however, several issues that remain to be addressed or even to be understood by the users of these plots.



Based on SPT Results (b) Based on CPT Results
Figure 3. Typical Plots of CSR versus Field Data, from Idriss and Boulanger (2008)

One of the most important is the meaning of the probability of misclassification in classic discriminant or logistic regression analysis. The probability in those analyses is the probability that, if a case belongs to one class, the data will indicate that it belongs in the other. In other words, it is the probability that, if the site actually liquefies, the data indicate that the site should be safe. What the engineer wants is the reverse of this. The engineer wants to know, given a set of data that indicate a safe site, the probability that the site will liquefy. Originally (e. g., Christian and Swiger 1975) plotted the probabilities of misclassification, that is, without Bayesian updating. More recently Juang et al. (2002, 2006), Cetin et al. (2002, 2004), Moss et al. (2006), Kayen et al. (2013), and others have applied Bayesian updating. This requires a prior estimate of the probability of liquefaction and an estimate of the probability that a site that does not liquefy has data indicating that it will. Since researchers cannot have an estimate of the prior probability of liquefaction for the abstract exercise of creating general plots, a common procedure is to assume a non-informative prior. This is, in fact, what was done in most of the published analyses, and it means that the analysts assumed equal prior probability of liquefaction and non-liquefaction, or 0.50. Figure 4 is a generic figure, similar to those presented by the above researchers, showing the shape of the typical probability curves, without specifying the parameters for either axis, but, in the more recent work, representing Bayesian updating.

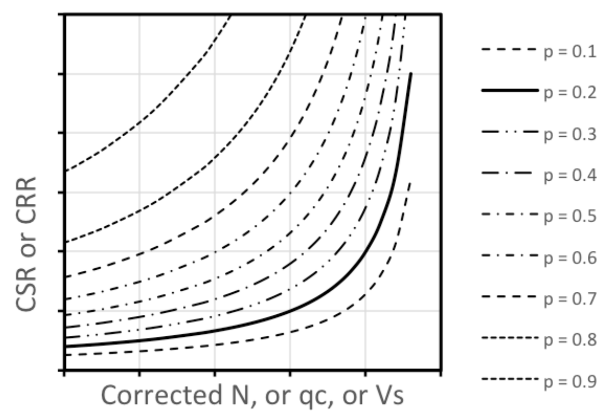


Figure 4. Generic Plot of Probabilities of Liquefaction Reflecting Bayesian Updating.

If Bayes's theorem is expressed in terms of odds, it becomes a statement that the posterior odds of an event are equal to the prior odds multiplied by the Likelihood Ratio, which is the probability of observing the data in the case of liquefaction divided by the probability of observing the data in the case of no liquefaction:

$$\frac{P[liquefaction|data]}{P[no\ liquefaction|data]} =$$

$$\frac{P_0[liquefaction]}{P_0[no\ liquefaction]} \times \frac{P[data|\square iquefaction]}{P[data|no\ liquefaction]} =$$

$$\frac{P_0[liquefaction]}{P_0[no\ liquefaction]} \times LR \tag{6}$$

For a non-informative prior, the prior odds are 0.5/0.5 = 1.0, and equation (6) reduces to

$$LR = \frac{P}{1-\ P} \tag{7}$$

in which P is the probability for any of the lines in Figure 4. Let us suppose that the data for a particular site fall on the heavy line with 20% probability. The odds are 0.2 / (1-0.2) = 0.25., and this is the value of LR as well, and plots like Figure 4 can be converted into plots of LR. It might be more convenient if reports of Bayesian analyses of liquefaction data presented the results as LR plots instead of probabilities with assumed non-informative priors.

