Chapter 5

Reliability Analysis within a Bayesian Framework

5.1 Abstract

A Bayesian framework allows for careful and thorough treatment of all types of uncertainties associated with the vagaries of observed liquefaction/non-liquefaction. Using a statistical framework and parameter estimation technique of this type permits the formulation and optimization of the model to be based on the underlying physics of the problem. This chapter outlines procedures for parameter estimation using CPT (cone penetration test) data, and the development of probabilistic triggering correlations. The results are curves of equal probability of seismic liquefaction triggering within normalized load vs. resistance space, for CPT field measurements, which can be used in performance-based engineering decisions

5.2 Introduction

Bayesian estimation techniques have enormous utility in engineering statistics. Bayes rule, which has a colored history in classical statistical fields, has found a home in the engineering decision process. Some advantages of the statistical approach used in this study are; full consideration of all types of uncertainties, the formulation of the limit-state with respect to the underlying physics of the problem, the facility to incorporate subjective information, and the ease of updating parameter estimation. These advantages not only produce a good estimation of the model parameters but also lead to a deeper understanding of the problem at hand.

5.3 Bayes Rule

Bayes rule is derived from simple rules of conditional probability, yet the simplicity portends little of the power of the Bayesian technique. Bayes rule can be written as (Box & Tiao, 1992),

$$f(\Theta) = c \cdot L(\Theta) \cdot p(\Theta) \tag{5.1}$$

where $L(\Theta)$ = Likelihood function containing an unknown set of parameters Θ

$p(\Theta)$ = prior distribution

$c = [\int L(\Theta) \cdot p(\Theta) \cdot d(\Theta)]^{-1}$ = normalizing constant

$f(\Theta)$ = posterior distribution.

The likelihood function is proportional to the conditional probability of observing a particular event given a set of parameters. The likelihood function incorporates the objective information that, in this case, is the statistical measurements associated with liquefaction/non-liquefaction. The prior distribution can include subjective information known about the distributions of the model parameters. The posterior distribution incorporates both the objective and subjective information into the distributions of the model parameters. The process of performing Bayesian updating involves formulating the likelihood function, selecting a prior, calculating the normalizing constant, and then calculating the posterior statistics.

The prior distribution tends to be the most controversial issue for detractors of Bayesian methods. Box & Tiao (1992) have shown that the use of a non-informative prior can lead to a wholly unbiased estimate of model parameters. A non-informative prior allows the data, through the likelihood function, to fully dictate the posterior distribution. A non-informative prior has no effect on the shape of the prior distribution, and is used when no prior information about the parameters is available.

5.4 Uncertainties in Liquefaction/Non-Liquefaction Data

In all studies that involve empirical data, the quality of the data and rigor of data processing is the dominant factor that affects the accuracy of the model. In a probabilistic approach, carefully quantifying the uncertainty is of utmost importance. The earthquake phenomenon of liquefaction/non-liquefaction is particularly suited for a probabilistic analysis because of the various associated uncertainties.

The uncertainty of a variable can be assessed in many ways. If the variable is directly measured (e.g. CPT tip and sleeve resistance) then some form of inherent variability and measurement error comprises the overall uncertainty. If the variable is a function of other independent variables (e.g. effective stress) then a combination of inherent variability, measurement error, and model error is present. Implicit in the quantification of all these forms of uncertainty is some aspect of human error. All these different types of uncertainty have been accommodated in this study. Examples of how the uncertainty is estimated for different cases are described below.

5.4.1 Data Treatment

The cyclic stress that may (or may not) have caused liquefaction includes measured variables such as the intensity of the ground shaking at the site and the total and effective stress conditions in the ground. The simplified procedure for evaluating stresses induced by earthquakes (Seed & Idriss, 1971) was used in this study,

$$CSR = \frac{\tau_{avg}}{\sigma'_v} = 0.65 \cdot \frac{a_{\max}}{g} \cdot \frac{\sigma_v}{\sigma'_v} \cdot r_d \tag{5.2}$$

Each of these variables has an associated uncertainty. The peak ground acceleration at the site is based on the best estimation of ground shaking possible (e.g. strong motion recordings, site response, calibrated attenuation relationships, adjustment of estimated site PGA through general site response modeling, etc.). A coefficient of variation (δ) ranging from 0.1 to 0.5, is prescribed depending on the method that is used to estimate a_{max}. This is a subjective determination of the variance of the ground shaking but is based on typical uncertainty bands from general attenuation relationships (Abrahamson & Silva, 1997).

The uncertainty associated with the nonlinear shear mass participation factor, r_d, was based on a brute force statistical method where 2,153 site response analyses were run using 50 sites and 42 ground motions covering a comprehensive suite of motions and soil profiles (Cetin et al., 2003). The variance was then taken from the dispersion of these simulations.

The uncertainties associated with total and vertical effective stresses, σ_v and σ_v', are functions of the uncertainties of the independent variables from which they are determined. This lends to a first order approximation of the uncertainty about the mean using a Taylor series expansion (Ang & Tang, 1975). These two variables tend to be cross-correlated, so their covariance was taken into account in estimating the uncertainty. The variance of the cyclic stress ratio, CSR, is also determined using a series expansion.

The cyclic strength, or the cyclic resistance, of the soil in question is based on index tests. The CPT index test is an *in situ* procedure that has been thoroughly studied. Because this index test have been carefully calibrated, the measurement error can be reasonably estimated (Kulhawy & Trautman, 1996).

The results of the estimated means and standard deviations can then be plotted as in Figure 5.1. This figure shows the variability associated with each site depicted as one standard deviation error bars about the mean. The relative spread or dispersion of the error indicates how much informational content each site holds.

5.5 Limit-State and Model Formulation

A general reliability formulation is based on a threshold called the limit-state function. The limit-state function goes to zero where stress and strength are equal. This defines two regions, one a failure region and the other a safe region. The probability of failure is the integration of the joint probability density of the stress and the strength over the failure region (Der Kiureghian, 1999).

This limit-state can be defined in any manner that corresponds to the observations. Some parameter estimation methods such as System Identification and Artificial Neural Networks define the limit-state using a black box approach, optimizing a randomly determined mathematical function for the given database. This may give a good fit to a specific database but offers little or no insight into the fundamental phenomena controlling the results. The approach used in these studies was to define the limit-state using an understanding of the physics of liquefaction based on the theoretical underpinnings of critical state theory, the knowledge garnered from laboratory experiments, and past deterministic and probabilistic studies.

Using a limit-state that is grounded in an understanding of soil mechanics then makes the model fitting a numerical experiment in liquefaction. The limit-state is a generalized mathematical model for separating liquefaction from non-liquefaction. By incorporating all pertinent variables and using the database as the litmus test, an optimum function can be defined that then may give further insight into the soil behavior.

The limit-state function (g) is,

$$g = \hat{g} + \varepsilon \tag{5.3}$$

$$\hat{g} = q_{c,1} \cdot (1+\theta_1 \cdot R_f) + (\theta_2 \cdot R_f) + c \cdot (1+\theta_3 \cdot R_f) - \theta_4 \cdot \ln(CSR) - \theta_5 \cdot \ln(M_w) - \theta_6 \cdot \ln(\sigma_v') - \theta_7 \tag{5.4}$$

where CSR is the simplified cyclic stress ratio, M_w is the moment magnitude, σ_v' is the effective stress, $q_{c,1}$ is the corrected CPT tip resistance, R_f is the friction ratio, c is the CPT normalization exponent, the θ's are model parameters, and the ε is the model error term. The limit-state is the threshold between liquefaction (g≤0) and non-liquefaction (g>0). We are solving this generalized limit-state function for the unknown θ's and ε.

Including a large number of variables can aid in capturing interactions that may otherwise be opaque. Including the moment magnitude allows for regression of the duration weighting factor, DWF_M (a.k.a. magnitude scaling factor, MSF) directly from the database. Including friction ratio gives insight into how fine material affects a soil's liquefiability. By including effective stress the importance and functional shape of K_σ can be assessed.

5.6 Bayesian Updating

The process of model parameter estimation, or testing the limit-state function against the data, is accomplished through Bayesian updating. According to Equation 5.1, to solve for the posterior distribution of the model parameters and model error term we must define the likelihood function, determine a prior distribution, and solve for the normalizing constant.

The likelihood function of seismic soil liquefaction initiation is the product of the probabilities of observing k liquefied sites and n-k non-liquefied sites. Thus,

$$L(\mathrm{X},\Theta,\varepsilon) \propto P\left[\bigcap_{i=1}^{k}\{\hat{g}(\mathrm{X}_i,\Theta_i)+\varepsilon_i \le 0\}\bigcap_{i=k+1}^{n}\{\hat{g}(\mathrm{X}_i,\Theta_i)+\varepsilon_i > 0\}\right] \tag{5.5}$$

where X are the observable or measurable variables and Θ are the unknown model parameters.