This is of more than academic interest when dealing with a case for which there is an informed prior. For example, Prof. Baise and her colleagues have developed methods for estimating the probability of liquefaction on the basis of geological, meteorological, and historical data (Brankman and Baise 2008, Zhu et al. 2014). Consider what happens for a ssite where their procedure identified an overall probability of liquefaction of 0.8. This is then a prior probability of liquefaction. If the field data for this site falls on the 20% line in Figure 4, corresponding to LR = 0.25, the posterior probability of liquefaction is found from

$$\frac{P[liquefaction|data]}{1-P[no\ liquefaction|data]} =$$

$$\frac{P_0[liquefaction]}{1-P_0[no\ liquefaction]} \times \mathrm{LR} = \frac{0.8}{0.2} 0.25 = 1.0$$

$$P[liquefaction|data] = 0.5 \tag{8}$$

This compares with the values of 0.20 for the non-informative prior. The prior probability can have a big effect.

3. COMBINED EFFECTS

Idriss and Boulanger (2010) give an explicit form to the CRR curve and give the following equation for the probability of liquefaction, that is, the probability of misclassifying the data for a particular site:

$$P_L N_{1,60cs} CSR_{M=7.5,\sigma'_v=1atm} = \left[\frac{\left(\frac{N_{1,60cs}}{14.1}\right)+\left(\frac{N_{1,60cs}}{126}\right)^2+\left(\frac{N_{1,60cs}}{23.6}\right)^3+\left(\frac{N_{1,60cs}}{125.4}\right)^4 - 2.67 - CSR_{MM=7.5,\sigma'_v=1atm=7.5,}}{\sigma_{lnR}}\right] \quad (9)$$

The numerator of equation (9) is the function embedded in the Idriss and Boulanger (2008) spreadsheet, and R is shortened notation for CRR. Idriss and Boulanger (2010) recommend that $\sigma_{\ln(R)} = 0.13$.

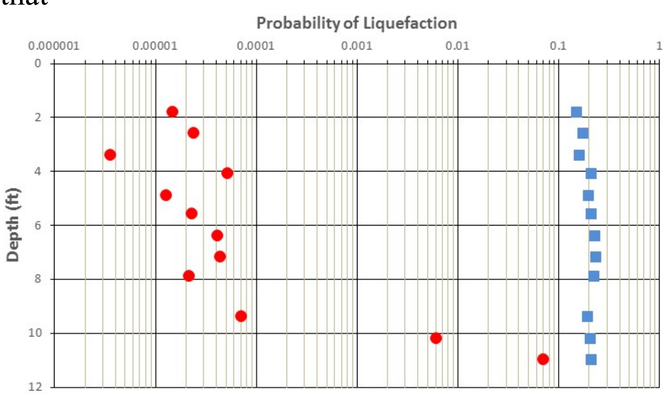


Figure 5. Probabilities of liquefaction from uncertainty in location of CSR/CRR curves (red circles) and data uncertainty (blue squares)

The same parametric values as in Figure 1 are input to equation (9), and the results are plotted in a modified version of Figure 2, shown as Figure 5. The red points represent the probabilities of liquefaction from equation (9), given that the values of the input data for a specific point are known precisely. In other words, they represent the uncertainty in the location, with respect to the observed case study data, of the CSR/CRR curves. The blue dots represent the probabilities of failure assuming that equation (9) is absolutely true but the values of the input data are uncertain. The simplest way to combine the two results is to assume they are independent. For most of the points in Figure 5 the uncertainty in equation (9) is so small that the scatter dominates, but in the case of the lowest point the probabilities are approximately 0.07 and 0.21, so the combined probability is 0.265.

4. CONCLUSIONS

This paper has examined only two aspects of the uncertainties in liquefaction evaluations. However, even this limited exploration reveals some points:

The uncertainties in the parameters in the conventional computations of FS are large enough that there is significant uncertainty in the result. A case with N values of 15 to 20 and estimated FS = 1.5 may have 20% probability of liquefaction even though the probability associated with the location of the CSR/CRR curves is small.

One of the principal reasons for the large uncertainty in FS is the number of multiplied correction factors, each contributing to the overall uncertainty. Attempts to refine the precision of the analyses may have increased their uncertainty.