If we combine the uncertainty from the important variables and model error into a cumulative error term, σ_Σ, the likelihood function can be rewritten as,

$$L(\mathrm{X},\Theta,\sigma_\Sigma) \propto \prod_{i=1}^{k}\Phi\left[-\frac{\hat{g}(\mathrm{X}_i,\Theta_i)}{\sigma_{\Sigma i}}\right]\cdot\prod_{i=k+1}^{n}\Phi\left[\frac{\hat{g}(\mathrm{X}_i,\Theta_i)}{\sigma_{\Sigma i}}\right] \tag{5.6}$$

where Φ is the standard normal cumulative distribution function.

A non-informative prior has been employed in this study, which allows likelihood function to dominate the posterior distribution. A non-informative prior distribution, by definition, will have as little affect on the posterior distribution as possible. The non-informative prior for a non-negative model parameter, should be uniform for ln(θ), which results in $p(\theta) \propto \theta^{-1}$ (Box & Tiao, 1992).

Importance sampling was used to calculate the normalizing constant integral. This is a Monte Carlo type simulation technique that efficiently approximates the complex integral numerically.

The posterior distribution describes the vital statistics of the model parameters and the error term. Using the error term as a gage for model fit, various permutations of the model within the limits of the described physics were tried. The objective was to minimize the model error, thereby giving the best fit to the data.

5.7 Sampling Bias

It has been recognized that a bias in the number of liquefied vs. non-liquefied data points exists. This bias can impact the results, resulting in an unbiased prediction. Cetin et al. (2002) explored this bias and presented a consistent method to account for what is called choice-based sampling bias as applied to the problem of liquefaction triggering using SPT (standard penetration test) data, which is an analog to this study.

The approach was based on Bayesian updating optimization, expert consensus, and sensitivity studies. The likelihood function (Equation 5.6) is modified for the data imbalance using a weighting factor (w),

$$L(\mathrm{X},\Theta,\sigma_{\Sigma}) \propto \prod_{i=1}^{k} \Phi\left[-\frac{\hat{g}(\mathrm{X}_i,\Theta_i)}{\sigma_{\Sigma i}}\right]^{w_{liquefied}} \prod_{i=k+1}^{n} \Phi\left[\frac{\hat{g}(\mathrm{X}_i,\Theta_i)}{\sigma_{\Sigma i}}\right]^{w_{non-liquefied}} \tag{5.7}$$

The weighting factor used in this study is $w_{\text{non-liquefied}} / w_{\text{liquefied}} = 1.5$, based on Cetin et al. (2002) and consensus of the expert review panel that reviewed the CPT database.

5.8 Reliability Analysis

The probability of liquefaction is calculated from the posterior distribution determined in the Bayesian updating. It should be noted that the probability associated with Bayesian methods is considered an expression of *degree-of-belief*, whereas in classical methods the probability is considered a measure of *relative frequency*. The probability of liquefaction can be estimated by a summation of the probabilities of all possible combinations of parameters that will define liquefaction. For any given set of variables, X, this requires integration over the liquefaction domain (g ≤ 0). The result is,

$$P[\hat{g}(\mathrm{X},\Theta)+\varepsilon \le 0] = \int_{\hat{g}(\mathrm{X},\Theta)+\varepsilon \le 0} \varphi(\varepsilon|\sigma_{\varepsilon}) \cdot f(\Theta,\sigma_{\varepsilon}) \cdot d\varepsilon \cdot d\Theta \cdot d\sigma_{\varepsilon} \qquad \text{Eq. 5.7}$$

The solution of this requires multi-fold integration. Good approximations of the resultant probabilities were achieved using a MVFOSM (mean value first order second moment) approach, because the failure surface was well behaved or not too strongly non-linear. These results were validated using FORM (first order reliability method), SORM (second order reliability method), and Monte Carlo simulations as coded in the program CALREL (Liu et al., 1989).

Contours of equal probability for the CPT are shown in Chapter 6. These probabilistic correlations are unbiased and appropriate for use in performance-based analyses.

5.9 Summary

The Bayesian framework used in this study allowed for careful and thorough treatment of all types of uncertainties associated with seismic soil liquefaction. The underlying physics of the problem was incorporated into the limit-state fuction. A process of numerical experimentation was performed to find the optimum formulation within the given physical bounds. Presented herein, are the concepts and procedures for estimating model parameters with Bayesian updating using CPT (cone penetration test) data, and the development of probabilistic liquefaction triggering correlations.

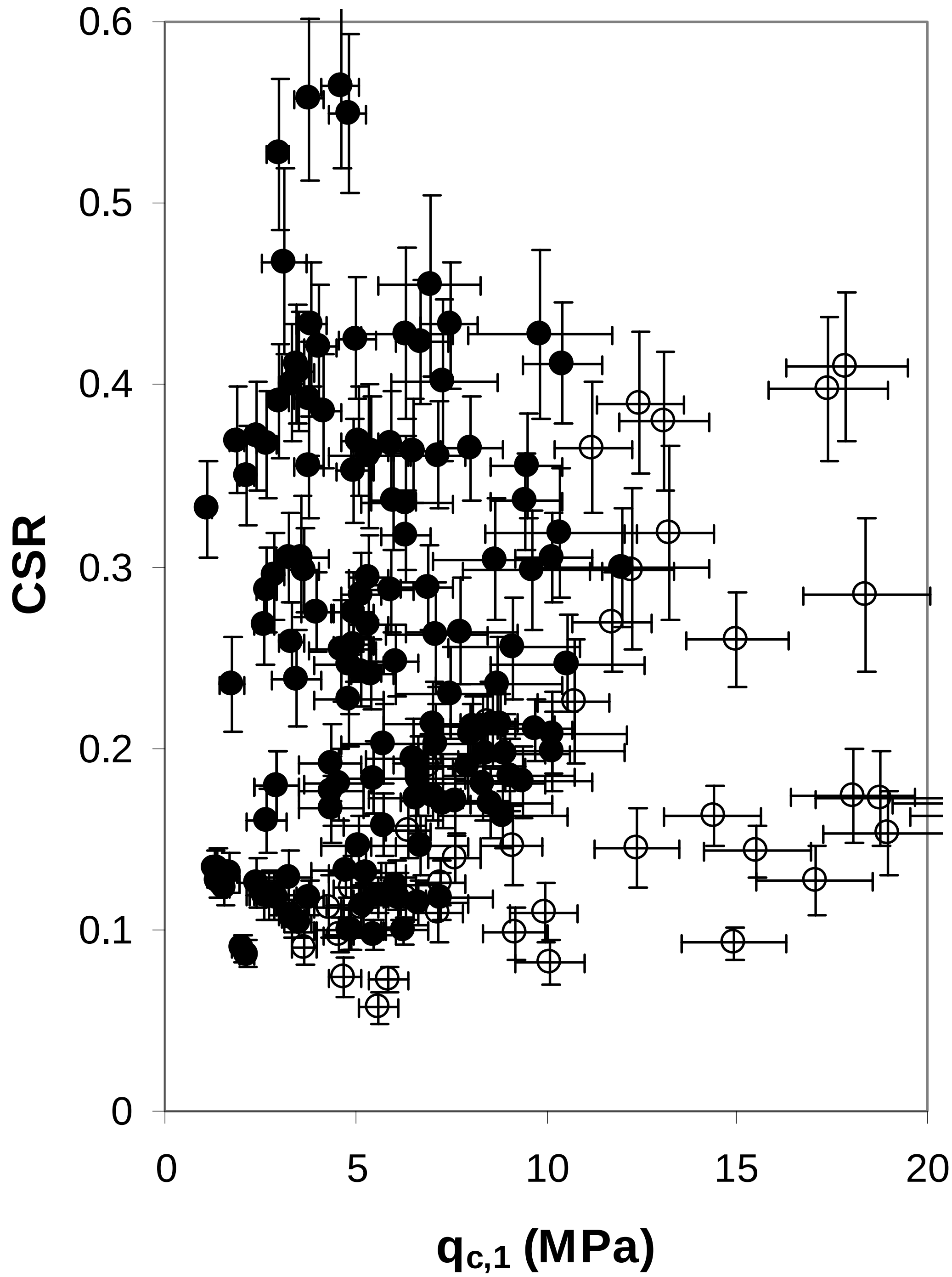

Figure 5.1 Plot showing liquefaction (dots) and non-liquefaction (circles) data points with ±1 standard deviation error bars.