People presenting plots of probabilities from empirical studies of liquefaction should state whether these are probabilities of liquefaction, of observing the data, or of something else. The explanation should be on the figure, or at least in its caption. It should not be buried in the text.

Many papers describe liquefaction probabilities computed by Bayesian updating using non-informative priors, which imply that the prior odds are 1.0 and that the likelihood ratio can be computed from the reported posterior probabilities. However, plots of the actual likelihood ratios may be more useful.

Although non-informative priors may be necessary in generic studies, prior probabilities often do exist for actual sites, and they have a major effect on the computed posterior probability of liquefaction.

5. REFERENCES

Brankman, C. M., and Baise, L. G. (2008) “Liquefaction susceptibility mapping in Boston, Massachusetts,” Environmental and Engineering Geoscience, 14, 1–16.

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

Cetin, K. O., Seed, R. B., Der Kiureghian, A., Tokimatsu, K., Harder, L. F., Jr., Kayen, R. E., and Moss, R. E. S. (2004) “SPT-based probabilistic and deterministic assessment of seismic soil liquefaction potential, Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 130(12): 1314-1340.

Christian, J. T., and Swiger, W. F. (1975) "Statistics of Liquefaction and SPT Results," Journal of the Geotechnical Engineering Division, ASCE, 101(GT11): 1135-1150, also closure (1976) 102(GT12):1279-1281.

Fisher, R. A. (1936) “The use of multiple measurements in taxonomic problems,” Annals of Eugenics 7: 179-188.

Idriss, I. M., and Boulanger, R. W. (2008) Soil Liquefaction during Earthquakes, Earthquake Engineering Research Institute, MNO-12, Oakland, CA, 243 pp.

Idriss, I. M., and Boulanger, R. W. (2010) “SPT-Based Liquefaction Triggering Procedures,” Report UCD/CGM-10/02, Department of Civil and Environmental Engineering, University of California, Davis.

Jha, S. K., and Suzuki, K. (2009) "Liquefaction Potential Index Considering Parameter Uncertainty," Engineering Geology, 107:55-60.

Juang, H. C., Jiang, T., and Andrus, R. D. (2002) "Assessing Probability-based Methods for Liquefaction Potential Evaluation," Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 128(7): 580-589.

Juang, H. C., Fang, S. Y., and Khor, E. H. (2006) "First-Order Reliability Method for Probabilistic Liquefaction Triggering Analysis," Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 132(3): 337-350.

Kayen, R. E., Moss, R. E. S., Thompson, E. M., Seed, Cetin, K. O., Der Keureghian, A., Tanaka, Y., and Tokimatsu, K. (2013) "Shear wave velocity-based probabilistic and deterministic assessment of seismic soil liquefaction potential," Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 139(3): 407-419.

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

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

National Research Council/National Academy of Engineering (NRC/NAE) (1985) Liquefaction of Soils during Earthquakes, National Academy Press, Washington, DC.

Richter, C. F. (1958) Elementary Seismology, W. H. Freeman & Co., San Francisco, CA, 768 pp.

Seed, H. B., and Lee, K. L. (1966) “Liquefaction of saturated sands during cyclic loading,” Journal of the Soil Mechanics and Foundations Division, ASCE, 92(SM6): 105-134.

Seed, H. B., and Idriss, I. M. (1971) “Simplified procedure for calculating soil liquefaction potential,” Journal of the Soil Mechanics and Foundations Division, ASCE, 97(SM9): 1249-1273.

Seed, H. B., and Idriss, I. M. (1982) Ground Motions and Soil Liquefaction during Earthquakes, Earthquake Engineering Research Institute, MNO-5, Berkeley, CA, 134 pp.

Youd, T. L., and 20 others (2001) “Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils,” Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 127(10): 817-833.

Zhu, J., Daley, D., Baise, L. G., Thompson, E. M., Wald, D. J., Knudsen, K. L. (2014) “A geospatial liquefaction model for rapid response and loss estimation,” accepted